Graphical Models for Multivariate Stationary Processes in Continuous Time

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Lea Helen Schenk, M.Sc.

Tag der mündlichen Prüfung:	16. April 2024
Referentin:	Prof. Dr. Vicky Fasen-Hartmann
Korreferentin:	Prof. Dr. Claudia Klüppelberg

Für meine Eltern, die immer an mich glauben.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen bedanken, die mich auf meinem Weg unterstützt und meine Promotionszeit am Institut für Stochastik und in der Arbeitsgruppe Didaktik der Mathematik geprägt haben.

Zunächst möchte ich mich bei Prof. Dr. Vicky Fasen-Hartmann für die hervorragende Betreuung in den letzten Jahren bedanken. Ohne die gemeinsamen fachlichen Diskussionen, ihre vielfältigen Anregungen und ihr Engagement wäre diese Arbeit sicher nicht entstanden. Sie hat sich immer viel Zeit für unsere Gespräche genommen und stand mir jederzeit mit einem offenen Ohr für Fragen und Probleme zur Verfügung. Sie hat mich in den richtigen Momenten unterstützt, wenn ich mich in Details festgefahren habe, und mich gleichzeitig selbstständig arbeiten lassen, wenn ich mich in eine Fragestellung vertieft habe.

Mein Dank gilt auch Prof. Dr. Claudia Klüppelberg, die sich bereit erklärt hat, meine Korreferentin zu sein. Auch wenn wir uns erst bei der Verteidigung persönlich kennenlernen, hat sie mich mit ihren positiven und zuversichtlichen Nachrichten wiederholt im richtigen Moment ermutigt.

Ich möchte mich außerdem ganz herzlich bei allen aktuellen und ehemaligen Mitgliedern des Instituts für Stochastik und der Arbeitsgruppe Didaktik der Mathematik bedanken. Danke für die wunderbaren gemeinsamen Jahre, die gemeinsamen Kaffee- und Mittagspausen und die erholsamen Nachmittagsspaziergängen, die geprägt waren von fachlichem und persönlichem Austausch. Ich habe mich während der gesamten Zeit an meinem Arbeitsplatz sehr wohl gefühlt und bin dankbar für die neuen Freunde, die ich hier kennenlernen durfte. Einige Personen möchte ich besonders hervorheben.

Aus der Arbeitsgruppe Didaktik der Mathematik möchte ich mich zunächst bei Ingrid bedanken, die mir meine Promotionsstelle ermöglicht hat. In diversen gemeinsamen Lehrveranstaltungen konnte ich viel von ihr lernen. Gleichzeitig hat sie mir in den gemeinsamen Seminaren große Gestaltungsfreiheit gelassen, Vertrauen in mich gesetzt und mich in meinen Entscheidungen bestärkt. Ebenso möchte ich mich bei Prof. Dr. Sebastian Bauer für die Unterstützung, die Zusammenarbeit und die Diskussionen über die Didaktik der Stochastik bedanken, von denen ich ebenfalls viel gelernt habe und weiter lernen darf. An dieser Stelle möchte ich außerdem Wolf nennen, der mich im letzten Jahr in der Lehre unterstützt hat, wenn es mal wieder zu viele Aufgaben gleichzeitig gab. Nicht zuletzt möchte ich Stephanie, Stephan und Peter erwähnen, deren Freundschaft ich nicht missen möchte und die immer Zeit und ein offenes Ohr hatten, wenn ich einen Perspektivwechsel benötigte.

Aus dem Institut für Stochastik möchte ich zunächst Bruno erwähnen, der mich, wie so viele, im richtigen Moment meines Studiums abgefangen und mein Interesse an der Stochastik geweckt hat. Außerdem möchte ich mich bei Lucas bedanken, der mich regelmäßig zu den dringend notwendigen Spaziergängen animiert hat, und bei Maximilian für seine Ratschläge und motivierenden Worte. Natürlich dürfen auch unsere wöchentliche Institutssportgruppe und die Boulderfreunde nicht unerwähnt bleiben, die über die letzten Jahre hinweg ebenfalls für den unverzichtbaren Ausgleich gesorgt haben.

Ganz besonders möchte ich mich noch bei Tamara und Steffen bedanken. Beide durfte ich bereits während meines Studiums kennenlernen und sie haben mich bei der Entscheidung zur Promotion und auf dem Weg dorthin unterstützt. Ohne sie wäre die Promotionszeit nur halb so schön gewesen. Bei Steffen bedanke ich mich vor allem für sein Interesse an fachlichen Detaildiskussionen, für seine anspornenden und aufmunternden Worte und natürlich für seine sportliche Unterstützung. Bei Tamara möchte ich mich für die unzähligen fachlichen und nichtfachlichen Kaffeegespräche bedanken. Ob es darum ging, Ideen zu finden, Beweise zu diskutieren, oder meine Gedanken zu sortieren, wenn ich mir selbst nicht mehr traute, bei Tamara konnte ich jederzeit am Büro stehen, egal was gerade los war. Besonders in den letzten Monaten war ihre moralische Unterstützung von unschätzbarem Wert.

Nicht zuletzt möchte ich mich bei meiner Familie für ihre liebevolle Unterstützung und Ermutigung bedanken. Danke, dass ihr immer an mich geglaubt habt und mir vor allem in der finalen Phase der Promotionszeit bewusstgemacht habt, wie viel ich schon erreicht habe.

Karlsruhe, Februar 2024 *Lea Helen Schenk*

Abstract

Graphical models are important and widely used tools for visualising dependency structures in various contexts. Despite their significance, graphical models for multivariate stationary stochastic processes in continuous time have received little research attention. This thesis addresses the gap by introducing three graphical models for multivariate stationary stochastic processes in continuous time: the causality graph, the local causality graph, and the partial correlation graph. The (local) causality graph is a mixed graph in which the edges represent appropriate concepts of Granger causality and contemporaneous correlation. In contrast, the partial correlation graph is an undirected graph, with edges representing partial correlations. The three graphical models are introduced in the following steps. First, we define and analyse the key concepts mentioned above, which form the basis of the edges in the graphical models. We then analyse the properties of the resulting graphs, with particular emphasis on Markov properties, in order to demonstrate the soundness of the models. We also derive various characterisations of the edges. Finally, we apply the graphical models to output processes of state space models such as MCAR processes. Here, our attention is directed towards the characterisations of the edges by model parameters, which in turn offer insightful interpretations.

Prior Publications

Substantial parts of the work presented in this thesis have resulted in preprints available on arXiv and submitted to peer-reviewed journals. Direct quotations from these publications appear throughout the thesis without being explicitly marked as such, to ensure better readability. As the publications cover related topics based on the same principles, the results have been merged, rearranged, and partly extended to ensure a consistent thesis. The publications are listed in order of the date of publication:

- Fasen-Hartmann and Schenk (2023b). Mixed causality graphs for continuous-time stationary processes. Preprint available at `https://doi.org/10.48550/arXiv.2308.08890`
- Fasen-Hartmann and Schenk (2023a). Mixed causality graphs for continuous-time state space models and orthogonal projections. Preprint available at `https://doi.org/10.48550/arXiv.2311.04478`
- Fasen-Hartmann and Schenk (2024). Partial correlation graphs for continuous-parameter time series. Preprint available at `https://doi.org/10.48550/arXiv.2401.16970`

The preprints Fasen-Hartmann and Schenk (2023b) and Fasen-Hartmann and Schenk (2023a) form the basis of Part I, while the preprint Fasen-Hartmann and Schenk (2024) forms the basis of Part II. To be more specific, Chapter 2 already presents some fundamental aspects of all three papers. Chapters 3, 4, 5, 6, 7, and 8 then largely consist of the preprint Fasen-Hartmann and Schenk (2023b). Chapter 9 is based on the preprint Fasen-Hartmann and Schenk (2023a). In addition, Chapters 10, 11, and 12 are based on the preprint Fasen-Hartmann and Schenk (2024).

Contents

CHAPTER 1

Introduction

Graphical models are a widely used tool for the visualisation of complex relationships in various disciplines and are particularly well suited for visualising dependency structures in stochastic processes. This thesis aims to introduce and analyse three powerful graphical models for stationary multivariate processes in *continuous time* and the dependency structures they require, as there is very little research on graphical models for continuous-time stochastic processes.

In this opening chapter, we provide an introduction to this thesis by first outlining the key aspects of the current state of research on graphical models, leading up to the research gap. We then explain our contributions by highlighting the research problems and objectives of this thesis, as well as the significance of our research and its limitations. Finally, we provide an outline of the thesis and a brief explanation of the notations used.

State of the art

The attempt to use graphical models to visualise and analyse dependency structures has been around for over a decade. It has its origins in several scientific fields, such as statistical physics (Gibbs, 1902) and genetics (Wright, 1921, 1934). Over the years, the theory and methodology of graphical models have been developed and extended in many directions, and graphical models have been applied in fields as diverse as biology, economics, engineering, finance, forensics, neuroscience, and psychology, to name just a few. There are several books on the state of the art of graphic models. However, the field is growing rapidly and is interdisciplinary by nature, with important contributions from a range of disciplines, so that no single author can cover its entire scope. We therefore refer to Whittaker (2008) and Edwards (2000) for the ideas behind graphical modelling and many examples, and to the comprehensive overviews of Lauritzen (2004) and Maathuis, Drton, Lauritzen, and Wainwright (2019) for the mathematical and statistical aspects of graphical models. For more application-oriented literature, we refer to Sinoquet and Mourad (2014) for graphical models from a biological perspective, and to the recent book by Sucar (2020) for an engineering perspective.

One of the main reasons for the interest in graphical modelling is the increasing number of complex multivariate datasets in a wide variety of fields. Extracting, analysing, and interpreting the information contained in such massive datasets can hardly be done by experience alone. It is becoming increasingly important to develop methods that present the key aspects in an easily understandable and clear way. Graphical models, in which the vertices symbolise the variables of interest and the edges illustrate hypothetical relationships between them, provide an appropriate, intuitive, and therefore widely used tool. They reduce complex data to the aspects of interest, extract key information, visualise the essential dependency structures, and provide a simple and clear illustration of the data structure. Such graphical models give an idea of how dependency structures between different variables look like, can be analysed and interpreted by experts, and can be easily communicated in a way that is not possible with the raw data. It should also be noted that graphs are a natural data structure for modern digital computers and can be easily implemented – another advantage of graphical models.

Looking specifically at graphical models suitable for stochastic processes, there has been a surge of interest in the last 25 years, with many publications devoted to their theory and application. Below, we review the key ideas relevant to this work, distinguishing between the discrete-time and continuous-time contexts. While most of the literature is cited in the main body of this thesis, where it can be discussed in detail, the most important literature is mentioned right away.

In the discrete-time setting, two graphical models are of particular interest for this thesis – the *path diagram* of Eichler (2007) and the *partial correlation graph* of Dahlhaus (2000). Both authors define and analyse graphical models in which the vertices represent the components $\mathcal{Z}_a = (Z_a(t))_{t\in\mathbb{Z}}$, $a \in V = \{1, \dots, k\}$, of a k-dimensional stationary process $\mathcal{Z}_V = (\mathcal{Z}_a)_{a\in V} = (Z_V(t))_{t\in\mathbb{Z}}$ in discrete time. However, the definition of and motivation behind the edges is quite different.

The *path diagram* of Eichler (2007) is a mixed graph, that is, a graph with directed edges represented by arrows ($\longrightarrow$) and undirected edges represented by dashed lines ($---$). Figure 1.1a displays an example of a mixed graph.

The directed edges in the path diagram represent (linear) Granger causal relationships between the components. Roughly speaking, the idea is that the component $\mathcal{Z}_a$ is not Granger causal for the component $\mathcal{Z}_b$, i.e., there is no directed edge from a to b, if, for all $t \in \mathbb{Z}$, the prediction of $Z_b(t+1)$ given all the (linear) information provided by $(Z_V(s))_{s\leq t}$ is equal to the prediction of $Z_b(t+1)$ given only the partial (linear) information provided by $(Z_{V\setminus\{a\}}(s))_{s\leq t}$. Note that a mathematical definition of causality was already given in the seminal works of Granger (1969) and Sims (1972). The definition has since been extended in numerous ways and enjoys popularity in many disciplines. We go into this topic in more detail in the introduction to Part I.

Eichler (2007) further defines the undirected edges in the path diagram by (linear) contemporaneous correlation. This concept addresses the problem that dependencies do not always have a temporally ordered cause-effect structure, which raises the need for modelling dependencies that occur contemporaneously. The idea is that two components $\mathcal{Z}_a$ and $\mathcal{Z}_b$ are contemporaneously uncorrelated, i.e., there is no undirected edge between a and b, if, for all $t \in \mathbb{Z}$, the random variables $Z_a(t+1)$ and $Z_b(t+1)$ are uncorrelated given the (linear) information provided by $(Z_V(s))_{s\leq t}$.

(a) A mixed graph (b) An undirected graph

Figure 1.1.: Two graphical models

Unlike the path diagram, the *partial correlation graph* of Dahlhaus (2000) is an undirected graph, as illustrated in Figure 1.1b. Here, the undirected edges are represented by lines (——) and stand for (linear) partial correlations. The idea is that $\mathcal{Z}_a$ and $\mathcal{Z}_b$ are partially uncorrelated, i.e., there is no undirected edge between a and b, if, for all $t \in \mathbb{Z}$, the random variables $Z_a(t)$ and $Z_b(t)$ are uncorrelated given the (linear) information provided by $\mathcal{Z}_{V\setminus\{a,b\}} = (Z_{V\setminus\{a,b\}}(t))_{t\in\mathbb{Z}}$.

Both the partial correlation graph and the causality graph have been analysed extensively and used in various practical applications, such as for air pollution data and for human tremor data (Dahlhaus, 2000; Dahlhaus & Eichler, 2003), to name just a few. These examples demonstrate in practice that the given graphical representations facilitate the understanding of relationships in multivariate time series.

Of course, the literature provides many other concepts for modelling dependency structures in discrete-time processes. These concepts are often based on conditional independence (e.g., Chamberlain, 1982; Eichler, 2012; Florens & Mouchart, 1982), which is the independence of processes given certain information, rather than on conditional uncorrelatedness. However, to our knowledge, Eichler (2012) is the only one to visualise his dependency structures in a mixed graph.

Although data is observed in discrete time, in many situations it is more appropriate to specify the underlying stochastic process in *continuous time*. This approach is especially necessary for high-frequency data, data with irregular intervals, or data with missing observations, which is often the case in finance, econometrics, and turbulence. In addition, many physical models are formulated in continuous time (e.g., Deck, 2006; Mahnke, Kaupužs, & Lubaševskij, 2009), so the use of continuous-time models is often more natural.

Given the above advantages of graphical models for visualising dependency structures in stochastic processes and the relevance of continuous-time stochastic processes, it is of interest to develop graphical representations for continuous-time multivariate stochastic processes as well. Suitable dependency concepts, especially causality concepts based on conditional independence, can be found, for instance, in the works of Comte and Renault (1996), Florens and Fougère (1996), and Petrovic and Dimitrijevic (2012). However, to our knowledge, these concepts have not yet been visualised in graphical models. An exception is the concept of local conditional independence, a dependency model for capturing asymmetric causal structures. For marked point processes, local conditional independence has been visualised in a directed graph, called the local independence graph, by Didelez (2006, 2008). This model was recently revisited and adapted to Itô processes by Mogensen and Hansen (2020, 2022).

Overall, however, there is very little research on graphical models for stationary continuous-time stochastic processes, and the many possibilities of graphical models have not yet been fully explored. On the one hand, the local independence graph proposed in the literature often makes use of the semi-martingale property of Itô processes (Mogensen & Hansen, 2020, 2022), which does not seem to be an appropriate tool for stationary stochastic processes. On the other hand, we were not able to find any mixed graph like the discrete-time path diagram in Eichler (2007), and thus, of course, none that visualises concepts of Granger causality and contemporaneous correlation simultaneously in one graph. Additionally, we were not able to find any undirected graphs in the literature that are as versatile and easy to use as the discrete-time partial correlation graph in Dahlhaus (2000). Furthermore, the concepts for modelling dependency structures in continuous-time processes proposed so far are incomplete in the sense that they are based on conditional independence rather than conditional uncorrelatedness. The concept of conditional uncorrelatedness should also be investigated since even in discrete time the graphical models using conditional independence require more technical and unwieldy assumptions (Eichler, 2012) than those using conditional uncorrelatedness (Eichler, 2007). This is also to be expected in continuous time, which may prevent broad applicability.

The research gap is thus the lack of availability of graphical models for stationary multivariate processes in continuous time, especially graphical models based on conditional correlation. Therefore, it is of utmost importance to establish linear dependency concepts suitable for stationary continuous-time processes and corresponding powerful graphical models that represent the essential dependency structure of the process in a clear and meaningful way.

Our contributions

In this thesis, we construct three graphical models for stationary multivariate stochastic processes in continuous time. In each graphical model, the vertices represent the different components $\mathcal{Y}_a = (Y_a(t))_{t\in\mathbb{R}}$, $a \in V = \{1,\ldots,k\}$, of the underlying process $\mathcal{Y}_V = (\mathcal{Y}_a)_{a\in V} = (Y_V(t))_{t\in\mathbb{R}}$. However, the edges are defined differently and visualise different dynamic relationships between the components. These relations, of course, must first be defined appropriately for stationary multivariate continuous-time processes.

In the *causality graph* $G_{CG} = (V, E_{CG})$ and the *local causality graph* $G^0_{CG} = (V, E^0_{CG})$, which we define and further analyse in Part I, the focus is on the graphical representation of appropriate (local) Granger causal relations. Compared to Eichler (2007), we face the challenge of defining what it means to take one step into the future in continuous time, and therefore propose two different approaches.

In the causality graph, the directed edges are defined using *(linear) Granger causality*. This means that there is no directed edge from a to b if, for all $t \in \mathbb{R}$ and $0 \leq h \leq 1$, the prediction of $Y_b(t+h)$ given all the (linear) information provided by $(Y_V(s))_{s\leq t}$ is equal to the prediction of $Y_b(t+h)$ given only the partial (linear) information provided by $(Y_{V\setminus\{a\}}(s))_{s\leq t}$. In comparison, in the local causality graph, the directed edges are defined by *local (linear) Granger causality*. Roughly speaking, the idea is to consider the equality of predictions in the limit as the size of the time step h tends to zero. In order to obtain non-trivial definitions, it is not the predictions of the components of the process $\mathcal{Y}_V$ themselves that must be considered, but the predictions of the difference quotients of their maximum derivatives.

The definition of the undirected edges follows the same idea. In the causality graph, two components $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are *(linearly) contemporaneously uncorrelated*, i.e., there is no undirected edge between a and b if, for all $t \in \mathbb{R}$ and $0 \leq h, h' \leq 1$, the random variables $Y_a(t+h)$ and $Y_b(t+h')$ are uncorrelated given the (linear) information provided by $(Y_V(s))_{s\leq t}$. In the local causality graph, we use local (linear) contemporaneous uncorrelatedness, where we again consider the limiting case and study the maximum derivatives of the components.

In summary, both the causality graph and the local causality graph act as a continuous-time counterpart to the path diagram of Eichler (2007).

In contrast to the (local) causality graph, the *partial correlation graph* $G_{PC} = (V, E_{PC})$, as introduced in Part II, is an undirected graph. Here, the undirected edges are defined by *partial correlations* between the components of the process. The idea is that $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are partially uncorrelated if, for all $t \in \mathbb{R}$, the random variables $Y_a(t)$ and $Y_b(t)$ are uncorrelated given the (linear) information provided by $\mathcal{Y}_{V\setminus\{a,b\}} = (Y_{V\setminus\{a,b\}}(t))_{t\in\mathbb{R}}$. The result is a graphical model that has the advantage of being simple and easy to use. This graphical model can then, of course, be seen as a continuous-time counterpart to the partial correlation graph of Dahlhaus (2000).

Let us conclude by highlighting some of the key aspects of our graphical models that emphasise the significance and usefulness of our approaches.

First of all, an advantage of all three graphical models is that they can be defined under fairly *general assumptions*. In the partial correlation graph, we only require that $\mathcal{Y}_V$ is wide-sense stationary, mean-square continuous, and has a positive definite spectral density function. In the more complex (local) causality graph, we restrict the spectral density a bit more and assume that the given process is purely non-deterministic. The general setting allows for the application to a wide range of continuous-time processes, including most Gaussian or Lévy-driven Ornstein-Uhlenbeck processes, multivariate continuous-time autoregressive (MCAR) processes, and even output processes of general linear state space models.

In fact, the graphical models could be defined under even weaker assumptions. However, these assumptions are sufficient for the validity of various *Markov properties*, so the models are meaningfully defined in this sense. Indeed, the graphs themselves are defined only by pairwise relationships between the components of the process. The various Markov properties then allow relationships between multivariate subprocesses to be inferred. They even provide graphical criteria for inferring relationships between multivariate subprocesses given only partial information, i.e., the dependency structures in subprocesses – a desirable property of graphical models.

Furthermore, for most output processes of linear state space models, the edges in the graphs can be characterised in terms of the model parameters of the process. This results in suitable criteria that provide a user-friendly way of visualising the complex dependency structures in graphical models and that can be easily implemented. Furthermore, the characterisations allow for insightful interpretations and comparisons to the existing (discrete-time) literature, further supporting the soundness of the definition of our graphs.

It should be noted, however, that the dependency concepts we propose are only able to capture linear relationships correctly and are based on the relationship of conditional uncorrelatedness. In particular, we do not study the widely used concept of conditional independence. Furthermore, our dependency concepts are tailored to stationary processes and their properties. Finally, the focus of this thesis is on the theoretical aspects of the graphical modelling approach. We neither discuss statistical methods for estimating the edges in the graphs from discrete-time data nor do we provide examples based on real data. Examples of specific processes are always on the theoretical level.

Outline of the thesis

This thesis is structured as follows. In Chapter 2, we lay the necessary foundations for the processes used in the thesis. We establish wide-sense stationary, mean-square continuous multivariate processes in continuous time that have a spectral density function, and state important properties of these processes. Subsequently, we define state space models in general, which are used as examples throughout the thesis, before moving on to the controller canonical state space model and the multivariate continuous-time autoregressive moving average (MCARMA) model. Finally, we discuss the invertibility of controller canonical state space models and establish invertible controller canonical state space (ICCSS) processes.

In Part I, we then introduce the first two graphical models, the *causality graph* and the *local causality graph*. To define these graphs, the necessary foundations on linear spaces, orthogonal projections, and conditional orthogonality (conditional uncorrelatedness) are covered in Chapter 3. Then, in Chapters 4 and 5, we introduce different concepts of Granger causality and contemporaneous correlation for stationary continuous-time processes to model different dependencies between the components of multivariate processes. Several equivalent characterisations are given for the different definitions, in particular by equality of orthogonal projections. Building on these different concepts, we define the causality graph and the local causality graph in Chapter 6 and discuss various notions of Markov properties in Chapter 7. In Chapter 8 and 9, we derive (local) causality graphs for continuous-time state space processes as an example. Chapter 8 is focused on multivariate continuous-time autoregressive (MCAR) processes, while Chapter 9 treats invertible controller canonical state space (ICCSS) processes. We show that most state space processes satisfy the assumptions of the (local) causality graph, resulting in well-defined graphical models that satisfy various notions of Markov properties. We compute the relevant orthogonal projections, which lead to meaningful characterisations of the edges in the (local) causality graph in terms of the model parameters. Throughout this part, we draw comparisons to the current literature.

In Part II, we introduce the third graphical model, the *partial correlation graph* for multivariate stationary stochastic processes in continuous time. To define this graph, we first establish the concept of partial correlation for continuous-time processes and give some characterisations and important properties in Chapter 10. Then, in Chapter 11, we define the partial correlation graph, show that it satisfies the usual Markov properties, and provide simple edge characterisations. We also give a comparison to the causality graph. As an example, we explicitly characterise and interpret the edges in the partial correlation graph for the popular MCAR processes in Chapter 12. Furthermore, we continue the comparison to the causality graph and provide an additional comparison to the local causality graph.

Finally, in Chapter 13, we complete this thesis with a conclusion and an outlook on the research questions arising from this thesis.

Notations

Throughout this thesis, we denote the space of all real or complex $(k \times k)$-dimensional matrices by $M_k(\mathbb{R})$ and $M_k(\mathbb{C})$. Similarly, $M_{k,d}(\mathbb{R})$ and $M_{k,d}(\mathbb{C})$ stand for the real or complex matrices of dimension $k \times d$. For a matrix $M \in M_{k,d}(\mathbb{C})$, the matrix entries are denoted by $[M]_{ab}$ for $a \in \{1, \dots, k\}$ and $b \in \{1, \dots, d\}$. Submatrices of M are denoted by $[M]_{AB}$ for non-empty $A \subseteq \{1, \dots, k\}$ and $B \subseteq \{1, \dots, d\}$. Further, we write $\#A$ for the cardinality of the set A. Some frequently used matrices are the identity matrix $I_k \in M_k(\mathbb{R})$ and the zero matrix $0_{k \times d} \in M_{k \times d}(\mathbb{R})$. The expression 0_k denotes either the $(k \times k)$-dimensional zero matrix or the k-dimensional zero vector, which should be clear from the context and is sometimes emphasised by $0_k \in M_k(\mathbb{R})$ or $0_k \in \mathbb{R}^k$. The vector $e_a \in \mathbb{R}^k$ represents the a-th unit vector and, similarly, we write

$$\mathbf{E}_j = \begin{pmatrix} 0_{k(j-1) \times k} \\ I_k \\ 0_{k(p-j) \times k} \end{pmatrix} \in M_{kp \times k}(\mathbb{R}), \quad j = 1, \dots, p. \tag{1.1}$$

For $p > q$, we also write

$$\overline{\overline{\mathbf{E}}} = \begin{pmatrix} I_{kq} \\ 0_{k(p-q) \times kq} \end{pmatrix} \in M_{kp \times kq}(\mathbb{R}) \quad \text{and} \quad \underline{\underline{\mathbf{E}}} = \begin{pmatrix} 0_{k \times k(q-1)} \\ I_k \end{pmatrix} \in M_{kq \times k}(\mathbb{R}). \tag{1.2}$$

The two-line marker indicates the position of non-zero entries in the matrix. A double overline represents entries at the top of the matrix, while a double underline represents entries at the bottom of the matrix. Moreover, $M^\top$ is the transpose of $M \in M_{k,d}(\mathbb{C})$ and $\|M\|$ denotes some matrix norm of M. $\sigma(M)$ is the set of eigenvalues of $M \in M_k(\mathbb{C})$ and $\det(M)$ is the determinant. For Hermitian matrices $M, N \in M_k(\mathbb{C})$, the notation $M \geq N$ represents the Loewner order and means that $M - N$ is positive semi-definite, for short, $M - N \geq 0$. Similarly, $M > 0$ indicates that M is positive definite. For a polynomial matrix $P(z) \in M_k(\mathbb{C})$, $z \in \mathbb{C}$, we denote by $\mathcal{N}(P) = \{z \in \mathbb{C} : \det(P(z)) = 0\}$ the roots of the univariate polynomial $\det(P(z))$, $z \in \mathbb{C}$, and $\deg(\det(P(z)))$ denotes its degree. Finally, for a function $f : \mathbb{R} \to M_k(\mathbb{C})$ with $f(\lambda) > 0$ for $\lambda \in \mathbb{R}$, we denote by $g : \mathbb{R} \to M_k(\mathbb{C})$ the function with $g(\lambda) = f(\lambda)^{-1}$, $\lambda \in \mathbb{R}$.

CHAPTER 2

FUNDAMENTALS

This chapter deals with the basics of stochastic processes, which are crucial to all parts of this thesis and which the reader should be familiar with before proceeding. For better readability, the fundamentals that are only relevant for one of the graphical models considered in this thesis are given in the preliminaries of the respective parts in Chapters 3 and 10.

In the chapter at hand, we first give a brief introduction to *wide-sense stationary* and *mean-square continuous* multivariate stochastic processes in continuous time. Building on this, we establish the spectral representation of such processes. We also provide well-known theory of covariance functions, spectral density functions, and the differentiability of wide-sense stationary and mean-square continuous multivariate processes. Most of these key results date back to Khintchine (1934) and Cramér (1940), and have been summarised in comprehensive overviews by Doob (1953), Rozanov (1967), and Yaglom (1987), among others.

As an important example of wide-sense stationary and mean-square continuous processes, we study time-invariant linear *state space models* $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$, which are characterised by matrices $\mathbf{A}^* \in M_{kp}(\mathbb{R})$, $\mathbf{B}^* \in M_{kp\times k}(\mathbb{R})$, and $\mathbf{C}^* \in M_{k\times kp}(\mathbb{R})$ as well as a driving $\mathbb{R}^k$-valued Lévy process $L = (L(t))_{t\in\mathbb{R}}$. The continuous-time linear state space model then consists of a state equation $dX(t) = \mathbf{A}^*X(t)dt + \mathbf{B}^*dL(t)$ and an observation equation $Y_V(t) = \mathbf{C}^*X(t)$. The $\mathbb{R}^{kp}$-valued process $\mathcal{X} = (X(t))_{t\in\mathbb{R}}$ is the input process and the $\mathbb{R}^k$-valued process $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$, $V = \{1, \dots, k\}$, is the output process of $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$.

State space models are important tools in many disciplines, we refer, for example, to Andresen, Benth, Koekebakker, and Zakamulin (2014), Benth, Klüppelberg, Müller, and Vos (2014), and Benth and Saltyte Benth (2009) for successful applications of state space models in economics and mathematical finance. The strength and the reason for the great interest in state space models lies, on one hand, in their applicability to irregularly spaced observations and high-frequency data. On the other hand, the driving Lévy process and the matrices $\mathbf{A}^*$, $\mathbf{B}^*$, $\mathbf{C}^*$ can be chosen quite freely. Even the occurrence of jumps in the sample paths is possible, making state space models a powerful and flexible tool for modelling phenomena from many different fields. The output processes $\mathcal{Y}_V$ of multivariate state space models are therefore used as examples throughout this thesis. In particular, we apply our graphical models to output processes of state space models to provide graphical visualisations of dependency structures between the univariate components.

A subclass of state space models are *multivariate continuous-time autoregressive moving average* (MCARMA) models, which are present if the matrices $\mathbf{A}^*$, $\mathbf{B}^*$, $\mathbf{C}^*$ have a special structure. MCARMA models are the continuous-time counterpart to the well-known discrete-time autoregressive moving average (ARMA) models (for a comprehensive introduction see Brockwell & Davis, 1991). The term MCARMA process describes the output process of the respective state space model. Special cases of MCARMA processes include multivariate continuous-time autoregressive (MCAR) processes, multivariate continuous-time Ornstein-Uhlenbeck processes, or Gaussian MCARMA processes, where the driving Lévy process is a Brownian motion. MCARMA processes were introduced by Marquardt and Stelzer (2007) to extend the concept of univariate continuous-time autoregressive moving average (CARMA) processes to the multivariate setting and to be able to model the joint behaviour of several possibly dependent univariate stochastic processes in continuous time. However, early papers dealing with the properties and the statistical analysis of univariate Gaussian CARMA processes already include those of Bartlett (1946), Doob (1944, 1953), and Phillips (1959). Lévy-driven CARMA processes were popularised by Peter Brockwell at the beginning of this century, see Brockwell (2014) for an overview. Very early Gaussian MCAR processes have previously been studied in the economics literature, e.g., in Harvey and Stock (1985a, 1985b, 1989), and were further explored in the well-known paper by Bergstrom (1997). Since their introduction, MCARMA processes have enjoyed great popularity and have stimulated a considerable amount of research in the last decade, see, e.g., Basse-O'Connor, Nielsen, Pedersen, and Rohde (2019), Brockwell and Schlemm (2013), Fasen-Hartmann and Mayer (2022), Schlemm and Stelzer (2012a, 2012b), and the references therein.

A second type of state space models of particular interest to this thesis are state space models in *controller canonical form*. Similar to MCARMA models, these are state space models with matrices $\mathbf{A}^*$, $\mathbf{B}^*$, $\mathbf{C}^*$ of a special structure. Controller canonical state space models generalise the definition of a univariate CARMA process in Brockwell (2014), and the matrix structure is well suited for computing edge characterisations in our (local) causality graph.

For the output process $\mathcal{Y}_V$ of some state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$, there are numerous equivalent state space models $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$ with the same output process $\mathcal{Y}_V$. Since our primary interest is in this output process, we can transition between these models, and we call them equivalent. In this introductory chapter, we focus on the existence of an equivalent controller canonical state space model $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$ for a given state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$, and we demonstrate that if such a representation exists, it is unique. While the equivalence of controller canonical state space models and MCARMA models is addressed in Brockwell and Schlemm (2013), and the equivalence of MCARMA models and general state space models is addressed in Schlemm and Stelzer (2012a), a question of existence remains open in both articles.

Finally, a special feature of the controller canonical state space model is that, under mild assumptions, we can recover the kp-dimensional input process $\mathcal{X}$ from the k-dimensional output process $\mathcal{Y}_V$, even though the matrix $\mathbf{C}^*$ in the observation equation $Y_V(t) = \mathbf{C}^* X(t)$ is usually rectangular and not invertible. This recovery property was previously known only for univariate CARMA processes (Brockwell & Lindner, 2015, Theorem 2.2) and is key to the calculation of orthogonal projections and edge characterisations for the (local) causality graph. If the input process can be recovered from the output process, the controller canonical state space model is called an invertible controller canonical state space model. The output process $\mathcal{Y}_V$ is referred to as an *invertible controller canonical state space process*, abbreviated as ICCSS process. Note that the invertible controller canonical form of MCARMA models was already discussed by Brockwell and Schlemm (2013).

The chapter is structured as follows. In Section 2.1, we establish some theory and properties of wide-sense stationary and mean-square continuous multivariate stochastic processes in continuous time. We briefly introduce Lévy processes in Section 2.2, enabling us to define state space models and discuss the properties established in Section 2.1. Following this introduction, in Section 2.3, we present the controller canonical form of a state space model. In the case of its existence, we prove its uniqueness. We also shortly introduce MCARMA models, along with the special cases of MCAR models and Ornstein-Uhlenbeck models. We further discuss the relationship between the controller canonical state space model and the MCARMA model. Concluding the chapter, in Section 2.4, we introduce the invertible controller canonical state space model and its corresponding output process, the ICCSS process. Finally, we establish that, for ICCSS processes, we can recover the input process from the output process.

2.1. Wide-sense stationary and mean-square continuous processes

We begin this section with an introduction to wide-sense stationary and mean-square continuous multivariate stochastic processes in continuous time. We refer to the books by Doob (1953), Rozanov (1967), and Yaglom (1987) for further insights and proofs of these well-known results. From now on, we denote a k-dimensional stochastic process in continuous time by $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ with index set $V = \{1, \ldots, k\}$. The one-dimensional subprocesses $\mathcal{Y}_a = (Y_a(t))_{t\in\mathbb{R}}$, $a \in V$, are called components of $\mathcal{Y}_V$. Furthermore, we denote multivariate subprocesses of $\mathcal{Y}_V$ by $\mathcal{Y}_A = (Y_A(t))_{t\in\mathbb{R}}$ for $A \subseteq V$, without always emphasising that $A \neq \emptyset$.

Definition 2.1. Suppose that $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ satisfies $\mathbb{E}[|Y_a(t)|^2] < \infty$ for all $a \in V$ and $t \in \mathbb{R}$. Then $\mathcal{Y}_V$ is (centred and) *wide-sense stationary* if, for all $a, b \in V$ and $s, t \in \mathbb{R}$, $\mathbb{E}[Y_a(t)] = 0$ and $\mathbb{E}[Y_a(t)\overline{Y_b(s)}]$ is a function of the difference $t - s$. We denote

$$c_{Y_aY_b}(t-s) := \mathbb{E}\left[Y_a(t)\overline{Y_b(s)}\right].$$

The process $\mathcal{Y}_V$ is *mean-square continuous* if, for all $a \in V$,

$$\lim_{t-s\to 0} \mathbb{E}\left[|Y_a(t) - Y_a(s)|^2\right] = 0.$$

We emphasise that when we discuss wide-sense stationary processes, we always include that the process has zero expectation. We further refer to $c_{Y_aY_a}(t)$, $t \in \mathbb{R}$, as the *covariance function* of the component $\mathcal{Y}_a$ and to $c_{Y_aY_b}(t)$, $t \in \mathbb{R}$, as the *cross-covariance function* of the components $\mathcal{Y}_a$ and $\mathcal{Y}_b$. Furthermore, for $s, t \in \mathbb{R}$, the *covariance function* of $\mathcal{Y}_V$ is denoted by

$$c_{Y_VY_V}(t-s) = (c_{Y_aY_b}(t-s))_{a,b\in V} = \left(\mathbb{E}\left[Y_a(t)\overline{Y_b(s)}\right]\right)_{a,b\in V} = \mathbb{E}\left[Y_V(t)\overline{Y_V(s)}^\top\right].$$

We establish properties of the covariance function of $\mathcal{Y}_V$, including an alternative characterisation of continuity in the mean square, which is easily verifiable for the example processes considered in this thesis (cf. Remark 2.13).

Lemma 2.2. *Suppose that $\mathcal{Y}_V$ is wide-sense stationary. Then, for all $t \in \mathbb{R}$ and $a, b \in V$, it applies that*

$$c_{Y_aY_b}(t) = \overline{c_{Y_bY_a}(-t)}, \quad |c_{Y_aY_b}(t)| \leq \sqrt{c_{Y_aY_a}(0)c_{Y_bY_b}(0)}, \quad \textit{and} \quad c_{Y_VY_V}(t) \geq 0.$$

Furthermore, $\mathcal{Y}_V$ is mean-square continuous if and only if, for all $a \in V$,

$$\lim_{t \to 0} c_{Y_aY_a}(t) = c_{Y_aY_a}(0).$$

A key property of wide-sense stationary and mean-square continuous processes is their spectral representation.

Proposition 2.3. *Suppose that $\mathcal{Y}_V$ is wide-sense stationary and mean-square continuous. Then, for $t \in \mathbb{R}$, we have*

$$Y_V(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \Phi_V(d\lambda), \tag{2.1}$$

with respect to a random orthogonal measure $\Phi_V = (\Phi_1, \ldots, \Phi_k)^\top$, where

$$\mathbb{E}\left[\Phi_V(d\lambda)\overline{\Phi_V(d\mu)}^\top\right] = \delta_{\lambda=\mu} F_{Y_VY_V}(d\lambda), \quad \mathbb{E}\left[\Phi_V(d\lambda)\right] = 0_k \in \mathbb{R}^k, \tag{2.2}$$

and $\delta_{\lambda=\mu}$ is the Kronecker Delta. The relation (2.1) is called spectral representation of $\mathcal{Y}_V$ and $F_{Y_VY_V}(\lambda) = (F_{Y_aY_b}(\lambda))_{a,b \in V}$, $\lambda \in \mathbb{R}$, is called spectral distribution function of $\mathcal{Y}_V$.

Stochastic integrals of deterministic Lebesgue-measurable functions with respect to a random orthogonal measure Φ_V are defined in the usual L^2-sense. For details on the definition and properties of such integrals, we refer to Rozanov (1967) and Doob (1953).

In this thesis, we discuss processes $\mathcal{Y}_V$, for which there exist *spectral density functions* given by

$$f_{Y_aY_a}(\lambda) = \frac{dF_{Y_aY_a}(\lambda)}{d\lambda},$$

for $\lambda \in \mathbb{R}$ and $a \in V$. In particular, this implies the existence of *cross-spectral density functions*

$$f_{Y_aY_b}(\lambda) = \frac{dF_{Y_aY_b}(\lambda)}{d\lambda},$$

for $\lambda \in \mathbb{R}$ and $a, b \in V$ with $a \neq b$. The *spectral density function* of the process $\mathcal{Y}_V$ is denoted by $f_{Y_VY_V}(\lambda) = (f_{Y_aY_b}(\lambda))_{a,b \in V}$ for $\lambda \in \mathbb{R}$. The following lemma summarises some of its properties.

Lemma 2.4. *Suppose that $\mathcal{Y}_V$ is wide-sense stationary and mean-square continuous with spectral density function $f_{Y_VY_V}(\lambda)$, $\lambda \in \mathbb{R}$. Then the following statements hold for all $a, b \in V$ and $\lambda, t \in \mathbb{R}$.*

(a) $\int_{-\infty}^{\infty} |f_{Y_aY_b}(\lambda)|\, d\lambda < \infty$,

(b) $c_{Y_aY_b}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_aY_b}(\lambda) d\lambda$,

(c) $f_{Y_VY_V}(\lambda) \geq 0$ *and* $f_{Y_VY_V}(\lambda) = \overline{f_{Y_VY_V}(\lambda)}^\top$.

Remark 2.5. The upper bound of the covariance function from Lemma 2.2 implies integrability only over a finite interval. To obtain a one-to-one relationship between $c_{Y_V Y_V}(t)$ and $f_{Y_V Y_V}(\lambda)$ by Fourier transformation, additional integrability assumptions on the covariance function are required. For example, suppose that $\int_{-\infty}^{\infty} |c_{Y_a Y_b}(t)|dt < \infty$ for all $a, b \in V$. Then $\mathcal{Y}_V$ has a spectral density function given by

$$f_{Y_V Y_V}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} c_{Y_V Y_V}(t)dt, \quad \lambda \in \mathbb{R}. \tag{2.3}$$

In this thesis, we only require the existence of a spectral density function and not the relation (2.3). Therefore, although our example processes always have integrable covariance functions, see Remark 2.14, long memory processes are also covered in this thesis.

In addition to the spectral density function, we introduce the spectral coherence function of $\mathcal{Y}_A$ and $\mathcal{Y}_B$, which is obtained by rescaling the cross-spectral density function of $\mathcal{Y}_A$ and $\mathcal{Y}_B$.

Definition 2.6. Suppose that $\mathcal{Y}_V$ is wide-sense stationary and mean-square continuous with spectral density function $f_{Y_V Y_V}(\lambda)$, $\lambda \in \mathbb{R}$. The *spectral coherence function* of $\mathcal{Y}_A$ and $\mathcal{Y}_B$ is defined as

$$R_{Y_A Y_B}(\lambda) := \Big(f_{Y_A Y_A}(\lambda)\Big)^{-1/2} f_{Y_A Y_B}(\lambda) \Big(f_{Y_B Y_B}(\lambda)\Big)^{-1/2}, \quad \lambda \in \mathbb{R},$$

where $(f_{Y_A Y_A}(\lambda))^{-1/2}$ denotes the inverse of the positive square root of $f_{Y_A Y_A}(\lambda)$. If $f_{Y_A Y_A}(\lambda)$ or $f_{Y_B Y_B}(\lambda)$ is singular for some $\lambda \in \mathbb{R}$, we set $R_{Y_A Y_B}(\lambda) := 0_{\alpha \times \beta}$ with $\alpha = \#A$ and $\beta = \#B$.

Remark 2.7. For $A = B$, we have $R_{Y_A Y_A}(\lambda) = I_\alpha$ if $f_{Y_A Y_A}(\lambda) > 0$ and $R_{Y_A Y_A}(\lambda) = 0_\alpha$ if $f_{Y_A Y_A}(\lambda)$ is singular. Furthermore, $f_{Y_{A\cup B} Y_{A\cup B}}(\lambda) > 0$ is sufficient for $f_{Y_A Y_A}(\lambda) > 0$ and $f_{Y_B Y_B}(\lambda) > 0$.

Finally, the following multivariate version of Doob (1953), XI, §9, Example 1, is essential to compute the mean-square derivative of a wide-sense stationary and mean-square continuous process $\mathcal{Y}_V$. The proof in that book is directly applicable.

Proposition 2.8. *Suppose that $\mathcal{Y}_V$ is wide-sense stationary and mean-square continuous with spectral density function $f_{Y_V Y_V}(\lambda)$, $\lambda \in \mathbb{R}$, and spectral representation* (2.1)*. Then the mean-square limit*

$$\underset{h\to 0}{l.i.m.} \frac{Y_V(t) - Y_V(t-h)}{h}$$

exists if and only if $\int_{-\infty}^{\infty} \lambda^2 \|f_{Y_V Y_V}(\lambda)\| d\lambda < \infty$. In this case, $\mathcal{Y}_V$ is called mean-square differentiable with derivative

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

Furthermore, the limit holds also $\mathbb{P}$-a.s., i.e.,

$$D^{(1)} Y_V(t) = \lim_{h\to 0} \frac{Y_V(t) - Y_V(t-h)}{h} \quad \mathbb{P}\text{-}a.s.$$

Remark 2.9. Suppose that $\mathcal{Y}_a$, $a \in V$, is mean-square differentiable. If we define the (mean-square) closed linear space $\mathcal{L}_{Y_a}(t)$ generated by $(Y_a(s))_{s\leq t}$ for $t \in \mathbb{R}$ by

$$\mathcal{L}_{Y_a}(t) := \overline{\left\{ \sum_{i=1}^{n} \gamma_{a,i} Y_a(t_i) : \gamma_{a,i} \in \mathbb{C},\, -\infty < t_1 \leq \ldots \leq t_n \leq t,\, n \in \mathbb{N} \right\}},$$

then, we have

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

Furthermore, by recursion, Proposition 2.8 also provides higher derivatives. One can show by induction that, if $\mathcal{Y}_a$ is j_a-times mean-square differentiable, then

$$\int_{-\infty}^{\infty} (i\lambda)^j \, e^{i\lambda t} \Phi_a(d\lambda) = D^{(j)} Y_a(t) \in \mathcal{L}_{Y_a}(t)$$

for $j = 1, \ldots, j_a$ and $t \in \mathbb{R}$, where $D^{(j)} Y_a(t)$ denotes the j-th mean-square derivative.

2.2. General state space models

To define state space models, we first need to define multivariate Lévy processes, which are the source of randomness in a state space model. Many common stochastic processes are Lévy processes. Examples are the Brownian motion, the Poisson process, and the compound Poisson process (Stelzer, 2011, p. 4). For more details on Lévy processes, we refer to Applebaum (2011), Protter (2005), and Sato (2007), and we start with the definition of a one-sided Lévy process.

Definition 2.10. A one-sided $\mathbb{R}^k$-valued *Lévy process* $(L(t))_{t\geq 0}$ is a stochastic process with stationary and independent increments, that is continuous in probability, and that satisfies $L(0) = 0_k \in \mathbb{R}^k$ $\mathbb{P}$-a.s.

Throughout this thesis, we work with two-sided k-dimensional Lévy processes $L = (L(t))_{t\in\mathbb{R}}$. These processes are obtained from two independent copies $(L_1(t))_{t\geq 0}$ and $(L_2(t))_{t\geq 0}$ of a one-sided Lévy process via the construction

$$L(t) = \begin{cases} L_1(t) & \text{if } t \geq 0, \\ -\lim_{s\nearrow -t} L_2(s) & \text{if } t < 0. \end{cases}$$

A common assumption for Lévy processes, often made in the analysis of state space models, is the following.

Assumption 1. *The two-sided 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* $\Sigma_L := \mathbb{E}[L(1)L(1)^\top]$.

The assumption that $L(1)$ has finite variance is non-trivial. However, it is essential for this thesis, as it provides finite second moments and thus, covariance functions of the input and the output processes of state space models. Although some of the statements in the present chapter can be made without Assumption 1 (cf. Fasen-Hartmann & Schenk, 2023a), for the sake of readability and consistency of the thesis, we require that the assumption always holds, and no longer point it out.

Based on the two-sided Lévy processes, we can now define linear time-invariant state space models with equal dimensions of the driving Lévy process L and the output process $\mathcal{Y}_V$. Note that it is possible to extend this definition to state space models where the driving Lévy process and the output process have different dimensions (Schlemm & Stelzer, 2012a, Definition 3.2).

Definition 2.11. An $\mathbb{R}^k$-valued continuous-time *state space model* $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ of dimension kp is characterised by an $\mathbb{R}^k$-valued Lévy process $L = (L(t))_{t\in\mathbb{R}}$, a state transition matrix $\mathbf{A}^* \in M_{kp}(\mathbb{R})$, an input matrix $\mathbf{B}^* \in M_{kp\times k}(\mathbb{R})$, and an observation matrix $\mathbf{C}^* \in M_{k\times kp}(\mathbb{R})$. It consists of a state equation

$$dX(t) = \mathbf{A}^* X(t)dt + \mathbf{B}^* dL(t) \tag{2.4}$$

and an observation equation

$$Y_V(t) = \mathbf{C}^* X(t). \tag{2.5}$$

The $\mathbb{R}^{kp}$-valued process $\mathcal{X} = (X(t))_{t\in\mathbb{R}}$ is the *input process*, and the process $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$, taking values in $\mathbb{R}^k$, is the *output process* of the state space model, also called *state space process*.

It is well known that every solution of the state equation (2.4) satisfies

$$X(t) = e^{\mathbf{A}^*(t-s)}X(s) + \int_s^t e^{\mathbf{A}^*(t-u)}\mathbf{B}^* dL(u), \quad s, t \in \mathbb{R},\ s < t. \tag{2.6}$$

A short introduction to such multivariate stochastic integrals with respect to Lévy processes is given in Marquardt and Stelzer (2007). A comprehensive overview is provided in the textbooks of Applebaum (2011) and Protter (2005).

In this thesis, we always work under the following standing assumption.

Assumption 2. *The state transition matrix* $\mathbf{A}^* \in M_{kp}(\mathbb{R})$ *satisfies* $\sigma(\mathbf{A}^*) \subseteq (-\infty, 0) + i\mathbb{R}$.

If Assumption 2 is satisfied, the state equation (2.4) has the unique causal, strictly stationary solution $\mathcal{X} = (X(t))_{t\in\mathbb{R}}$ given by (Sato & Yamazato, 1983, Theorem 5.1)

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

A solution $\mathcal{X}$ is called *causal* if, for all $t \in \mathbb{R}$, $X(t)$ is independent of the σ-algebra generated by $(L(s) - L(t))_{s>t}$. Of course, there also exists a causal, strictly stationary version of the output process $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$, which has the moving average representation (e.g., Schlemm & Stelzer, 2012a, equation (3.9))

$$Y_V(t) = \int_{-\infty}^{\infty} g(t-u)dL(u) \quad \text{with} \quad g(t) = \mathbf{C}^* e^{\mathbf{A}^* t}\mathbf{B}^* \mathbf{1}_{[0,\infty)}(t), \quad t \in \mathbb{R}. \tag{2.7}$$

Throughout this thesis, we assume that Assumption 2 is satisfied and work with these causal, strictly stationary versions of $\mathcal{X}$ and $\mathcal{Y}_V$, respectively. We now introduce well-known second-order properties of $\mathcal{X}$ and $\mathcal{Y}_V$.

Remark 2.12. Due to the finite second moments of $L(1)$, both $\mathcal{X}$ and $\mathcal{Y}_V$ have finite second moments (Brockwell & Schlemm, 2013, Lemma A.4). Since the input process $\mathcal{X}$ and the output process $\mathcal{Y}_V$ are strictly stationary, $\mathcal{X}$ and $\mathcal{Y}_V$ are wide-sense stationary.

The covariance function of the input process $\mathcal{X}$ is (Schlemm & Stelzer, 2012a, Proposition 3.2)

$$\begin{aligned} c_{XX}(t) &= e^{\mathbf{A}^* t} c_{XX}(0), && t \geq 0, \\ c_{XX}(t) &= c_{XX}(-t)^\top = c_{XX}(0) e^{-(\mathbf{A}^*)^\top t}, && t < 0, \end{aligned} \tag{2.8}$$

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

$$\mathbf{A}^* c_{XX}(0) + c_{XX}(0)(\mathbf{A}^*)^\top = -\mathbf{B}^* \Sigma_L (\mathbf{B}^*)^\top. \tag{2.9}$$

Due to the observation equation (2.5), the covariance function of the output process $\mathcal{Y}_V$ is

$$c_{Y_V Y_V}(t) = \mathbf{C}^* c_{XX}(t) (\mathbf{C}^*)^\top, \quad t \in \mathbb{R}. \tag{2.10}$$

Remark 2.13. Since the matrix exponential is continuous, we have $\lim_{t \to 0} c_{XX}(t) = c_{XX}(0)$. Lemma 2.2 then implies that the input process $\mathcal{X}$ is mean-square continuous. Due to relation (2.10), the output process $\mathcal{Y}_V$ is also mean-square continuous.

Remark 2.14. The covariance function $c_{Y_V Y_V}(t)$, $t \in \mathbb{R}$, is integrable as an exponentially decaying function. Thus, the one-to-one relation (2.3) between the spectral density function $f_{Y_V Y_V}(\lambda)$, $\lambda \in \mathbb{R}$, and the covariance function $c_{Y_V Y_V}(t)$, $t \in \mathbb{R}$, applies. The spectral density function of $\mathcal{Y}_V$ exists and is given by (Schlemm & Stelzer, 2012b, Proposition 3.4)

$$f_{Y_V Y_V}(\lambda) = \frac{1}{2\pi} \mathbf{C}^* \left(i\lambda I_{kp} - \mathbf{A}^* \right)^{-1} \mathbf{B}^* \Sigma_L \left(\mathbf{B}^* \right)^\top \left(-i\lambda I_{kp} - (\mathbf{A}^*)^\top \right)^{-1} \left(\mathbf{C}^* \right)^\top, \quad \lambda \in \mathbb{R}. \tag{2.11}$$

As a final important function for a state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$, we establish the rational matrix function

$$H(z) = \mathbf{C}^* \left(z I_{kp} - \mathbf{A}^* \right)^{-1} \mathbf{B}^*, \quad z \in \mathbb{C} \backslash \sigma(\mathbf{A}^*), \tag{2.12}$$

which is called the *transfer function*. The importance of the transfer function H goes beyond its role in the spectral density function. According to the spectral representation theorem (Lax, 2002, Theorem 17.5), we are able to recover the so-called kernel function $\mathbf{C}^* e^{\mathbf{A}^* t} \mathbf{B}^*$, $t \in \mathbb{R}$, of the output process $\mathcal{Y}_V$ via

$$\mathbf{C}^* \mathrm{e}^{\mathbf{A}^* t} \mathbf{B}^* = \frac{1}{2\pi i} \int_\Gamma \mathrm{e}^{zt} H(z) dz,$$

where Γ is a closed contour in the complex numbers that winds around each eigenvalue of $\mathbf{A}^*$

exactly once. The transfer function H even uniquely determines this kernel function (Schlemm & Stelzer, 2012b, Lemma 3.2). Since the representation (2.7) shows that the behaviour of the output process $\mathcal{Y}_V$ depends on the values of the matrices $\mathbf{A}^*$, $\mathbf{B}^*$, and $\mathbf{C}^*$ only through the kernel function, $\mathcal{Y}_V$ is also determined for a given Lévy process.

The triple $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ in the transfer function H is called an *algebraic realisation* of the latter. The dimension kp of the matrix $\mathbf{A}^*$ is the dimension of the algebraic realisation $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$. Every transfer function H has many algebraic realisations of various dimensions. A particularly convenient class are those of minimal dimension. The algebraic realisation $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ of dimension kp is said to be *minimal* if there is no other algebraic realisation of H with dimension less than kp. Minimality can be defined not only through the dimension of $\mathbf{A}^*$ but can also be characterised by the rank of two special matrices (Rugh, 1996, Theorem 10.13).

Theorem 2.15. *An algebraic realisation* $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ *is minimal if and only if it is controllable and observable, which means that the controllability matrix and the observability matrix*

$$\begin{aligned}\mathcal{C} &= \begin{pmatrix}\mathbf{B}^* & \mathbf{A}^*\mathbf{B}^* & (\mathbf{A}^*)^2\mathbf{B}^* & \cdots & (\mathbf{A}^*)^{kp-1}\mathbf{B}^*\end{pmatrix} \quad \textit{and}\\ \mathcal{O} &= \begin{pmatrix}\mathbf{C}^{*\top} & (\mathbf{C}^*\mathbf{A}^*)^\top & \cdots & \left(\mathbf{C}^*(\mathbf{A}^*)^{kp-1}\right)^\top\end{pmatrix}\end{aligned}$$

have full rank.

We say that a state space model is minimal, controllable, or observable if the corresponding algebraic realisation has this property. There is also the following connection between two minimal algebraic realisations (Hannan & Deistler, 2012, Lemma 2.3.4).

Lemma 2.16. *Two minimal algebraic realisations* $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ *and* $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ *are realisations of the same transfer function H if and only if there exists a non-singular matrix* $T \in M_{kp}(\mathbb{R})$*, such that*

$$\mathbf{A} = T\mathbf{A}^*T^{-1}, \quad \mathbf{B} = T\mathbf{B}^*, \quad \textit{and} \quad \mathbf{C} = \mathbf{C}^*T^{-1}.$$

Minimal realisations are therefore unique up to a transformation of the basis. Finally, the minimal nature of the algebraic realisation provides a useful property of the covariance function.

Remark 2.17. If $\Sigma_L > 0$ and the state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ is controllable, then, for $h > 0$, Schlemm and Stelzer (2012b) state in Corollary 3.9 that

$$\int_0^h e^{\mathbf{A}^*u}\mathbf{B}^*\Sigma_L(\mathbf{B}^*)^\top e^{(\mathbf{A}^*)^\top u}du > 0.$$

Therefore, Theorem 12.6.18 of Bernstein (2009) provides

$$c_{XX}(0) = \int_0^\infty e^{\mathbf{A}^*u}\mathbf{B}^*\Sigma_L(\mathbf{B}^*)^\top e^{(\mathbf{A}^*)^\top u}du > 0.$$

2.3. Controller canonical state space models

When transitioning between two algebraic realisations $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ and $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ of a transfer function H, we already know that the kernel function and the output process $\mathcal{Y}_V$ of the corresponding state space models remain unchanged. Given our primary interest in $\mathcal{Y}_V$, this raises the question of whether there are algebraic realisations of a state space model that are particularly well suited to a specific problem. One such algebraic realisation is the controller canonical algebraic realisation, which results in the *controller canonical state space model*. This model is not only simple but also suitable for predicting the output process $\mathcal{Y}_V$. The existence of the controller canonical state space model is the subject of this chapter.

To introduce the controller canonical state space model, it is necessary to revisit the transfer function. For the transfer function H of a state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ Kailath (1980) states in Lemma 6.3-8 that there exist $(k \times k)$-dimensional polynomial matrices P and Q, such that

$$H(z) = Q(z)P(z)^{-1}, \quad z \in \mathbb{C} \setminus \sigma(\mathbf{A}^*), \tag{2.13}$$

is a *coprime right polynomial fraction description*, which in turn means that the matrix $[P(z)^\top Q(z)^\top]^\top$ has full rank. In his Lemma 6.3-3, Kailath (1980) even gives a construction method for such a decomposition.

However, without any additional assumption, the coprime polynomials P and Q in (2.13) are not unique. For example, we can take any invertible matrix $S \in M_k(\mathbb{R})$ to see that $P(z)S$ and $Q(z)S$ also satisfy $H(z) = Q(z)SS^{-1}P(z)^{-1}$. Despite the many different coprime polynomials P and Q that satisfy (2.13), to the best of our knowledge, it remains unclear whether there *exists* a coprime right polynomial fraction description with

$$P(z) = I_k z^p + A_1 z^{p-1} + \ldots + A_p \quad \text{and} \quad Q(z) = C_0 + C_1 z + \ldots + C_q z^q, \tag{2.14}$$

$A_1, A_2, \ldots, A_p, C_0, C_1, \ldots, C_q \in M_k(\mathbb{R})$, and $p, q \in \mathbb{N}_0$, $p > q$. The focus is on ensuring that z^p is the highest power of P with a coefficient matrix of I_k. Even if we additionally assume that $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ is minimal, we just obtain the existence of coprime polynomials with $\deg(\det(P(z))) = kp$ (Rugh, 1996, Theorem 17.5). The construction method of Kailath (1980) often gives a polynomial P with higher powers than p, which are eliminated in the calculation of the determinant. Brockwell and Schlemm (2013), Theorem 3.2, and Schlemm and Stelzer (2012a), Corollary 3.4, implicitly assume such a coprime right (or left) polynomial fraction description with polynomials P and Q as in (2.14), 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 Example 2.24, we present some cases where this assumption is fulfilled.

For the purpose of this thesis, not only the existence but also the *uniqueness* of a coprime right polynomial fraction description of the transfer function H with polynomials P and Q as in (2.14) is important. In the next proposition, we derive that the uniqueness follows directly from the existence, and we introduce the controller canonical state space model.

Proposition 2.18. *Let $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ be a state space model with transfer function H. Suppose there exists a coprime right polynomial fraction description of H with polynomials P and Q as in* (2.14)*, such that*

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

Then P and Q are unique. Moreover, defining

$$\mathbf{A} = \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -A_p & -A_{p-1} & \cdots & \cdots & -A_1 \end{pmatrix} \in M_{kp}(\mathbb{R}), \quad \mathbf{B} = \begin{pmatrix} 0_k \\ \vdots \\ 0_k \\ I_k \end{pmatrix} \in M_{kp\times k}(\mathbb{R}), \tag{2.15}$$
$$\mathbf{C} = \begin{pmatrix} C_0 & C_1 & \cdots & C_q & 0_k & \cdots & 0_k \end{pmatrix} \in M_{k\times kp}(\mathbb{R}),$$

then $\sigma(\mathbf{A}^*) = \sigma(\mathbf{A})$ and

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

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

Proof. Suppose that there exist two coprime right polynomial fraction descriptions of the transfer function H with polynomial matrices as in (2.14), so

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

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

$$P_1(z) = P_2(z)U(z).$$

Both polynomials $P_1(z)$ and $P_2(z)$ have the highest power $I_k z^p$, so a comparison of the coefficients gives $U(z) = I_k$. Hence, $P_1(z) = P_2(z)$ and $Q_1(z) = Q_2(z)$, which results in the uniqueness of the decomposition. The fact that $H(z)$ is equal to $\mathbf{C}\left(zI_{kp} - \mathbf{A}\right)^{-1}\mathbf{B}$ follows from the proof of Theorem 3.2 in Brockwell and Schlemm (2013). Furthermore, the algebraic realisations $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ and $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ are minimal, because the polynomials P and Q are right coprime and $\deg(\det(P(z))) = kp$ obviously holds (Kailath, 1980, Theorem 6.5-1). Then a consequence of Lemma 2.16 is the existence of a non-singular matrix $T \in M_k(\mathbb{R})$, such that

$$\mathbf{A} = T\mathbf{A}^*T^{-1}, \quad \mathbf{B} = T\mathbf{B}^*, \quad \text{and} \quad \mathbf{C} = \mathbf{C}^*T^{-1}.$$

Together with Proposition 11.2.8.(v) of Bernstein (2009), for $s,t \in \mathbb{R}$, $s<t$, we obtain

$$Y_V(t) = \mathbf{C}^* e^{\mathbf{A}^*(t-s)} X(s) + \int_s^t \mathbf{C}^* e^{\mathbf{A}^*(t-u)} \mathbf{B}^* dL(u) = \mathbf{C} e^{\mathbf{A}(t-s)} (TX(s)) + \int_s^t \mathbf{C} e^{\mathbf{A}(t-u)} \mathbf{B} dL(u).$$

Thus, $\mathcal{Y}_V$ is a solution of the state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ if and only if it is a solution of $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$. Finally, $\sigma(\mathbf{A}) = \sigma(T\mathbf{A}^*T^{-1}) = \sigma(\mathbf{A}^*)$ (Schlemm & Stelzer, 2012b, p. 2219). ■

Remark 2.19. The proof was done without Assumption 2 and the integral representation (2.7) for generality. Requiring Assumption 2 simplifies the proof since $\mathcal{Y}_V$ depends on the matrices only through the kernel function due to equation (2.7). Since $\mathbf{C}^* \left(zI_{kp} - \mathbf{A}^*\right)^{-1} \mathbf{B}^* = \mathbf{C} \left(zI_{kp} - \mathbf{A}\right)^{-1} \mathbf{B}$ for $z \in \mathbb{C}$ implies $\mathbf{C}^* e^{\mathbf{A}^* t} \mathbf{B}^* = \mathbf{C} e^{\mathbf{A}t} \mathbf{B}$ for $t \in \mathbb{R}$ (Schlemm & Stelzer, 2012b, Lemma 3.2), $\mathcal{Y}_V$ is a solution of $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ if and only if it is a solution of $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$.

In particular, Proposition 2.18 implies that there exists no other minimal state space representation with matrices of the same structure as in (2.15), this representation is unique. Since the output process $\mathcal{Y}_V$ of the state space models $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ and $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$ is equal, we can henceforth assume under the existence assumption that, without loss of generality, a state space model is given as the unique controller canonical state space model $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$ with matrices as in (2.15).

Another example of a state space model, characterised by matrices $\mathbf{A}^*$, $\mathbf{B}^*$, and $\mathbf{C}^*$ of a special structure, is the *MCARMA model*. The output process $\mathcal{Y}_V$ is called *MCARMA process*. The definition of an MCARMA process is motivated by the idea that $\mathcal{Y}_V$ should solve the differential equation (Marquardt & Stelzer, 2007, Remark 3.19)

$$P^*(D) Y_V(t) = Q^*(D) DL(t), \tag{2.16}$$

for $t \in \mathbb{R}$, where D is the differential operator with respect to t,

$$P^*(z) = I_k z^{p^*} + P_1 z^{p^*-1} + \ldots + P_{p^*}, \quad \text{and} \quad Q^*(z) = Q_0 z^{q^*} + Q_1 z^{q^*-1} + \ldots + Q_{q^*} \tag{2.17}$$

with $p^*, q^* \in \mathbb{N}_0$, $p^* > q^*$, $P_1, P_2, \ldots, P_{p^*}$, $Q_0, Q_1, \ldots, Q_{q^*} \in M_k(\mathbb{R})$. P^* is the autoregressive (AR) polynomial and Q^* the moving average (MA) polynomial. This differential equation is the continuous-time analogue of the discrete-time ARMA equation. However, a Lévy process is not differentiable, so $\mathcal{Y}_V$ cannot be formally defined by (2.16). The proper definition as a state space model is as follows (Marquardt & Stelzer, 2007, Theorem 3.12).

Definition 2.20. Define

$$\mathbf{A}^* = \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -P_{p^*} & -P_{p^*-1} & \cdots & \cdots & -P_1 \end{pmatrix} \in M_{kp^*}(\mathbb{R}), \quad \mathbf{B}^* = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p^*} \end{pmatrix} \in M_{kp^* \times k}(\mathbb{R}),$$

$$\mathbf{C}^* = \begin{pmatrix} I_k & 0_k & \cdots & 0_k \end{pmatrix} \in M_{k \times kp^*}(\mathbb{R}),$$

where $\beta_1 = \ldots = \beta_{p^*-q^*-1} = 0_k \in M_k(\mathbb{R})$, $\beta_{p^*-j} = -\sum_{i=1}^{p^*-j-1} P_i \beta_{p^*-j-i} + Q_{q^*-j}$ for $j = q^*, \ldots, 0$, and $p^*, q^* \in \mathbb{N}_0$, $p^* > q^*$. Then the state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ is called a *multivariate continuous-time autoregressive moving average model* of order (p^*, q^*), or MCARMA(p^*, q^*) model for short. The output process $\mathcal{Y}_V$ is called a *multivariate continuous-time autoregressive moving average process* of order (p^*, q^*). In short, an MCARMA(p^*, q^*) process.

Remark 2.21. An important special case of the MCARMA(p^*, q^*) process (model) is the MCARMA$(p^*, 0)$ =MCAR(p^*) process (model), which we also call the multivariate continuous-time autoregressive process (model). The input matrix $\mathbf{B}^*$ is simplified to

$$(\mathbf{B}^*)^\top = \begin{pmatrix} 0_k & \cdots & 0_k & I_k \end{pmatrix} \in M_{k \times kp^*}(\mathbb{R})$$

and the trivial MA polynomial is $Q^*(z) = I_k$ for $z \in \mathbb{C}$. Another important special case is the MCARMA$(1, 0)$ process (model), also known as the Ornstein-Uhlenbeck process (model). Here, the matrices and polynomials are simplified to $\mathbf{A}^* = -P_1$, $\mathbf{B}^* = I_k$, and $\mathbf{C}^* = I_k$, as well as $P^*(z) = I_k z + P_1$ and $Q^*(z) = I_k$ for $z \in \mathbb{C}$.

Due to the specific structure of the matrices $\mathbf{A}^*$, $\mathbf{B}^*$, and $\mathbf{C}^*$, the MCARMA representation of a state space model is in the discrete-time literature often referred to as the observer canonical state space model (Kailath, 1980). In the following, we relate the MCARMA model and the controller canonical state space model.

Remark 2.22. For the MCAR(p^*) model, a comparison of the algebraic realisations shows that $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ is already in controller canonical form $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ with $p = p^*$. Of course, this fact also applies to the Ornstein-Uhlenbeck model. For MCARMA(p^*, q^*) models, Schlemm and Stelzer (2012a), Corollary 3.4, state the equivalence between the classes of state space models and MCARMA models, and Brockwell and Schlemm (2013), Theorem 3.2, state the equivalence between the classes of MCARMA models and controller canonical state space models. However, as mentioned above, both implicitly assume the existence of a coprime left or right polynomial fraction description (2.13) with polynomials as in (2.14), which does not seem obvious to us. For univariate state space models ($k = 1$), the existence of a coprime right polynomial fractional description is apparent (see the proof of Proposition 2.30), so that the classes of univariate state space models and univariate CARMA models are equivalent. Particularly, any output process of a univariate state space model has a representation in controller canonical form.

A peculiarity of MCARMA models is that the AR polynomial P^* and the MA polynomial Q^* provide a left polynomial fraction description of the transfer function H, i.e., $H(z) = P^*(z)^{-1}Q^*(z)$ (Marquardt & Stelzer, 2007; Brockwell & Schlemm, 2013, Lemma 3.1). Therefore, the spectral density function of an MCARMA process is (Marquardt & Stelzer, 2007, (3.43))

$$f_{Y_V Y_V}(\lambda) = \frac{1}{2\pi} P^*(i\lambda)^{-1} Q^*(i\lambda) \Sigma_L Q^*(-i\lambda)^\top \left(P^*(-i\lambda)^{-1}\right)^\top, \quad \lambda \in \mathbb{R}. \tag{2.18}$$

Furthermore, if the MCARMA model is minimal, the left polynomial fraction description $H(z) = P^*(z)^{-1}Q^*(z)$ is coprime (Kailath, 1980, Theorem 6.5-1). The connection to the coprime right polynomial fraction description (2.13) with polynomials P and Q as in (2.14) is given in the next lemma.

Lemma 2.23. *Let* $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ *be an MCARMA*(p^*, q^*) *model with polynomials* P^* *and* Q^* *as in* (2.17). *Suppose* $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ *is minimal and there exists a coprime right polynomial fraction description* (2.13) *of the transfer function* H *with polynomials* P *and* Q *as in* (2.14). *Then* P *and* Q *are unique,* $p^* = p$, $q^* = q$, $Q_0 = C_q$, $\mathcal{N}(P^*) = \mathcal{N}(P)$, *and* $\mathcal{N}(Q^*) = \mathcal{N}(Q)$.

Proof. The uniqueness follows from Proposition 2.18. Furthermore, $p^* = p$ holds by Lemma 6.5-6 of Kailath (1980), which provides that $kp^* = \deg(\det(P^*(z))) = \deg(\det(P(z))) = kp$. Since $P^*(z)^{-1}Q^*(z) = Q(z)P(z)^{-1}$, we have $p^* - q^* = p - q$ and therefore $q^* = q$. Comparing the highest order coefficient on both sides of $Q^*(z)P(z) = P^*(z)Q(z)$ gives $Q_0 = C_q$. Finally, Rugh (1996) states in Theorem 16.19 that $\mathcal{N}(P^*) = \mathcal{N}(P)$ and $\mathcal{N}(Q^*) = \mathcal{N}(Q)$. ■

To conclude this chapter, we provide examples of coprime left and right polynomial fraction descriptions.

Example 2.24.

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

$$P^*(z) = \begin{pmatrix} (z+2)^2 & 0 \\ 0 & (z+2)^2 \end{pmatrix}, \quad Q^*(z) = \begin{pmatrix} z+1 & 0 \\ 0 & z+1 \end{pmatrix},$$

$z \in \mathbb{C}$. Since P^* and Q^* are diagonal polynomial matrices and are right coprime, the unique coprime right polynomial fraction description of the transfer function H is given by $P(z) = P^*(z)$ and $Q(z) = Q^*(z)$.

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

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

$z \in \mathbb{C}$. Then the unique coprime right polynomial fractional description of the transfer function H is given by

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

(c) For MCAR(p) models, $Q^*(z) = I_k$ holds for $z \in \mathbb{C}$. Therefore, $P(z) = P^*(z)$ and $Q(z) = I_k$ always provide the unique coprime right polynomial fractional description of the transfer function H.

In particular, these examples show the existence of coprime right polynomial fraction descriptions (2.13) with polynomial matrices P and Q as in (2.14). Since the analysis of polynomial fraction descriptions does not belong to the central topics of interest of this thesis, we do not investigate this problem further and move on to the topic of the invertibility of a state model.

2.4. Invertible controller canonical state space models

Suppose that $\mathcal{Y}_V$ is the output process of a state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ with controller canonical representation $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$. Due to the observation equation $Y_V(t) = \mathbf{C}X(t)$, we obtain the output process $\mathcal{Y}_V$ directly from the input process $\mathcal{X}$. However, recovering the input process from the output process is not as obvious because, in general, $\mathbf{C}$ is rectangular and not invertible. One special case in which the input process is relatively easy to recover is the MCAR(p) process. The state equation (2.4) and the specific structure of the matrices $\mathbf{A}$ and $\mathbf{B}$ yield

$$D^{(1)}X^{(j)}(t) = X^{(j+1)}(t)$$

for $j = 1, \ldots, p-1$ and $t \in \mathbb{R}$, where

$$X^{(j)}(t) = \begin{pmatrix} X_{(j-1)k+1}(t) & \cdots & X_{jk}(t) \end{pmatrix}^\top \tag{2.19}$$

for $j = 1, \ldots, p$ is the j-th k-block of $X(t)$, i.e., $X(t) = (X^{(1)}(t)^\top \ \cdots \ X^{(p)}(t)^\top)^\top$ for $t \in \mathbb{R}$. Together with the observation equation (2.5) and the simple structure of the matrix $\mathbf{C}$, we obtain

$$D^{(j)}Y_V(t) = D^{(j)}X^{(1)}(t) = X^{(j+1)}(t) \tag{2.20}$$

inductively for $j = 1, \ldots, p-1$ and $t \in \mathbb{R}$. In this case, it is possible to recover $X(t)$ from $Y_V(t)$ and the derivatives via

$$X(t) = \begin{pmatrix} Y_V(t)^\top & D^{(1)}Y_V(t)^\top & \cdots & D^{(p-1)}Y_V(t)^\top \end{pmatrix}^\top,$$

for $t \in \mathbb{R}$, which implies that the input process $\mathcal{X}$ can be recovered from the output process $\mathcal{Y}_V$.

Remark 2.25.

(a) The relation (2.20) implies that the MCAR(p) process $\mathcal{Y}_V$ and its components $\mathcal{Y}_a$, $a \in V$, are $(p-1)$-times mean-square differentiable. Since the Lévy process and therefore $\mathcal{X}^{(p)} = (X^{(p)}(t))_{t\in\mathbb{R}}$ is not differentiable and the controller canonical state space representation is unique, the p-th derivative of $\mathcal{Y}_V$ and its components $\mathcal{Y}_a$, $a \in V$, does not exist. They are exactly $(p-1)$-times mean-square differentiable. Proposition 2.8 further provides that $\mathcal{Y}_V$ and its components are not only $(p-1)$-times mean-square differentiable but the limit holds $\mathbb{P}$-a.s. as well.

(b) For $a \in V$ and $t \in \mathbb{R}$, it follows that $D^{(j)}Y_a(t) \in \mathcal{L}_{Y_a}(t)$ for $j = 1, \ldots, p-1$ due to Remark 2.9.

Although the MCAR model is a controller canonical state space model, we cannot use the same approach to recover the input process from the output process for controller canonical state space models with $q > 0$. In fact, the state equation (2.4) still yields

$$D^{(1)}X^{(j)}(t) = X^{(j+1)}(t),$$

for $j = 1, \ldots, p-1$ and $t \in \mathbb{R}$. However, due to the more complex structure of $\mathbf{C}$, we have

$$Y_V(t) = \mathbf{C}X(t) = \sum_{i=0}^{q} C_i X^{(i+1)}(t).$$

In addition, we obtain that, for $j = 1, \dots, p-q-1$ and $t \in \mathbb{R}$,

$$D^{(j)} Y_V(t) = \sum_{i=0}^{q} C_i X^{(i+j+1)}(t). \tag{2.21}$$

Consequently, the p k-blocks can generally not be recovered from these $(p-q)$ equations.

Remark 2.26.

(a) In particular, the relation (2.21) implies that $\mathcal{Y}_V$ and its components $\mathcal{Y}_a$, $a \in V$, are $(p-q-1)$-times mean-square differentiable with the stated representations of the derivatives, that also hold $\mathbb{P}$-a.s. due to Proposition 2.8. In Lemma 9.5 and Remark 9.6, we discuss the maximal differentiability of the components.

(b) For $a \in V$ and $t \in \mathbb{R}$, we have $D^{(j)} Y_a(t) \in \mathcal{L}_{Y_a}(t)$ for $j = 1, \dots, p-q-1$ due to Remark 2.9.

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

$$\begin{aligned} \overline{\overline{\mathbf{C}}} &:= \mathbf{C} = \begin{pmatrix} C_0 & \cdots & C_q & 0_k & \cdots & 0_k \end{pmatrix} \quad \text{and} \\ \underline{\underline{\mathbf{C}}} &:= \begin{pmatrix} 0_k & \cdots & 0_k & C_0 & \cdots & C_q \end{pmatrix} \in M_{k \times kp}(\mathbb{R}). \end{aligned} \tag{2.22}$$

From the observation equation (2.5) and equation (2.21), we then get, for $t \in \mathbb{R}$, the shorthand expressions

$$Y_V(t) = \overline{\overline{\mathbf{C}}} X(t) \quad \text{and} \quad D^{(p-q-1)} Y_V(t) = \underline{\underline{\mathbf{C}}} X(t).$$

For controller canonical state space models with $q > 0$, we overcome the challenge of recovering the input process from the output process under mild assumptions, which leads to the subclass of *invertible controller canonical state space models*.

Definition 2.27. Let $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ be a state space model with controller canonical representation $(\mathbf{A}, \mathbf{B}, \overline{\overline{\mathbf{C}}}, L)$ as in (2.15) and polynomials P and Q as in (2.14) with $p > q > 0$. Suppose

$$\operatorname{rank}(C_q) = k, \quad \mathcal{N}(Q) \subseteq (-\infty, 0) + i\mathbb{R}, \quad \text{and} \quad \mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}. \tag{2.23}$$

Then $(\mathbf{A}, \mathbf{B}, \overline{\overline{\mathbf{C}}}, L)$ is called a (causal) *invertible controller canonical state space model* of order (p, q), an ICCSS(p, q) model for short. The output process $\mathcal{Y}_V = (Y_V(t))_{t \in \mathbb{R}}$ of the ICCSS(p, q) model is called *invertible controller canonical state space process* of order (p, q). In short, an ICCSS(p, q) process.

The assumption on P is the usual causal stationarity Assumption 2 because of $\mathcal{N}(P) = \sigma(\mathbf{A})$ (Marquardt & Stelzer, 2007, Corollary 3.8). The mild assumptions on Q are necessary to recover the input process $\mathcal{X}$ from the output process $\mathcal{Y}_V$ and to motivate the name ICCSS model, as we see in the remainder of this section. Of course, $q > 0$ excludes the class of MCAR(p) models. The assumptions in (2.23) are, for example, satisfied in Example 2.24(a,b).

Under the assumptions in (2.23), Brockwell and Schlemm (2013), Lemma 4.1, derive a stochastic differential equation for the first kq components of $\mathcal{X}$. This result follows simply from combining the first q k-blocks of the state equation (2.4) with the observation equation (2.5), having the special structure of $\mathbf{A}$, $\mathbf{B}$, and $\overline{\overline{\mathbf{C}}}$ in mind.

Lemma 2.28. *Let $\mathcal{Y}_V$ be a ICCSS(p,q) process with $p > q > 0$. Define the (kq)-dimensional upper truncated state vector $\mathcal{X}^q = (X^q(t))_{t\in\mathbb{R}}$ of $\mathcal{X}$ via*

$$X^q(t) = \begin{pmatrix} X^{(1)}(t) \\ \vdots \\ X^{(q)}(t) \end{pmatrix}, \quad t \in \mathbb{R}, \tag{2.24}$$

where $X^{(1)}(t), \dots, X^{(q)}(t)$ are the k-dimensional random vectors defined in (2.19)*. Then $\mathcal{X}^q$ satisfies*

$$dX^q(t) = \mathbf{\Lambda} X^q(t)dt + \mathbf{\Theta} Y_V(t)dt, \tag{2.25}$$

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

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

Remark 2.29.

(a) The assumptions in (2.23) correspond to the miniphase assumption in classical time series analysis (Hannan & Deistler, 2012, (1.3.15)) and imply Assumption A2 in Brockwell and Schlemm (2013), who even allow for rectangular matrices $C_0, \dots, C_q$. To see this connection, observe that the assumptions in (2.23) yield

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

which is one of their assumptions. Furthermore, it applies that $\sigma(\mathbf{\Lambda}) = \mathcal{N}(C_q^{-1}Q)$ (Marquardt & Stelzer, 2007, Lemma 3.8). Thus, $\mathbf{\Lambda}$ has full rank and, due to the structure of $\mathbf{\Lambda}$, 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 P^* and the MA polynomial Q^* of an MCARMA model are left coprime, the assumptions in (2.23) can analogously be made for P^* and Q^*, respectively. Indeed, $\mathcal{N}(Q^*) = \mathcal{N}(Q)$ and $\mathcal{N}(P^*) = \mathcal{N}(P)$ by Lemma 2.23. Further, comparing the highest power coefficients on both sides of $Q^*(z)P(z) = P^*(z)Q(z)$ gives $Q_0 I_k = I_k C_q$. Then Q_0 has full rank if and only if C_q has full rank.

The differential equation (2.25) is readily integrated to (Brockwell & Schlemm, 2013, (4.3))

$$X^q(t) = e^{\mathbf{\Lambda}(t-s)} X^q(s) + \int_s^t e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du, \quad s < t, \ s, t \in \mathbb{R}. \tag{2.26}$$

Hence, we can compute $X^q(t)$ based on the knowledge of the initial value $X^q(s)$ and $(Y_V(u))_{s \leq u \leq t}$. In Proposition 2.30 below, we show that $X^q(t)$ is even uniquely determined by the entire past $(Y_V(s))_{s \leq t}$, implying that the truncated state vector $\mathcal{X}^q$ can be recovered from $\mathcal{Y}_V$. This result is the multivariate generalisation of Theorem 2.2 by Brockwell and Lindner (2015).

Proposition 2.30. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$. Then, for all $t \in \mathbb{R}$,*

$$X^q(t) = \int_{-\infty}^t e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du \quad \mathbb{P}\text{-}a.s.$$

Proof. The proof is divided into four steps. In the first three steps, we derive auxiliary results that lead to the proof of the statement in the fourth step.

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

$$\lim_{|u| \to \infty} e^{-\varepsilon |u|} |Y_v(u)| = 0 \quad \mathbb{P}\text{-a.s.} \tag{2.27}$$

Therefore, we relate (2.27) back to Proposition 2.6 of Brockwell and Lindner (2015), 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. For this purpose, let $\varepsilon > 0$ and $v \in V$. Note that, for $t \in \mathbb{R}$,

$$Y_v(t) = \int_{-\infty}^t e_v^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}(t-u)} \mathbf{B} dL(u) = \sum_{\ell=1}^k \int_{-\infty}^t e_v^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}(t-u)} \mathbf{B} e_\ell dL_\ell(u) = \sum_{\ell=1}^k Y_v^\ell(t).$$

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

$$dX(t) = \mathbf{A} X(t) dt + \mathbf{B} e_\ell dL_\ell(t) \quad \text{and} \quad Y_v^\ell(t) = e_v^\top \overline{\overline{\mathbf{C}}} X(t),$$

and has the transfer function H_v^ℓ given by

$$H_v^\ell(z) = e_v^\top \overline{\overline{\mathbf{C}}} (z I_{kp} - \mathbf{A})^{-1} \mathbf{B} e_\ell, \quad z \in \mathbb{C} \setminus \sigma(\mathbf{A}).$$

Then Kailath (1980) provides in Lemma 6.3-8 the existence of (right) coprime polynomials $P_v^\ell(z)$ and $Q_v^\ell(z)$ as in (2.14), 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 polynomial fraction description of the form (2.14) does not arise. Indeed, $1 \cdot p = \deg(\det(P_v^\ell(z))) = \deg(P_v^\ell(z))$ follows immediately, and the coefficient of 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 with a 1 as the leading coefficient. The classes of univariate CARMA processes and univariate causal continuous-time state space models are equivalent (Schlemm & Stelzer, 2012a, Corollary 3.4) implying that $\mathcal{Y}_v^\ell$ 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, Theorem 12.9.16 of Bernstein (2009) gives that the poles of $H_v^\ell(z)$ are a subset of $\sigma(\mathbf{A})$, resulting in

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

which means that the AR polynomial $P_v^\ell(z)$ has no zeros on the imaginary axis. Thus, $\mathcal{Y}_v^\ell$ satisfies the assumptions by Brockwell and Lindner (2015), Proposition 2.6, and we obtain

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

for $\ell = 1, \ldots, k$. Finally, the claim (2.27) follows by

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

Step 2: Next, we show that the limit

$$\lim_{s\to-\infty} \int_s^t e^{-\lambda(t-u)} \left|Y_v(u)\right| du \tag{2.28}$$

exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$ and $\lambda > 0$. First, from relation (2.27), we obtain that there exists a 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 some $u_0(\omega) < 0$ with

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

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

$$\begin{aligned}
\int_s^t e^{-\lambda(t-u)} \left|Y_v(\omega, u)\right| du &= \int_{u_0(\omega)}^t e^{-\lambda(t-u)} \left|Y_v(\omega, u)\right| du + \int_s^{u_0(\omega)} e^{-\lambda(t-u)} \left|Y_v(\omega, u)\right| du \\
&\le \int_{u_0(\omega)}^t e^{-\lambda(t-u)} \left|Y_v(\omega, u)\right| du + \int_s^{u_0(\omega)} e^{-\lambda(t-u)} e^{-\frac{\lambda}{2}u} \gamma du \\
&= \int_{u_0(\omega)}^t e^{-\lambda(t-u)} \left|Y_v(\omega, u)\right| du + \gamma e^{-\lambda t} \frac{2}{\lambda} \left(e^{\frac{\lambda}{2}u_0(\omega)} - e^{\frac{\lambda}{2}s}\right) \\
&\le \int_{u_0(\omega)}^t e^{-\lambda(t-u)} \left|Y_v(\omega, u)\right| du + \gamma e^{-\lambda t} \frac{2}{\lambda}.
\end{aligned}$$

Thus, by dominated convergence, the limit in (2.28) exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$ and $\lambda > 0$.
Step 3: Eventually, we derive that not only the univariate integral (2.28) exists but also

$$\lim_{s\to-\infty} \int_s^t e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du \tag{2.29}$$

exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$. First, the assumptions in (2.23) provide 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. Hence, 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

$$\left\| e^{\mathbf{\Lambda} t} \right\| \leq c_1 e^{-\lambda t} \quad \forall\, t \geq 0. \tag{2.30}$$

Now, we obtain

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

Due to (2.28), the limit of each of those summands exists, so (2.29) exists $\mathbb{P}$-a.s. for $t \in \mathbb{R}$.
Step 4: Finally, we can prove the statement of the proposition. Recall that, due to equation (2.26), for $s, t \in \mathbb{R}$ with $s < t$,

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

Since we assume that $\mathcal{X}$ is the unique causal, strictly stationary solution of the stochastic differential equation, $\mathcal{X}^q$ 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 the assumptions in (2.23) that $\sigma(\mathbf{\Lambda}) \subseteq (-\infty, 0) + i\mathbb{R}$. These properties lead to

$$\lim_{s \to -\infty} e^{\mathbf{\Lambda}(t-s)} X^q(s) = 0_{kq} \in \mathbb{R}^{kq}$$

in distribution and probability by Slutsky's lemma, since the limit is a degenerate random vector. In combination with the limit result (2.29), we receive, for $t \in \mathbb{R}$,

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

Note that in Proposition 9.4, we show that the integral representation in Proposition 2.30 also holds in L^2. Thus, $X^q(t) \in \mathcal{L}_{Y_V}(t)$, $t \in \mathbb{R}$, and $\mathcal{X}^q$ can also be recovered in the L^2-sense. We are going to provide the evidence there, where we can better emphasise its relevance.

The remaining k-blocks $\mathcal{X}^{(q+j)} = (X^{(q+j)}(t))_{t \in \mathbb{R}}$, $j = 1, \dots, p-q$, are obtained from $\mathcal{X}^q$ and $\mathcal{Y}_V$ by differentiation, as in Lemma 4.2 of Brockwell and Schlemm (2013).

Lemma 2.31. *Let $\mathcal{Y}_V$ be a ICCSS(p,q) process with $p > q > 0$. Then*

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

Note that there is a duplication of notation in Brockwell and Schlemm (2013), which can be seen by recalculating their base case

$$X^{(q+1)}(t) = D^{(1)} X^{(q)}(t) = \underline{\underline{\mathbf{E}}}^\top D^{(1)} X^q(t) = \underline{\underline{\mathbf{E}}}^\top \left(\mathbf{\Lambda} X^q(t) + \mathbf{\Theta} Y_V(t) \right), \quad t \in \mathbb{R}.$$

In summary, for $t \in \mathbb{R}$, we can compute not only the truncated state vector $X^q(t)$ but also the full state vector $X(t)$ based on the knowledge of $(Y_V(s))_{s \leq t}$. This fact justifies calling the ICCSS process $\mathcal{Y}_V$ *invertible* if the assumptions in (2.23) are satisfied.

Part I.

(Local) Causality graphs

Part I. (Local) Causality Graphs

The main goal of this part is to introduce two mixed graphical models for stationary multivariate processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ in continuous time, called the *causality graph* and the *local causality graph*. In these graphical models, the vertices represent the different components $\mathcal{Y}_a = (Y_a(t))_{t\in\mathbb{R}}$, $a \in V = \{1, \ldots, k\}$, of the underlying process $\mathcal{Y}_V$. Furthermore, the vertices are connected by directed and undirected edges, which visualise dynamic relationships between the components. A directed edge represents Granger causality and an undirected edge represents contemporaneous correlation, respectively.

The mathematical notion of *causality* was popularised by Clive W. J. Granger and Christopher A. Sims. In his original work, Granger (1969) uses a linear vector autoregressive (VAR) model, while Sims (1972) uses a moving average (MA) model to understand the causal effects in a bivariate model. Since then, their ideas have been extended in various ways and have been applied in diversified fields, such as econometrics (Imbens, 2022), environmental science (Cox & Popken, 2015), genomics (Heerah, Molinari, Guerrier, & Marshall-Colon, 2021), neuroscience (Bergmann & Hartwigsen, 2021), and social systems (Kuzma, Cruickshank, & Carley, 2022). The recent publication by Shojaie and Fox (2022) provides a comprehensive review of Granger causality and its advances. A detailed discussion of the relationships between Granger and Sims causality is given in Kuersteiner (2010), see also Dufour and Renault (1998) and Eichler (2013).

A fundamental aspect of Granger causality is that the cause precedes the effect. However, not every interesting relationship between two component series is necessarily ordered in time and therefore a causal relationship and directed. Well-known examples include the correlation between aggressive behaviour and the amount of time spent playing computer games each day (Lemmens, Valkenburg, & Peter, 2011), the correlation between the increase in the stork population and the increase in out-of-hospital births (Höfer, Przyrembel, & Verleger, 2004), and the correlation between the number of infants who sleep with the light on and the number of people who develop myopia in later life (Zadnik et al., 2000). An example widely discussed in psychology without a common underlying cause is presented by Jung (1969). It is hence also important to model contemporaneous relationships of an undirected nature, e.g., by *contemporaneous correlation*.

To define the two mixed graphical models, we need concepts of Granger causality and contemporaneous correlation suitable for stationary multivariate stochastic processes in continuous time. Thus, in this part, we first define Granger causality and contemporaneous correlation for such processes by orthogonal projections onto linear spaces, resulting in *conditional orthogonality relations*. For discrete-time processes, this approach was already studied in Dufour and Renault (1998), Eichler (2007), and Florens and Mouchart (1985).

In contrast to the other papers, Eichler (2007) even represents conditional orthogonality relations of a discrete-time VAR process in a graph, the so-called path diagram. In this graphical model, an appropriate concept of Granger causality based on conditional orthogonality models the directed influences. In addition, a concept of contemporaneous correlation, also based on conditional orthogonality, models the undirected influences. The path diagram of Eichler (2007) is used in many applications, such as for human tremor data, for air pollution data (Dahlhaus & Eichler, 2003), for neuronal spike train data, and to study of relationships between the alpha rhythm in EEG data and blood oxygenation level dependent responses in fMRI data (Eichler, 2006). These applications highlight the importance of the path diagram.

An alternative approach is to consider *conditional independence relations* using conditional expectations given σ-fields generated by subprocesses, rather than concepts based on conditional orthogonality. We refer to the approaches of Chamberlain (1982), Eichler (2012), and Florens and Mouchart (1982) for discrete-time processes, and to Comte and Renault (1996), Florens and Fougère (1996), and Petrovic and Dimitrijevic (2012) for continuous-time processes and especially for semi-martingales. Eichler (2012) even defines a graphical model for time series in discrete time, representing certain conditional independence relations. He uses strong Granger causality for the directed edges and contemporaneous conditional dependence for the undirected edges. Of particular interest to this thesis is also the work of Comte and Renault (1996), who propose to model directed influences by global Granger causality and local Granger causality, and undirected influences by global instantaneous causality and local instantaneous causality in continuous time, but their results are not related to graphical models.

For Gaussian random vectors, conditional independence and conditional orthogonality are equivalent, and the standard literature on graphical models for random vectors is based on conditional independence (Lauritzen, 2004). In non-Gaussian time series models, however, conditional expectations are much more difficult to compute than orthogonal projections. This is also reflected in the fact that the assumptions in Eichler (2012) are much more technical and difficult to verify than those in Eichler (2007).

Note that an extension of conditional independence is the concept of *local independence* for Markov processes going back to Schweder (1970), which has been extended to stochastic processes that admit a Doob-Meyer decomposition by Aalen (1987). It is a widespread dependence model to capture asymmetric causal structures in a graph, particularly in the context of composable finite Markov processes (Didelez, 2007), of point processes (Didelez, 2006, 2008; Eichler, Dahlhaus, & Dueck, 2017), and in physical systems (Commenges & Gégout-Petit, 2009). The definitions by Didelez (2006) were recently revisited and adapted by Mogensen and Hansen (2020, 2022), who study local independence graphs for Itô processes. The semi-martingale property of an Itô process is important to the results of these papers, but semi-martingales do not seem to be the right tool for stationary time series models, especially for non-Gaussian models.

For the above reasons, in this part, we restrict ourselves to *linear influences* between the components of a multivariate process and use conditional orthogonality relations to define Granger causality and contemporaneous correlation for continuous-time processes. We also give several equivalent characterisations of the concepts and relate them to other definitions in the literature. In addition, we define local versions of Granger causality and contemporaneous correlation that are less strong and are not necessary in discrete time. Based on the different definitions, we then introduce two mixed graphs, the *causality graph* and the *local causality graph*.

The (local) causality graph encodes (local) Granger causality and (local) contemporaneous correlation between the components of the process $\mathcal{Y}_V$. Conversely, a mixed graph can be associated with a set of conditional orthogonality constraints imposed on the stochastic process

$\mathcal{Y}_V$. Such a set of conditional orthogonality relations encoded by a graph is commonly known as a *Markov property* for the graph (cf. Lauritzen, 2004; Whittaker, 2008). Eichler (2007, 2012) discusses Markov properties for his mixed graphical models, namely the pairwise, the local, the block-recursive, and two global Markov properties. For the global Markov properties, he uses the graph theoretic concepts of m-separation (Richardson, 2003) and p-separation (Levitz, Perlman, & Madigan, 2001), respectively. For the local independence graph, Didelez (2006) develops and investigates an asymmetric notion of separation, called δ-separation, and discusses different levels of Markov properties. In addition, Mogensen and Hansen (2022) show that the multivariate Ornstein-Uhlenbeck process driven by a Brownian motion is the only process that satisfies their global Markov property. As the above literature shows, the derivation of global Markov properties can be quite challenging and is often only valid under additional or even restrictive assumptions.

In our (local) causality graph, we show the pairwise, local, and block-recursive Markov property, and then discuss the global Markov properties for both graphs. Importantly, we find that the causality graph satisfies the global AMP Markov property, which combines the purely graph-theoretic concept of m-separation with conditional orthogonality. Since the notion of m-separation is strong, we present less restrictive alternatives in the global Granger-causal Markov property, that imply Granger non-causality and contemporaneous uncorrelatedness. Note that the assumptions we make, in particular for the validity of the Markov properties, are quite general. We require only a wide-sense stationary, mean-square continuous stochastic process in continuous time, which is purely non-deterministic, and has a spectral density function that satisfies a mild additional restriction. Although the local causality graph satisfies the pairwise, local, and block-recursive Markov properties, not surprisingly, stronger assumptions are required for global Markov properties.

Finally, for a comprehensive understanding of the (local) causality graph and the dependency structures it contains, we derive the graphical structure for the popular *state space processes*, as introduced in Chapter 2. We start with the causal multivariate continuous-time autoregressive (MCAR) process, where we show that the (local) causality graph is well defined and satisfies the desired Markov properties. We compute orthogonal projections that allow us to explicitly characterise the different edge types by the model parameters of the process. However, for general output processes of state space models, the methods suitable for MCAR processes to compute the orthogonal projections are no longer applicable. Therefore, we consider invertible controller canonical state space (ICCSS) processes, for which the orthogonal projections can be computed relying on the invertibility assumption. Despite this distinction, we derive interpretatively meaningful characterisations of the edge types for both MCAR and ICCSS processes, which can be compared with each other and with the literature. At the same time, we obtain clear and easily communicable representations of the intrinsic dependency structure for these important process classes.

Structure of the part

In Chapter 3, we lay the foundation by introducing the conditional orthogonality relation and the corresponding linear spaces generated by multivariate stochastic processes in continuous time. In Chapters 4 and 5, we then define, discuss, and relate various directed and undirected relationships between the components of stationary continuous-time processes. This work results in the definition of the causality graph and the local causality graph in Chapter 6. For these graphical models, we discuss several Markov properties in Chapter 7. The application to causal MCAR processes and ICCSS processes is given in Chapters 8 and 9, respectively, where the edges are characterised by the model parameters.

CHAPTER 3

Preliminaries

In this chapter, we treat the following topics. In Section 3.1, we introduce basic notions, well-known results for orthogonal projections onto linear spaces, and, most importantly, the conditional orthogonality relation with its graphoid property. In Section 3.2, we then define linear spaces generated by multivariate stochastic processes. Via the conditional orthogonality of such linear spaces, the directed and undirected linear relations between subprocesses can be defined in Chapters 4 and 5. Furthermore, keeping in mind the graphoid property, in Section 3.2, we also discuss properties of these linear spaces, such as additivity, separability, and conditional linear separation under certain assumptions on $\mathcal{Y}_V$. Finally, we introduce the assumption that a process is purely non-deterministic. These considerations are crucial for the Markov properties for (local) causality graphs in Chapter 7.

3.1. Conditional orthogonality relation

Let $L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$ be the Hilbert space of square-integrable real- or complex-valued random variables on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. As usual, the inner product is denoted by $\langle X, Y \rangle_{L^2} = \mathbb{E}[X\overline{Y}]$ for $X, Y \in L^2$. Orthogonality with respect to this inner product is denoted by $X \perp Y$. We set $\|X\|_{L^2} = \sqrt{\langle X, X \rangle_{L^2}}$ for $X \in L^2$ and identify random variables that are equal $\mathbb{P}$-a.s. Note that if $\underset{n\to\infty}{\text{l.i.m.}}\, X_n = X$ and $\underset{n\to\infty}{\text{l.i.m.}}\, Y_n = Y$, where $X_n, Y_n, X, Y \in L^2$, $n \in \mathbb{N}$, and $\underset{n\to\infty}{\text{l.i.m.}}$ denotes the limit in the mean square, then

$$\lim_{n\to\infty} \mathbb{E}\left[X_n Y\right] = \mathbb{E}\left[XY\right] \quad \text{and} \quad \lim_{n\to\infty} \mathbb{E}\left[X_n Y_n\right] = \mathbb{E}\left[XY\right], \tag{3.1}$$

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

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

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

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

Even when $\mathcal{L}_1$ and $\mathcal{L}_2$ are closed subspaces, this sum may fail to be closed if both are infinite dimensional. A classic example of this can be found in Halmos (1957), p. 28. Hence, the *closed sum* is denoted by

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

We further denote the *orthogonal projection* of $X \in L^2$ on the closed linear subspace $\mathcal{L}_1 \subseteq L^2$ by $P_{\mathcal{L}_1}(X) = P_{\mathcal{L}_1}X$. Reviews of the properties of orthogonal projections can be found, e.g., in Brockwell and Davis (1991), Lindquist and Picci (2015), and Weidmann (1980). Some of the common properties, which we frequently need in this thesis and therefore list below, are the following.

Lemma 3.1. *Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$ be closed linear subspaces of L^2. Then the subsequent properties apply:*

(a) $P_{\mathcal{L}_1}X = P_{\mathcal{L}_1}P_{\mathcal{L}_2}X = P_{\mathcal{L}_2}P_{\mathcal{L}_1}X$ *for* $X \in L^2$.

(b) $P_{\mathcal{L}_1}(\alpha X + \beta Y) = \alpha P_{\mathcal{L}_1}X + \beta P_{\mathcal{L}_1}Y$ *for* $X, Y \in L^2$, $\alpha, \beta \in \mathbb{C}$.

(c) If $\underset{n\to\infty}{\text{l.i.m.}}\, X_n = X$ *then* $\underset{n\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_1}X_n = P_{\mathcal{L}_1}X$ *for* $X_n, X \in L^2$, $n \in \mathbb{N}$.

With those notations and properties in mind, we define the conditional orthogonality relation as in Eichler (2007), p. 347.

Definition 3.2. Let $\mathcal{L}_i$, $i = 1, \ldots, 3$, be closed linear subspaces of L^2. Then $\mathcal{L}_1$ and $\mathcal{L}_2$ are *conditionally orthogonal* given $\mathcal{L}_3$ if and only if

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

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

Note that Definition 3.2 reduces to the usual orthogonality when $\mathcal{L}_3 = \{0\}$. For a more detailed discussion of the conditional orthogonality relation, we refer to Florens and Mouchart (1985), who study conditional orthogonality in terms of general Hilbert spaces. We only summarise the *graphoid property* of the conditional orthogonality relation as given in Eichler (2007), Proposition A.1.

Lemma 3.3. *Let $\mathcal{L}_i$, $i = 1, \ldots, 4$, be closed linear subspaces of L^2. Then the conditional orthogonality relation defines a semi-graphoid, i.e., it satisfies the following properties:*

(C1) Symmetry: $\mathcal{L}_1 \perp \mathcal{L}_2 \,|\, \mathcal{L}_3 \Rightarrow \mathcal{L}_2 \perp \mathcal{L}_1 \,|\, \mathcal{L}_3$.

(C2) (De-) Composition: $\mathcal{L}_1 \perp \mathcal{L}_2 \,|\, \mathcal{L}_4$ *and* $\mathcal{L}_1 \perp \mathcal{L}_3 \,|\, \mathcal{L}_4 \Leftrightarrow \mathcal{L}_1 \perp \mathcal{L}_2 \vee \mathcal{L}_3 \,|\, \mathcal{L}_4$.

(C3) Weak union: $\mathcal{L}_1 \perp \mathcal{L}_2 \vee \mathcal{L}_3 \,|\, \mathcal{L}_4 \Rightarrow \mathcal{L}_1 \perp \mathcal{L}_2 \,|\, \mathcal{L}_3 \vee \mathcal{L}_4$.

(C4) Contraction: $\mathcal{L}_1 \perp \mathcal{L}_2 \,|\, \mathcal{L}_4$ *and* $\mathcal{L}_1 \perp \mathcal{L}_3 \,|\, \mathcal{L}_2 \vee \mathcal{L}_4 \Rightarrow \mathcal{L}_1 \perp \mathcal{L}_2 \vee \mathcal{L}_3 \,|\, \mathcal{L}_4$.

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

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

Remark 3.4. If $(\mathcal{L}_2 \vee \mathcal{L}_4) \cap (\mathcal{L}_3 \vee \mathcal{L}_4) = \mathcal{L}_4$ holds, we say that $\mathcal{L}_2$ are $\mathcal{L}_3$ *conditionally linearly separated* by $\mathcal{L}_4$ (cf. Eichler, 2007, p. 348).

3.2. Generated linear subspaces

The conditional orthogonality relation is predestined for the analysis of linear relations between two subprocesses $\mathcal{Y}_A$ and $\mathcal{Y}_B$ *given* another subprocess $\mathcal{Y}_C$. For this purpose, however, we must first define suitable closed linear spaces generated by subprocesses $\mathcal{Y}_A$. For an in-depth analysis of such linear spaces, we refer to the numerous publications of Cramér (e.g., Cramér, 1961; Cramér & Leadbetter, 1967) as well as to the comprehensive works by Gladyshev (1958), Matveev (1961), and Robertson (1968).

Definition 3.5. Let $\mathcal{Y}_V$ be wide-sense stationary and let $A \subseteq V$, $s, t \in \mathbb{R}$, $s < t$. Then define

$$\begin{aligned} \ell_{Y_A}(s,t) &:= \left\{ \sum_{i=1}^{n} \sum_{a \in A} \gamma_{a,i} Y_a(t_i) : \gamma_{a,i} \in \mathbb{C},\ s \leq t_1 \leq \ldots \leq t_n \leq t,\ n \in \mathbb{N} \right\}, \\ \mathcal{L}_{Y_A}(s,t) &:= \overline{\ell_{Y_A}(s,t)}, \end{aligned}$$

which is the closed linear subspace generated by $(Y_A(r))_{s \leq r \leq t}$. Furthermore,

$$\begin{aligned} \ell_{Y_A}(t) &:= \left\{ \sum_{i=1}^{n} \sum_{a \in A} \gamma_{a,i} Y_a(t_i) : \gamma_{a,i} \in \mathbb{C},\ -\infty < t_1 \leq \ldots \leq t_n \leq t,\ n \in \mathbb{N} \right\}, \\ \mathcal{L}_{Y_A}(t) &:= \overline{\ell_{Y_A}(t)}, \end{aligned}$$

which is the closed linear subspace generated by $(Y_A(s))_{s \leq t}$, as well as

$$\begin{aligned} \ell_{Y_A}(t,\infty) &:= \left\{ \sum_{i=1}^{n} \sum_{a \in A} \gamma_{a,i} Y_a(t_i) : \gamma_{a,i} \in \mathbb{C},\ t \leq t_1 \leq \ldots \leq t_n < \infty,\ n \in \mathbb{N} \right\}, \\ \mathcal{L}_{Y_A}(t,\infty) &:= \overline{\ell_{Y_A}(t,\infty)}, \end{aligned}$$

which is the closed linear subspace generated by $(Y_A(s))_{s \geq t}$. In addition,

$$\begin{aligned} \ell_{Y_A} &:= \left\{ \sum_{i=1}^{n} \sum_{a \in A} \gamma_{a,i} Y_a(t_i) : \gamma_{a,i} \in \mathbb{C},\ -\infty < t_1 \leq \ldots \leq t_n < \infty,\ n \in \mathbb{N} \right\}, \\ \mathcal{L}_{Y_A} &:= \overline{\ell_{Y_A}}, \end{aligned}$$

which is the closed linear space generated by the entire process $\mathcal{Y}_A$, and

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

called the remote past of the process $\mathcal{Y}_A$. Finally, the closed linear subspace generated by $Y_A(t)$ for fixed $t \in \mathbb{R}$ is denoted by

$$L_{Y_A}(t) := \left\{ \sum_{a \in A} \gamma_a Y_a(t) : \gamma_a \in \mathbb{C} \right\}.$$

Remark 3.6. If $A = \emptyset$, the linear spaces are taken as $\{0\}$. Further, for the definition of the linear spaces, wide-sense stationarity is obviously not necessary. As we rely on this assumption throughout the analysis of (local) causality graphs, we already require it throughout this section for the sake of simplicity.

Below we discuss properties of the linear spaces from Definition 3.5, noting that these properties are trivially valid if $A = \emptyset$ or $B = \emptyset$. First of all, the linear spaces increase both in the time domain and in the index set by definition. For example,

$$\mathcal{L}_{Y_A}(-\infty) \subseteq \mathcal{L}_{Y_A}(s) \subseteq \mathcal{L}_{Y_A}(t) \subseteq \mathcal{L}_{Y_A}, \quad \mathcal{L}_{Y_A}(t) \subseteq \mathcal{L}_{Y_B}(t),$$

whenever $s, t \in \mathbb{R}$, $s \leq t$, $A \subseteq B \subseteq V$. Furthermore, we give an auxiliary result, which we use throughout this thesis and which shows that the linear spaces are additive both in the time domain and in the index set, increasing in time towards a linear limit space.

Lemma 3.7. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq V$, $s, t \in \mathbb{R}$, $s < t$. Then the following statements apply:*

(a) $\mathcal{L}_{Y_A}(s) \vee \mathcal{L}_{Y_A}(s,t) = \mathcal{L}_{Y_A}(t)$ *$\mathbb{P}$-a.s.*

(b) $\mathcal{L}_{Y_A}(s,t) \vee \mathcal{L}_{Y_B}(s,t) = \mathcal{L}_{Y_{A\cup B}}(s,t)$ *$\mathbb{P}$-a.s.*

(c) $\mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_{A\cup B}}(t)$ *$\mathbb{P}$-a.s.*

(d) $\overline{\bigcup_{n\in\mathbb{N}} \mathcal{L}_{Y_A}(n)} = \mathcal{L}_{Y_A}$ *$\mathbb{P}$-a.s.*

Proof. Let $A, B \subseteq V$ and $s, t \in \mathbb{R}$, $s < t$.

(a) First of all, $\mathcal{L}_{Y_A}(s) \subseteq \mathcal{L}_{Y_A}(t)$ and $\mathcal{L}_{Y_A}(s,t) \subseteq \mathcal{L}_{Y_A}(t)$ hold by definition of the linear spaces. Hence, we have $\mathcal{L}_{Y_A}(s) + \mathcal{L}_{Y_A}(s,t) \subseteq \mathcal{L}_{Y_A}(t)$, since $\mathcal{L}_{Y_A}(t)$ is a linear space. As $\mathcal{L}_{Y_A}(t)$ is closed, the first direction $\mathcal{L}_{Y_A}(s) \vee \mathcal{L}_{Y_A}(s,t) \subseteq \mathcal{L}_{Y_A}(t)$ follows. For the opposite subset relation, let $Y^A \in \ell_{Y_A}(t)$. Then there are coefficients $\gamma_{a,i} \in \mathbb{C}$ and time points $-\infty < t_1 \leq \ldots \leq t_n \leq t$, $n \in \mathbb{N}$, such that $\mathbb{P}$-a.s.

$$\begin{aligned} Y^A = \sum_{i=1}^{n} \sum_{a\in A} \gamma_{a,i} Y_a(t_i) &= \sum_{t_i \leq s} \sum_{a\in A} \gamma_{a,i} Y_a(t_i) + \sum_{t_i > s} \sum_{a\in A} \gamma_{a,i} Y_a(t_i) \\ &\in \ell_{Y_A}(s) + \ell_{Y_A}(s,t) \subseteq \mathcal{L}_{Y_A}(s) \vee \mathcal{L}_{Y_A}(s,t). \end{aligned}$$

Therefore, $\ell_{Y_A}(t) \subseteq \mathcal{L}_{Y_A}(s) \vee \mathcal{L}_{Y_A}(s,t)$ is valid. As $\mathcal{L}_{Y_A}(s) \vee \mathcal{L}_{Y_A}(s,t)$ is closed, the relation $\mathcal{L}_{Y_A}(t) \subseteq \mathcal{L}_{Y_A}(s) \vee \mathcal{L}_{Y_A}(s,t)$ follows.

(b,c) If one decomposes the elements of $\ell_{Y_{A\cup B}}(s,t)$ and $\ell_{Y_{A\cup B}}(t)$, respectively, in the index set instead of in the time domain, the proof is very similar to the proof of (a) and is therefore skipped.

(d) First, $\mathcal{L}_{Y_A}(n) \subseteq \mathcal{L}_{Y_A}$, $n \in \mathbb{N}$, applies by definition. Hence, we have $\overline{\bigcup_{n\in\mathbb{N}} \mathcal{L}_{Y_A}(n)} \subseteq \mathcal{L}_{Y_A}$, as $\mathcal{L}_{Y_A}$ is closed. For the opposite subset relation, let $Y_A \in \ell_{Y_A}$. Then there are again coefficients $\gamma_{a,i} \in \mathbb{C}$ and time points $-\infty < t_1 \leq \ldots \leq t_m < \infty$, $m \in \mathbb{N}$, such that

$$Y_A = \sum_{i=1}^{m} \sum_{a\in A} \gamma_{a,i} Y_a(t_i) \quad \mathbb{P}\text{-a.s.}$$

Therefore, for some $n_0 \geq t_m$, $n_0 \in \mathbb{N}$, we have $Y^A \in \mathcal{L}_{Y_A}(t_m) \subseteq \mathcal{L}_{Y_A}(n_0) \subseteq \overline{\bigcup_{n\in\mathbb{N}} \mathcal{L}_{Y_A}(n)}$. Hence, $\ell_{Y_A} \subseteq \overline{\bigcup_{n\in\mathbb{N}} \mathcal{L}_{Y_A}(n)}$ and $\mathcal{L}_{Y_A} \subseteq \overline{\bigcup_{n\in\mathbb{N}} \mathcal{L}_{Y_A}(n)}$, as $\overline{\bigcup_{n\in\mathbb{N}} \mathcal{L}_{Y_A}(n)}$ is closed. ■

Now, given the intersection property (C5) in Lemma 3.3, we prove that all linear spaces used in this thesis are *separable* for a wide-sense stationary and mean-square continuous process $\mathcal{Y}_V$. Note that instead of the continuity in the mean square, it is sufficient to assume that the right and left limits in the mean square exist (cf. Cramér, 1961, Lemma 1).

Lemma 3.8. *Let $\mathcal{Y}_V$ be wide-sense stationary and mean-square continuous. Further, let $A \subseteq V$, $s,t \in \mathbb{R}$, $s < t$. Then $\mathcal{L}_{Y_A}$, $\mathcal{L}_{Y_A}(t)$, and $\mathcal{L}_{Y_A}(s,t)$ are separable.*

Proof. Let $A \subseteq V$ and $s,t \in \mathbb{R}$, $s < t$. We refer to Cramér (1961), Lemma 1, for the proof of ℓ_{Y_A} being separable. If M_A is a countable dense subset of ℓ_{Y_A}, it is also a countable dense subset of $\mathcal{L}_{Y_A}$, which can be explained as follows. Let $Y \in \mathcal{L}_{Y_A}$ be the limit in the mean square of a sequence $Y_n \in \ell_{Y_A}$, $n \in \mathbb{N}$, and let $\varepsilon > 0$. Then there exists some $n_0 \in \mathbb{N}$ such that $\|Y - Y_n\|_{L^2} < \varepsilon/2$ for $n \geq n_0$. Furthermore, we can chose $m_A \in M_A$ such that $\|Y_{n_0} - m_A\|_{L^2} < \varepsilon/2$, since M_A is dense in ℓ_{Y_A}. Then

$$\|Y - m_A\|_{L^2} \leq \|Y - Y_{n_0}\|_{L^2} + \|Y_{n_0} - m_A\|_{L^2} < \varepsilon.$$

Thus, M_A is a countable dense subset of $\mathcal{L}_{Y_A}$ and $\mathcal{L}_{Y_A}$ is separable. Similarly, we obtain that $\mathcal{L}_{Y_A}(t)$ and $\mathcal{L}_{Y_A}(s,t)$ are separable using, e.g., $P_{\mathcal{L}_{Y_A}(t)} M_A$ and $P_{\mathcal{L}_{Y_A}(s,t)} M_A$ as countable dense subsets of $\mathcal{L}_{Y_A}(t)$ and $\mathcal{L}_{Y_A}(s,t)$, respectively. ■

Furthermore, in view of the intersection property (C5) again, we require that $\mathcal{L}_{Y_A}(t)$ and $\mathcal{L}_{Y_B}(t)$ are *conditionally linearly separated* by $\mathcal{L}_{Y_C}(t)$ for $t \in \mathbb{R}$ and disjoint subsets $A, B, C \subseteq V$. This assumption is a lot more intricate because it is very abstract and difficult to verify. In order to understand conditional linear separation more clearly, we introduce a first sufficient criterion.

Lemma 3.9. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $t \in \mathbb{R}$. Suppose that, for all $A, B \subseteq V$ with $A \cap B = \emptyset$, we have*

$$\mathcal{L}_{Y_A}(t) \cap \mathcal{L}_{Y_B}(t) = \{0\} \quad \textit{and} \quad \mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t) \quad \mathbb{P}\text{-}a.s. \tag{3.2}$$

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

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

Proof. Let $t \in \mathbb{R}$ and $A, B, C \subseteq V$ be disjoint subsets. Then $\mathcal{L}_{Y_C}(t) \subseteq \mathcal{L}_{Y_{A\cup C}}(t) \cap \mathcal{L}_{Y_{B\cup C}}(t)$ is immediately true. For the second subset relation $\mathcal{L}_{Y_{A\cup C}}(t) \cap \mathcal{L}_{Y_{B\cup C}}(t) \subseteq \mathcal{L}_{Y_C}(t)$, suppose that $Y \in \mathcal{L}_{Y_{A\cup C}}(t) \cap \mathcal{L}_{Y_{B\cup C}}(t)$. Then assumption (3.2) provides

$$Y \in \mathcal{L}_{Y_{A\cup C}}(t) = \mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_C}(t) \quad \text{and} \quad Y \in \mathcal{L}_{Y_{B\cup C}}(t) = \mathcal{L}_{Y_B}(t) + \mathcal{L}_{Y_C}(t).$$

Therefore, there exist $Y^A \in \mathcal{L}_{Y_A}(t)$, $Z^B \in \mathcal{L}_{Y_B}(t)$, and $Y^C, Z^C \in \mathcal{L}_{Y_C}(t)$, such that we have $Y = Y^A + Y^C = Z^B + Z^C$ $\mathbb{P}$-a.s. This equality yields

$$Y^A - Z^B = Z^C - Y^C \in \mathcal{L}_{Y_{A\cup B}}(t) \cap \mathcal{L}_{Y_C}(t),$$

where $\mathcal{L}_{Y_{A\cup B}}(t) \cap \mathcal{L}_{Y_C}(t) = \{0\}$ $\mathbb{P}$-a.s. by assumption (3.2). Finally,

$$Y^A = Z^B \in \mathcal{L}_{Y_A}(t) \cap \mathcal{L}_{Y_B}(t) = \{0\} \quad \mathbb{P}\text{-a.s.},$$

where we apply assumption (3.2) again. As claimed, $Y = Y^C \in \mathcal{L}_{Y_C}(t)$ $\mathbb{P}$-a.s. ■

The first assumption in (3.2) is the linear independence of the two linear spaces. The second assumption is the closedness of the sum. It makes little sense to formulate these two properties as assumptions on $\mathcal{Y}_V$, as they are still too abstract and difficult to verify. Therefore, we aim to provide an easy-to-use criterion.

The problem of the linear independence and closedness of the sum of a pair of subspaces is the subject of numerous publications, see, e.g., Feshchenko (2012), Hongke (2008), Schochetman, Smith, and Tsui (2001), and the literature therein. According to Proposition 2.3 by Feshchenko (2012), an equivalent characterisation of (3.2) is the existence of an $\varepsilon > 0$, such that

$$\left\|Y^A + Y^B\right\|_{L^2}^2 \geq \varepsilon \left(\left\|Y^A\right\|_{L^2}^2 + \left\|Y^B\right\|_{L^2}^2\right)$$

for all $Y^A \in \mathcal{L}_{Y_A}(t)$, $Y^B \in \mathcal{L}_{Y_B}(t)$, $t \in \mathbb{R}$, and disjoint subsets $A, B \subseteq V$. Replacing Y^A and Y^B with their spectral representations (2.1) leads to a sufficient and manageable condition constraining the spectral density function $f_{Y_V Y_V}(\lambda)$, $\lambda \in \mathbb{R}$, of processes $\mathcal{Y}_V$.

Assumption 3. *Suppose that $\mathcal{Y}_V$ has a spectral density function $f_{Y_V Y_V}(\lambda) > 0$, $\lambda \in \mathbb{R}$, and that there exists an $0 < \varepsilon < 1$, such that*

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

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

For $A = \{a\}$, the function $d_{AB}(\lambda)$, $\lambda \in \mathbb{R}$, is called multiple coherence and we refer to Koopmans (1964a, 1964b) and Goodman (1963) for further reading. Assumption 3 is satisfied, e.g., for most Ornstein-Uhlenbeck processes, MCAR processes, and even most state space processes (for details see Chapters 8 and 9). Assumption 3 seems to be a reasonable assumption and is indeed sufficient for conditional linear separation.

Proposition 3.10. *Let $\mathcal{Y}_V$ be wide-sense stationary, mean-square continuous, and satisfy Assumption 3. Then, for all $t \in \mathbb{R}$ and disjoint subsets $A, B, C \subseteq V$, we have*

$$\mathcal{L}_{Y_A}(t) \cap \mathcal{L}_{Y_B}(t) = \{0\}, \quad \mathcal{L}_{Y_A}(t) + \mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t),$$

and

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

Proof. Let $A, B \subseteq V$ be disjoint with $\#A = \alpha$ and $\#B = \beta$. First, according to Assumption 3, there exists an $0 < \varepsilon < 1$, such that

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

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

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

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

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

for almost all $\lambda \in \mathbb{R}$. Since $(1-\widetilde{\varepsilon}) f_{Y_B Y_B}(\lambda) > 0$, Bernstein (2009), Proposition 8.2.4, provides

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

respectively,

$$\begin{pmatrix} f_{Y_A Y_A}(\lambda) & f_{Y_A Y_B}(\lambda) \\ f_{Y_B Y_A}(\lambda) & f_{Y_B Y_B}(\lambda) \end{pmatrix} \geq \widetilde{\varepsilon} \begin{pmatrix} f_{Y_A Y_A}(\lambda) & 0_{\alpha \times \beta} \\ 0_{\beta \times \alpha} & f_{Y_B Y_B}(\lambda) \end{pmatrix} \tag{3.3}$$

for almost all $\lambda \in \mathbb{R}$. With this preliminary work, we can now provide the actual proof of the assertion. So let $Y^A \in \mathcal{L}_{Y_A}(t)$ and $Y^B \in \mathcal{L}_{Y_B}(t)$, $t \in \mathbb{R}$. Then, we have $Y^A \in \mathcal{L}_{Y_A}$ and $Y^B \in \mathcal{L}_{Y_B}$. Due to Rozanov (1967), I, (7.2), the spectral representations

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

hold, where Φ_A and Φ_B are the random spectral measures form the spectral representation (2.1) of the subprocesses $\mathcal{Y}_A$ and $\mathcal{Y}_B$. Furthermore, $\varphi(\lambda) \in C^{1\times\alpha}$ and $\psi(\lambda) \in C^{1\times\beta}$, $\lambda \in \mathbb{R}$, are measurable vector functions that satisfy

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

Using relation (3.3) and the monotonicity of the integral, we obtain

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

Then Feshchenko (2012), Proposition 2.3, provides that, for $t \in \mathbb{R}$,

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

Thus, Lemma 3.9 yields the final statement $\mathcal{L}_{Y_{A\cup C}}(t) \cap \mathcal{L}_{Y_{B\cup C}}(t) = \mathcal{L}_{Y_C}(t)$ $\mathbb{P}$-a.s. ■

The wide-sense stationarity, mean-square continuity, and Assumption 3 now ensure, as desired, that the conditional orthogonality relation satisfies the *property of intersection (C5)* in Lemma 3.3 for suitable linear subspaces. We comment on Assumption 3 and its interpretation below.

Remark 3.11. Assumption 3 is not necessary for (C5) to hold, the conditional linear separation is sufficient. Furthermore, the proof of Proposition 3.10 remains the same for $Y^A \in \mathcal{L}_{Y_A}(s)$ and $Y^B \in \mathcal{L}_{Y_B}(t)$ for $s, t \in \mathbb{R}$ and disjoint subsets $A, B \subseteq V$. Thus, Assumption 3 even provides

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

Interpretation 3.12. The boundedness $d_{AB}(\lambda) \leq I_\alpha$ is valid without Assumption 3. Indeed, suppose Φ_B is the random spectral measure from the spectral representation of $\mathcal{Y}_B$ in (2.1), then the spectral density function of

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

is (cf. Lemma 10.6)

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

and it is non-negative definite according to Lemma 2.4(c). Furthermore, Assumption 3 especially forbids certain purely linear relationships between the components, which can be seen as follows. Assume that $d_{AB}(\lambda) = I_\alpha$ for almost all $\lambda \in \mathbb{R}$. Then $f_{\varepsilon_{A|B}\varepsilon_{A|B}}(\lambda) = 0_\alpha \in M_k(\mathbb{R})$ for almost all $\lambda \in \mathbb{R}$ and thus, $c_{\varepsilon_{A|B}\varepsilon_{A|B}}(t) = 0_\alpha$ for all $t \in \mathbb{R}$. Therefore, $\varepsilon_{A|B}(t) = 0_\alpha$ $\mathbb{P}$-a.s. and $Y_A(t)$ is already a linear transformation of $\mathcal{Y}_B$. Somewhat loosely, one could say that Assumption 3 not only forbids a purely linear relationship between $\mathcal{Y}_A$ and $\mathcal{Y}_B$ but already requires some kind of distance between the subprocesses due to the uniform boundedness. This also fits with Brillinger (2001), equation (8.3.10), who calls the matrix function $d_{AB}(\lambda)$, $\lambda \in [-\pi, \pi]$, in discrete-time a measure of the linear association of $\mathcal{Y}_A$ and $\mathcal{Y}_B$ at frequency λ.

Remark 3.13. Eichler (2007) proposes a comparable assumption on the spectral density function, also with the aim that the property of intersection (C5) is valid for the conditional orthogonality relation, but for discrete-time processes $\mathcal{Z}_V = (Z_V(t))_{t\in\mathbb{Z}}$. Eichler (2007), equation (2.1), requires the existence of a constant $c > 1$, such that the spectral density function satisfies

$$\frac{1}{c} I_k \leq f_{Z_V Z_V}(\lambda) \leq c I_k \tag{3.4}$$

for all $\lambda \in [-\pi, \pi]$. If this assumption is satisfied, matrix algebra calculations as in the proof of Lemma 8.6 yield that for any disjoint subsets $A, B \subseteq V$

$$f_{Z_A Z_A}(\lambda) - f_{Z_A Z_B}(\lambda) f_{Z_B Z_B}(\lambda)^{-1} f_{Z_B Z_A}(\lambda) \geq \frac{1}{c} I_\alpha \geq \frac{1}{c^2} f_{Z_A Z_A}(\lambda).$$

Thus, for $\lambda \in [-\pi, \pi]$, Assumption 3 is satisfied with $\varepsilon = 1/c^2$. However, we point out that we cannot generalise his assumption directly to continuous-time processes by assuming (3.4) in continuous-time for almost all $\lambda \in \mathbb{R}$. This requirement is too strict and is not even met for simple Ornstein-Uhlenbeck processes, as we can see in the following example.

Example 3.14. Let $\mathcal{Y}_V$ be a two-dimensional causal Ornstein-Uhlenbeck process (cf. Remark 2.21), such that the driving two-dimensional Lévy process satisfies Assumption 1 and that is specified by

$$\mathbf{A} = \begin{pmatrix} -3 & 1 \\ 1 & -3 \end{pmatrix} \quad \text{and} \quad \Sigma_L = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}.$$

The eigenvalues of $f_{Y_V Y_V}(\lambda)$ are

$$\lambda_1(\lambda) = \frac{1}{4\pi\,(\lambda^2+16)} \quad \text{and} \quad \lambda_2(\lambda) = \frac{3}{4\pi\,(\lambda^2+4)}.$$

Both eigenvalue functions tend to zero for $\lambda \to \pm\infty$. Furthermore, due to Bernstein (2009), Lemma 8.4.1, we have

$$\frac{1}{4\pi\,(\lambda^2+16)} I_2 = \lambda_{min}(\lambda) I_2 \leq f_{Y_V Y_V}(\lambda) \leq \lambda_{max}(\lambda) I_2 = \frac{3}{4\pi\,(\lambda^2+4)} I_2.$$

Therefore, we can not find a positive uniform lower bound $1/c\, I_2$ for all $\lambda \in \mathbb{R}$. Nevertheless, straightforward calculations provide that

$$\frac{f_{Y_1Y_2}(\lambda) f_{Y_2Y_1}(\lambda)}{f_{Y_1Y_1}(\lambda) f_{Y_2Y_2}(\lambda)} = \frac{(\lambda^2+22)^2}{4(\lambda^2+13)^2} \leq \frac{(0+22)^2}{4(0+13)^2} \leq 1 - \frac{1}{5}.$$

Since this expression is symmetric under the interchange of the index sets, we have examined all the functions required to be bounded. Thus, Assumption 3 is satisfied with the explicit bound $\varepsilon = 1/5$. In Section 8.2, we show more generally that Assumption 3 is satisfied for Ornstein-Uhlenbeck processes with $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ and $\Sigma_L > 0$.

For the proof of the global AMP Markov property for our causality graph, we need another assumption about the linear space of the remote past of the process, which we introduce at this point. Any process that is wide-sense stationary can be uniquely decomposed in a deterministic and a purely non-deterministic process, that are mutually orthogonal (Gladyshev, 1958, Theorem 1). From the point of view of applications, deterministic processes are not important. Therefore, it is reasonable to assume that the given process is purely non-deterministic.

Assumption 4. *Let $\mathcal{Y}_V$ be purely non-deterministic, that is $\mathcal{L}_{Y_V}(-\infty) = \{0\}$ $\mathbb{P}$-a.s.*

Loosely speaking, we may say that any information present in the process must have entered as a new impulse at some instant in the past (Cramér, 1971, p. 7). From another viewpoint, a process being purely non-deterministic means that the prediction of the infinitely removed future only consists of the knowledge of the mean, i.e.,

$$\underset{h\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_V}(t)} Y_a(t+h) = 0 \quad \mathbb{P}\text{-a.s.} \tag{3.5}$$

for all $a \in V$ and $t \in \mathbb{R}$ (Rozanov, 1967, III, Theorem 2.1). Further necessary and sufficient conditions for processes being purely non-deterministic can be found, e.g., in Gladyshev (1958), Theorem 3, Rozanov (1967), III, Theorem 2.4, and Matveev (1961), Theorem 1.

Finally, we can deduce the following property from Assumptions 3 and 4, that stands in analogy to assumption (M) on σ-fields in Eichler (2012) and equation (2.4) in Eichler (2001). This is the property we require for the proof of the global AMP Markov property for the causality graph.

Lemma 3.15. *Let $\mathcal{Y}_V$ be wide-sense stationary, mean-square continuous, and satisfy Assumptions 3 and 4. Further, let $A \subseteq V$ and $t \in \mathbb{R}$. Then*

$$\bigcap_{k\in\mathbb{N}} \left(\mathcal{L}_{Y_A}(t-k) \vee \mathcal{L}_{Y_{V\setminus A}}(t)\right) = \mathcal{L}_{Y_{V\setminus A}}(t) \quad \mathbb{P}\text{-}a.s. \tag{3.6}$$

Proof. Let $t \in \mathbb{R}$ and $A \subseteq V$. Obviously, the subset relation $\supseteq$ is valid. For the subset relation $\subseteq$, suppose that

$$Y \in \bigcap_{k\in\mathbb{N}} \left(\mathcal{L}_{Y_A}(t-k) \vee \mathcal{L}_{Y_{V\setminus A}}(t)\right).$$

Then $Y \in \mathcal{L}_{Y_A}(t-k) \vee \mathcal{L}_{Y_{V\setminus A}}(t) = \mathcal{L}_{Y_A}(t-k) + \mathcal{L}_{Y_{V\setminus A}}(t)$ for $k \in \mathbb{N}$ due to Assumption 3 and Proposition 3.10, respectively. Hence, there exist $Y^A_{t-k} \in \mathcal{L}_{Y_A}(t-k)$ and $Y^{V\setminus A}_{t-k} \in \mathcal{L}_{Y_{V\setminus A}}(t)$, such that $Y = Y^A_{t-k} + Y^{V\setminus A}_{t-k}$ $\mathbb{P}$-a.s. for $k \in \mathbb{N}$. Furthermore,

$$Y^A_{t-1} - Y^A_{t-k} = Y^{V\setminus A}_{t-k} - Y^{V\setminus A}_{t-1} \in \mathcal{L}_{Y_A}(t-1) \cap \mathcal{L}_{Y_{V\setminus A}}(t-1) = \{0\} \quad \mathbb{P}\text{-a.s.}$$

due to Proposition 3.10 again. Therefore,

$$Y^A_{t-1} = Y^A_{t-k} \in \mathcal{L}_{Y_A}(t-1) \cap \mathcal{L}_{Y_A}(t-k) = \mathcal{L}_{Y_A}(t-k) \subseteq \mathcal{L}_{Y_V}(t-k) \quad \mathbb{P}\text{-a.s.}$$

Since $k \in \mathbb{N}$ is arbitrary and due to Assumption 4

$$Y^A_{t-1} \in \bigcap_{k\in\mathbb{N}} \mathcal{L}_{Y_V}(t-k) = \mathcal{L}_{Y_V}(-\infty) = \{0\} \quad \mathbb{P}\text{-a.s.}$$

But then, as claimed, $Y = Y^{V\setminus A}_{t-1} \in \mathcal{L}_{Y_{V\setminus A}}(t)$ $\mathbb{P}$-a.s. ■

Remark 3.16. Assumptions 3 and 4 are not necessary assumptions for the global AMP Markov properties for the causality graph to hold, but are easier to verify. Sufficient and weaker assumptions are the conditional linear separation and relation (3.6).

CHAPTER 4

Directed influences: Granger causality concepts

In this chapter, we study Granger-causal influences between components of multivariate processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ in continuous time. In discrete time, it is common to say that a component $\mathcal{Z}_a = (Z_a(t))_{t\in\mathbb{Z}}$ is Granger non-causal for another component $\mathcal{Z}_b = (Z_b(t))_{t\in\mathbb{Z}}$ if the prediction of $Z_b(t+1)$ based on the (linear) information available at time t provided by $(Z_V(s))_{s\le t}$ is equal to the prediction of $Z_b(t+1)$ based on the reduced (linear) information provided by $(Z_{V\setminus\{a\}}(s))_{s\le t}$. We transfer this approach to the continuous-time setting. For this purpose, we need to specify the objects that are the continuous-time counterpart of the random variable $Z_b(t+1)$, since there are no fixed time steps in continuous time. We further restrict ourselves to linear Granger-causal influences and introduce different concepts of Granger causality. We emphasise that Granger causality has a temporal order, the cause always precedes the effect. The concepts are therefore directed and we are going to apply them to define the directed edges in the (local) causality graph in Chapter 6.

We structure the chapter at hand as follows. In Section 4.1, we define Granger causality by considering the prediction of $Y_b(t+h)$ on the entire time interval $0 \le h \le 1$. Then, in Section 4.2, we discuss the limiting case $h \to 0$. In Section 4.3, we study global Granger causality, i.e., the prediction of $Y_b(t+h)$ for the entire future $h \ge 0$. Finally, in Section 4.4, we examine relations between the different concepts.

4.1. Granger causality

A first approach to Granger causality is the direct continuous-time counterpart to Definition 2.2 by Eichler (2007), considering a conditional orthogonality relation of linear spaces on the time interval $0 \le h \le 1$.

Definition 4.1. Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A$ is *Granger non-causal* for $\mathcal{Y}_B$ with respect to $\mathcal{Y}_S$ if and only if, for all $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \nrightarrow\!\!\!\!\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$.

Note that we are studying Granger causal relations between subprocesses of $\mathcal{Y}_V$, so A and B are taken to be non-empty. We present equivalent characterisations of Granger non-causality.

Lemma 4.2. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then the following statements are equivalent:*

(a) $\mathcal{Y}_A \nrightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$,

(b) $\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ *for all* $t \in \mathbb{R}$,

(c) $\ell_{Y_B}(t, t+1) \perp \ell_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ *for all* $t \in \mathbb{R}$,

(d) $L_{Y_b}(s) \perp L_{Y_a}(s') \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ *for all* $a \in A$, $b \in B$, $t \le s \le t+1$, $s' \le t$, *and* $t \in \mathbb{R}$.

Proof. Let $A, B \subseteq S \subseteq V$ with $A \cap B = \emptyset$.

(a) $\Rightarrow$ (b): Suppose that $\mathcal{Y}_A \nrightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$. That is, $\mathcal{L}_{Y_B}(t, t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$.

Step 1: Let $Y^B \in \mathcal{L}_{Y_B}(t, t+1)$. Then, we obtain on assumption that, for $Y^A \in \mathcal{L}_{Y_A}(t)$,

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

Step 2: Let $Y^B \in \mathcal{L}_{Y_B}(t)$. Then $Y^B \in \mathcal{L}_{Y_{S\setminus A}}(t)$ and we obtain that $P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y^B = Y^B$, such that, for $Y^A \in \mathcal{L}_{Y_A}(t)$,

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

Step 3: Let $Y^B \in \mathcal{L}_{Y_B}(t+1)$. We receive $\mathcal{L}_{Y_B}(t+1) = \mathcal{L}_{Y_B}(t) \vee \mathcal{L}_{Y_B}(t, t+1)$ due to Lemma 3.7(a). Therefore, there exists a sequence $Y_n^B \in \mathcal{L}_{Y_B}(t) + \mathcal{L}_{Y_B}(t, t+1)$, $n \in \mathbb{N}$, such that we have $\lim_{n\to\infty} \|Y^B - Y_n^B\|_{L^2} = 0$. Lemma 3.1(c) provides

$$\lim_{n\to\infty} \left\| P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y^B - P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y_n^B \right\|_{L^2} = 0.$$

Together with relation (3.1), we obtain, for $Y^A \in \mathcal{L}_{Y_A}(t)$,

$$\begin{aligned}
&\mathbb{E}\left[\left(Y^B - P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y^B\right) \overline{\left(Y^A - P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y^A\right)}\right] \\
&\quad = \lim_{n\to\infty} \mathbb{E}\left[\left(Y_n^B - P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y_n^B\right) \overline{\left(Y^A - P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y^A\right)}\right].
\end{aligned}$$

Since $Y_n^B \in \mathcal{L}_{Y_B}(t) + \mathcal{L}_{Y_B}(t, t+1)$, $n \in \mathbb{N}$, and by Step 1 and Step 2, the right-hand side is zero. So the left-hand side is also zero. Finally, $\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ is valid for all $t \in \mathbb{R}$.

(b) $\Rightarrow$ (a): Suppose that $\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$. Because of the subset relation $\mathcal{L}_{Y_B}(t, t+1) \subseteq \mathcal{L}_{Y_B}(t+1)$, it follows that $\mathcal{L}_{Y_B}(t, t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$.

Similarly, we can conclude by subset arguments that (a) $\Rightarrow$ (c) and (c) $\Rightarrow$ (d).

(c) $\Rightarrow$ (a): Suppose that $\ell_{Y_B}(t, t+1) \perp \ell_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$. Let $Y^B \in \mathcal{L}_{Y_B}(t, t+1)$. Then there exists a sequence $Y_n^B \in \ell_{Y_B}(t, t+1)$, $n \in \mathbb{N}$, such that $\lim_{n\to\infty} \|Y^B - Y_n^B\|_{L^2} = 0$. For $Y^A \in \ell_{Y_A}(t)$, Lemma 3.1(c) and relation (3.1) yield

$$\mathbb{E}\left[\left(Y^B - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y^B\right)\overline{\left(Y^A - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y^A\right)}\right]$$
$$= \lim_{n\to\infty}\mathbb{E}\left[\left(Y_n^B - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y_n^B\right)\overline{\left(Y^A - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y^A\right)}\right].$$

We apply the assumption (c) to obtain that the expression on the right-hand side is zero. In conclusion, $\mathcal{L}_{Y_B}(t,t+1) \perp \ell_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$. In a second step, one can show analogously that $\mathcal{L}_{Y_B}(t,t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$.

(d) $\Rightarrow$ (c): Suppose that $L_{Y_b}(s) \perp L_{Y_a}(s') \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $a \in A$, $b \in B$, $t \leq s \leq t+1$, $s' \leq t$, and $t \in \mathbb{R}$. Let $Y^B \in \ell_{Y_B}(t,t+1)$. Then there are coefficients $\gamma_{b,i} \in \mathbb{C}$ and time points $t \leq t_1 \leq \ldots \leq t_n \leq t+1$, $n \in \mathbb{N}$, such that

$$Y^B = \sum_{i=1}^{n}\sum_{b\in B}\gamma_{b,i}Y_b(t_i) \quad \mathbb{P}\text{-a.s.}$$

For $Y^a \in L_{Y_a}(s')$, we obtain by Lemma 3.1(b) and by linearity of the expectation that

$$\mathbb{E}\left[\left(Y^B - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y^B\right)\overline{\left(Y^a - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y^a\right)}\right]$$
$$= \sum_{i=1}^{n}\sum_{b\in B}\gamma_{b,i}\,\mathbb{E}\left[\left(Y_b(t_i) - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y_b(t_i)\right)\overline{\left(Y^a - P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y^a\right)}\right].$$

Finally, we apply assumption (d) to obtain that the expectation on the right-hand side is zero. Thus, $\ell_{Y_B}(t,t+1) \perp L_{Y_a}(s') \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $a \in A$, $s' \leq t$, and $t \in \mathbb{R}$. In a second step, one can show analogously that $\ell_{Y_B}(t,t+1) \perp \ell_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$. ■

Remark 4.3. The characterisation in Lemma 4.2(b) is similar to Definition 2.2 by Eichler (2012), which uses conditional independence instead of conditional orthogonality. The other characterisations are useful for verifying Granger non-causality. Furthermore, it is obviously always possible to just rewrite one of the linear spaces, that is, for example,

$$\begin{aligned} \mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S \quad &\Leftrightarrow \quad L_{Y_b}(t+h) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t) \quad \forall\, b \in B,\ 0 \leq h \leq 1,\ t \in \mathbb{R}, \\ &\Leftrightarrow \quad L_{Y_B}(t+h) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t) \quad \forall\, 0 \leq h \leq 1,\ t \in \mathbb{R}. \end{aligned}$$

The former relation provides another natural way to define Granger non-causality as a direct counterpart to Eichler (2012).

From the characterisations derived so far, the idea of Granger non-causality as an equality of two predictions, as given, for example, in the introduction and Dufour and Renault (1998) for discrete-time processes, is not yet clear. Therefore, we provide another characterisation.

Theorem 4.4. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$ if and only if, for all $b \in B$, $0 \leq h \leq 1$, and $t \in \mathbb{R}$,*

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

Proof. Due to Proposition 2.4.2 by Lindquist and Picci (2015), $\mathcal{L}_{Y_B}(t,t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ is equivalent to $P_{\mathcal{L}_{Y_S}(t)}Y^B = P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y^B$ $\mathbb{P}$-a.s. for all $Y^B \in \mathcal{L}_{Y_B}(t,t+1)$. Due to the linearity and continuity of orthogonal projections, see Lemma 3.1(b,c), this statement is in turn equivalent to $P_{\mathcal{L}_{Y_S}(t)}Y_b(t+h) = P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y_b(t+h)$ $\mathbb{P}$-a.s. for all $b \in B$, $0 \leq h \leq 1$, and $t \in \mathbb{R}$. ■

In other words, the linear information provided by $(Y_A(s))_{s\leq t}$ can be forgotten without any consequences for the linear prediction of $Y_B(t+h)$ on the time interval $0 \leq h \leq 1$. The characterisations in Lemma 4.2 and Theorem 4.4, furthermore, provide important properties of Granger non-causality. First, in Lemma 4.2(d), we implicitly show that

$$\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S \quad \Leftrightarrow \quad \mathcal{Y}_A \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_S \quad \forall\, b \in B. \tag{4.1}$$

We now discuss the transition from the subset A to the individual elements $a \in A$.

Proposition 4.5. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then*

$$\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S \quad \Rightarrow \quad \mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_S \quad \forall\, a \in A,\, b \in B. \tag{4.2}$$

If $\mathcal{Y}_V$ is additionally mean-square continuous and satisfies Assumption 3, then

$$\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S \quad \Leftrightarrow \quad \mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_S \quad \forall\, a \in A,\, b \in B. \tag{4.3}$$

Proof. For relation (4.2), assume that $\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$. Because of Theorem 4.4, we obtain that $P_{\mathcal{L}_{Y_S}(t)}Y_b(t+h) = P_{\mathcal{L}_{Y_{S\setminus A}}(t)}Y_b(t+h)$ $\mathbb{P}$-a.s. for all $b \in B$, $0 \leq h \leq 1$, and $t \in \mathbb{R}$. Together with Lemma 3.1(a), it follows that, for all $a \in A$, $b \in B$, $0 \leq h \leq 1$, and $t \in \mathbb{R}$,

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

Due to Theorem 4.4, that is $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_S$ for all $a \in A$ and $b \in B$.

For relation (4.3), assume that $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_S$ for all $a \in A$, $b \in B$. Then, we receive, due to Theorem 4.4, that $P_{\mathcal{L}_{Y_S}(t)}Y_b(t+h) = P_{\mathcal{L}_{Y_{S\setminus\{a\}}}(t)}Y_b(t+h)$ $\mathbb{P}$-a.s. for all $a \in A$, $b \in B$, $0 \leq h \leq 1$, and $t \in \mathbb{R}$. This equality implies that

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

Now, from Proposition 3.10, we conclude that

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

implying, due to Lemma 3.1(a), that

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

for all $b \in B$, $t \in \mathbb{R}$, and $0 \leq h \leq 1$. We apply Theorem 4.4 and obtain $\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$. ■

Remark 4.6. The assumption of wide-sense stationarity is, of course, not necessary for the definition and characterisation of Granger causality. However, it is relevant for relation (4.3), which requires Proposition 3.10, hence Assumption 3, and in turn wide-sense stationarity.

4.2. Local Granger causality

Florens and Fougère (1996), Definition 2.1, and Comte and Renault (1996), Definition 1, take a different approach to defining Granger causality in continuous time, using conditional expectations instead of orthogonal projections and generated σ-fields instead of generated linear spaces as information sets. Comte and Renault (1996), Definition 2, further define a local version of Granger causality in the context of semi-martingales. In their Proposition 1, the authors relate this concept to the definition of Renault and Szafarz (1991), who study first-order stochastic differential equations. Instead of considering the whole time interval $0 \leq h \leq 1$, Comte and Renault (1996) examine the transition $h \to 0$ and, to get non-trivial limits, they use difference quotients. The authors also note that the highest existing derivative of the process must always be examined in order to obtain a non-trivial criterion. Therefore, in the style of their characterisation of local Granger causality and our Theorem 4.4, we define the following linear version of local Granger causality.

Definition 4.7. Let $\mathcal{Y}_V$ be wide-sense stationary. Suppose that $\mathcal{Y}_v$ 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 $D^{(j_v)}\mathcal{Y}_v$, where, for $j_v = 0$, we define $D^{(0)}\mathcal{Y}_v = \mathcal{Y}_v$. Further, let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A$ is *locally Granger non-causal* for $\mathcal{Y}_B$ with respect to $\mathcal{Y}_S$ if and only if, for all $b \in B$ and $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \not\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S$.

Remark 4.8. Since $\mathcal{Y}_b$ is by assumption not (j_b+1)-times differentiable, the L^2-limit of $(D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t))/h$ does not exist. However, it is still possible that the L^2-limits of

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

exist, and only then local Granger non-causality is possible. In Chapters 8 and 9, we develop that for causal MCAR processes and ICCSS processes these limits indeed exist.

In the following lemma, we relate local Granger non-causality to a kind of local conditional orthogonality relation, in analogy to Definition 4.1 of Granger non-causality.

Proposition 4.9. *Let $\mathcal{Y}_V$ be wide-sense stationary. Suppose that $\mathcal{Y}_v$ is j_v-times mean-square differentiable, but the (j_v+1)-derivative does not exist for $v \in V$. Further, let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S$ implies that, for all $b \in B$, $t \in \mathbb{R}$, and $Y^A \in \mathcal{L}_{Y_A}(t)$,*

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

Proof. The assumption $\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S$ states that, for all $b \in B$ and $t \in \mathbb{R}$,

$$\begin{aligned} &\operatorname*{l.i.m.}_{h\to 0} P_{\mathcal{L}_{Y_S}(t)} \left(\frac{D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t)}{h}\right) \\ &\quad = \operatorname*{l.i.m.}_{h\to 0} P_{\mathcal{L}_{Y_{S\setminus A}}(t)} \left(\frac{D^{(j_b)}Y_b(t+h) - D^{(j_b)}Y_b(t)}{h}\right) \quad \mathbb{P}\text{-a.s.} \end{aligned} \tag{4.4}$$

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

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

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

$$\begin{aligned} \lim_{h\to 0} \Bigg(&\frac{1}{h}\, \mathbb{E}\left[\left(D^{(j_b)}Y_b(t+h) - P_{\mathcal{L}_{Y_{S\setminus A}}(t)} D^{(j_b)}Y_b(t+h)\right) \overline{Y^A}\right] \\ &+ \frac{1}{h}\, \mathbb{E}\left[\left(P_{\mathcal{L}_{Y_{S\setminus A}}(t)} D^{(j_b)}Y_b(t+h) - P_{\mathcal{L}_{Y_S}(t)} D^{(j_b)}Y_b(t+h)\right) \overline{Y^A}\right]\Bigg) = 0. \end{aligned} \tag{4.5}$$

In the second summand, we apply Remark 2.9, i.e., $D^{(j_b)}Y_b(t) \in \mathcal{L}_{Y_{S\setminus A}}(t) \subseteq \mathcal{L}_{Y_S}(t)$ to get the difference of the two orthogonal projections from relation (4.4). Then the relations (4.4) and (3.1) imply that the second summand in (4.5) is zero in the limit. So the first summand is zero, i.e.,

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

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

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

Adding the limit, the last two equations yield, as claimed,

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

Remark 4.10.

(a) The converse direction requires additional assumptions about linear spaces that go beyond the scope of this thesis. While the relation in Proposition 4.9 could also define a local concept of Granger non-causality, Definition 4.7 better captures the idea of Granger non-causality as the equality of two projections. Moreover, this definition is consistent with the literature.

(b) To see the similarity of Proposition 4.9 to Definition 4.1, we recall that $\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$ if and only if, for all $b \in B$, $0 \leq h \leq 1$, $t \in \mathbb{R}$, and $Y^A \in \mathcal{L}_{Y_A}(t)$,

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

Below, we study properties of local Granger non-causality. First of all, by definition, we receive

$$\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S \quad \Leftrightarrow \quad \mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_S \quad \forall\, b \in B.$$

However, as for Granger non-causality, the transition from a subset A to individual elements $a \in A$ is not immediately clear and is covered in the next proposition, requiring Assumption 3.

Proposition 4.11. *Let $\mathcal{Y}_V$ be wide-sense stationary. Suppose that $\mathcal{Y}_v$ is j_v-times mean-square differentiable, but the (j_v+1)-derivative does not exist for $v \in V$. Let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then*

$$\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S \quad \Rightarrow \quad \mathcal{Y}_a \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_S \quad \forall\, a \in A,\ b \in B. \tag{4.6}$$

Assume additionally that $\mathcal{Y}_V$ is mean-square continuous and Assumption 3 is satisfied. Then

$$\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S \quad \Leftrightarrow \quad \mathcal{Y}_a \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_S \quad \forall\, a \in A,\ b \in B. \tag{4.7}$$

Proof. For relation (4.6), we assume that $\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S$, i.e.,

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

$\mathbb{P}$-a.s. for all $b \in B$ and $t \in \mathbb{R}$. Together with Lemma 3.1(a,c), this equality yields

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

for all $a \in A$, $b \in B$, and $t \in \mathbb{R}$. That is, $\mathcal{Y}_a \not\rightarrow\!\!\!\rightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_S$ for all $a \in A$, $b \in B$.

For relation (4.7), we assume that $\mathcal{Y}_a \nrightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_S$ for all $a \in A$, $b \in B$, i.e.,

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

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

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

Proposition 3.10, which requires Assumption 3, yields

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

Together with Lemma 3.1(a,c), it follows that

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

for all $b \in B$ and $t \in \mathbb{R}$. That is, $\mathcal{Y}_A \nrightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_S$. ∎

Finally, we show the left decomposition property of local Granger non-causality, which is important for the discussion of global Markov properties for the local causality graph in Section 7.3.

Lemma 4.12. *Let $\mathcal{Y}_V$ be wide-sense stationary. Suppose that $\mathcal{Y}_v$ is j_v-times mean-square differentiable, but the (j_v+1)-derivative does not exist for $v \in V$. Let $A, B, C, D \subseteq V$ be disjoint subsets. Then, we have*

$$\mathcal{Y}_{A\cup B} \nrightarrow_0 \mathcal{Y}_C \mid \mathcal{Y}_{A\cup B\cup C\cup D} \quad \Rightarrow \quad \mathcal{Y}_A \nrightarrow_0 \mathcal{Y}_C \mid \mathcal{Y}_{A\cup C\cup D}.$$

Proof. The assumption $\mathcal{Y}_{A\cup B} \nrightarrow_0 \mathcal{Y}_C \mid \mathcal{Y}_{A\cup B\cup C\cup D}$ states that, for all $c \in C$ and $t \in \mathbb{R}$,

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

Together with Lemma 3.1(a,c), it follows that

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

That is, $\mathcal{Y}_A \nrightarrow_0 \mathcal{Y}_C \mid \mathcal{Y}_{A\cup C\cup D}$. ∎

Remark 4.13.

(a) For Granger non-causality, the left decomposition property applies without elaborate proof, since $\mathcal{Y}_{A\cup B} \nrightarrow \mathcal{Y}_C \mid \mathcal{Y}_{A\cup B\cup C\cup D}$ if and only if, for all $t \in \mathbb{R}$,

$$\mathcal{L}_{Y_C}(t, t+1) \perp \mathcal{L}_{Y_{A\cup B}}(t) \mid \mathcal{L}_{Y_{C\cup D}}(t).$$

The property of decomposition (C2) from Lemma 3.3 and $\mathcal{L}_{Y_{A\cup B}}(t) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_B}(t)$ imply, for all $t \in \mathbb{R}$,

$$\mathcal{L}_{Y_C}(t, t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{C\cup D}}(t),$$

which is $\mathcal{Y}_A \nrightarrow \mathcal{Y}_C \mid \mathcal{Y}_{A\cup C\cup D}$.

(b) The right decomposition property of local Granger non-causality, i.e.,

$$\text{“}\mathcal{Y}_A \nrightarrow_0 \mathcal{Y}_{B\cup C} \mid \mathcal{Y}_{A\cup B\cup C\cup D} \quad\Rightarrow\quad \mathcal{Y}_A \nrightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup D}\text{”},$$

is not to be expected. It also applies to Granger non-causality only under additional conditions, for example, using the graphoid properties (C4) and (C5) from Lemma 3.3. The lack of right decomposability can be explained as follows: $\mathcal{Y}_A$ may be (local) Granger non-causal for $\mathcal{Y}_{B\cup C}$ given $\mathcal{Y}_{A\cup B\cup C\cup D}$, since the corresponding information of $\mathcal{Y}_A$ is already present in $\mathcal{Y}_C$. However, if $\mathcal{Y}_C$ is omitted, there may be a (local) Granger causal influence of $\mathcal{Y}_A$ on $\mathcal{Y}_B$. This topic has been addressed, for example, by Didelez (2006) in the context of directed local independence graphs.

4.3. Global Granger causality

A third concept of directed influence is to consider Granger causality up to an arbitrary, possibly infinite, horizon. In discrete time, the concept of global Granger causality goes back to the seminal work of Sims (1972) and is also called Sims causality. We introduce the following definition of Granger causality up to a fixed horizon and the definition of global Granger causality as a counterpart to the discrete-time Definition 4.4 of Eichler (2007), using conditional orthogonality relations. As usual, we restrict ourselves to linear causal influences between the components.

Definition 4.14. Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A$ is *Granger non-causal* for $\mathcal{Y}_B$ with respect to $\mathcal{Y}_S$ *up to horizon* h, $h \in \mathbb{R}$, if and only if, for all $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \not\rightarrow\!\!\rightarrow_h \mathcal{Y}_B \mid \mathcal{Y}_S$. Furthermore, $\mathcal{Y}_A$ is *global Granger non-causal* for $\mathcal{Y}_B$ with respect to $\mathcal{Y}_S$ if and only if, for all $h \geq 0$ and $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \not\rightarrow\!\!\rightarrow_\infty \mathcal{Y}_B \mid \mathcal{Y}_S$.

The analysis of such long-run Granger non-causality relations is a useful complement to understanding the relationships between the components of $\mathcal{Y}_V$. For example, by comparing global Granger non-causality and Granger non-causality, it is possible to distinguish between long-run and short-run effects.

Remark 4.15. The characterisations of Granger causality up to a fixed horizon and global Granger causality are similar to those for Granger causality in Lemma 4.2 and Theorem 4.4, with analogous proof. Furthermore, analogue relationships as in (4.1) and Proposition 4.5 are valid. In particular, for $h \geq 0$,

$$\begin{aligned}
\mathcal{Y}_A \not\rightarrow\!\!\rightarrow_h \mathcal{Y}_B \mid \mathcal{Y}_S \quad &\Leftrightarrow \quad L_{Y_b}(s) \perp L_{Y_a}(s') \mid \mathcal{L}_{Y_{S\setminus A}}(t) \\
&\qquad\; \forall\, a \in A,\ b \in B,\ t \leq s \leq t+h,\ s' \leq t,\ t \in \mathbb{R}, \\
&\Leftrightarrow \quad P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h') = P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y_b(t+h') \quad \mathbb{P}\text{-a.s.} \\
&\qquad\; \forall\, b \in B,\ 0 \leq h' \leq h,\ t \in \mathbb{R}, \\
\mathcal{Y}_A \not\rightarrow\!\!\rightarrow_\infty \mathcal{Y}_B \mid \mathcal{Y}_S \quad &\Leftrightarrow \quad L_{Y_b}(s) \perp L_{Y_a}(s') \mid \mathcal{L}_{Y_{S\setminus A}}(t) \\
&\qquad\; \forall\, a \in A,\ b \in B,\ t \leq s \leq t+h,\ h \geq 0,\ s' \leq t,\ t \in \mathbb{R}, \\
&\Leftrightarrow \quad P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h') = P_{\mathcal{L}_{Y_{S\setminus A}}(t)} Y_b(t+h') \quad \mathbb{P}\text{-a.s.} \\
&\qquad\; \forall\, b \in B,\ 0 \leq h' \leq h,\ h \geq 0,\ t \in \mathbb{R}.
\end{aligned}$$

These characterisations allow, as usual, for an interpretation of the Granger non-causality concepts as an equality of linear predictions.

For global Granger causality, there is one more natural characterisation that emphasises the global perspective and provides another natural way to define the concept.

Lemma 4.16. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A \not\rightarrow\!\!\rightarrow_\infty \mathcal{Y}_B \mid \mathcal{Y}_S$ if and only if, for all $t \in \mathbb{R}$,*

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

Proof. Suppose that $\mathcal{Y}_A \nrightarrow_\infty \mathcal{Y}_B \,|\, \mathcal{Y}_S$. That is, due to Remark 4.15, equivalent to the conditional orthogonality relation $L_{Y_b}(s) \perp L_{Y_a}(s') \,|\, \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $a \in A$, $b \in B$, $t \leq s \leq t+h$, $h \geq 0$, $s' \leq t$, and $t \in \mathbb{R}$. This statement is the same as $L_{Y_b}(s) \perp L_{Y_a}(s') \,|\, \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $a \in A$, $b \in B$, $s \geq t$, $s' \leq t$, and $t \in \mathbb{R}$. The latter conditional orthogonality relation is equivalent to $\mathcal{L}_{Y_B}(t,\infty) \perp \mathcal{L}_{Y_A}(t) \,|\, \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$, as in the proof of Lemma 4.2. ∎

4.4. Relations between the Granger causality concepts

Given the similarity of the different definitions of Granger causality, we expect *relations between the different concepts*. Above all, we expect that from global Granger non-causality follows Granger non-causality, and from Granger non-causality follows local Granger non-causality. In the following lemma, we address such relationships. We refer to Dufour and Renault (1998), Eichler (2013), and Kuersteiner (2010) for relations between the different definitions for discrete-time processes.

Proposition 4.17. *Let $\mathcal{Y}_V$ be wide-sense stationary. Suppose that $\mathcal{Y}_v$ is j_v-times mean-square differentiable, but the (j_v+1)-derivative does not exist. Let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then the following implications hold:*

(a) $\mathcal{Y}_A \nrightarrow_\infty \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad \Rightarrow \quad \mathcal{Y}_A \nrightarrow \mathcal{Y}_B \,|\, \mathcal{Y}_S$.

(b) $\mathcal{Y}_A \nrightarrow_\infty \mathcal{Y}_{S\setminus A} \,|\, \mathcal{Y}_S \quad \Leftrightarrow \quad \mathcal{Y}_A \nrightarrow \mathcal{Y}_{S\setminus A} \,|\, \mathcal{Y}_S$.

(c) $\mathcal{Y}_A \nrightarrow \mathcal{Y}_{S\setminus A} \,|\, \mathcal{Y}_S \quad \Rightarrow \quad \mathcal{Y}_A \nrightarrow_\infty \mathcal{Y}_B \,|\, \mathcal{Y}_S$.

(d) $\mathcal{Y}_A \nrightarrow \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad \Rightarrow \quad \mathcal{Y}_A \nrightarrow_0 \mathcal{Y}_B \,|\, \mathcal{Y}_S$.

Proof.

(a) This implication is obvious by definition.

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

$$\mathcal{L}_{Y_{S\setminus A}}(t+k) \perp \mathcal{L}_{Y_A}(t) \,|\, \mathcal{L}_{Y_{S\setminus A}}(t) \quad \forall\, t \in \mathbb{R},\ k \in \mathbb{N}. \tag{4.8}$$

First, we note that the assumption $\mathcal{Y}_A \nrightarrow \mathcal{Y}_{S\setminus A} \,|\, \mathcal{Y}_S$ and Lemma 4.2(b) yield the *base case*

$$\mathcal{L}_{Y_{S\setminus A}}(t+1) \perp \mathcal{L}_{Y_A}(t) \,|\, \mathcal{L}_{Y_{S\setminus A}}(t) \quad \forall\, t \in \mathbb{R}. \tag{4.9}$$

Now, replacing t by $t+1$ in the *induction hypothesis* (4.8) gives

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

Since we have $\mathcal{L}_{Y_A}(t+1) = \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_A}(t,t+1)$ by Lemma 3.7, the property (C2) from Lemma 3.3 implies

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

which is, by Lemma 3.7 again,

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

This result, the base case (4.9), and the property (C4) yield

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

Finally, $\mathcal{L}_{Y_{S\setminus A}}(t+k+1) \vee \mathcal{L}_{Y_{S\setminus A}}(t,t+1) = \mathcal{L}_{Y_{S\setminus A}}(t+k+1)$ gives the *induction step*

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

To bring the proof to an end, let $h \geq 0$ and denote by $\lceil h \rceil$ the upper Gaussian bracket. Then we have $\mathcal{L}_{Y_{S\setminus A}}(t+h) \subseteq \mathcal{L}_{Y_{S\setminus A}}(t+\lceil h \rceil)$. Relation (4.8), Lemma 3.7, and the property (C2) imply

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

which is $\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_{S\setminus A} \mid \mathcal{Y}_S$ by definition.

(c) This statement follows directly due to (b), $B \subseteq S \setminus A$, Lemma 3.7, and the property (C2).

(d) Let $\mathcal{Y}_A \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_S$. That is $\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{S\setminus A}}(t)$ for all $t \in \mathbb{R}$ due to Lemma 4.2(b). Then, as in the proof of Theorem 4.4 (cf. Proposition 2.4.2 by Lindquist & Picci, 2015), we have

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

for all $Y^B \in \mathcal{L}_{Y_B}(t+1)$ and $t \in \mathbb{R}$. Furthermore, Remark 2.9 provides that, for $b \in B$ and $0 \leq h \leq 1$, we have $D^{(j_b)}Y_b(t+h) \in \mathcal{L}_{Y_B}(t+h) \subseteq \mathcal{L}_{Y_B}(t+1)$. All together yield

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

Since, in addition, $D^{(j_b)}Y_b(t) \in \mathcal{L}_{Y_{S\setminus A}}(t) \subseteq \mathcal{L}_{Y_S}(t)$ by Remark 2.9 again, we have

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

Letting $h \to 0$, we receive the statement. ■

Remark 4.18. Dufour and Renault (1998), p. 1106, explain the lack of equivalence between Granger causality and global Granger causality as follows: If there are auxiliary components, $\mathcal{Y}_A$ might not help to predict $\mathcal{Y}_B$ given $\mathcal{Y}_S$ one step ahead, but $\mathcal{Y}_A$ might help to predict $\mathcal{Y}_B$ given $\mathcal{Y}_S$ several periods ahead. For example, the values of $\mathcal{Y}_A$ up to time t help to predict $\mathcal{L}_{Y_B}(t+1,t+2)$, even though the values are useless to predict $\mathcal{L}_{Y_B}(t,t+1)$, as $\mathcal{Y}_A$ helps to predict the environment one period ahead, which in turn influences $\mathcal{Y}_A$ at a subsequent period. Thus, it is also not surprising that we have equivalence in the case without environment in Proposition 4.17(b), which applies in particular to bivariate processes, i.e., $\mathcal{Y}_a \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_{\{a,b\}}$ if and only if $\mathcal{Y}_a \not\rightarrow\!\!\!\rightarrow_\infty \mathcal{Y}_b \mid \mathcal{Y}_{\{a,b\}}$.

The similarities and differences between the various definitions of Granger causality can also be observed in Example 4.19, where we study Ornstein-Uhlenbeck processes. In particular, we find that, generally, the opposite direction of Proposition 4.17(d) does not apply.

Example 4.19. Let $\mathcal{Y}_V$ be a k-dimensional Ornstein-Uhlenbeck process (cf. Remark 2.21), such that the driving k-dimensional Lévy process satisfies Assumption 1, $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$, and $\Sigma_L > 0$. For this process, we derive in Corollaries 8.15(b) and 8.24(a) that

$$\begin{aligned} \mathcal{Y}_a \nrightarrow_\infty \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \mathcal{Y}_a \nrightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad &\Leftrightarrow \quad [\mathbf{A}^\alpha]_{ba} = 0 \quad \forall\, \alpha = 1, \ldots, k-1, \\ \mathcal{Y}_a \nrightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad &\Leftrightarrow \quad [\mathbf{A}]_{ba} = 0. \end{aligned}$$

Of course,

$$\begin{aligned} \mathcal{Y}_a \nrightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad &\Rightarrow \quad [\mathbf{A}^\alpha]_{ba} = 0 \quad \forall\, \alpha = 1, \ldots, k-1 \quad \Rightarrow \quad [\mathbf{A}]_{ba} = 0 \\ &\Rightarrow \quad \mathcal{Y}_a \nrightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V, \end{aligned}$$

but, generally, the opposite direction does not hold only if, for example, $\mathbf{A}$ is a diagonal matrix. It is clear from the example that the definition of Granger non-causality is much stronger than the definition of local Granger non-causality. In general, there is no equivalence. However, Granger non-causality and global Granger non-causality are equivalent for Ornstein-Uhlenbeck processes, and the case in Remark 4.18 does not occur.

CHAPTER 5

Undirected influences: Contemporaneous correlation concepts

Although the main interest is usually in Granger causality, it is also fruitful to think about undirected relations between components of multivariate processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ in continuous time. One such approach is the concept of contemporaneous correlation. In discrete time, two components $\mathcal{Z}_a = (Z_a(t))_{t\in\mathbb{Z}}$ and $\mathcal{Z}_b = (Z_b(t))_{t\in\mathbb{Z}}$ are contemporaneously uncorrelated if and only if, given the amount of information provided by $(Z_V(s))_{s\leq t}$, the random variables $Z_a(t+1)$ and $Z_b(t+1)$ are uncorrelated. To adapt this approach to the continuous-time setting, we restrict ourselves to linear influences, and we have to specify the objects corresponding to $Z_a(t+1)$ and $Z_b(t+1)$ in continuous time. In doing so, we introduce various concepts of contemporaneous correlation. We also give characterisations of our definitions both via the conditional orthogonality relation and via the equality of orthogonal projections. Note that the term undirected means that there is no temporal order, as opposed to Granger causality. We will therefore use these concepts to define the undirected edges in our (local) causality graph in Chapter 6.

The chapter at hand is structured as follows. First, in Section 5.1, we introduce a definition of contemporaneous correlation by considering the correlation of the random variables $Y_a(t+h)$ and $Y_b(t+h')$ on the entire time interval $0 \leq h, h' \leq 1$. In Section 5.2, we then discuss the limiting case, where the size of the time step tends to zero. In Section 5.3, we study global contemporaneous correlation, i.e., consider the correlation of the random variables on the entire future $h, h' \geq 0$. Finally, in Section 5.4, we examine relations between the different concepts.

5.1. Contemporaneous correlation

Let us start with the first concept of an undirected influence, motivated in discrete time by Eichler (2007) in Definition 2.2, by considering the entire time interval $0 \leq h \leq 1$.

Definition 5.1. Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A$ and $\mathcal{Y}_B$ are *contemporaneously uncorrelated* with respect to $\mathcal{Y}_S$ if and only if, for all $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \nsim \mathcal{Y}_B \mid \mathcal{Y}_S$.

Note that again we are studying relations between subprocesses of $\mathcal{Y}_V$, so A and B are taken to be non-empty. By analogy with Lemma 4.2, we obtain the following equivalent characterisations. Since the proof is very similar, it is omitted.

Lemma 5.2. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then the following statements are equivalent:*

(a) $\mathcal{Y}_A \nsim \mathcal{Y}_B \,|\, \mathcal{Y}_S$,

(b) $\mathcal{L}_{Y_A}(t+1) \perp \mathcal{L}_{Y_B}(t+1) \,|\, \mathcal{L}_{Y_S}(t)$ *for all* $t \in \mathbb{R}$,

(c) $\ell_{Y_A}(t, t+1) \perp \ell_{Y_B}(t, t+1) \,|\, \mathcal{L}_{Y_S}(t)$ *for all* $t \in \mathbb{R}$,

(d) $L_{Y_a}(s) \perp L_{Y_b}(s') \,|\, \mathcal{L}_{Y_S}(t)$ *for all* $a \in A$, $b \in B$, $t \leq s, s' \leq t+1$, *and* $t \in \mathbb{R}$.

Remark 5.3.

(a) In Lemma 5.2(d), we implicitly show that

$$\mathcal{Y}_A \nsim \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_S \quad \forall\, a \in A,\ b \in B, \tag{5.1}$$

which is useful for verifying if two subprocesses are contemporaneously uncorrelated. Unlike for Granger non-causality in relation (4.3), no additional assumption has to be made.

(b) Rewriting Lemma 5.2(d), we obtain that

$$\begin{aligned} \mathcal{Y}_A \nsim \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad &\Leftrightarrow \quad L_{Y_a}(t+h) \perp L_{Y_b}(t+h') \,|\, \mathcal{L}_{Y_S}(t) \\ &\qquad\ \forall\, a \in A,\ b \in B,\ 0 \leq h, h' \leq 1,\ t \in \mathbb{R}, \\ &\Leftrightarrow \quad L_{Y_A}(t+h) \perp L_{Y_B}(t+h') \,|\, \mathcal{L}_{Y_S}(t) \\ &\qquad\ \forall\, 0 \leq h, h' \leq 1,\ t \in \mathbb{R}. \end{aligned}$$

The former relation provides another natural way to define contemporaneous correlation as a continuous-time counterpart to Eichler (2007), Definition 2.2.

(c) In view of Eichler (2007), Definition 2.2, it is further plausible to define that $\mathcal{Y}_A$ and $\mathcal{Y}_B$ are contemporaneously uncorrelated with respect to $\mathcal{Y}_S$ by

$$L_{Y_A}(t+h) \perp L_{Y_B}(t+h) \,|\, \mathcal{L}_{Y_S}(t) \quad \forall\, 0 \leq h \leq 1,\ t \in \mathbb{R}.$$

In this case, however, no global Markov property (cf. Chapter 7) can be shown in the associated causality graph. The proofs rely heavily on Definition 5.1 and Lemma 3.3.

Similar to Granger causality, a characterisation of contemporaneous correlation can be given as the equality of two orthogonal projections. The proof is similar to that of Theorem 4.4 and we skip the details.

Proposition 5.4. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A \nsim \mathcal{Y}_B \,|\, \mathcal{Y}_S$ if and only if, for all $b \in B$, $0 \leq h \leq 1$, and $t \in \mathbb{R}$,*

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

Analogously, $\mathcal{Y}_A \nsim \mathcal{Y}_B \,|\, \mathcal{Y}_S$ if and only if, for all $a \in A$, $0 \leq h \leq 1$, and $t \in \mathbb{R}$,

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

In other words, the linear prediction of $\mathcal{Y}_B$ ($\mathcal{Y}_A$) in the near future based on $\mathcal{L}_{Y_S}(t)$ cannot be improved by adding additional linear information about $\mathcal{Y}_A$ ($\mathcal{Y}_B$) in the near future. Note again that the stationarity assumption is not necessary for the definition of contemporaneous correlation and its characterisations. But it is essential for the relation (4.3), so we always include the assumption for simplicity.

5.2. Local contemporaneous correlation

So far, contemporaneous correlation is defined by considering the entire time interval $0 \leq h \leq 1$. To define a local version of this concept, note that the characterisation of subprocesses being contemporaneously uncorrelated in Lemma 5.2(b) means that, for any $Y^A \in \mathcal{L}_{Y_A}(t+1)$ and $Y^B \in \mathcal{L}_{Y_B}(t+1)$, $t \in \mathbb{R}$, we have

$$\mathbb{E}\left[\left(Y^A - P_{\mathcal{L}_{Y_S}(t)}Y^A\right)\overline{\left(Y^B - P_{\mathcal{L}_{Y_S}(t)}Y^B\right)}\right] = 0. \tag{5.2}$$

The motivation for the local version is that instead of taking the whole linear spaces $\mathcal{L}_{Y_A}(t+1)$, we use only the highest derivative $D^{(j_a)}Y_a(t+h)$ for each $a \in A$ and consider the limiting case $h \to 0$, similarly for $\mathcal{L}_{Y_B}(t+1)$. For non-trivial limits, we also have to divide by h. Then, we get the following definition.

Definition 5.5. Let $\mathcal{Y}_V$ be wide-sense stationary. Suppose that $\mathcal{Y}_v$ is j_v-times mean-square differentiable, but the (j_v+1)-derivative does not exist for $v \in V$. Let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A$ and $\mathcal{Y}_B$ are *locally contemporaneously uncorrelated* with respect to $\mathcal{Y}_S$ if and only if, for all $a \in A$, $b \in B$, and $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \nsim_0 \mathcal{Y}_B \,|\, \mathcal{Y}_S$.

Remark 5.6.

(a) By definition, without any additional assumptions, we get

$$\mathcal{Y}_A \nsim_0 \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_S \quad \forall\, a \in A,\ b \in B, \tag{5.3}$$

which is useful for verifying local contemporaneous correlation.

(b) Definition 5.5 is similar to the characterisation of local instantaneous non-causality for semimartingales in Proposition 3 by Comte and Renault (1996), using orthogonal projections instead of conditional expectations and linear spaces instead of σ-fields.

(c) Due to Remark 2.9, we have $\mathcal{Y}_A \nsim_0 \mathcal{Y}_B \mid \mathcal{Y}_S$ if and only if, for all $a \in A$, $b \in B$, and $t \in \mathbb{R}$,

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

A sufficient condition in analogy to Definition 4.7 of local Granger non-causality and to the characterisation of contemporaneous correlation in Proposition 5.4 is therefore

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

for all $a \in A$, $b \in B$, and $t \in \mathbb{R}$. This statement is shown analogously to Proposition 4.9.

5.3. GLOBAL CONTEMPORANEOUS CORRELATION

Finally, we introduce the concept of contemporaneous correlation up to a fixed horizon and the concept of global contemporaneous correlation, analogous to Granger causality up to a fixed horizon and global Granger causality. These concepts make it again possible to distinguish between short-term and long-term effects.

Definition 5.7. Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A$ and $\mathcal{Y}_B$ are *contemporaneously uncorrelated* with respect to $\mathcal{Y}_S$ *up to horizon* h, $h \geq 0$, if and only if, for all $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \nsim_h \mathcal{Y}_B \mid \mathcal{Y}_S$. Furthermore, $\mathcal{Y}_A$ and $\mathcal{Y}_B$ are *globally contemporaneously uncorrelated* with respect to $\mathcal{Y}_S$ if and only if, for all $h \geq 0$ and $t \in \mathbb{R}$,

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

We write $\mathcal{Y}_A \nsim_\infty \mathcal{Y}_B \mid \mathcal{Y}_S$.

The usual characterisations of contemporaneous correlation analogous to Lemma 5.2 and Proposition 5.4 apply. Since the proofs are in turn similar to those of Lemma 4.2 and Theorem 4.4, respectively, they are omitted. In particular, we have

$$
\begin{aligned}
\mathcal{Y}_A \nsim_h \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad &\Leftrightarrow \quad L_{Y_a}(s) \perp L_{Y_b}(s') \,|\, \mathcal{L}_{Y_S}(t) && (5.4)\\
&\qquad \forall\, a \in A,\ b \in B,\ t \leq s, s' \leq t+h,\ t \in \mathbb{R},\\
&\Leftrightarrow \quad P_{\mathcal{L}_{Y_S}(t) \vee \mathcal{L}_{Y_A}(t,t+h)} Y_b(t+h') = P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h') \quad \mathbb{P}\text{-a.s.}\\
&\qquad \forall\, b \in B,\ 0 \leq h' \leq h,\ t \in \mathbb{R}.\\
\mathcal{Y}_A \nsim_\infty \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad &\Leftrightarrow \quad L_{Y_a}(s) \perp L_{Y_b}(s') \,|\, \mathcal{L}_{Y_S}(t) && (5.5)\\
&\qquad \forall\, a \in A,\ b \in B,\ t \leq s, s' \leq t+h,\ h \geq 0,\ t \in \mathbb{R},\\
&\Leftrightarrow \quad P_{\mathcal{L}_{Y_S}(t) \vee \mathcal{L}_{Y_A}(t,t+h)} Y_b(t+h') = P_{\mathcal{L}_{Y_S}(t)} Y_b(t+h') \quad \mathbb{P}\text{-a.s.}\\
&\qquad \forall\, b \in B,\ 0 \leq h' \leq h,\ h \geq 0,\ t \in \mathbb{R}.
\end{aligned}
$$

Remark 5.8. Because of the relations (5.4) and (5.5), without any further assumptions, we obtain that, for $h \geq 0$,

$$
\begin{aligned}
\mathcal{Y}_A \nsim_h \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad &\Leftrightarrow \quad \mathcal{Y}_a \nsim_h \mathcal{Y}_b \,|\, \mathcal{Y}_S \qquad \forall\, a \in A,\ b \in B,\\
\mathcal{Y}_A \nsim_\infty \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad &\Leftrightarrow \quad \mathcal{Y}_a \nsim_\infty \mathcal{Y}_b \,|\, \mathcal{Y}_S \qquad \forall\, a \in A,\ b \in B.
\end{aligned}
$$

For global contemporaneous correlation, there is one more important characterisation that emphasises the global perspective and provides another natural way to define this concept. The proof is similar to that of Lemma 4.16, using relation (5.5), so we skip the details.

Lemma 5.9. *Let $\mathcal{Y}_V$ be wide-sense stationary and let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then $\mathcal{Y}_A \nsim_\infty \mathcal{Y}_B \,|\, \mathcal{Y}_S$ if and only if, for all $t \in \mathbb{R}$,*

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

5.4. Relations between the contemporaneous correlation concepts

Given the similarity of the different definitions of contemporaneous correlation, *relationships between the concepts* are expected. Two subprocesses $\mathcal{Y}_A$ and $\mathcal{Y}_B$ being globally contemporaneously uncorrelated given $\mathcal{Y}_S$ should also be contemporaneously uncorrelated, and two subprocesses $\mathcal{Y}_A$ and $\mathcal{Y}_B$ being contemporaneously uncorrelated given $\mathcal{Y}_S$ should also be locally contemporaneously uncorrelated.

Proposition 5.10. *Let $\mathcal{Y}_V$ be wide-sense stationary. Suppose that $\mathcal{Y}_v$ is j_v-times mean-square differentiable, but the (j_v+1)-derivative does not exist for $v \in V$. Let $A, B \subseteq S \subseteq V$, $A \cap B = \emptyset$. Then the following implications hold:*

(a) $\mathcal{Y}_A \nsim_\infty \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad \Rightarrow \quad \mathcal{Y}_A \nsim \mathcal{Y}_B \,|\, \mathcal{Y}_S.$

(b) $\mathcal{Y}_A \nsim \mathcal{Y}_B \,|\, \mathcal{Y}_S \quad \Rightarrow \quad \mathcal{Y}_A \nsim_0 \mathcal{Y}_B \,|\, \mathcal{Y}_S.$

These relations follow by definition, because of Remark 2.9, and the relation (5.2). The similarities and differences between the various definitions also become apparent when we study the example of an Ornstein-Uhlenbeck process. In particular, we find that the opposite direction in Proposition 5.10(b) does not hold in general.

Example 5.11. Suppose that $\mathcal{Y}_V$ is the Ornstein-Uhlenbeck process from Example 4.19. Then we derive in Corollaries 8.17(b) and 8.24(b) that

$$\begin{aligned} \mathcal{Y}_a \nsim_\infty \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad &\Leftrightarrow \quad \left[\mathbf{A}^\alpha \Sigma_L (\mathbf{A}^\top)^\beta\right]_{ab} = 0 \quad \forall\, \alpha, \beta = 0, \dots, k-1, \\ \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad &\Leftrightarrow \quad [\Sigma_L]_{ab} = 0. \end{aligned}$$

Of course, we obtain that

$$\begin{aligned} \mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad &\Rightarrow \quad \left[\mathbf{A}^\alpha \Sigma_L (\mathbf{A}^\top)^\beta\right]_{ab} = 0 \quad \forall\, \alpha, \beta = 0, \dots, k-1 \quad \Rightarrow \quad [\Sigma_L]_{ab} = 0 \\ &\Rightarrow \quad \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V, \end{aligned}$$

but, generally, the opposite does not hold only if, for example, $\mathbf{A}$ is a diagonal matrix. It is clear from the example that it is much more restrictive for subprocesses to be contemporaneously uncorrelated than for them to be locally contemporaneously uncorrelated. However, contemporaneous correlation and global contemporaneous correlation are equivalent for Ornstein-Uhlenbeck processes.

Chapter 6

Definition of (local) causality graphs

In this section, we finally introduce our first two graphical models for stationary multivariate processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ in continuous time. In these graphs, the vertices represent the different components $\mathcal{Y}_a = (Y_a(t))_{t\in\mathbb{R}}$, $a \in V = \{1,\ldots,k\}$, of the underlying multivariate process. Furthermore, the vertices are connected by directed and undirected edges, representing concepts of Granger causality and contemporaneous correlation. Both graphs allow us to visualise and communicate the corresponding complex dependency structures between the components of the process $\mathcal{Y}_V$ simply and clearly, which is an important goal of the definition of the graphs.

In principle, it is possible to define a graph with any of the concepts of Granger causality and contemporaneous correlation. However, our goal is to define graphs with concepts that are as strong as necessary, but as weak as possible, so that the usual Markov properties for mixed graphs are satisfied. In addition, for Ornstein-Uhlenbeck processes and even for MCAR processes and ICCSS processes, global Granger causality and Granger causality as well as global contemporaneous correlation and contemporaneous correlation coincide (see Chapters 8 and 9). For these reasons, we do not discuss a global graph and define the causality graph and the local causality graph as follows.

Definition 6.1. Let $\mathcal{Y}_V$ be wide-sense stationary, mean-square continuous, and satisfy Assumptions 3 and 4.

(a) If we define $V = \{1,\ldots,k\}$ as the vertices and the edges E_{CG} via
   (i) $a \longrightarrow b \notin E_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V$,
   (ii) $a \,\text{---}\, b \notin E_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V$,

   for $a, b \in V$, $a \neq b$, then $G_{CG} = (V, E_{CG})$ is called *(mixed) causality graph* for $\mathcal{Y}_V$.

(b) If we define $V = \{1,\ldots,k\}$ as the vertices and the edges E^0_{CG} via
   (i) $a \longrightarrow b \notin E^0_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V$,
   (ii) $a \,\text{---}\, b \notin E^0_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V$,

   for $a, b \in V$, $a \neq b$, then $G^0_{CG} = (V, E^0_{CG})$ is called *local (mixed) causality graph* for $\mathcal{Y}_V$.

In words, in both graphs, each vertex $a \in V$ represents one component $\mathcal{Y}_a$. Two vertices a and b are joined by a directed edge $a \longrightarrow b$, whenever $\mathcal{Y}_a$ is (locally) Granger causal for $\mathcal{Y}_b$ with respect to $\mathcal{Y}_V$, and by an undirected edge $a \,\text{---}\, b$, whenever $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are (locally) contemporaneously correlated given $\mathcal{Y}_V$.

Remark 6.2.

(a) The motivation for the name *causality* graph comes, of course, from the Granger causality relations in the directed edges, and also from the fact that the concept of contemporaneous correlation used for the undirected edges is sometimes called local causality. They are further named *mixed* causality graphs because (local) causality graphs can contain two types of edges. The orientation of the directed edge makes a difference and multiple edges of the same type and orientation are not allowed. Thus, two vertices a and b can be connected by up to three edges, namely $a \longrightarrow b$, $a \longleftarrow b$, and $a \dashrightarrow b$. Since, usually, we do not consider purely directed or undirected graphs in the present part of this thesis, we omit the prefix mixed for ease of notation.

(b) The assumptions are not necessary for the definition of the graphs but are essential for the usual Markov properties for the graphical models to hold (cf. Chapter 7). Wide-sense stationarity and continuity in the mean-square are basic requirements, otherwise, for example, Assumption 3 is not well defined, which is in turn a sufficient assumption for conditional linear separation and relation (4.3). The relation (4.3) is already used for the first time in the proof of the local Markov property. Assumption 4 is only required in the proof of the global AMP Markov property for the causality graph. Since we show global Markov properties for the local causality graph only in special cases, Assumption 4 is not necessary there.

(c) From Proposition 4.17 and 5.10, we already know that $a \longrightarrow b \notin E_{CG}$ implies $a \longrightarrow b \notin E^0_{CG}$ and, similarly, $a \mathrel{-\,-\,-} b \notin E_{CG}$ gives $a \mathrel{-\,-\,-} b \notin E^0_{CG}$. In summary, $E^0_{CG} \subseteq E_{CG}$. The graph defined by the local versions of Granger causality and contemporaneous correlation has fewer edges than the graph G_{CG} based on the classical Granger causality and contemporaneous correlation and, in general, the graphs are not equal. An advantage of the local causality graph G^0_{CG} is that it allows the modelling of more general graphs than the causality graph G_{CG}, see Remark 8.30.

Chapter 7

Markov properties for (local) causality graphs

The (local) causality graph encodes (local) Granger causality and (local) contemporaneous correlation between the different components of the process $\mathcal{Y}_V$. Conversely, the resulting graphs can be associated with a set of conditional orthogonality relations imposed on $\mathcal{Y}_V$, commonly known as Markov properties (cf. Lauritzen, 2004; Whittaker, 2008). These Markov properties allow, for instance, to infer (local) Granger non-causality and (local) contemporaneous uncorrelatedness between multivariate subprocesses by analysing certain paths in the graphical models. They even provide graphical criteria for inferring such relationships between multivariate subprocesses given only partial information. In this way, the graphical models provide additional information about dependency structures in subprocesses, which is a desirable property of graphical models.

In this chapter, we introduce various levels of Markov properties, structured in the following way. We start with the motivation and validity of the pairwise, local, and block-recursive causal Markov property for the (local) causality graph in Section 7.1. We then move on to two global Markov properties, namely the global AMP Markov property and the global Granger-causal Markov property for the causality graph in Section 7.2. Finally, we discuss global Markov properties for the local causality graph in Section 7.3.

7.1. Pairwise, local, and block-recursive Markov property

Let us start with simple Markov properties that we would intuitively expect to be satisfied in any reasonably defined graphical model. First of all, the (local) causality graph visualises the pairwise relationships between the components of a process $\mathcal{Y}_V$ by definition. This is the *pairwise causal Markov property.*

Proposition 7.1.

(a) Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the pairwise causal Markov property with respect to G_{CG}, i.e., for all $a, b \in V$, $a \neq b$,

(i) $a \longrightarrow b \notin E_{CG} \quad \Rightarrow \quad \mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V,$

(ii) $a \dashrightarrow\!\!\!\!\!\!- - - \, b \notin E_{CG} \quad \Rightarrow \quad \mathcal{Y}_a \nsim \mathcal{Y}_b \mid \mathcal{Y}_V.$

(b) Let $G^0_{CG} = (V, E^0_{CG})$ be the local causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the pairwise causal Markov property with respect to G^0_{CG}.

Further, in a mixed graph $G = (V, E)$, we denote by

$$\mathrm{pa}(a) = \{v \in V \mid v \longrightarrow a \in E\} \quad \text{and} \quad \mathrm{ne}(a) = \{v \in V \mid v \dashrightarrow a \in E\}$$

the set of *parents* and the set of *neighbours* of $a \in V$, respectively. If we then consider a vertex $a \in V$ in the causality graph, all vertices $b \in V \setminus (\mathrm{pa}(a) \cup \{a\})$ are Granger non-causal for a due to $b \longrightarrow a \notin E_{CG}$, i.e., $\mathcal{Y}_b \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_a \mid \mathcal{Y}_V$. A direct consequence of relation (4.3) is that $\mathcal{Y}_{V \setminus (\mathrm{pa}(a) \cup \{a\})} \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_a \mid \mathcal{Y}_V$. Similarly, for a vertex $a \in V$, all vertices $b \in V \setminus (\mathrm{ne}(a) \cup \{a\})$ are contemporaneously uncorrelated to a due to $b \dashrightarrow a \notin E_{CG}$, i.e., $\mathcal{Y}_b \nsim \mathcal{Y}_a \mid \mathcal{Y}_V$. Relation (5.1) yields $\mathcal{Y}_{V \setminus (\mathrm{ne}(a) \cup \{a\})} \nsim \mathcal{Y}_a \mid \mathcal{Y}_V$. These two properties are called the *local causal Markov property* for the causality graph. The same reasoning works for the local causality graph, using relations (4.7) and (5.3) respectively, resulting in the following statement.

Proposition 7.2.

(a) Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the local causal Markov property with respect to G_{CG}, i.e., for all $a \in V$,

(i) $\mathcal{Y}_{V \setminus (pa(a) \cup \{a\})} \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_a \mid \mathcal{Y}_V$,

(ii) $\mathcal{Y}_{V \setminus (ne(a) \cup \{a\})} \nsim \mathcal{Y}_a \mid \mathcal{Y}_V$.

(b) Let $G^0_{CG} = (V, E^0_{CG})$ be the local causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the local causal Markov property with respect to G^0_{CG}.

Remark 7.3. We emphasise that we already use wide-sense stationarity, mean-square continuity, and Assumption 3 in the application of the relations (4.3) and (4.7), which justifies that we include the assumptions in the definition of the graph.

Furthermore, in a mixed graph $G = (V, E)$, we denote by

$$\mathrm{pa}(A) = \bigcup_{a \in A} \mathrm{pa}(a) \quad \text{and} \quad \mathrm{ne}(A) = \bigcup_{a \in A} \mathrm{ne}(a)$$

the set of *parents* and *neighbours*, respectively, of vertices in $A \subseteq V$. Again, we expect components that are not parents of A to be Granger non-causal for A. Components that are not neighbours of A should be contemporaneously uncorrelated to A. These two properties are the *block-recursive causal Markov property.* The above reasoning for the local causal Markov property carries over, using again the relations (4.3), (4.7), (5.1), and (5.3).

Proposition 7.4.

(a) Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the block-recursive causal Markov property with respect to G_{CG}, i.e., for all $A \subseteq V$,

(i) $\mathcal{Y}_{V \setminus (pa(A) \cup A)} \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_A \mid \mathcal{Y}_V$,

(ii) $\mathcal{Y}_{V \setminus (ne(A) \cup A)} \nsim \mathcal{Y}_A \mid \mathcal{Y}_V$.

(b) Let $G^0_{CG} = (V, E^0_{CG})$ be the local causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the block-recursive causal Markov property with respect to G^0_{CG}.

Remark 7.5. In the (local) causality graph, the three Markov properties are satisfied because of the assumptions we make. Although we expect these Markov properties to hold in any reasonable graphical model, this is not self-evident. Eichler (2012) therefore proposes to specify mixed graphs that satisfy the block-recursive causal Markov property as *graphical time series models* in his Theorem 2.1 and Definition 2.3.

Finally, to illustrate the various Granger-causal Markov properties for a (local) causality graph, we present Example 2.1 of Eichler (2012), which is constructed from a 5-dimensional VAR(1) process in discrete time. This example can easily be obtained as the local causality graph of an appropriate 5-dimensional Ornstein-Uhlenbeck process (cf. Proposition 8.28). However, likely, it cannot be reproduced as a causality graph (cf. Remark 8.30), so this example must be taken as hypothetical. Since similar causality graphs can be constructed, and because of the illustrative nature of this example, we include it anyway.

Example 7.6. Let $\mathcal{Y}_V$ be a 5-dimensional process in continuous time, whose associated (local) causality graph is given in Figure 7.1.

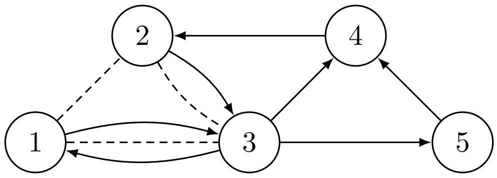


Figure 7.1.: (Local) causality graph for Example 7.6

Since $1 \longrightarrow 2 \notin E_{CG}$ ($1 \longrightarrow 2 \notin E^0_{CG}$) and $2 {-}{-}{-} 4 \notin E_{CG}$ ($2 {-}{-}{-} 4 \notin E^0_{CG}$), the pairwise causal Markov property implies $\mathcal{Y}_1 \not\twoheadrightarrow \mathcal{Y}_2 \,|\, \mathcal{Y}_V$ ($\mathcal{Y}_1 \not\twoheadrightarrow_0 \mathcal{Y}_2 \,|\, \mathcal{Y}_V$) and $\mathcal{Y}_2 \nsim \mathcal{Y}_4 \,|\, \mathcal{Y}_V$ ($\mathcal{Y}_2 \nsim_0 \mathcal{Y}_4 \,|\, \mathcal{Y}_V$). Moreover, using the local causal Markov property, it follows that $\mathcal{Y}_{\{1,3,5\}} \not\twoheadrightarrow \mathcal{Y}_2 \mid \mathcal{Y}_V$ ($\mathcal{Y}_{\{1,3,5\}} \not\twoheadrightarrow_0 \mathcal{Y}_2 \mid \mathcal{Y}_V$), since the vertex 2 has only vertex 4 as a parent. Similarly, $\mathcal{Y}_{\{4,5\}} \nsim \mathcal{Y}_2 \mid \mathcal{Y}_V$ ($\mathcal{Y}_{\{4,5\}} \nsim_0 \mathcal{Y}_2 \mid \mathcal{Y}_V$), since the vertex 2 has both vertices 1 and 3 as neighbours. Finally, the block-recursive causal Markov property provides $\mathcal{Y}_1 \not\twoheadrightarrow \mathcal{Y}_{\{2,4\}} \mid \mathcal{Y}_V$ ($\mathcal{Y}_1 \not\twoheadrightarrow_0 \mathcal{Y}_{\{2,4\}} \mid \mathcal{Y}_V$), since only vertex 1 is neither in the set $\{2, 4\}$ nor a parent of vertex 2 or 4. Analogously, $\mathcal{Y}_{\{4,5\}} \nsim \mathcal{Y}_{\{1,2\}} \mid \mathcal{Y}_V$ ($\mathcal{Y}_{\{4,5\}} \nsim_0 \mathcal{Y}_{\{1,2\}} \mid \mathcal{Y}_V$), since the vertices 4 and 5 are neither in the set $\{1, 2\}$ nor a neighbour of vertex 1 or 2.

7.2. Global Markov properties for causality graphs

The three Markov properties we discussed so far only encode Granger non-causality and contemporaneous uncorrelatedness with respect to $\mathcal{Y}_V$. For a better understanding of the dependency structure of $\mathcal{Y}_V$, we are also interested in relations with respect to partial information, i.e., the dependency structure of subprocesses $\mathcal{Y}_C$, $C \subseteq V$. To obtain such relations from the graphical models, the following approaches are available.

The first approach is based on the notion of *m-separation*, a path-oriented separation concept in mixed graphs. Based on this concept, we discuss the *global AMP Markov property*, which relates m-separation to conditional orthogonality of linear subspaces generated by subprocesses over the entire time span. The second approach is based on the concept of *augmentation separation*, which constructs an undirected graph from the mixed graph by augmentation and then considers classical separation in the resulting undirected graph. This concept leads to the alternative augmentation separation criterion for the global AMP Markov property. Finally, we introduce the *global Granger-causal Markov property*, which provides sufficient criteria for Granger non-causality and contemporaneous uncorrelatedness with respect to partial information. The graphical criteria used are less restrictive than the separation requirements in the global AMP Markov property. Since global Markov properties for the local causality graph require additional assumptions, the results for the local model are presented in the separate Section 7.3.

7.2.1. Global AMP Markov property

We start with the concept of *m-separation* since the augmentation separation approach is not straightforward in the sense that the graph is modified during the test. To define m-separation, we give definitions from graph theory, which can be found in Eichler (2007, 2009, 2012).

Definition 7.7. Let $G = (V, E)$ be a mixed graph and $a, b \in V$. A *path* π between two vertices a and b is a sequence $\pi = \langle e_1, \dots, e_n \rangle$ of edges $e_i \in E$, such that e_i is an edge between v_{i-1} and v_i for some sequence of vertices $v_0 = a, v_1, \dots, v_{n-1}, v_n = b$ in V. We say that a and b are the *endpoints* of the path π, while $v_1, \dots, v_{n-1}$ are *intermediate vertices*. We call n the *length* of π. An intermediate vertex v_i on π is said to be a *collider* on the path if the edges preceding and succeeding v_i on π both have an arrowhead or a dashed tail at v_i, i.e., $\longrightarrow v_i \longleftarrow$, $\longrightarrow v_i \dashrightarrow$ ---, --- $v_i \longleftarrow$, or --- v_i --- . Otherwise, the vertex v_i is said to be a *non-collider* on the path. A path π between a and b is said to be *m-connecting* given a set $C \subseteq V$ if

(a) every non-collider on π is not in C, and

(b) every collider on π is in C.

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

The concept of m-separation is the natural extension of d-separation for directed graphs (Pearl, 1994) and the classical separation for undirected graphs (cf. Definition 7.17) to mixed graphs (Richardson, 2003). It was earlier also called d-separation by Spirtes, Richardson, Meek, Scheines, and Glymour (1998) and Koster (1999). Since we are considering mixed graphs, which are generally not directed, we prefer the notion of m-separation. Note that condition (a) in Definition 7.7 differs from the original definition of m-connecting paths by Richardson (2003), and takes into account that we are considering paths that may intersect themselves, as in Eichler (2007, 2011). Nevertheless, the concept of m-separation in Definition 7.7 is equivalent to the definition of Richardson (2003). In contrast, Eichler (2012) uses another natural extension of d-separation, called p-separation. This concept was introduced by Levitz et al. (2001) for chain graphs and, in contrast to m-separation, --- v_i --- is considered to be a non-collider.

Now we present the main result, the so-called *global AMP Markov property.*

Theorem 7.8. *Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the global AMP Markov property with respect to G_{CG}, i.e., for all disjoint subsets $A, B, C \subseteq V$, we have*

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

In other words, Theorem 7.8 states that if the sets A and B are m-separated given C, the random variables $Y^A \in \mathcal{L}_{Y_A}$ and $Y^B \in \mathcal{L}_{Y_B}$ are uncorrelated after removing all of the information provided by $\mathcal{L}_{Y_C}$.

Remark 7.9. Similar statements can be found, e.g., in Eichler (2001), Theorem 4.8, Eichler (2007), Theorem 3.1, and Eichler (2012), Theorem 4.1. However, the graphs defined in these references are based on discrete-time edge definitions. The definition of the undirected edges in Eichler (2012) further differs from our definition. The linear continuous-time analogue of his definition is $\mathcal{L}_{Y_a}(t, t+1) \perp \mathcal{L}_{Y_b}(t, t+1) \mid \mathcal{L}_{Y_V}(t) \vee \mathcal{L}_{V \setminus \{a,b\}}(t, t+1)$. However, most of the proof steps from the aforementioned literature apply and we give the references where appropriate.

We divide the proof of the global AMP Markov property into four auxiliary lemmata that build on each other and culminate in the proof of Theorem 7.8. We begin with a purely graph-theoretic lemma, for which we introduce the following notations. In a mixed graph $G = (V, E)$,

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

denote the set of *children* and the *district* of $a \in V$. Furthermore, for $A \subseteq V$,

$$\text{ch}(A) = \bigcup_{a \in A} \text{ch}(a) \quad \text{and} \quad \text{dis}(A) = \bigcup_{a \in A} \text{dis}(a)$$

denote the set of *children* and the *district* of $A \subseteq V$, respectively. We further note that the proof of the first lemma does not depend on the specific edge definition in a mixed graph. The proof of Lemma B.1 of Eichler (2007) thus carries over without changes and we get the following result.

Lemma 7.10. *Let $G = (V, E)$ be a mixed graph and $A, B \subseteq V$ be disjoint subsets. Then*

$$A \bowtie_m B \mid V \setminus (A \cup B) \;\; [G] \quad \Rightarrow \quad \mathit{dis}\,(A \cup \mathit{ch}\,(A)) \cap \mathit{dis}\,(B \cup \mathit{ch}(B)) = \emptyset.$$

The second lemma establishes a first relation between m-separation and conditional orthogonality.

Lemma 7.11. *Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Suppose that $A, B \subseteq V$ are disjoint subsets, let $t \in \mathbb{R}$, and $k \in \mathbb{N}$. Then*

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

The proof of Lemma 7.11 is a step-by-step adaptation of the proof of Lemma 4.1 by Eichler (2012). It is an induction over k, using Lemma 7.10 and the properties of a semi-graphoid given in Lemma 3.3. For the reader's convenience, we provide the details.

Proof. For the *base case*, we note that, for $a \in \mathrm{dis}(A)$ and $b \in V \setminus \mathrm{dis}(A)$, $a \dashrightarrow\!\!\!\!\!\!- b \notin E_{CG}$. So $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V$ and relation (5.1) gives $\mathcal{Y}_{\mathrm{dis}(A)} \nsim \mathcal{Y}_{V\setminus\mathrm{dis}(A)} \,|\, \mathcal{Y}_V$. Then, by Lemma 5.2(b),

$$\mathcal{L}_{Y_{\mathrm{dis}(A)}}(t) \perp \mathcal{L}_{Y_{V\setminus\mathrm{dis}(A)}}(t) \,|\, \mathcal{L}_{Y_V}(t-1)$$

for $t \in \mathbb{R}$. In addition, $A \subseteq \mathrm{dis}(A)$ by definition of the district and $B \subseteq V \setminus \mathrm{dis}(A)$ by Lemma 7.10. Hence, using the properties (C3) and (C1) from Lemma 3.3, we obtain

$$\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \,|\, \mathcal{L}_{Y_V}(t-1) \vee \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t).$$

A suitable combination of linear spaces yields, as claimed,

$$\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \,|\, \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-1)$$

for $t \in \mathbb{R}$. Now let us assume the *induction hypothesis*

$$\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \,|\, \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k) \tag{7.1}$$

for $t \in \mathbb{R}$ and $k \in \mathbb{N}$. To obtain the *induction step*, we find that, for vertices $b \in B$ and $a \in V \setminus (B \cup \mathrm{ch}(B))$, we have $b \longrightarrow a \notin E_{CG}$. Therefore, $\mathcal{Y}_b \not\twoheadrightarrow \mathcal{Y}_a \,|\, \mathcal{Y}_V$ and the relation (4.3) yields $\mathcal{Y}_B \not\twoheadrightarrow \mathcal{Y}_{V\setminus(B\cup\mathrm{ch}(B))} \,|\, \mathcal{Y}_V$. Then, by Lemma 4.2(b),

$$\mathcal{L}_{Y_{V\setminus(B\cup\mathrm{ch}(B))}}(t+1) \perp \mathcal{L}_{Y_B}(t) \,|\, \mathcal{L}_{Y_{V\setminus B}}(t)$$

for $t \in \mathbb{R}$. The property (C3) with $\mathcal{L}_{Y_B}(t) = \mathcal{L}_{Y_B}(t) \vee \mathcal{L}_{Y_B}(t-k)$ and a suitable combination of linear spaces yield

$$\mathcal{L}_{Y_{V\setminus(B\cup\mathrm{ch}(B))}}(t+1) \perp \mathcal{L}_{Y_B}(t) \,|\, \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k)$$

for $t \in \mathbb{R}$ and $k \in \mathbb{N}$. The induction hypothesis (7.1), the previous conditional orthogonality, and properties (C4) and (C1) imply

$$\mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_{V\setminus(B\cup\mathrm{ch}(B))}}(t+1) \perp \mathcal{L}_{Y_B}(t) \,|\, \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k)$$

for $t \in \mathbb{R}$ and $k \in \mathbb{N}$. Property (C3) yields the first auxiliary statement

$$\begin{gathered}\mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \perp \mathcal{L}_{Y_B}(t) \,|\\ \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k) \vee \mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))\setminus(B\cup\mathrm{ch}(B))}}(t+1).\end{gathered} \tag{7.2}$$

In addition, Lemma 7.10 states that $\mathrm{ch}(A) \cap \mathrm{dis}(B \cup \mathrm{ch}(B)) = \emptyset$. Therefore, for $a \in A$ and $b \in \mathrm{dis}(B \cup \mathrm{ch}(B))$, we have $a \longrightarrow b \notin E_{CG}$. Hence, $\mathcal{Y}_a \not\twoheadrightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V$ and relation (4.3) yields $\mathcal{Y}_A \not\twoheadrightarrow \mathcal{Y}_{\mathrm{dis}(B\cup\mathrm{ch}(B))} \,|\, \mathcal{Y}_V$. Then, by Lemma 4.2(b),

$$\mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \perp \mathcal{L}_{Y_A}(t) \,|\, \mathcal{L}_{Y_{V\setminus A}}(t)$$

for $t \in \mathbb{R}$. Consequently, property (C3) yields the second auxiliary statement

$$\mathcal{L}_{Y_{B\cup\mathrm{ch}(B)}}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{V\setminus A}}(t) \vee \mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))\setminus(B\cup\mathrm{ch}(B))}}(t+1). \tag{7.3}$$

Moreover, the definition of the district provides $a \dashrightarrow b \notin E_{CG}$ for $a \in \mathrm{dis}(B \cup \mathrm{ch}(B))$ and $b \in V \setminus \mathrm{dis}(B \cup \mathrm{ch}(B))$. Therefore, $\mathcal{Y}_a \nsim \mathcal{Y}_b \mid \mathcal{Y}_V$ and $\mathcal{Y}_{\mathrm{dis}(B\cup\mathrm{ch}(B))} \nsim \mathcal{Y}_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))} \mid \mathcal{Y}_V$ due to relation (5.1). Then, by Lemma 5.2(b),

$$\mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \perp \mathcal{L}_{Y_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \mid \mathcal{L}_{Y_V}(t)$$

for $t \in \mathbb{R}$. Property (C3) and appropriate decomposition of linear spaces yield

$$\mathcal{L}_{Y_{B\cup\mathrm{ch}(B)}}(t+1) \perp \mathcal{L}_{Y_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \mid \mathcal{L}_{Y_{V\setminus A}}(t) \vee \mathcal{L}_{Y_A}(t) \vee \mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))\setminus(B\cup\mathrm{ch}(B))}}(t+1).$$

In conjunction with relation (7.3), this conditional orthogonality and property (C4) provide

$$\mathcal{L}_{Y_{B\cup\mathrm{ch}(B)}}(t+1) \perp \mathcal{L}_{Y_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \vee \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{V\setminus A}}(t) \vee \mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))\setminus(B\cup\mathrm{ch}(B))}}(t+1).$$

By suitable decomposition of linear spaces and property (C3), we obtain

$$\begin{gathered}\mathcal{L}_{Y_{B\cup\mathrm{ch}(B)}}(t+1) \perp \mathcal{L}_{Y_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \vee \mathcal{L}_{Y_A}(t) \mid \\ \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_B}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k) \vee \mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))\setminus(B\cup\mathrm{ch}(B))}}(t+1).\end{gathered}$$

Together with relation (7.2), this conditional orthogonality and property (C4) imply

$$\begin{gathered}\mathcal{L}_{Y_{B\cup\mathrm{ch}(B)}}(t+1) \vee \mathcal{L}_{Y_B}(t) \perp \mathcal{L}_{Y_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \vee \mathcal{L}_{Y_A}(t) \mid \\ \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k) \vee \mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))\setminus(B\cup\mathrm{ch}(B))}}(t+1).\end{gathered}$$

Since $B \subseteq B \cup \mathrm{ch}(B)$ and $A \subseteq V \setminus \mathrm{dis}(B \cup \mathrm{ch}(B))$, we have

$$\begin{gathered}\mathcal{L}_{Y_{B\cup\mathrm{ch}(B)}}(t+1) \perp \mathcal{L}_{Y_{V\setminus\mathrm{dis}(B\cup\mathrm{ch}(B))}}(t+1) \mid \\ \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k) \vee \mathcal{L}_{Y_{\mathrm{dis}(B\cup\mathrm{ch}(B))\setminus(B\cup\mathrm{ch}(B))}}(t+1).\end{gathered}$$

The property (C3), together with $B \subseteq B \cup \mathrm{ch}(B)$ and $A \subseteq V \setminus \mathrm{dis}(B \cup \mathrm{ch}(B))$ again, provides

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t+1) \mid \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k) \vee \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t+1).$$

If we finally summarise the linear spaces in the condition, we obtain

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t+1) \mid \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t+1) \vee \mathcal{L}_{Y_{A\cup B}}(t-k).$$

Since this relation is valid for all $t \in \mathbb{R}$, we can substitute t by $t-1$ to obtain the induction step

$$\mathcal{L}_{Y_B}(t) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k-1)$$

for $t \in \mathbb{R}$ and $k \in \mathbb{N}$. ■

In the third lemma, we eliminate the linear space $\mathcal{L}_{Y_{A\cup B}}(t-k)$ in the condition by considering the transition $k \to \infty$.

Lemma 7.12. *Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Suppose that $A, B \subseteq V$ are disjoint subsets and let $t \in \mathbb{R}$. Then*

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

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

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

due to Lemma 3.15. Additionally, Theorems 4.31(b) and 4.32(c) of Weidmann (1980) yield

$$\underset{k\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_{A\cup B}}(t-k)\vee\mathcal{L}_{Y_{V\setminus(A\cup B)}}(t)} Y = P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}(t)} Y \quad \text{for } Y \in L^2. \tag{7.4}$$

Let $Y^A \in \mathcal{L}_{Y_A}(t) \subseteq L^2$ and $Y^B \in \mathcal{L}_{Y_B}(t) \subseteq L^2$. Then, using the relations (3.1) and (7.4), we obtain

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

The expression on the right-hand side is zero, since $\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t) \vee \mathcal{L}_{Y_{A\cup B}}(t-k)$ for $t \in \mathbb{R}$ and $k \in \mathbb{N}$, due to Lemma 7.11. Thus, the expression on the left-hand side is zero as well and $\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t)$. ■

Remark 7.13. In the proof of Lemma 7.12, we apply Assumption 4, respectively Lemma 3.15, for the first time.

In the last of the four lemmata, we consider the transition $t \to \infty$.

Lemma 7.14. *Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Suppose that $A, B \subseteq V$ are disjoint subsets. Then*

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

Proof. Note that $\mathcal{L}_{Y_S}(n) \subseteq \mathcal{L}_{Y_S}(n+1)$ for $n \in \mathbb{N}$, and Lemma 3.7(d) provides $\overline{\bigcup_{n\in\mathbb{N}} \mathcal{L}_{Y_S}(n)} = \mathcal{L}_{Y_S}$ $\mathbb{P}$-a.s. for any $S \subseteq V$. Then Theorems 4.31(b) and 4.32(b) of Weidmann (1980) imply that, for $Y^A \in \mathcal{L}_{Y_A}$ and $Y^B \in \mathcal{L}_{Y_B}$, we have

$$\begin{aligned}
&\underset{n\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_A}(n)} Y^A = P_{\mathcal{L}_{Y_A}} Y^A = Y^A, &&\underset{n\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_B}(n)} Y^B = P_{\mathcal{L}_{Y_B}} Y^B = Y^B, \\
&\underset{n\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_{V\setminus(A\cup B)}(n)} Y^A = P_{\mathcal{L}_{V\setminus(A\cup B)}} Y^A, &&\underset{n\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_{V\setminus(A\cup B)}(n)} Y^B = P_{\mathcal{L}_{V\setminus(A\cup B)}} Y^B.
\end{aligned} \tag{7.5}$$

Furthermore, triangle inequality, Pythagorean theorem, and the relations (7.5) imply

$$\begin{aligned}
&\left\| P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}(n)} P_{\mathcal{L}_{Y_A}(n)} Y^A - P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}} Y^A \right\|_{L^2} \\
&\leq \left\| P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}(n)} \left(P_{\mathcal{L}_{Y_A}(n)} Y^A - Y^A \right) \right\|_{L^2} + \left\| P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}(n)} Y^A - P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}} Y^A \right\|_{L^2} \\
&\leq \left\| P_{\mathcal{L}_{Y_A}(n)} Y^A - Y^A \right\|_{L^2} + \left\| P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}(n)} Y^A - P_{\mathcal{L}_{Y_{V\setminus(A\cup B)}}} Y^A \right\|_{L^2} \to 0
\end{aligned}$$

as $n \to \infty$. The analogous result holds if A is replaced by B. In other words,

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

Finally, relation (3.1) yields

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

The expression on the right-hand side is zero, since $\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_{V\setminus(A\cup B)}}(t)$ for $t \in \mathbb{R}$, due to Lemma 7.12. Thus, the left-hand side is zero as well and $\mathcal{L}_{Y_A} \perp \mathcal{L}_{Y_B} \mid \mathcal{L}_{Y_{V\setminus(A\cup B)}}$. ■

For the final proof of Theorem 7.8, we establish additional notations. In a mixed graph $G = (V, E)$,

$$\text{an}(a) = \{v \in V | v \longrightarrow \cdots \longrightarrow a \text{ or } v = a\} \quad \text{and} \quad \text{an}(A) = \bigcup_{a\in A} \text{an}(a)$$

denote the set of *ancestors* of $a \in V$ and $A \subseteq V$, respectively. Further, for $A \subseteq V$, $G_A = (A, E_A)$ denotes the *induced subgraph* of the mixed graph, where E_A contains all edges $e \in E$ that have both endpoints in A. Then the proof of Theorem 3.1 of Eichler (2007) carries over to continuous time, we include the details so that the proof of the global AMP Markov property is complete.

Proof of Theorem 7.8. By Corollary 1 and Proposition 2 of Koster (1999), we have

$$A \bowtie_m B \mid C \;\; [G_{CG}] \quad \Leftrightarrow \quad A \bowtie_m B \mid C \;\; [G_{CG,\text{an}(A\cup B\cup C)}] \quad \Leftrightarrow \quad A' \bowtie_m B' \mid C \;\; [G_{CG,\text{an}(A\cup B\cup C)}]$$

for disjoint subsets A' and B' with $A \subseteq A'$, $B \subseteq B'$, and $A' \cup B' \cup C = \text{an}(A\cup B\cup C)$. Lemma 7.14 yields $\mathcal{L}_{Y_{A'}} \perp \mathcal{L}_{Y_{B'}} \mid \mathcal{L}_{Y_C}$ and an application of the properties (C1) and (C2) from Lemma 3.3 implies $\mathcal{L}_{Y_A} \perp \mathcal{L}_{Y_B} \mid \mathcal{L}_{Y_C}$. ■

We include a visualisation of the m-separation criterion and the global AMP Markov property, continuing with our Example 7.6 and Example 3.2 by Eichler (2007), respectively.

Example 7.15. First, consider the path $1 \longrightarrow 3 \text{ - - - } 2 \longleftarrow 4 \longleftarrow 3 \longrightarrow 5$ in the causality graph. This path is m-blocked given $\{3,4\}$ since, for example, 2 is a collider but not in $\{3,4\}$. Alternatively, 4 is a non-collider but in $\{3,4\}$. In fact, every path between 1 and 5 is m-blocked given $\{3,4\}$, which we explain as follows. Every path between 1 and 5 contains either $3 \longrightarrow 5$, $3 \longrightarrow 4 \longleftarrow 5$, or $2 \longleftarrow 4 \longleftarrow 5$. The first two paths are m-blocked by vertex 3, which is a non-collider but in $\{3,4\}$. The last path is blocked by vertex 4, which is also a non-collider but in $\{3,4\}$. So $\{1\} \bowtie_m \{5\} \mid \{3,4\}$ applies and the global AMP Markov property provides $\mathcal{L}_{Y_1} \perp \mathcal{L}_{Y_5} \mid \mathcal{L}_{Y_{\{3,4\}}}$.

To establish that two sets A and B are m-separated given C, one must show that there is no m-connecting path between A and B given C. Since paths may be self-intersecting, the number of paths between A and B may be infinite. Although the search for m-connecting paths can be restricted to paths where no edges occur twice with the same orientation, an algorithmic implementation of such a search does not seem to be straightforward (cf. Eichler, 2011). This is one reason to study the *augmentation separation* approach. Additionally, this approach has the advantage of allowing us to establish connections to the (undirected) partial correlation graph in Section 11.2. In the following, we apply the concept of augmentation separation to the causality graph. For further applications of the augmentation separation approach, see Eichler (2007, 2011) and Richardson (2003), where the following definitions are provided.

Definition 7.16. Let $G = (V, E)$ be a mixed graph and $a, b \in V$. Two vertices a and b are said to be *collider connected* if they are connected by a *pure collider path*, which is a path on which every intermediate vertex is a collider. Then the undirected *augmented graph* $G^a = (V, E^a)$ is derived from a mixed graph $G = (V, E)$ via

$$a \longrightarrow b \notin E^a \quad \Leftrightarrow \quad a \text{ and } b \text{ are not collider connected in } G.$$

Note that every single edge is trivially considered to be a collider path. Thus, every directed and undirected edge in the causality graph corresponds to an undirected edge in the augmented causality graph, and the augmented causality graph generally has more edges than the causality graph. We further define the concept of separation in undirected graphs as follows.

Definition 7.17. Let $G = (V, E)$ be an undirected graph. For $A, B, C \subseteq V$, we say that C *separates* A and B if every path from a vertex $a \in A$ to a vertex $b \in B$ contains at least one vertex from the separating set C. We write $A \bowtie B \mid C$ for short.

Based on these definitions, we obtain an alternative form of the global AMP Markov property. This result follows from Theorem 7.8 and Eichler (2011), who, in Theorem 3.1, states in general mixed graphs the equivalence

$$A \bowtie_m B \mid C \;\; [G] \quad \Leftrightarrow \quad A \bowtie B \mid C \;\; [(G_{\mathrm{an}(A \cup B \cup C)})^a].$$

Proposition 7.18. *Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Then, for all disjoint subsets $A, B, C \subseteq V$, we have*

$$A \bowtie B \mid C \;\; [(G_{CG,an(A \cup B \cup C)})^a] \quad \Rightarrow \quad \mathcal{L}_{Y_A} \perp \mathcal{L}_{Y_B} \mid \mathcal{L}_{Y_C}.$$

As usual, we give an example for visualisation purposes and continue with our Example 7.6 analogous to Example 3.5 provided by Eichler (2007).

Example 7.19. In Figure 7.1, we find that the vertices 1 and 4 are connected by the pure collider path $1 \dashrightarrow\!\!\!\! - 2 \longleftarrow 4$. Hence, the vertices 1 and 4 are connected by an additional edge in the augmented causality graph $(G_{CG})^a$. There are no other edges that need to be added, so $(G_{CG})^a$ is given in Figure 7.2.

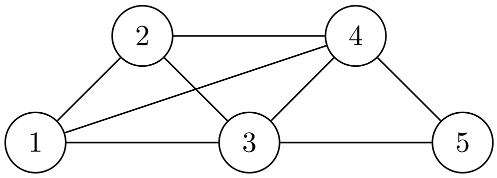


Figure 7.2.: Augmented causality graph for Example 7.19

Since $\mathrm{an}(\{1,3,4,5\}) = \{1,2,3,4,5\} = V$ in the original mixed graph and $1 \bowtie 5 \mid \{3,4\}\ [(G_{CG})^a]$ in the undirected augmented graph, we obtain that $\mathcal{L}_{Y_1} \perp \mathcal{L}_{Y_5} \mid \mathcal{L}_{Y_{\{3,4\}}}$ due to Proposition 7.18. This relation is, of course, the same conditional orthogonality relation as in Example 7.15.

7.2.2. Global Granger-causal Markov property

In this section, we derive sufficient graphical conditions for Granger non-causality and for subprocesses to be contemporaneously uncorrelated. An obvious idea is to start again with m-separation as in the global AMP Markov property. However, this condition is much stronger than necessary. In fact, it is sufficient to require that only those paths are m-blocked, that point in the "right" direction. A motivating example to justify this demand is given by Eichler (2007) on p. 341. Below, we introduce graph-theoretic notions and then present the so-called *global Granger-causal Markov property.*

Definition 7.20. Let $G = (V, E)$ be a mixed graph and $a, b \in V$. A path π between vertices a and b is called *b-pointing* if it has an arrowhead at the endpoint b. More generally, a path π between subsets A and B, $A, B \subseteq V$, is said to be *B-pointing* if it is b-pointing for some $b \in B$. Furthermore, a path π between vertices a and b is said to be *bi-pointing* if it has an arrowhead at both endpoints a and b.

Theorem 7.21. *Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies the global Granger-causal Markov property with respect to G_{CG}, i.e., for all disjoint subsets $A, B, C \subseteq V$, the following conditions hold.*

(a) If every B-pointing path in G_{CG} between A and B is m-blocked given $B \cup C$, then

$$\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_{A \cup B \cup C}.$$

(b) If $a \dashrightarrow\!\!\!\! - b \notin E_{CG}$ for all $a \in A$ and $b \in B$, and if every bi-pointing path in G_{CG} between A and B is m-blocked given $A \cup B \cup C$, then

$$\mathcal{Y}_A \not\sim \mathcal{Y}_B \mid \mathcal{Y}_{A \cup B \cup C}.$$

A similar result in discrete time can be found in Eichler (2007), Theorems 4.1 and 4.2, and Eichler (2012), Theorem 4.2. The proof of Theorem 7.21 is, as expected, similar to the discrete-time proof in Eichler (2007, 2012). It is based on the properties of a semi-graphoid (cf. Lemma 3.3), the block-recursive causal Markov property (cf. Proposition 7.4), and on the auxiliary Lemma 7.12. As usual, we include the details for completeness.

Proof. As a preliminary consideration, we note that by Lemma 7.12, analogously to the proof of Theorem 7.8 (cf. Eichler, 2007, Theorem 3.1), it follows that, for disjoint subsets $A, B, C \subseteq V$,

$$A \bowtie_m B \mid C \;\; [G_{CG}] \quad \Rightarrow \quad \mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_B}(t) \mid \mathcal{L}_{Y_C}(t) \tag{7.6}$$

for $t \in \mathbb{R}$. The proof of *statement (a)* is divided into two steps.

Step 1: We show by contradiction that if every B-pointing path in G_{CG} between A and B is m-blocked given $B \cup C$, then it follows that $A \bowtie_m \mathrm{pa}(B) \setminus (B \cup C) \mid B \cup C \;\; [G_{CG}]$. So suppose $A \not\bowtie_m \mathrm{pa}(B) \setminus (B \cup C) \mid B \cup C \;\; [G_{CG}]$. Then there are vertices $a \in A$ and $u \in \mathrm{pa}(B) \setminus (B \cup C)$, as well as an m-connecting path π given $B \cup C$ between a and u. Since $u \in \mathrm{pa}(B)$, there is a vertex $b \in B$, such that $u \longrightarrow b \in E_{CG}$. We connect the path π and the edge $u \longrightarrow b$ to a new path $\widetilde{\pi}$ from a to b. The path $\widetilde{\pi}$ is m-connecting given $B \cup C$, because, on the one hand, π is m-connecting given $B \cup C$. On the other hand, the path $\widetilde{\pi}$ has only one new intermediate vertex u, which is a non-collider since the succeeding edge is of the form $u \longrightarrow b$. In addition, $u \notin B \cup C$. In summary, $\widetilde{\pi}$ is a B-pointing path between A and B and m-connecting given $B \cup C$, which contradicts the premise in statement (a).

Step 2: Based on the preliminary consideration (7.6) and Step 1, we prove statement (a). First, $A \cap (\mathrm{pa}(B) \setminus (B \cup C)) = \emptyset$ respectively $A \cap \mathrm{pa}(B) = \emptyset$, else there are vertices $a \in A$ and $b \in B$ such that $a \longrightarrow b \in E_{CG}$ is a B-pointing path between A and B and m-connecting given $B \cup C$. Then, since by Step 1 $A \bowtie_m \mathrm{pa}(B) \setminus (B \cup C) \mid B \cup C \;\; [G_{CG}]$, and the subsets A, $\mathrm{pa}(B) \setminus (B \cup C)$, and $B \cup C$ are disjoint, the relation (7.6) provides

$$\mathcal{L}_{Y_A}(t) \perp \mathcal{L}_{Y_{\mathrm{pa}(B) \setminus (B \cup C)}}(t) \mid \mathcal{L}_{Y_{B \cup C}}(t) \tag{7.7}$$

for $t \in \mathbb{R}$. Further, Proposition 7.4 states $\mathcal{Y}_{V \setminus (\mathrm{pa}(B) \cup B)} \not\twoheadrightarrow \mathcal{Y}_B \mid \mathcal{Y}_V$. By Lemma 4.2(b), that is

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_{V \setminus (\mathrm{pa}(B) \cup B)}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(B) \cup B}}(t).$$

Using the property (C3) from Lemma 3.3, we obtain

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_{V \setminus (\mathrm{pa}(B) \cup B \cup C)}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(B) \cup B \cup C}}(t).$$

Then the property (C2), $A \cap (\mathrm{pa}(B) \cup B \cup C) = \emptyset$, and a suitable combination of linear spaces yield

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{B \cup C}}(t) \vee \mathcal{L}_{Y_{\mathrm{pa}(B) \setminus (B \cup C)}}(t). \tag{7.8}$$

We bring the relations (7.7) and (7.8) together with the property (C4), to obtain

$$\mathcal{L}_{Y_{\mathrm{pa}(B)\setminus(B\cup C)}}(t) \vee \mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{B\cup C}}(t).$$

Finally, the property (C2) yields

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_A}(t) \mid \mathcal{L}_{Y_{B\cup C}}(t)$$

for $t \in \mathbb{R}$. By Lemma 4.2(b), this is $\mathcal{Y}_A \not\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup C}$.

The proof of *statement (b)* is divided into three steps and, for better readability, we abbreviate $S = A \cup B \cup C$.

Step 1: We prove that $\mathrm{pa}(B) \setminus S \subseteq V \setminus (\mathrm{pa}(A) \cup S)$ and by symmetry $\mathrm{pa}(A) \setminus S \subseteq V \setminus (\mathrm{pa}(B) \cup S)$, i.e., we show by contradiction that

$$(\mathrm{pa}(B) \setminus S) \cap (\mathrm{pa}(A) \cup S) = (\mathrm{pa}(B) \cap \mathrm{pa}(A)) \setminus S = \emptyset.$$

So suppose this intersection is not empty. Then there is a vertex $u \in (\mathrm{pa}(B) \cap \mathrm{pa}(A)) \setminus S$ and we find vertices $a \in A$ and $b \in B$ together with a path $a \longleftarrow u \longrightarrow b$. This path is m-connecting given S since u is a non-collider and $u \notin S$. At the same time, this path is bi-pointing. Thus, the path contradicts the premise in statement (b).

Step 2: We prove that $\mathrm{pa}(A) \setminus S \bowtie_m \mathrm{pa}(B) \setminus S \mid S\ [G_{CG}]$ by contradiction. Therefore, suppose that $\mathrm{pa}(A) \setminus S \not\bowtie_m \mathrm{pa}(B) \setminus S \mid S\ [G_{CG}]$. Then there are vertices $u \in \mathrm{pa}(A) \setminus S$ and $v \in \mathrm{pa}(B) \setminus S$, as well as an m-connecting path π given S between u and v. Since $u \in \mathrm{pa}(A)$, there is a vertex $a \in A$ and an edge $a \longleftarrow u$. Analogously, there is a vertex $b \in B$ and an edge $v \longrightarrow b$, since $v \in \mathrm{pa}(B)$. We complete the path π to a new path $\widetilde{\pi}$ from a to b with these two edges. Obviously, $\widetilde{\pi}$ is bi-pointing. Furthermore, $\widetilde{\pi}$ is m-connecting given S, since π is m-connecting given S, u and v are non-colliders, and $u, v \notin S$. This result contradicts the premise in statement (b).

Step 3: Based on Steps 1 and 2, as well as the preliminary consideration (7.6), we prove statement (b). On assumption, $a \dashrightarrow b \notin E_{CG}$, i.e., we have $\mathcal{Y}_a \nsim \mathcal{Y}_b \mid \mathcal{Y}_V$ for $a \in A$ and $b \in B$. Relation (5.1) yields $\mathcal{Y}_A \nsim \mathcal{Y}_B \mid \mathcal{Y}_V$ and Lemma 5.2(b) implies

$$\mathcal{L}_{Y_A}(t+1) \perp \mathcal{L}_{Y_B}(t+1) \mid \mathcal{L}_{Y_V}(t)$$

for $t \in \mathbb{R}$. Furthermore, Proposition 7.4 states that $\mathcal{Y}_{V\setminus(\mathrm{pa}(A)\cup A)} \not\rightarrow \mathcal{Y}_A \mid \mathcal{Y}_V$. By Lemma 4.2(b), that is

$$\mathcal{L}_{Y_A}(t+1) \perp \mathcal{L}_{Y_{V\setminus(\mathrm{pa}(A)\cup A)}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(A)\cup A}}(t)$$

for $t \in \mathbb{R}$. The two previous conditional orthogonality relations, the property (C4), and a suitable combination of linear spaces give

$$\mathcal{L}_{Y_A}(t+1) \perp \mathcal{L}_{Y_B}(t+1) \vee \mathcal{L}_{Y_{V\setminus(\mathrm{pa}(A)\cup A)}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(A)\cup A}}(t).$$

Then the property (C3) yields

$$\mathcal{L}_{Y_A}(t+1) \perp \mathcal{L}_{Y_B}(t+1) \vee \mathcal{L}_{Y_{V\setminus(\mathrm{pa}(A)\cup S)}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(A)\cup S}}(t).$$

Since by Step 1 $\mathrm{pa}(B) \setminus S \subseteq V \setminus (\mathrm{pa}(A) \cup S)$, the property (C2) implies

$$\mathcal{L}_{Y_A}(t+1) \perp \mathcal{L}_{Y_B}(t+1) \vee \mathcal{L}_{Y_{\mathrm{pa}(B)\setminus S}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(A)\cup S}}(t). \tag{7.9}$$

In addition, Step 2 yields $\mathrm{pa}(A) \setminus S \bowtie_m \mathrm{pa}(B) \setminus S \mid S\ [G_{CG}]$, where $\mathrm{pa}(A) \setminus S$, $\mathrm{pa}(B) \setminus S$, and S are disjoint. Thus, the relation (7.6) provides

$$\mathcal{L}_{Y_{\mathrm{pa}(A)\setminus S}}(t) \perp \mathcal{L}_{Y_{\mathrm{pa}(B)\setminus S}}(t) \mid \mathcal{L}_{Y_S}(t) \tag{7.10}$$

for $t \in \mathbb{R}$. Again, Proposition 7.4 states that $\mathcal{Y}_{V\setminus(\mathrm{pa}(B)\cup B)} \not\rightarrow \mathcal{Y}_B \,|\, \mathcal{Y}_V$. That is, by Lemma 4.2(b),

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_{V\setminus(\mathrm{pa}(B)\cup B)}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(B)\cup B}}(t)$$

for $t \in \mathbb{R}$. The property (C3) gives

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_{V\setminus(\mathrm{pa}(B)\cup S)}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(B)\cup S}}(t).$$

Step 1, i.e., $\mathrm{pa}(A) \setminus S \subseteq V \setminus (\mathrm{pa}(B) \cup S)$ and the property (C2) yield

$$\mathcal{L}_{Y_B}(t+1) \perp \mathcal{L}_{Y_{\mathrm{pa}(A)\setminus S}}(t) \mid \mathcal{L}_{Y_{\mathrm{pa}(B)\cup S}}(t).$$

In conjunction with relation (7.10), property (C4) gives

$$\mathcal{L}_{Y_{\mathrm{pa}(A)\setminus S}}(t) \perp \mathcal{L}_{Y_B}(t+1) \vee \mathcal{L}_{Y_{\mathrm{pa}(B)\setminus S}}(t) \mid \mathcal{L}_{Y_S}(t).$$

Then, together with relation (7.9), the property (C3) yields

$$\mathcal{L}_{Y_A}(t+1) \vee \mathcal{L}_{Y_{\mathrm{pa}(A)\setminus S}}(t) \perp \mathcal{L}_{Y_B}(t+1) \vee \mathcal{L}_{Y_{\mathrm{pa}(B)\setminus S}}(t) \mid \mathcal{L}_{Y_S}(t).$$

Finally, the property (C2) gives

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

for $t \in \mathbb{R}$. By Lemma 5.2(b), this statement is $\mathcal{Y}_A \nsim \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup C}$. ■

As a consequence of the global Granger-causal Markov property, we find that the m-separation $A \bowtie_m B \,|\, C\,[G_{CG}]$ is indeed quite strong, implying Granger non-causality between $\mathcal{Y}_A$ and $\mathcal{Y}_B$ in both directions, as well as $\mathcal{Y}_A$ and $\mathcal{Y}_B$ being contemporaneously uncorrelated given $\mathcal{Y}_{A\cup B\cup C}$.

Corollary 7.22. *Let $G_{CG} = (V, E_{CG})$ be the causality graph for $\mathcal{Y}_V$ and let $A, B, C \subseteq V$ be disjoint subsets. Then*

$$A \bowtie_m B \mid C\,[G_{CG}] \quad \Rightarrow \quad \mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup C}, \quad \mathcal{Y}_B \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_A \mid \mathcal{Y}_{A\cup B\cup C}, \quad \mathcal{Y}_A \nsim \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup C}.$$

Proof. We refer to Eichler (2012), Corollary 4.1, and Eichler (2007), Corrolary 4.3, for the proof of $\mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup C}$ and by symmetry $\mathcal{Y}_B \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_A \mid \mathcal{Y}_{A\cup B\cup C}$. We include the proof of $\mathcal{Y}_A \nsim \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup C}$, which is not carried out in these references but is based on the same ideas. Suppose that $A \bowtie_m B \mid C$, i.e., every path between A and B is m-blocked given C. First, $a \dashrightarrow b \notin E_{CG}$ for $a \in A$ and $b \in B$, otherwise a --- b is an m-connecting path given C between A and B. Second, in particular, every bi-pointing path π between A and B is m-blocked given C. If π contains no intermediate vertices in $A \cup B$, then π is also m-blocked given $A \cup B \cup C$. If π contains intermediate vertices in $A \cup B$, then it can be partitioned as $\pi = \langle \pi_1, \pi_2 \rangle$, where π_1 is a path between A and some $b \in B$ with no intermediate points in B, i.e., b is the first vertex on the path π that is in B. Further, π_1 can be decomposed as $\pi_1 = \langle \pi_{11}, \pi_{12} \rangle$, where π_{12} is a path between some $a \in A$ and b with no intermediate points in A, i.e., a is the last vertex on π_1 that is in A. Because of the assumption, π_{12} is m-blocked given C. Since it has no intermediate vertices in $A \cup B$, it is also m-blocked given $A \cup B \cup C$. Then π is m-blocked given $A \cup B \cup C$ and Theorem 7.21 implies $\mathcal{Y}_A \nsim \mathcal{Y}_B \mid \mathcal{Y}_{A\cup B\cup C}$. ■

To visualise the global Granger-causal Markov property, we continue with Example 7.6, see also Example 4.2 by Eichler (2012).

Example 7.23. In Example 7.15, we find that $\{1\} \bowtie_m \{5\} \mid \{3,4\}$. Therefore, in the causality graph, Corollary 7.22 gives $\mathcal{Y}_1 \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_5 \mid \mathcal{Y}_{\{1,3,4,5\}}$, $\mathcal{Y}_5 \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_1 \mid \mathcal{Y}_{\{1,3,4,5\}}$, and $\mathcal{Y}_1 \nsim \mathcal{Y}_5 \mid \mathcal{Y}_{\{1,3,4,5\}}$. We further analyse the relations between $\mathcal{Y}_1$ and $\mathcal{Y}_4$ with respect to $\mathcal{Y}_{\{1,3,4\}}$. We first notice that $\{1\}$ and $\{4\}$ are not m-separated given $\{3\}$, since the path $1 \longrightarrow 3 \longleftarrow 2 \longleftarrow 4$ is m-connecting given $\{3\}$. Therefore, we cannot use Corollary 7.22. We have to apply the global Granger-causal Markov property from Theorem 7.21. If we look at the 1-pointing paths between the vertices 4 and 1, we find that a path of this form always ends with a directed edge $3 \longrightarrow 1$. Since in this case 3 is a non-collider, such a path is always m-blocked given $\{1,3\}$. Hence, $\mathcal{Y}_4 \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_1 \mid \mathcal{Y}_{\{1,3,4\}}$. Similarly, we can analyse all bi-pointing paths between vertices 1 and 4. Again, a path of this form always ends with a directed edge $3 \longrightarrow 1$ at vertex 1. Since vertex 3 is a non-collider, such a path is always m-blocked given $\{1,3,4\}$. Furthermore, we have 1 --- $4 \notin E_{CG}$. Hence, $\mathcal{Y}_1 \nsim \mathcal{Y}_4 \mid \mathcal{Y}_{\{1,3,4\}}$. Finally, all 4-pointing paths end with either $3 \longrightarrow 4$, $3 \longrightarrow 5 \longrightarrow 4$, or $2 \longleftarrow 4 \longleftarrow 5 \longrightarrow 4$. These paths are m-blocked given $\{3,4\}$, since the first two are m-blocked by the non-collider 3 and the last one is m-blocked by the non-collider 4. Hence, $\mathcal{Y}_1 \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_4 \mid \mathcal{Y}_{\{1,3,4\}}$.

Remark 7.24. We note that the concept of m-separation is connected to Granger non-causality up to horizon h and global Granger non-causality, which shows that causality graphs also provide a natural framework for discussing indirect effects. Replacing the definitions of Eichler (2007) with our definitions, the proof is analogous to that of his Theorem 4.5, so we skip the details. It applies that, for disjoint $A, B, C \subseteq V$ and $\kappa \in \mathbb{N}$,

$$A \cap \mathrm{an}_\kappa(B) = \emptyset, \quad A \bowtie_m \mathrm{an}_\kappa(B) \setminus (B \cup C) \mid B \cup C \, [G_{CG}] \quad \Rightarrow \quad \mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow_\kappa \mathcal{Y}_B \mid \mathcal{Y}_{A \cup B \cup C},$$
$$A \cap \mathrm{an}(B) = \emptyset, \quad A \bowtie_m \mathrm{an}(B) \setminus (B \cup C) \mid B \cup C \, [G_{CG}] \quad \Rightarrow \quad \mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow_\infty \mathcal{Y}_B \mid \mathcal{Y}_{A \cup B \cup C}.$$

In these relations, $\mathrm{an}_\kappa(A)$ denotes the set of ancestors of degree $\kappa \in \mathbb{N}$ of A, which are given by vertices $v \in V$ such that $v \in A$ or v is connected to some $a \in A$ by a path $v \longrightarrow \cdots \longrightarrow a$ of length less than or equal to κ. We set $\mathrm{an}_0(A) = A$.

As a summary of this section, we highlight the main message: We define the causality graph to illustrate pairwise relationships between variables. However, the associated causality graph can be analysed using various levels of Markov properties. These properties allow us to infer new Granger non-causality relations and knowledge about subprocesses being contemporaneously uncorrelated, even with respect to partial information.

7.3. Global Markov properties for local causality graphs

For the local causality graph, global Markov properties are, as expected, much more difficult to derive due to the weaker definition of the edges. However, at least under additional assumptions, the property of m-separation implies local Granger non-causality in both directions and local contemporaneous uncorrelatedness. We start with the first special case, where $C = V \setminus (A \cup B)$.

Proposition 7.25. *Let $G^0_{CG} = (V, E^0_{CG})$ be the local causality graph for $\mathcal{Y}_V$ and let $A, B \subseteq V$ be disjoint subsets. Then*

$$A \bowtie_m B \mid V \setminus (A \cup B) \, [G^0_{CG}] \quad \Rightarrow \quad \mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_V, \quad \mathcal{Y}_B \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_A \mid \mathcal{Y}_V, \quad \mathcal{Y}_A \not\sim_0 \mathcal{Y}_B \mid \mathcal{Y}_V.$$

Proof. Due to Lemma 7.10, the m-separation implies that $\mathrm{ch}(A) \cap B = \emptyset$, $A \cap \mathrm{ch}(B) = \emptyset$, and $\mathrm{ne}(A) \cap B = \emptyset$. Using the relations (4.7) and (5.3), it follows that $\mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_V$, $\mathcal{Y}_B \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_A \mid \mathcal{Y}_V$, and $\mathcal{Y}_A \not\sim_0 \mathcal{Y}_B \mid \mathcal{Y}_V$. ■

In the second special case, we assume that $\mathrm{pa}(A) \cup \mathrm{pa}(B) \subseteq A \cup B \cup C$. Then, the block-recursive Markov property and the left decomposition property result in the desired relations.

Proposition 7.26. *Let $G^0_{CG} = (V, E^0_{CG})$ be the local causality graph for $\mathcal{Y}_V$ and let $A, B, C \subseteq V$ be disjoint subsets. Suppose that $pa(A) \cup pa(B) \subseteq A \cup B \cup C$. Then*

$$A \bowtie_m B \mid C \, [G^0_{CG}] \quad \Rightarrow \quad \mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_{A \cup B \cup C}, \quad \mathcal{Y}_B \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_A \mid \mathcal{Y}_{A \cup B \cup C}, \quad \mathcal{Y}_A \not\sim_0 \mathcal{Y}_B \mid \mathcal{Y}_{A \cup B \cup C}.$$

Proof. First of all, Proposition 7.4 states that $\mathcal{Y}_{V \setminus (B \cup \mathrm{pa}(B))} \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_V$. By assumption, we have $B \cup \mathrm{pa}(B) \subseteq A \cup B \cup C$. Furthermore, $A \cap \mathrm{pa}(B) = \emptyset$. Otherwise, there are vertices $a \in A$ and $b \in B$ such that $a \longrightarrow b \in E^0_{CG}$ is an m-connecting path between A and B given C. This path is a contradiction to $A \bowtie_m B \mid C \, [G^0_{CG}]$. Thus, $B \cup \mathrm{pa}(B) \subseteq B \cup C$ and relation (4.7) yields $\mathcal{Y}_{V \setminus (B \cup C)} \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_V$. Then the left decomposition property (cf. Lemma 4.12) gives $\mathcal{Y}_A \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_B \mid \mathcal{Y}_{A \cup B \cup C}$. By symmetry of m-separation, $\mathcal{Y}_B \not\rightarrow\!\!\!\!\rightarrow_0 \mathcal{Y}_A \mid \mathcal{Y}_{A \cup B \cup C}$ also follows.

It remains to show that $\mathcal{Y}_A \nsim_0 \mathcal{Y}_B \,|\, \mathcal{Y}_{A\cup B\cup C}$. Proposition 7.4 provides $\mathcal{Y}_{V\setminus(B\cup\mathrm{ne}(B))} \nsim_0 \mathcal{Y}_B \,|\, \mathcal{Y}_V$. Note that $A \cap \mathrm{ne}(B) = \emptyset$. Else there are vertices $a \in A$ and $b \in B$ such that $a \,\text{---}\, b \in E^0_{CG}$ is an m-connecting path between A and B given C. This path is again a contradiction to $A \bowtie_m B \mid C \; [G^0_{CG}]$. So relation (5.3) yields $\mathcal{Y}_A \nsim_0 \mathcal{Y}_B \,|\, \mathcal{Y}_V$. By Definition 5.5 and due to Remark 2.9, we get

$$
\begin{aligned}
0 &= \lim_{h\to 0} \frac{1}{h}\, \mathbb{E}\Big[\Big(D^{(j_a)} Y_a(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(j_a)} Y_a(t+h) \Big) \\
&\qquad \cdot \overline{\Big(D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(j_b)} Y_b(t+h) \Big)} \Big] \\
&= \lim_{h\to 0} \mathbb{E}\Bigg[\Bigg(\frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} - P_{\mathcal{L}_{Y_V}(t)} \Bigg(\frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \Bigg) \Bigg) \\
&\qquad \cdot \overline{\Big(D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(j_b)} Y_b(t+h) \Big)} \Bigg]
\end{aligned}
\tag{7.11}
$$

for $t \in \mathbb{R}$, $a \in A$, and $b \in B$. Due to Proposition 7.4 and $A \cup \mathrm{pa}(A) \subseteq A \cup B \cup C$, we receive, as in the first part of this proof, $\mathcal{Y}_{V\setminus(A\cup B\cup C)} \not\twoheadrightarrow_0 \mathcal{Y}_A \,|\, \mathcal{Y}_V$, which means that

$$
\begin{aligned}
&\underset{h\to 0}{\mathrm{l.i.m.}}\; P_{\mathcal{L}_{Y_V}(t)} \Bigg(\frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \Bigg) \\
&\quad = \underset{h\to 0}{\mathrm{l.i.m.}}\; P_{\mathcal{L}_{Y_{A\cup B\cup C}}(t)} \Bigg(\frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \Bigg) \qquad \mathbb{P}\text{-a.s.},
\end{aligned}
$$

for $t \in \mathbb{R}$ and $a \in A$. Relation (7.11) and similar arguments as in the proof of Proposition 4.9 imply

$$
\begin{aligned}
0 &= \lim_{h\to 0} \mathbb{E}\Bigg[\Bigg(\frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} - P_{\mathcal{L}_{Y_{A\cup B\cup C}}(t)} \Bigg(\frac{D^{(j_a)} Y_a(t+h) - D^{(j_a)} Y_a(t)}{h} \Bigg) \Bigg) \\
&\qquad \cdot \overline{\Big(D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(j_b)} Y_b(t+h) \Big)} \Bigg] \\
&= \lim_{h\to 0} \frac{1}{h}\, \mathbb{E}\Big[\Big(D^{(j_a)} Y_a(t+h) - P_{\mathcal{L}_{Y_{A\cup B\cup C}}(t)} D^{(j_a)} Y_a(t+h) \Big) \\
&\qquad \cdot \overline{\Big(D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(j_b)} Y_b(t+h) \Big)} \Big]
\end{aligned}
$$

for $t \in \mathbb{R}$, $a \in A$, and $b \in B$. In a second step, Proposition 7.4 and $B \cup \mathrm{pa}(B) \subseteq A \cup B \cup C$ imply $\mathcal{Y}_{V\setminus(A\cup B\cup C)} \not\twoheadrightarrow_0 \mathcal{Y}_B \,|\, \mathcal{Y}_V$, which analogously yields

$$
\begin{aligned}
0 = \lim_{h\to 0} \frac{1}{h}\, \mathbb{E}\Big[&\Big(D^{(j_a)} Y_a(t+h) - P_{\mathcal{L}_{Y_{A\cup B\cup C}}(t)} D^{(j_a)} Y_a(t+h) \Big) \\
&\cdot \overline{\Big(D^{(j_b)} Y_b(t+h) - P_{\mathcal{L}_{Y_{A\cup B\cup C}}(t)} D^{(j_b)} Y_b(t+h) \Big)} \Big]
\end{aligned}
$$

for $t \in \mathbb{R}$, $a \in A$, and $b \in B$. That is, $\mathcal{Y}_A \nsim_0 \mathcal{Y}_B \,|\, \mathcal{Y}_{A\cup B\cup C}$. ■

Remark 7.27. The assumption $\mathrm{pa}(B) \subseteq A \cup B \cup C$ is sufficient for $\mathcal{Y}_A \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \,|\, \mathcal{Y}_{A\cup B\cup C}$. Further, $\mathrm{an}(A \cup B \cup C) = A \cup B \cup C$ implies $\mathrm{pa}(A) \cup \mathrm{pa}(B) \subseteq A \cup B \cup C$, so we also have a condition for local Granger non-causality and local contemporaneous uncorrelatedness for ancestral subsets.

Remark 7.28. The lack of the right decomposability of local Granger non-causality (cf. Remark 4.13) is the key issue when trying to derive a global Markov property without additional assumptions. In the case $A \cup B \cup C \subset V$, Corollary 1 and Proposition 2 of Koster (1999) yield

$$A \bowtie_m B \mid C \;\; [G^0_{CG}] \quad \Leftrightarrow \quad A' \bowtie_m B' \mid C \;\; [G^0_{CG,\mathrm{an}(A\cup B\cup C)}]$$

for disjoint subsets A' and B' with $A \subseteq A'$, $B \subseteq B'$, and $A' \cup B' \cup C = \mathrm{an}(A \cup B \cup C)$, as in the proof of Theorem 7.8. According to Proposition 7.25, we can conclude that

$$\mathcal{Y}_{A'} \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_{B'} \,|\, \mathcal{Y}_{A'\cup B'\cup C}, \quad \mathcal{Y}_{B'} \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_{A'} \,|\, \mathcal{Y}_{A'\cup B'\cup C}, \quad \text{and} \quad \mathcal{Y}_{A'} \nsim_0 \mathcal{Y}_{B'} \,|\, \mathcal{Y}_{A'\cup B'\cup C}$$

in $[G^0_{CG,\mathrm{an}(A\cup B\cup C)}]$. These statements also hold in $[G^0_{CG}]$, since the definitions of local Granger non-causality and local contemporaneous uncorrelatedness do not depend on whether we choose the subgraph with vertices in $A' \cup B' \cup C$ or the whole graph with vertices in V. To get, e.g., $\mathcal{Y}_A \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \,|\, \mathcal{Y}_{A\cup B\cup C}$, we not only need left decomposability but also right decomposability.

In the causality graph, we solve this problem by conducting a (discrete) induction into the past using the conditional orthogonality relation, which satisfies properties (C1) to (C5) and thus provides much flexibility in the proofs. Didelez (2006) was able to give sufficient and satisfiable criteria for right decomposability in the local conditional independence graph in her Proposition 4.9. In our local causality graph, her result corresponds to the implication

$$\text{“}\mathcal{Y}_A \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_{B\cup C\cup D_1} \,|\, \mathcal{Y}_{A\cup B\cup C\cup D} \quad \text{and} \quad \mathcal{Y}_{B\cup C} \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_{A\cup D_2} \,|\, \mathcal{Y}_{A\cup B\cup C\cup D}$$
$$\Rightarrow \quad \mathcal{Y}_A \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_B \,|\, \mathcal{Y}_{A\cup B\cup D}\text{”},$$

where $D = D_1 \cup D_2$. However, this statement does not seem to be valid in our model. Further research on sufficient and satisfiable assumptions for the right decomposition property and for the global Markov properties is required beyond the scope of this thesis.

To visualise the above results for the local causality graph, we continue with Example 7.6.

Example 7.29. In Example 7.15, we establish that $\{1\} \bowtie_m \{5\} \mid \{3,4\}$, which does not depend on whether we consider the causality graph or the local causality graph. Moreover, $\mathrm{pa}(\{1\}) = \{3\}$ and $\mathrm{pa}(\{5\}) = \{4\}$, so $\mathrm{pa}(\{1\}) \cup \mathrm{pa}(\{5\}) \subseteq \{1,3,4,5\}$. Therefore, in the local causality graph, Proposition 7.26 implies the relations $\mathcal{Y}_1 \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_5 \,|\, \mathcal{Y}_{\{1,3,4,5\}}$, $\mathcal{Y}_5 \nrightarrow\!\!\!\rightarrow_0 \mathcal{Y}_1 \,|\, \mathcal{Y}_{\{1,3,4,5\}}$, and $\mathcal{Y}_1 \nsim_0 \mathcal{Y}_5 \,|\, \mathcal{Y}_{\{1,3,4,5\}}$, analogous to the relations in the causality graph in Example 7.23. However, since $\{1\} \not\bowtie_m \{4\} \,|\, \{3\}$, and we do not have a global Granger-causal Markov property, we cannot derive the same relations between $\mathcal{Y}_1$ and $\mathcal{Y}_4$ with respect to $\mathcal{Y}_{\{1,3,4\}}$ from Example 7.23.

In summary, even if we cannot derive global Markov properties for the local causality graph with the desired generality, we can still make statements about local Granger non-causality relations and subprocesses being locally contemporaneously uncorrelated given partial information.

CHAPTER 8

(Local) Causality graphs for MCAR processes

In this chapter, we gain a deeper understanding of the causality graph and the local causality graph. We find that the graph definitions are not only reasonable in terms of Markov properties but even interpretatively meaningful in examples. For this purpose, we apply the graphical models to causal *multivariate continuous-time autoregressive processes* $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ of order $p \geq 1$ (MCAR(p) processes) and to their special case of Ornstein-Uhlenbeck processes. In this way, we also provide a simple visual representation of the corresponding dependency structures for this important class of processes.

MCAR(p) processes were introduced in Remark 2.21 as state space models with matrices of a certain simple structure. MCAR models are thus already state space models in controller canonical form and are therefore denoted by $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$ throughout the present chapter. As in Chapter 2, we simply assume that the k-dimensional driving Lévy process satisfies Assumption 1 and that $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ applies. We do not explicitly restate these assumptions.

We open this chapter with the derivation of *alternative representations* of MCAR processes and their highest derivative processes. These representations are important for computing *orthogonal projections* on linear spaces, which in turn are essential for characterising the edges in the (local) causality graph. To the best of our knowledge, such orthogonal projections have not yet been treated in the existing literature. Rozanov (1967), III, 5, is devoted to the topic of prediction of general stationary processes. He touches briefly on univariate processes with a spectral density function that is a rational function, but the representations in that book are based on a specific maximal decomposition of the spectral density function. This decomposition is generally not expressible as a simple formula, so he only considers univariate examples. His methods are therefore not suitable for our purposes. Furthermore, as elaborated on in Chapter 9, Brockwell and Lindner (2015) only consider the prediction of univariate CARMA processes and the multivariate generalisation in Basse-O'Connor et al. (2019) is inconsistent with the univariate result.

For MCAR processes, not only do we compute the necessary orthogonal projections, but MCAR processes with $\Sigma_L > 0$ also satisfy the assumptions of the (local) causality graph. Thus, the graphical models are *well defined* and satisfy the preferred *Markov properties*. The representations of the orthogonal projections then lead to the main result, the *characterisation of the edge types* in the (local) causality graph by the model parameters of the MCAR process. These characterisations are interpretatively meaningful and comparable to the discrete-time analogue of VAR models in Eichler (2007). They are even clear enough to discuss the existence of (local) causality graphs.

It is noteworthy that the Gaussian MCAR process and the Gaussian Ornstein-Uhlenbeck process considered in Comte and Renault (1996) and Mogensen and Hansen (2022) are special cases of

Lévy-driven MCAR processes. The characterisations of the directed and undirected influences of Comte and Renault (1996) in a bivariate Gaussian CAR model align with our characterisations in the local causality graph. Furthermore, the characterisations of the directed and undirected edges in the canonical local independence graph for Gaussian Ornstein-Uhlenbeck processes in Mogensen and Hansen (2022) are also consistent with our characterisations in the local causality graph. However, the Graph Ornstein-Uhlenbeck process in Courgeau and Veraart (2022) is a network without defining the direction of the edges and is therefore not comparable to our model.

We structure the chapter as follows. In Section 8.1, we derive alternative representations and orthogonal projections of MCAR processes. In Section 8.2, we show that MCAR processes satisfy the assumptions of the (local) causality graph, so that both graphical models are well defined and the desired Markov properties are satisfied. Then, in Section 8.3, we come to the main results of the chapter, the characterisations of the directed and undirected edges by model parameters of MCAR processes. For Ornstein-Uhlenbeck processes, the characterisations are included as a special case. In this section, we also give a detailed interpretation of the obtained characterisations and draw parallels to the literature. Note that in Chapter 9, we consider the application to invertible controller canonical state space (ICCSS) processes, so we examine the relationships between the results for MCAR processes and those for ICCSS processes in that chapter.

8.1. Orthogonal projections of MCAR processes

The orthogonal projections of MCAR processes are of great importance throughout this chapter. We need them primarily to characterise the directed and undirected edges for MCAR processes in the (local) causality graph, but also to check that the graphical models are well defined. Therefore, in this section, we compute the required orthogonal projections of MCAR processes and their $(p-1)$-th derivatives on different subspaces. To do this, we first give an *alternative representation* of $Y_a(t+h)$ and $D^{(p-1)}Y_a(t+h)$, $a \in V$, some step $h \geq 0$ into the future, that is suitable for the prediction purpose. Note that we consider $D^{(p-1)}Y_a(t+h)$ because, according to Remark 2.25, it is the highest existing derivative of the MCAR process, which we require for the definition of local Granger causality and local contemporaneous correlation, respectively.

Lemma 8.1. *Let $\mathcal{Y}_V$ be an MCAR(p) process. Further, let $t \in \mathbb{R}$, $h \geq 0$, and $a \in V$. Then*

$$Y_a(t+h) = e_a^\top \mathbf{C} e^{\mathbf{A}h} \sum_{j=1}^{p} \mathbf{E}_j D^{(j-1)} Y_V(t) + e_a^\top \mathbf{C} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.} \quad \text{and}$$

$$D^{(p-1)}Y_a(t+h) = e_a^\top \mathbf{E}_p^\top e^{\mathbf{A}h} \sum_{j=1}^{p} \mathbf{E}_j D^{(j-1)} Y_V(t) + e_a^\top \mathbf{E}_p^\top \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.},$$

where $\mathbf{E}_j$, $j = 1, \ldots, p$, are defined in equation (1.1)*, and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ in Remark 2.21.*

Proof. Let $t \in \mathbb{R}$, $h \geq 0$, and $a \in V$. First of all, because of relation (2.6),

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

We define the j-th k-block $\mathcal{X}^{(j)} = (X^{(j)}(t))_{t\in\mathbb{R}}$ of $\mathcal{X} = (X(t))_{t\in\mathbb{R}}$ as in equation (2.19). Then, together with relation (2.20), it follows that

$$\begin{aligned} Y_a(t+h) &= e_a^\top \mathbf{C} e^{\mathbf{A}h} \sum_{j=1}^{p} \mathbf{E}_j X^{(j)}(t) + e_a^\top \mathbf{C} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \\ &= e_a^\top \mathbf{C} e^{\mathbf{A}h} \sum_{j=1}^{p} \mathbf{E}_j D^{(j-1)} Y_V(t) + e_a^\top \mathbf{C} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u). \end{aligned}$$

For the proof of the representation of $D^{(p-1)}Y_a(t+h)$, we note that the MCAR(p) process $\mathcal{Y}_V$ is $(p-1)$-times differentiable with $D^{(p-1)}Y_V(t+h) = X^{(p)}(t+h) = \mathbf{E}_p^\top X(t+h)$ by relation (2.20). Replacing $\mathbf{C}$ with $\mathbf{E}_p^\top$ in the previous calculation gives the assertion. ■

From the alternative representations of $Y_a(t+h)$ and $D^{(p-1)}Y_a(t+h)$ in Lemma 8.1, we conclude that, on the one hand, the past $(Y_V(s))_{s\leq t}$ of all components and, on the other hand, the future of the Lévy process $(L(s)-L(t))_{t\leq s\leq t+h}$ is relevant for $Y_a(t+h)$ and $D^{(p-1)}Y_a(t+h)$. With this knowledge, we can specify the *orthogonal projections*.

Proposition 8.2. *Let $\mathcal{Y}_V$ be an MCAR(p) process. Further, let $t\in\mathbb{R}$, $h\geq 0$, $S\subseteq V$, and $a\in V$. Then*

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

Proof. Let $h\geq 0$, $t\in\mathbb{R}$, $S\subseteq V$, and $a\in V$. First, for $v\in S$, $Y_v(t), D^{(1)}Y_v(t), \dots, D^{(p-1)}Y_v(t)$ are already in $\mathcal{L}_{Y_S}(t)$ by Remark 2.9 and are therefore projected onto themselves. Thus,

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

In addition, $(Y_S(s))_{s\leq t}$ and $(L(s)-L(t))_{t\leq s\leq t+h}$ are independent (Marquardt & Stelzer, 2007, Theorem 3.12), so $e_a^\top \mathbf{C} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u)$ is projected to zero. Together with Lemma 8.1, the representation of $P_{\mathcal{L}_{Y_S}(t)} Y_a(t+h)$ follows. The representation of $P_{\mathcal{L}_{Y_S}(t)} D^{(p-1)} Y_a(t+h)$ again follows by replacing $\mathbf{C}$ with $\mathbf{E}_p^\top$ in the previous calculation. ■

To apply local Granger causality and local contemporaneous correlation to MCAR processes, we also need the orthogonal projections in the following theorem. Note that, indeed, the limits exist, which is essential for the well-definedness of local Granger causality in Definition 4.7.

Theorem 8.3. *Let $\mathcal{Y}_V$ be an MCAR(p) process. Further, let $t \in \mathbb{R}$, $S \subseteq V$, and $a \in V$. Then*

$$\begin{aligned}&\underset{h\to 0}{l.i.m.}\, P_{\mathcal{L}_{Y_S}(t)}\left(\frac{D^{(p-1)}Y_a(t+h) - D^{(p-1)}Y_a(t)}{h}\right)\\ &= e_a^\top \mathbf{E}_p^\top \mathbf{A} \sum_{v\in S}\sum_{j=1}^{p} \mathbf{E}_j e_v D^{(j-1)}Y_v(t) + e_a^\top \mathbf{E}_p^\top \mathbf{A} \sum_{v\in V\setminus S}\sum_{j=1}^{p} \mathbf{E}_j e_v P_{\mathcal{L}_{Y_S}(t)} D^{(j-1)}Y_v(t) \quad \mathbb{P}\text{-}a.s.\end{aligned}$$

and, for $h \geq 0$,

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

Proof. Let $t \in \mathbb{R}$, $h \geq 0$, $S \subseteq V$, and $a \in V$. From Proposition 8.2, we already know that

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

Now $\lim_{h\to 0}(e^{\mathbf{A}h} - I_{kp})/h = \mathbf{A}$ implies the first assertion

$$\begin{aligned}&\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)}\left(\frac{D^{(p-1)}Y_a(t+h) - D^{(p-1)}Y_a(t)}{h}\right)\\ &= e_a^\top \mathbf{E}_p^\top \mathbf{A} \sum_{v\in S}\sum_{j=1}^{p} \mathbf{E}_j e_v D^{(j-1)}Y_v(t) + e_a^\top \mathbf{E}_p^\top \mathbf{A} \sum_{v\in V\setminus S}\sum_{j=1}^{p} \mathbf{E}_j e_v P_{\mathcal{L}_{Y_S}(t)} D^{(j-1)}Y_v(t) \quad \mathbb{P}\text{-a.s.}\end{aligned}$$

The second assertion follows directly from Lemma 8.1 and Proposition 8.2 with $S = V$. ■

Remark 8.4. For $S = V$, the orthogonal projections in the previous statements are simplified. For $t \in \mathbb{R}$, $h \geq 0$, and $a \in V$, we obtain the explicit representation

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

This result is the multivariate generalisation of Theorem 2.8 by Brockwell and Lindner (2015) for univariate CARMA processes. It also applies that

$$P_{\mathcal{L}_{Y_V}(t)} D^{(p-1)} Y_a(t+h) = e_a^\top \mathbf{E}_p^\top e^{\mathbf{A}h} X(t) \quad \mathbb{P}\text{-a.s.} \quad \text{and}$$
$$\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_V}(t)} \left(\frac{D^{(p-1)} Y_a(t+h) - D^{(p-1)} Y_a(t)}{h} \right) = e_a^\top \mathbf{E}_p^\top \mathbf{A} X(t) \quad \mathbb{P}\text{-a.s.}$$

If $S \subset V$, the representations of the orthogonal projections in Proposition 8.2 and Theorem 8.3 are only partially determined. For a more explicit representation in this case, we refer, as previously mentioned, to Rozanov (1967), III, 5. Since more explicit representations are not relevant for the derivation of the (local) causality graph, we do not explore this topic further.

8.2. Establishment of (local) causality graphs for MCAR processes

In the following, we establish that the (local) causality graph for the MCAR process $\mathcal{Y}_V$ is well defined. To achieve this, we verify that $\mathcal{Y}_V$ is wide-sense stationary, mean-square continuous, and satisfies Assumptions 3 and 4. This preliminary work then results in the establishment of the (local) causality graph for MCAR processes and the validity of the desired Markov properties. Of the aforementioned assumptions, the validity of both *wide-sense stationarity* and *continuity in the mean square* is immediately clear from Remarks 2.12 and 2.13. Therefore, we focus our discussion on the remaining two assumptions.

The proof of *Assumption 3* on the spectral density function of the MCAR process (see equations (2.11) and (2.18) for its representation) is elaborate, but the basic idea is simple. First, we observe that $\Sigma_L > 0$ implies $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Then, on the one hand, we prove that, for disjoint (and non-empty) subsets $A, B \subseteq V$, an epsilon bound for

$$d_{AB}(\lambda) := f_{Y_A Y_A}(\lambda)^{-1/2} f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda) f_{Y_A Y_A}(\lambda)^{-1/2}$$

can always be found on compact intervals. On the other hand, the matrix function $d_{AB}(\lambda)$, $\lambda \in \mathbb{R}$, converges to a boundary matrix as $|\lambda| \to \infty$, which can also be bounded. Together, these auxiliary results then give Assumption 3. We start with the simple part of Assumption 3, namely that $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$.

Lemma 8.5. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. Then $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$.*

Proof. The transfer function $P(i\lambda)^{-1}$, $\lambda \in \mathbb{R}$, has full rank and we have $\Sigma_L > 0$ by assumption. Therefore, the positive definiteness of $f_{Y_V Y_V}(\lambda) = 1/(2\pi) P(i\lambda)^{-1} \Sigma_L (P(-i\lambda)^{-1})^\top$ follows for $\lambda \in \mathbb{R}$. ■

Now to the second part of Assumption 3, where we claim that there exists an $0 < \varepsilon < 1$, such that $d_{AB}(\lambda) \leq (1-\varepsilon) I_\alpha$ for (almost) all $\lambda \in \mathbb{R}$ and for all disjoint subsets $A, B \subseteq V$ with $\#A = \alpha$. To prove this statement, we require several auxiliary lemmata. First, we show that there exists an epsilon bound on compact intervals.

Lemma 8.6. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. Further, let $A, B \subseteq V$, $A \cap B = \emptyset$, and denote $\#A = \alpha$. Then, for each compact interval $K \subset \mathbb{R}$, there exists an $0 < \varepsilon_K < 1$, such that*

$$d_{AB}(\lambda) \leq (1 - \varepsilon_K) I_\alpha \quad \forall\, \lambda \in K.$$

Proof. Since $\det(P(i\lambda))$ has no zeros due to $\mathcal{N}(P) = \sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$, the spectral density function $f_{Y_V Y_V}(\lambda) = 1/(2\pi) P(i\lambda)^{-1} \Sigma_L (P(-i\lambda)^{-1})^\top$, $\lambda \in \mathbb{R}$, is continuous. Then Bhatia (1997) states in Corollary VI.1.6 that there exist continuous functions $\sigma_1(\lambda), \ldots, \sigma_k(\lambda)$, $\lambda \in \mathbb{R}$, which are the eigenvalues of $f_{Y_V Y_V}(\lambda)$. Since $f_{Y_V Y_V}(\lambda)$ is Hermitian and positive definite (cf. Lemma 2.4 and Lemma 8.5), these eigenvalues are in $(0, \infty)$ and, in particular, the functions can be taken as ordered by $0 < \sigma_1(\lambda) \leq \ldots \leq \sigma_k(\lambda)$ for $\lambda \in \mathbb{R}$, see Bhatia (1997), p. 154. Furthermore, Bernstein (2009) provides in Lemma 8.4.1 that $\sigma_1(\lambda) I_k \leq f_{Y_V Y_V}(\lambda) \leq \sigma_k(\lambda) I_k$ and by Bernstein (2009), Proposition 8.1.2, we obtain

$$\sigma_1(\lambda) I_{\alpha+\beta} \leq f_{Y_{A\cup B} Y_{A\cup B}}(\lambda) \quad \text{and} \quad f_{Y_A Y_A}(\lambda) \leq \sigma_k(\lambda) I_\alpha \quad \forall\, \lambda \in \mathbb{R},$$

denoting $\#B = \beta$. Let $\lambda \in \mathbb{R}$. Using Proposition 8.1.2 by Bernstein (2009) gives

$$(f_{Y_{A\cup B} Y_{A\cup B}}(\lambda))^{-1} \leq \frac{1}{\sigma_1(\lambda)} I_{\alpha+\beta}$$

and, together with Bernstein (2009), Proposition 8.2.5, we receive

$$\left(f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda)\right)^{-1} \leq \frac{1}{\sigma_1(\lambda)} I_\alpha.$$

Now Bernstein (2009), Proposition 8.1.2, yields

$$\sigma_1(\lambda) I_\alpha \leq f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda).$$

If we combine this result with $f_{Y_A Y_A}(\lambda) \leq \sigma_k(\lambda) I_\alpha$ from above, we obtain

$$\frac{\sigma_1(\lambda)}{\sigma_k(\lambda)} f_{Y_A Y_A}(\lambda) \leq \sigma_1(\lambda) I_\alpha \leq f_{Y_A Y_A}(\lambda) - f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda).$$

Thus,

$$f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda) \leq \left(1 - \frac{\sigma_1(\lambda)}{\sigma_k(\lambda)}\right) f_{Y_A Y_A}(\lambda)$$

and Bernstein (2009), Proposition 8.1.2, finally provides

$$d_{AB}(\lambda) = f_{Y_A Y_A}(\lambda)^{-1/2} f_{Y_A Y_B}(\lambda) f_{Y_B Y_B}(\lambda)^{-1} f_{Y_B Y_A}(\lambda) f_{Y_A Y_A}(\lambda)^{-1/2} \leq \left(1 - \frac{\sigma_1(\lambda)}{\sigma_k(\lambda)}\right) I_\alpha.$$

We now distinguish two cases to prove the assertion. In the first case, let $\sigma_1(\lambda)/\sigma_k(\lambda) = 1$ for all $\lambda \in K$. Then

$$d_{AB}(\lambda) \leq 0_\alpha \in M_\alpha(\mathbb{R}) \quad \forall\, \lambda \in K$$

and the assertion holds for any $0 < \varepsilon_K < 1$. Without loss of generality, we set $\varepsilon_K = 1/2$. In the second case, let $\sigma_1(\lambda)/\sigma_k(\lambda) < 1$ for at least one $\lambda \in K$. Since the continuous function $\sigma_1(\lambda)/\sigma_k(\lambda)$, $\lambda \in \mathbb{R}$, achieves its minimum on the compact set K, we define $\varepsilon_K = \min_{\lambda\in K} \sigma_1(\lambda)/\sigma_k(\lambda)$. Thus, we obtain that $0 < \varepsilon_K < 1$ and

$$d_{AB}(\lambda) \leq (1-\varepsilon_K) I_\alpha \quad \forall\, \lambda \in K.$$

Since $\sigma_1(\lambda)/\sigma_k(\lambda) \leq 1$ for all $\lambda \in K$, these are all cases that occur and the assertion holds. ■

Second, we establish a relationship between the convergence of matrices in norm and the Loewner order, that we could not find in the literature. The result is similar to the epsilon definition of the convergence of sequences.

Lemma 8.7. *Let $F(\lambda) \in M_\alpha(\mathbb{R})$, $\lambda \in \mathbb{R}$, $\alpha \in \mathbb{N}$, and $M \in M_\alpha(\mathbb{R})$ be Hermitian matrices, such that $\lim_{|\lambda|\to\infty} \|F(\lambda) - M\| = 0$. Then, for any $\varepsilon^* > 0$, there exists a $\lambda^* \in \mathbb{R}$, such that*

$$F(\lambda) \leq M + \varepsilon^* I_\alpha \quad \forall\, |\lambda| \geq \lambda^*.$$

Proof. Let $\varepsilon^* > 0$. Due to $\lim_{|\lambda|\to\infty} \|F(\lambda) - M\| = 0$, it obviously holds that

$$\lim_{|\lambda|\to\infty} \left| [F(\lambda) - M]_{ij} \right| = 0$$

for $i, j = 1, \dots, \alpha$. It follows that, for $\varepsilon^* > 0$ and $k \geq \alpha$, there exists a $\lambda^* \in \mathbb{R}$, such that

$$\left| [F(\lambda) - M]_{ij} \right| \leq \frac{\varepsilon^*}{k}$$

for all $|\lambda| \geq \lambda^*$ and $i, j = 1, \dots, \alpha$. Now, for any $x \in \mathbb{R}^\alpha$ and $|\lambda| \geq \lambda^*$, we obtain

$$\begin{aligned}
x^\top (F(\lambda) - M)\, x &= \sum_{i=1}^{\alpha}\sum_{j=1}^{\alpha} x_i \,[F(\lambda) - M]_{ij}\, x_j \\
&= \frac{1}{4}\sum_{i=1}^{\alpha}\sum_{j=1}^{\alpha} \Big((x_i + x_j)^2 \,[F(\lambda) - M]_{ij} - (x_i - x_j)^2 \,[F(\lambda) - M]_{ij} \Big) \\
&\leq \frac{1}{4}\sum_{i=1}^{\alpha}\sum_{j=1}^{\alpha} \left((x_i + x_j)^2 \frac{\varepsilon^*}{k} + (x_i - x_j)^2 \frac{\varepsilon^*}{k} \right) \\
&= \frac{\varepsilon^*}{2k}\sum_{i=1}^{\alpha}\sum_{j=1}^{\alpha} \left(x_i^2 + x_j^2 \right) \\
&= \frac{\varepsilon^*}{2k}\sum_{i=1}^{\alpha} \left(\alpha x_i^2 + x^\top x \right) \\
&= \frac{\varepsilon^* \alpha}{k} x^\top x.
\end{aligned}$$

Thus, since $k \geq \alpha$, we obtain $F(\lambda) - M \leq \varepsilon^* I_\alpha$ and $F(\lambda) \leq M + \varepsilon^* I_\alpha$ for $|\lambda| \geq \lambda^*$. ■

Based on Lemma 8.7, we prove that $d_{AB}(\lambda)$ converges to a boundary matrix M as $|\lambda| \to \infty$.

Lemma 8.8. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. Further, let $A, B \subseteq V$, $A \cap B = \emptyset$, and denote $\#A = \alpha$. Define*

$$M := H_{AA}^{-1/2} H_{AB} H_{BB}^{-1} H_{BA} H_{AA}^{-1/2},$$

where, for $S, S_1, S_2 \subseteq V$, we denote

$$H_{S_1 S_2} = E_{S_1}^\top \mathbf{C}\mathbf{B}\Sigma_L \mathbf{B}^\top \mathbf{C}^\top E_{S_2} \quad \text{and} \quad [E_S]_{ij} = \begin{cases} 1, & \text{if } i = j \in S, \\ 0, & \text{else}. \end{cases}.$$

Then, for $\varepsilon^* > 0$, there exists a $\lambda^* > 0$, such that

$$d_{AB}(\lambda) \leq M + \varepsilon^* I_\alpha \quad \forall\, |\lambda| \geq \lambda^*.$$

Proof. Bernstein (2009) states in equation (4.4.23) that

$$(i\lambda I_{kp} - \mathbf{A})^{-1} = \sum_{n=0}^{kp-1} \frac{(i\lambda)^n}{\chi_{\mathbf{A}}(i\lambda)} \Delta_n,$$

where $\Delta_n \in M_{kp}(\mathbb{R})$, $\Delta_{kp-1} = I_{kp}$, and $\chi_{\mathbf{A}}(z) = z^{kp} + \gamma_{kp-1} z^{kp-1} + \dots + \gamma_1 z + \gamma_0$, $z \in \mathbb{C}$, is the characteristic polynomial of $\mathbf{A}$ with $\gamma_1, \dots, \gamma_{kp-1} \in \mathbb{R}$, see also Bernstein (2009), equation (4.4.3). Inserting this result into representation (2.11) of the spectral density function yields

$$f_{Y_V Y_V}(\lambda) = \frac{1}{2\pi} \sum_{m=0}^{kp-1} \sum_{n=0}^{kp-1} \frac{(i\lambda)^m}{\chi_{\mathbf{A}}(i\lambda)} \frac{(-i\lambda)^n}{\chi_{\mathbf{A}}(-i\lambda)} \mathbf{C}\Delta_m \mathbf{B}\Sigma_L \mathbf{B}^\top \Delta_n^\top \mathbf{C}^\top.$$

In particular, we have

$$f_{Y_a Y_b}(\lambda) = \frac{1}{2\pi \chi_{\mathbf{A}}(i\lambda)\chi_{\mathbf{A}}(-i\lambda)} \sum_{m=0}^{kp-1} \sum_{n=0}^{kp-1} (i\lambda)^{m+n} (-1)^n e_a^\top \mathbf{C}\Delta_m \mathbf{B}\Sigma_L \mathbf{B}^\top \Delta_n^\top \mathbf{C}^\top e_b$$

for $a, b \in V$. From this rational function, we can specify the asymptotic behaviour. The numerator contains a complex polynomial of maximum degree $2kp - 2$ with leading coefficient

$$e_a^\top \mathbf{C}\Delta_{kp-1} \mathbf{B}\Sigma_L \mathbf{B}^\top \Delta_{kp-1}^\top \mathbf{C}^\top e_b = e_a^\top \mathbf{C}\mathbf{B}\Sigma_L \mathbf{B}^\top \mathbf{C}^\top e_b,$$

which can be zero. The denominator is a complex polynomial of degree $2kp$ with leading coefficient 2π. Differentiating between $e_a^\top \mathbf{C}\mathbf{B}\Sigma_L \mathbf{B}^\top \mathbf{C}^\top e_b = 0$ and $e_a^\top \mathbf{C}\mathbf{B}\Sigma_L \mathbf{B}^\top \mathbf{C}^\top e_b \neq 0$ results in

$$\lim_{|\lambda| \to \infty} \left| 2\pi\lambda^2 f_{Y_a Y_b}(\lambda) - e_a^\top \mathbf{C}\mathbf{B}\Sigma_L \mathbf{B}^\top \mathbf{C}^\top e_b \right| = 0.$$

Finally, for $S_1, S_2 \subseteq V$, we receive

$$\lim_{|\lambda| \to \infty} \left\| 2\pi\lambda^2 f_{Y_{S_1} Y_{S_2}}(\lambda) - H_{S_1 S_2} \right\| = 0. \tag{8.1}$$

Since $2\pi\lambda^2 f_{Y_BY_B}(\lambda) > 0$ for $\lambda \neq 0$ as well as $H_{BB} = E_B^\top \mathbf{C}\mathbf{B}\Sigma_L\mathbf{B}^\top\mathbf{C}^\top E_B > 0$, Bühler and Salamon (2018) provide in Corollary 1.5.7(ii) the continuity of the formation of the inverse. It follows that

$$\lim_{|\lambda|\to\infty} \left\| \frac{1}{2\pi\lambda^2} f_{Y_BY_B}(\lambda)^{-1} - H_{BB}^{-1} \right\| = 0. \tag{8.2}$$

In addition, Bhatia (1997), Theorem X.1.1 and relation (X.2), respectively, provide the following inequality for induced matrix norms and $\lambda \neq 0$,

$$\left\| \sqrt{2\pi}|\lambda| f_{Y_AY_A}(\lambda)^{1/2} - H_{AA}^{1/2} \right\|_{ind} \leq \left\| 2\pi\lambda^2 f_{Y_AY_A}(\lambda) - H_{AA} \right\|_{ind}^{1/2}.$$

Because of the equivalence of matrix norms and because the right-hand side of the inequality converges to zero, we get

$$\lim_{|\lambda|\to\infty} \left\| \sqrt{2\pi}|\lambda| f_{Y_AY_A}(\lambda)^{1/2} - H_{AA}^{1/2} \right\| = 0.$$

Using the positive definiteness of the positive square root and Bühler and Salamon (2018), Corollary 1.5.7(ii), again, it follows that

$$\lim_{|\lambda|\to\infty} \left\| \frac{1}{\sqrt{2\pi}|\lambda|} f_{Y_AY_A}(\lambda)^{-1/2} - H_{AA}^{-1/2} \right\| = 0. \tag{8.3}$$

An application of the limit results (8.1), (8.2), and (8.3), as well as the submultiplicativity of the induced matrix norm, result in

$$\lim_{|\lambda|\to\infty} \Big\| f_{Y_AY_A}(\lambda)^{-1/2} f_{Y_AY_B}(\lambda) f_{Y_BY_B}(\lambda)^{-1} f_{Y_BY_A}(\lambda) f_{Y_AY_A}(\lambda)^{-1/2}$$
$$-H_{AA}^{-1/2} H_{AB} H_{BB}^{-1} H_{BA} H_{AA}^{-1/2} \Big\|_{ind} = 0.$$

Therefore, $\lim_{|\lambda|\to\infty} \|d_{AB}(\lambda) - M\| = 0$. Finally, Lemma 8.7 provides that, for each $\varepsilon^* > 0$, there exists a $\lambda^* \in \mathbb{R}$, such that

$$d_{AB}(\lambda) \leq M + \varepsilon^* I_\alpha \quad \forall\, |\lambda| \geq \lambda^*.$$ ■

As a final auxiliary result, we show that the matrix M can also be bounded.

Lemma 8.9. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. Further, let $A, B \subseteq V$, $A \cap B = \emptyset$, and denote $\#A = \alpha$. Then there exists an $0 < \varepsilon_M < 1$, such that*

$$M \leq (1 - \varepsilon_M) I_\alpha,$$

where M is defined as in Lemma 8.8.

Proof. First of all, analogous to the proof of Lemma 8.6, we obtain

$$M = H_{AA}^{-1/2} H_{AB} H_{BB}^{-1} H_{BA} H_{AA}^{-1/2} \leq \left(1 - \frac{\sigma_1}{\sigma_k}\right) I_\alpha,$$

where σ_1 is the smallest eigenvalue and σ_k is the largest eigenvalue of $\mathbf{C}\mathbf{B}\Sigma_L\mathbf{B}^\top\mathbf{C}^\top$. Note that the matrix $\mathbf{C}\mathbf{B}\Sigma_L\mathbf{B}^\top\mathbf{C}^\top$ is positive definite due to $\Sigma_L > 0$, as well as $\mathbf{C}$ and $\mathbf{B}$ being of full rank. Thus, the eigenvalues σ_1 and σ_k are positive. Again, we distinguish between two cases. In the first case, let $\sigma_1/\sigma_k = 1$. Then $M \leq 0_\alpha \in M_\alpha(\mathbb{R})$ and the assertion holds for any $0 < \varepsilon_M < 1$, we set $\varepsilon_M = 1/2$. In the second case, let $\sigma_1/\sigma_k < 1$. Then, we set $\varepsilon_M = \sigma_1/\sigma_k$ and obtain that $0 < \varepsilon_M < 1$ as well as $M \leq (1-\varepsilon_M)I_\alpha$. Since $\sigma_1/\sigma_k \leq 1$, these are already all cases that can occur and the assertion follows. ■

Based on the previous lemmata, we are finally in a position to prove Assumption 3.

Proposition 8.10. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. Then $\mathcal{Y}_V$ satisfies Assumption 3.*

Proof. First of all, Lemma 8.5 provides $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Furthermore, with the notation of Lemma 8.8 and Lemma 8.9, we choose $0 < \varepsilon^* < \varepsilon_M$. Lemma 8.8 provides that there exists a $\lambda^* \in \mathbb{R}$, such that

$$d_{AB}(\lambda) \leq M + \varepsilon^* I_\alpha \quad \forall\, |\lambda| \geq \lambda^*.$$

Furthermore, Lemma 8.9 yields

$$d_{AB}(\lambda) \leq M + \varepsilon^* I_\alpha \leq (1-\varepsilon_M)I_\alpha + \varepsilon^* I_\alpha = (1-(\varepsilon_M - \varepsilon^*))I_\alpha.$$

For $|\lambda| \geq \lambda^*$, we thus find the boundary matrix $(1-(\varepsilon_M-\varepsilon^*))I_\alpha$, where $0 < \varepsilon_M - \varepsilon^* < 1$, due to the choice of ε^*. On the compact interval $K = [-\lambda^*, \lambda^*]$, Lemma 8.6 states that there exists an $0 < \varepsilon_K < 1$, such that $d_{AB}(\lambda) \leq (1-\varepsilon_K)I_\alpha$. We set $\varepsilon_{AB} = \min\{\varepsilon_K, \varepsilon_M - \varepsilon^*\}$. Then $d_{AB}(\lambda) \leq (1-\varepsilon_{AB})I_\alpha$ for all $\lambda \in \mathbb{R}$. However, ε_{AB} still depends on A and B. Since there are only finitely many of such index sets, we set $\varepsilon = \min\{\varepsilon_{AB} : A, B \subseteq V, A \cap B = \emptyset\}$. We obtain that $0 < \varepsilon < 1$ and $d_{AB}(\lambda) \leq (1-\varepsilon)I_\alpha$ is satisfied for all $\lambda \in \mathbb{R}$ and disjoint subsets $A, B \subseteq V$. ■

For the (local) causality graph to be well defined, all that is missing is that $\mathcal{Y}_V$ satisfies *Assumption 4*. It is expected that the MCAR process satisfies this assumption, as in our case the driving Lévy process has no drift term. We show Assumption 4 based on the characterisation (3.5) of purely non-deterministic processes.

Proposition 8.11. *Let $\mathcal{Y}_V$ be an MCAR(p) process. Then $\mathcal{Y}_V$ satisfies Assumption 4.*

Proof. For Assumption 4, we first apply representation (2.6), the independence of $(X(s))_{s\leq t}$ and $(L(s)-L(t))_{t\leq s\leq t+h}$ (Marquardt & Stelzer, 2007, Theorem 3.12), to obtain

$$\underset{h\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_X(t)} X(t+h) = \underset{h\to\infty}{\text{l.i.m.}}\, P_{\mathcal{L}_X(t)} \left(e^{\mathbf{A}h} X(t) + \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B}\, dL(u) \right) = \underset{h\to\infty}{\text{l.i.m.}}\, e^{\mathbf{A}h} X(t).$$

The limit on the right side of the equation is the zero vector due to $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$, resulting in the input process $\mathcal{X}$ being purely non-deterministic. By relation (3.5), this limit result is equivalent to $\bigcap_{t\in\mathbb{R}} \mathcal{L}_X(t) = \{0\}$. Since $\bigcap_{t\in\mathbb{R}} \mathcal{L}_{Y_V}(t) \subseteq \bigcap_{t\in\mathbb{R}} \mathcal{L}_X(t)$, the process $\mathcal{Y}_V$ is purely non-deterministic as well. ■

Given that both Assumptions 3 and 4 are satisfied, a direct consequence of the results in Chapters 6 and 7 is the following.

Proposition 8.12. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. If we define $V = \{1, \ldots, k\}$ as the vertices and the edges E_{CG} via*

(a) $a \longrightarrow b \notin E_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \nrightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V,$

(b) $a \dashline b \notin E_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V,$

for $a, b \in V$ with $a \neq b$, then the causality graph $G_{CG} = (V, E_{CG})$ for the MCAR process $\mathcal{Y}_V$ is well defined and satisfies the pairwise, local, block-recursive, global AMP, and global Granger-causal Markov property.

Proposition 8.13. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. If we define $V = \{1, \ldots, k\}$ as the vertices and the edges E_{CG}^0 via*

(a) $a \longrightarrow b \notin E_{CG}^0 \quad \Leftrightarrow \quad \mathcal{Y}_a \nrightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V,$

(b) $a \dashline b \notin E_{CG}^0 \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V,$

for $a, b \in V$ with $a \neq b$, then the local causality graph $G_{CG}^0 = (V, E_{CG}^0)$ for the MCAR process $\mathcal{Y}_V$ is well defined and satisfies the pairwise, local, and block-recursive Markov property. Furthermore, the statements of Propositions 7.25 and 7.26 apply.

8.3. Edge characterisations for MCAR processes

In this section, we come to the main part of the chapter, the *characterisation* of the directed and undirected edges in the (local) causality graph based on the model parameters $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and Σ_L of the causal MCAR process. We present several characterisations, provide interpretations of the results, and draw parallels to the existing literature. We start by characterising the *directed edges* in the (local) causality graph.

Proposition 8.14. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. Further, let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \nrightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[\mathbf{C}e^{\mathbf{A}h}\mathbf{E}_j\right]_{ba} = \left[e^{\mathbf{A}h}\right]_{b\,k(j-1)+a} = 0 \quad \forall\, 0 \leq h \leq 1,\ j = 1, \ldots, p,$

(b) $\mathcal{Y}_a \nrightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[\mathbf{E}_p^\top \mathbf{A}\mathbf{E}_j\right]_{ba} = [-A_{p-j+1}]_{ba} = 0 \qquad \forall\, j = 1, \ldots, p.$

Proof.

(a) Due to Theorem 4.4, $\mathcal{Y}_a \nrightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V$ if and only if

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

for $0 \leq h \leq 1$ and $t \in \mathbb{R}$. From Proposition 8.2, we also know that

$$P_{\mathcal{L}_{Y_V}(t)} Y_b(t+h) = \sum_{j=1}^{p} \sum_{v \in V} e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_v D^{(j-1)} Y_v(t) \quad \mathbb{P}\text{-a.s.},$$

$$\begin{aligned}P_{\mathcal{L}_{Y_{V\setminus\{a\}}}(t)}Y_b(t+h) =& \sum_{j=1}^{p}\sum_{v\in V\setminus\{a\}} e_b^\top \mathbf{C}e^{\mathbf{A}h}\mathbf{E}_j e_v D^{(j-1)}Y_v(t) \\ &+ \sum_{j=1}^{p} e_b^\top \mathbf{C}e^{\mathbf{A}h}\mathbf{E}_j e_a P_{\mathcal{L}_{Y_{V\setminus\{a\}}}(t)} D^{(j-1)}Y_a(t) \quad \mathbb{P}\text{-a.s.}\end{aligned}$$

for $0 \le h \le 1$ and $t \in \mathbb{R}$. We equate the two orthogonal projections and remove the coinciding terms. Then, we receive $\mathcal{Y}_a \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V$ if and only if, for $0 \le h \le 1$ and $t \in \mathbb{R}$,

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

The expression on the left side of the equation is in $\mathcal{L}_{Y_a}(t)$ due to Remark 2.9. The expression on the right side is in $\mathcal{L}_{Y_{V\setminus\{a\}}}(t)$. Because of the equality, both are in $\mathcal{L}_{Y_{V\setminus\{a\}}}(t) \cap \mathcal{L}_{Y_a}(t) = \{0\}$, making use of Proposition 3.10. Thus, $\mathcal{Y}_a \not\rightarrow\!\!\!\!\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V$ if and only if

$$\sum_{j=1}^{p} e_b^\top \mathbf{C}e^{\mathbf{A}h}\mathbf{E}_j e_a D^{(j-1)}Y_a(t) = 0 \quad \mathbb{P}\text{-a.s.} \tag{8.4}$$

for $0 \le h \le 1$ and $t \in \mathbb{R}$. In the following, we show that equation (8.4) is equivalent to

$$e_b^\top \mathbf{C}e^{\mathbf{A}h}\mathbf{E}_j e_a = 0 \quad \forall\, 0 \le h \le 1,\ j = 1,\ldots,p. \tag{8.5}$$

Clearly, (8.5) implies (8.4). For the opposite direction, suppose that equation (8.4) is satisfied. Define the kp-dimensional vector $y = (y_1,\ldots,y_{kp})$ with entries

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

Then, the equations (8.4) and (2.20) imply

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

In particular,

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

However, in the case of an MCAR process, the matrix

$$\left(\mathbf{B}, \mathbf{AB}, \ldots, \mathbf{A}^{k-1}\mathbf{B}\right) = \begin{pmatrix} 0_k & \cdots & 0_k & I_k \\ \vdots & \iddots & \iddots & \star \\ 0_k & \iddots & \iddots & \vdots \\ I_k & \star & \cdots & \star \end{pmatrix}$$

is already of full rank kp. Therefore, the controllability matrix $\mathcal{C}$ (cf. Theorem 2.15) is also of full rank. Since $\Sigma_L > 0$, Remark 2.17 provides that $c_{XX}(0) > 0$. Therefore, y is the zero vector and relation (8.5) is valid.

(b) Note that all components of $\mathcal{Y}_V$ are $(p-1)$-times mean-square differentiable, but the p-th derivative does not exist, see Remark 2.25. Thus, we have $j_v = p-1$ for any $v \in V$. Then, based on Definition 4.7 and Theorem 8.3, the proof is analogous to the proof of (a). ■

For Ornstein-Uhlenbeck processes, setting $p=1$ in Proposition 8.14 gives the following simplified characterisations.

Corollary 8.15. *Let $\mathcal{Y}_V$ be an Ornstein-Uhlenbeck process with $\Sigma_L > 0$. Further, let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[e^{\mathbf{A}h}\right]_{ba} = 0 \quad \forall\, 0 \leq h \leq 1,$

(b) $\mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad [\mathbf{A}]_{ba} = 0.$

The characterisations of (local) Granger causality in Proposition 8.14 and Corollary 8.15 are convenient, since we no longer need to elaborately compute and compare orthogonal projections as in Theorem 4.4 and Definition 4.7. Moreover, the deterministic criteria depend only on the state transition matrix $\mathbf{A}$ and not on the driving Lévy process. Below, we also provide a characterisation of the *undirected edges* in the (local) causality graph by model parameters of the MCAR process.

Proposition 8.16. *Let $\mathcal{Y}_V$ be an MCAR(p) process. Further, let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[\int_0^{\min(h,\widetilde{h})} \mathbf{C} e^{\mathbf{A}(h-u)} \mathbf{B} \Sigma_L \mathbf{B}^\top e^{\mathbf{A}^\top(\widetilde{h}-u)} \mathbf{C}^\top du\right]_{ab} = 0 \quad \forall\, 0 \leq h, \widetilde{h} \leq 1,$

(b) $\mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad [\Sigma_L]_{ab} = 0.$

Proof.

(a) Let $a, b, v \in V$, $t \in \mathbb{R}$, and $0 \leq h, \widetilde{h} \leq 1$. Remark 8.4 and equation (2.6) result in

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

Thus, $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V$ if and only if, for all $0 \leq h, \widetilde{h} \leq 1$,

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

(b) Let $a, b, v \in V$, $t \in \mathbb{R}$, and $h \geq 0$. An application of Theorem 8.3 yields

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

So, similar to (a), we obtain

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

If we set $f(u) = \mathbf{E}_p^\top e^{\mathbf{A}u} \mathbf{B} \Sigma_L \mathbf{B}^\top e^{-\mathbf{A}^\top u} \mathbf{E}_p$, $0 \leq u \leq 1$, and F as its primitive function, we get

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

■

For Ornstein-Uhlenbeck processes, these results are simplified to the following.

Corollary 8.17. *Let $\mathcal{Y}_V$ be an Ornstein-Uhlenbeck process. Further, let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[\int_0^{\min(h,\widetilde{h})} e^{\mathbf{A}(h-u)} \Sigma_L e^{\mathbf{A}^\top (\widetilde{h}-u)} du\right]_{ab} = 0 \quad \forall\, 0 \leq h, \widetilde{h} \leq 1,$

(b) $\mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad [\Sigma_L]_{ab} = 0.$

The characterisations of components being (locally) contemporaneously uncorrelated in Proposition 8.16 and Corollary 8.17 are again convenient, as they depend only on the model parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and Σ_L. We also emphasise that the local concept depends only on Σ_L. Below, we relate the characterisations from Propositions 8.14 and 8.16 to the current *literature*. We start with results from Comte and Renault (1996) in continuous time.

Remark 8.18. Comte and Renault (1996) investigate non-stationary MCAR processes driven by Brownian motions on local Granger causality and local instantaneous causality. Their definitions are set in the context of semi-martingales, use conditional expectations instead of orthogonal projections, and thus differ from ours. However, Comte and Renault (1996) also obtain in Proposition 20 that $\mathcal{Y}_a$ does not locally Granger cause $\mathcal{Y}_b$ if and only if $[A_j]_{ba} = 0$ for $j = 1, \ldots, p$. Further, there is no local instantaneous causality between $\mathcal{Y}_a$ and $\mathcal{Y}_b$ if and only if $[\Sigma_L]_{ab} = 0$.

We further compare our results for the continuous-time MCAR(p) process with those for the discrete-time vector AR(p) (VAR(p)) process by Eichler (2007), whose article provided the basis for our considerations. We start with the local causality graph, as the comparison is obvious for this graphical model.

Remark 8.19. The k-dimensional VAR(p) process $\mathcal{Z}_V = (Z_V(t))_{t\in\mathbb{Z}}$ is defined by

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

where $\Phi_n \in M_k(\mathbb{R})$, $n = 1, \ldots, p$, are the VAR(p) coefficients and $(\varepsilon(t))_{t\in\mathbb{Z}}$ is a k-dimensional white noise process with covariance matrix $0 < \Sigma_\varepsilon \in M_k(\mathbb{R})$. Further, define the AR polynomial $\Phi(\lambda) = I_k - \Phi_1\lambda - \ldots - \Phi_p\lambda^p$, $\lambda \in \mathbb{C}$, and denote the back-shift operator by $\mathtt{B}$. Then

$$\Phi(\mathtt{B})Z_V(t) = \varepsilon(t),$$

which corresponds to the idea for an MCAR(p) process to be the solution of

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

where $P(\lambda) = I_k\lambda^p + A_1\lambda^{p-1} + \ldots + A_p$, $\lambda \in \mathbb{C}$. Let $G = (V, E)$ be the path diagram of $\mathcal{Z}_V$, as in Definition 2.1 by Eichler (2007).

(a) *Directed edges:* Lemma 2.3 and Definition 2.1 in Eichler (2007) state that the directed edges in the path diagram G for the discrete-time VAR(p) process $\mathcal{Z}_V$ satisfy

$$\mathcal{Z}_a \not\rightarrow \mathcal{Z}_b \mid \mathcal{Z}_V \quad \Leftrightarrow \quad a \longrightarrow b \notin E \quad \Leftrightarrow \quad [\Phi_j]_{ba} = 0 \quad \forall\, j = 1, \ldots, p.$$

This characterisation is directly analogous to the characterisation of the directed edges in the local causality graph G^0_{CG} for an MCAR(p) process, where

$$\mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_V \quad \Leftrightarrow \quad a \longrightarrow b \notin E^0_{CG} \quad \Leftrightarrow \quad [A_j]_{ba} = 0 \quad \forall\, j = 1, \ldots, p.$$

The continuous-time and discrete-time models have in common that there is no directed edge from vertex a to b if and only if the ba-th components of the autoregressive coefficients are zero.

(b) *Undirected edges:* For the undirected edges in the path diagram G for the VAR(p) process $\mathcal{Z}_V$, Lemma 2.3 and Definition 2.1 in Eichler (2007) state

$$\mathcal{Z}_a \nsim \mathcal{Z}_b \mid \mathcal{Z}_V \quad \Leftrightarrow \quad a \dashrightarrow b \notin E \quad \Leftrightarrow \quad [\Sigma_\varepsilon]_{ab} = 0.$$

This characterisation is again in analogy to the condition for the undirected edges in the local causality graph G^0_{CG} for an MCAR(p) process, where

$$\mathcal{Y}_a \nsim_0 \mathcal{Y}_b \mid \mathcal{Y}_V \quad \Leftrightarrow \quad a \dashrightarrow b \notin E^0_{CG} \quad \Leftrightarrow \quad [\Sigma_L]_{ab} = 0.$$

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

Next, we compare the path diagram for the VAR process with the causality graph for the MCAR process. To do this, we first give an additional interpretation for the causality graph in Interpretation 8.20. Then, we compare the two graphical models in Remark 8.21.

Interpretation 8.20. For the purpose of interpretation of the directed and undirected edges in the causality graph G_{CG}, we recall the representation of the component $Y_b(t+h)$ of the MCAR process $\mathcal{Y}_V$ from Lemma 8.1 as

$$Y_b(t+h) = \sum_{j=1}^{p} e_b^\top \mathbf{M}_j(h) D^{(j-1)} Y_V(t) + e_b^\top \varepsilon(t,h), \tag{8.6}$$

where

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

(a) *Directed edges:* For the directed edges in the causality graph G_{CG}, an application of Proposition 8.14 gives the characterisation

$$\mathcal{Y}_a \not\to \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad [\mathbf{M}_j(h)]_{ba} = 0 \quad \forall\, 0 \le h \le 1,\ j = 1, \dots, p. \tag{8.7}$$

This statement means that the components $Y_a(t)$, $DY_a(t)$,..., $D^{(p-1)}Y_a(t)$ in the representation (8.6) vanish, because the corresponding prefactors are zero. $Y_a(t)$ and its derivatives do not matter to calculate $Y_b(t+h)$ some time step h into the future.

(b) *Undirected edges:* A consequence of Proposition 8.16 is the following characterisation for the undirected edges in the causality graph G_{CG}

$$\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[\mathbb{E}\left[\varepsilon(t,h)\varepsilon(t,\widetilde{h})^\top\right]\right]_{ab} = \left[\mathbb{E}\left[\varepsilon(0,h)\varepsilon(0,\widetilde{h})^\top\right]\right]_{ab} = 0 \quad \forall\, 0 \le h, \widetilde{h} \le 1. \tag{8.8}$$

This statement means that the noise terms $e_a^\top \varepsilon(t,h)$ and $e_b^\top \varepsilon(t,\widetilde{h})$ of $Y_a(t+h)$ and $Y_b(t+\widetilde{h})$ are uncorrelated for any $t \ge 0$ and $0 \le h, \widetilde{h} \le 1$.

Remark 8.21. The characterisations of the directed and undirected edges in the causality graph in Interpretation 8.20 are well suited for comparison to the path diagram for VAR processes by Eichler (2007). The challenge is that in the representation of the continuous-time process in (8.6) derivatives appear, that have to be related to appropriate differences in the discrete-time process

$$Z_V(t+1) = \sum_{n=1}^{p} \Phi_n Z_V(t+1-n) + \varepsilon(t+1), \quad t \in \mathbb{Z}. \tag{8.9}$$

Thus, our first goal is to replace the back-shifts $Z_V(t+1-n)$, $n = 1, \dots, p$, in equation (8.9) by differences. To do this, we define a discrete-time difference operator iteratively by

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

$j = 1, \dots, p-1$, where we set $\mathtt{D}^{(0)} Z_V(t) = Z_V(t)$. By induction (cf. Lemma A.1), one can show

$$Z_V(t+1-n) = \sum_{j=1}^{n} \binom{n-1}{j-1} (-1)^{j-1} \mathtt{D}^{(j-1)} Z_V(t)$$

for $n = 1, \ldots, p$ and $t \in \mathbb{Z}$. Then, we obtain the representation of the VAR(p) process

$$\begin{aligned} Z_V(t+1) &= \sum_{n=1}^{p} \sum_{j=1}^{n} \binom{n-1}{j-1} (-1)^{j-1} \Phi_n \mathtt{D}^{(j-1)} Z_V(t) + \varepsilon(t+1) \\ &= \sum_{j=1}^{p} \sum_{n=j}^{p} \binom{n-1}{j-1} (-1)^{j-1} \Phi_n \mathtt{D}^{(j-1)} Z_V(t) + \varepsilon(t+1). \end{aligned}$$

Now define

$$\mathbf{M}_j \coloneqq \sum_{n=j}^{p} \binom{n-1}{j-1} (-1)^{j-1} \Phi_n, \quad j = 1, \ldots, p.$$

This gives the representation of the b-th component

$$Z_b(t+1) = \sum_{j=1}^{p} e_b^\top \mathbf{M}_j \mathtt{D}^{(j-1)} Z_V(t) + e_b^\top \varepsilon(t+1),$$

which is in analogy to representation (8.6) of the b-th component of an MCAR(p) processes.

(a) *Directed edges:* In Remark 8.19, we find that for the discrete-time VAR(p) process $\mathcal{Z}_V$, the directed edges in the path diagram G satisfy

$$\mathcal{Z}_a \not\rightarrow \mathcal{Z}_b \mid \mathcal{Z}_V \quad \Leftrightarrow \quad a \longrightarrow b \notin E \quad \Leftrightarrow \quad [\Phi_j]_{ba} = 0 \quad \forall\, j = 1, \ldots, p.$$

But,

$$[\Phi_j]_{ba} = 0 \quad \forall\, j = 1, \ldots, p \quad \Leftrightarrow \quad [\mathbf{M}_j]_{ba} = \sum_{n=j}^{p} \binom{n-1}{j-1} (-1)^{j-1} [\Phi_n]_{ba} = 0 \ \ \forall\, j = 1, \ldots, p.$$

This characterisation is again analogous to the characterisation of the directed edges in the causality graph G_{CG} for the MCAR(p) process in relation (8.7), where

$$\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V \quad \Leftrightarrow \quad a \longrightarrow b \notin E_{CG} \quad \Leftrightarrow \quad [\mathbf{M}_j(h)]_{ba} = 0 \quad \forall\, j = 1, \ldots, p,\ 0 \leq h \leq 1.$$

(b) *Undirected edges:* For the path diagram G for the VAR(p) process $\mathcal{Z}_V$, Remark 8.19 yields

$$\mathcal{Z}_a \nsim \mathcal{Z}_b \mid \mathcal{Z}_V \quad \Leftrightarrow \quad a \dashline b \notin E \quad \Leftrightarrow \quad \left[\mathbb{E}\left[\varepsilon(0)\varepsilon(0)^\top\right]\right]_{ab} = 0.$$

In this case, we have the similarity to the condition (8.8) for the undirected edges for the MCAR(p) process in the causality graph G_{CG}

$$\mathcal{Y}_a \nsim \mathcal{Y}_b \mid \mathcal{Y}_V \quad \Leftrightarrow \quad a \dashline b \notin E_{CG} \quad \Leftrightarrow \quad \left[\mathbb{E}\left[\varepsilon(0,h)\varepsilon(0,\widetilde{h})^\top\right]\right]_{ab} = 0 \quad \forall\, 0 \leq h, \widetilde{h} \leq 1.$$

As a final comparison of the causality graph for MCAR processes and the path diagram for VAR processes, consider the special case of the Ornstein-Uhlenbeck process. The essential feature is that a continuous-time Ornstein-Uhlenbeck process sampled at discrete and equidistant points in time is a discrete-time VAR(1) process. We can therefore establish relationships between the graphical models for the Ornstein-Uhlenbeck process and its sampled process.

Remark 8.22. Let $\mathcal{Y}_V$ be an Ornstein-Uhlenbeck process. Then $\mathcal{Y}_V$ sampled at time points of distance $0 \leq h \leq 1$ is a discrete-time VAR(1) process with representation

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

We denote this process by $\mathcal{Y}_V^{(h)} = (Y_V((k+1)h))_{k\in\mathbb{Z}}$ and the corresponding discrete-time path diagram by $G^{(h)} = (V, E^{(h)})$. Then, a direct consequence of Remark 8.19, Interpretation 8.20, Corollary 8.15, and Corollary 8.17, is that, for $a, b \in V$, $a \neq b$, and fixed $0 \leq h \leq 1$,

(a) $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Rightarrow \quad \left[e^{\mathbf{A}h}\right]_{ba} = 0 \quad \Rightarrow \quad \mathcal{Y}_a^{(h)} \not\rightarrow \mathcal{Y}_b^{(h)} \,|\, \mathcal{Y}_V^{(h)}$,

(b) $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Rightarrow \quad \left[\mathbb{E}[\varepsilon(0,h)\varepsilon(0,h)^\top]\right]_{ab} = 0 \quad \Rightarrow \quad \mathcal{Y}_a^{(h)} \nsim \mathcal{Y}_b^{(h)} \,|\, \mathcal{Y}_V^{(h)}$.

These implications state that a directed (undirected) edge $a \longrightarrow b \in E^{(h)}$ ($a \;\text{---}\; b \in E^{(h)}$) in the discrete-time model implies as well a directed (undirected) edge $a \longrightarrow b \in E_{CG}$ ($a \;\text{---}\; b \in E_{CG}$) in the continuous-time model. In general, however, the opposite is not true. In summary, $E^{(h)} \subseteq E_{CG}$ for every $0 \leq h \leq 1$. We believe that this result applies to general MCAR(p) processes.

This phenomenon is an advantage of the causality graph over the local causality graph, since, in the local causality graph, there is generally no relationship between the edges $E_{CG}^{(0)}$ and $E^{(h)}$.

The characterisations of the edges in Propositions 8.14 and 8.16 for the causality graph are nice for interpretation and comparison to the literature but cumbersome, because they depend on the lags $h, \widetilde{h}$. Below, we provide *simpler characterisations* of the directed and undirected edges in the causality graph, where the lags $h, \widetilde{h}$ no longer play a role. These characterisations allow us to obtain further relations between the causality graph and the local causality graph, as well as to discuss the existence of given mixed graphs as a (local) causality graph.

Theorem 8.23. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $\Sigma_L > 0$. Further, let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad [\mathbf{C}\mathbf{A}^\alpha \mathbf{E}_j]_{ba} = [\mathbf{A}^\alpha]_{b,k(j-1)+a} = 0 \quad \forall\, \alpha = 1, \ldots, kp-1,\ j = 1, \ldots, p,$

(b) $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[\mathbf{C}\mathbf{A}^\alpha \mathbf{B}\Sigma_L \mathbf{B}^\top (\mathbf{A}^\top)^\beta \mathbf{C}^\top\right]_{ab} = 0 \quad \forall\, \alpha, \beta = 0, \ldots, kp-1.$

Proof.
(a) $\Leftarrow$: Suppose that $e_b^\top \mathbf{C}\mathbf{A}^\alpha \mathbf{E}_j e_a = 0$ for all $\alpha = 1, \ldots, kp-1$ and $j = 1, \ldots, p$. Bernstein (2009) provides in equation (11.2.1) that, for $h \in \mathbb{R}$,

$$e^{\mathbf{A}h} = \sum_{\alpha=0}^{kp-1} \psi_\alpha(h)\mathbf{A}^\alpha, \quad \text{where} \quad \psi_\alpha(h) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{\chi_{\mathbf{A}}^{[\alpha+1]}(z)}{\chi_{\mathbf{A}}(z)} e^{tz} dz, \tag{8.10}$$

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

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

for $0 \leq h \leq 1$, such that Proposition 8.14(a) results in $\mathcal{Y}_a \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V$.

$\Rightarrow$: Assume that $\mathcal{Y}_a \not\rightarrow\!\!\!\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V$. Thus, $e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_a = 0$ for all $0 \leq h \leq 1$ and $j = 1, \ldots, p$ by Proposition 8.14(a). Further, for $h \in \mathbb{R}$, define the function f by

$$f(h) = e_b^\top \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_j e_a.$$

We differentiate f using Bernstein (2009), Proposition 11.1.4. Then

$$f^{(\alpha)}(h) = e_b^\top \mathbf{C}\mathbf{A}^\alpha e^{\mathbf{A}h} \mathbf{E}_j e_a \quad \forall\, h \in \mathbb{R},\ \alpha = 1, \ldots, kp-1,$$

where $f^{(\alpha)}$ denotes the α-th derivative. Since $f(h) = 0$ for $0 \leq h \leq 1$ and $f^{(\alpha)}$ is continuous, we obtain $f^{(\alpha)}(h) = 0$ for $0 \leq h \leq 1$. Putting $h = 0$, we get, as claimed,

$$0 = e_b^\top \mathbf{C}\mathbf{A}^\alpha \mathbf{E}_j e_a \quad \forall\, \alpha = 1, \ldots, kp-1,\ j = 1, \ldots, p.$$

(b) $\Leftarrow$: Assume that $e_a^\top \mathbf{C}\mathbf{A}^\alpha \mathbf{B}\Sigma_L \mathbf{B}^\top (\mathbf{A}^\top)^\beta \mathbf{C}^\top e_b = 0$ for all $\alpha, \beta = 0, \ldots, kp-1$. We apply the representation of the matrix exponential (8.10) and obtain

$$\begin{aligned} &e_a^\top \mathbf{C} \int_0^{\min(h,\widetilde{h})} e^{\mathbf{A}(h-s)} \mathbf{B}\Sigma_L \mathbf{B}^\top e^{\mathbf{A}^\top(\widetilde{h}-s)} ds\, \mathbf{C}^\top e_b \\ &= \sum_{\alpha=0}^{kp-1} \sum_{\beta=0}^{kp-1} \int_0^{\min(h,\widetilde{h})} \psi_\alpha(h-s)\varphi_\beta(\widetilde{h}-s) e_a^\top \mathbf{C}\mathbf{A}^\alpha \mathbf{B}\Sigma_L \mathbf{B}^\top \left(\mathbf{A}^\top\right)^\beta \mathbf{C}^\top e_b\, ds = 0 \end{aligned}$$

for $0 \leq h, \widetilde{h} \leq 1$, $t \in \mathbb{R}$, by assumption. Proposition 8.16(a) then gives $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V$.

$\Rightarrow$: Suppose that $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V$. Due to Proposition 5.4, we have, for $0 \leq h \leq 1$ and $t \in \mathbb{R}$,

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

In addition, Remark 8.4 provides that $P_{\mathcal{L}_{Y_V}(t)} Y_a(t+h) = e_a^\top \mathbf{C} e^{\mathbf{A}h} X(t)$. Both together yield

$$P_{\mathcal{L}_{Y_V}(t) \vee \mathcal{L}_{Y_b}(t,t+1)} Y_a(t+h) = e_a^\top \mathbf{C} e^{\mathbf{A}h} X(t) \quad \mathbb{P}\text{-a.s.} \tag{8.11}$$

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

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

Plugging in equations (8.11) and (2.5), it follows that

$$\begin{aligned}0 &= \mathbb{E}\left[\left(Y_a(t+h) - e_a^\top \mathbf{C} e^{\mathbf{A}h} X(t)\right) Y_b(t+\widetilde{h})\right] \\ &= e_a^\top \mathbf{C}\, \mathbb{E}\left[\left(X(t+h) - e^{\mathbf{A}h} X(t)\right) X(t+\widetilde{h})\right] \mathbf{C}^\top e_b \\ &= e_a^\top \mathbf{C} \left(c_{XX}(h-\widetilde{h}) - e^{\mathbf{A}h} c_{XX}(-\widetilde{h})\right) \mathbf{C}^\top e_b\end{aligned}$$

for $0 \le h, \widetilde{h} \le 1$. If we consider only the case $0 \le \widetilde{h} \le h \le 1$, equation (2.8) provides

$$\begin{aligned}0 &= e_a^\top \mathbf{C} \left(e^{\mathbf{A}(h-\widetilde{h})} c_{XX}(0) - e^{\mathbf{A}h} c_{XX}(0) e^{\mathbf{A}^\top \widetilde{h}}\right) \mathbf{C}^\top e_b \\ &= e_a^\top \mathbf{C} e^{\mathbf{A}h} \left(e^{-\mathbf{A}\widetilde{h}} c_{XX}(0) - c_{XX}(0) e^{\mathbf{A}^\top \widetilde{h}}\right) \mathbf{C}^\top e_b,\end{aligned}$$

using Bernstein (2009), Corollary 11.1.6. Now, for $h, \widetilde{h} \in \mathbb{R}$, we define the function γ by

$$\gamma(h, \widetilde{h}) = e_a^\top \mathbf{C} e^{\mathbf{A}h} \left(e^{-\mathbf{A}\widetilde{h}} c_{XX}(0) - c_{XX}(0) e^{\mathbf{A}^\top \widetilde{h}}\right) \mathbf{C}^\top e_b.$$

Differentiating γ several times according to Bernstein (2009), Proposition 11.1.4, provides

$$\frac{\partial^m}{\partial h^m} \frac{\partial^n}{\partial \widetilde{h}^n} \gamma(h, \widetilde{h}) = e_a^\top \mathbf{C} \mathbf{A}^m e^{\mathbf{A}h} \left((-\mathbf{A})^n e^{-\mathbf{A}\widetilde{h}} c_{XX}(0) - c_{XX}(0) \left(\mathbf{A}^\top\right)^n e^{\mathbf{A}^\top \widetilde{h}}\right) \mathbf{C}^\top e_b, \quad m, n \in \mathbb{N}_0.$$

Since $\gamma(h, \widetilde{h}) = 0$ for $0 \le \widetilde{h} \le h \le 1$, and due to the continuity of the functions under consideration, we obtain that the derivatives are zero for $0 \le \widetilde{h} \le h \le 1$. Plugging in $h = \widetilde{h} = 0$ yields

$$e_a^\top \mathbf{C} \mathbf{A}^m c_{XX}(0) \left(\mathbf{A}^\top\right)^n \mathbf{C}^\top e_b = e_a^\top \mathbf{C} \mathbf{A}^m (-\mathbf{A})^n c_{XX}(0) \mathbf{C}^\top e_b, \quad m, n \in \mathbb{N}_0. \tag{8.12}$$

Finally, equations (2.9) and (8.12) lead to

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

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

In the case of an Ornstein-Uhlenbeck process, the characterisation of the edges in the causality graph can be reduced to the following.

Corollary 8.24. *Let $\mathcal{Y}_V$ be an Ornstein-Uhlenbeck process with $\Sigma_L > 0$. Further, let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \not\to \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad [\mathbf{A}^\alpha]_{ba} = 0 \qquad \forall\, \alpha = 1, \dots, k-1,$

(b) $\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Leftrightarrow \quad \left[\mathbf{A}^\alpha \Sigma_L (\mathbf{A}^\top)^\beta\right]_{ab} = 0 \qquad \forall\, \alpha, \beta = 0, \dots, k-1.$

Building on the characterisations in Theorem 8.23 and Corollary 8.24, we make some additional remarks. We start with the *relationships* between global Granger causality, Granger causality, and local Granger causality. Analogously, we consider the relationships between global contemporaneous correlation, contemporaneous correlation, and local contemporaneous correlation.

Remark 8.25. The characterisations in Theorem 8.23 no longer depend on $h, \widetilde{h}$. Hence, Granger non-causality and global Granger non-causality, as well as contemporaneous uncorrelatedness and global contemporaneous uncorrelatedness are equivalent for MCAR processes.

Remark 8.26. In Proposition 4.17 and 5.10, we establish in a general setting that Granger non-causality implies local Granger non-causality and contemporaneous uncorrelatedness implies local contemporaneous uncorrelatedness. These relationships are now also evident in the characterisations for MCAR processes. To recognise this fact, observe that, due to the structure of $\mathbf{A}$, multiplying $\mathbf{A}$ by itself means deleting the first row of matrices in $\mathbf{A}$, moving all the other rows up one and adding a new row at the bottom. Inductively, it thus follows that

$$\mathbf{C}\mathbf{A}^{p-1} = \begin{pmatrix} 0_k & \cdots & 0_k & I_k \end{pmatrix} \quad \text{and} \quad \mathbf{C}\mathbf{A}^{p} = \begin{pmatrix} -A_p & \cdots & -A_1 \end{pmatrix}. \tag{8.13}$$

If we now choose $\alpha = p$, then relation (8.13), Proposition 8.14, and Theorem 8.23 yield

$$\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Rightarrow \quad 0 = [\mathbf{C}\mathbf{A}^p\mathbf{E}_j]_{ba} = -\left[A_{p-j+1}\right]_{ba} \quad \forall\, j = 1, \ldots, p \quad \Leftrightarrow \quad \mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V.$$

Similarly, choosing $\alpha = \beta = p-1$, relation (8.13), Proposition 8.16, and Theorem 8.23 give

$$\mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V \quad \Rightarrow \quad 0 = \left[\mathbf{C}\mathbf{A}^{p-1}\mathbf{B}\Sigma_L\mathbf{B}^\top(\mathbf{A}^{p-1})^\top\mathbf{C}^\top\right]_{ab} = [\Sigma_L]_{ab} \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V.$$

The desired relationships between the edges in the causality graph and the local causality graph are valid and in line with the theory. We have $E_{CG}^{(0)} \subseteq E_{CG}$ and, in general, the sets are not equal, because the opposite implications are usually not true.

We emphasise that if there is no directed (undirected) edge in the causality graph, not only is there no directed (undirected) edge in the local causality graph, but also the paths in the local causality graph are restricted. Such relations are the subject of the next remark, where we focus on Ornstein-Uhlenbeck processes.

Remark 8.27. Let $a, b \in V$ with $a \neq b$.

(a) *Directed edges:* Suppose $a \longrightarrow b \notin E_{CG}$. For $\alpha = 1$, Corollary 8.24(a) states, as expected, that $a \longrightarrow b \notin E_{CG}^0$. For $\alpha = 2$, Corollary 8.24(a) gives

$$0 = \left[\mathbf{A}^2\right]_{ba} = \sum_{c_1 \in \{1,\ldots,k\}} \left([\mathbf{A}]_{bc_1} [\mathbf{A}]_{c_1 a}\right).$$

That is, in the local causality graph, the indirect influence of component a on b via one intermediate component c_1 is cancelled out. The purely directed paths $a \longrightarrow c_1 \longrightarrow b$ in the local causality graph compensate each other. More generally, Corollary 8.24(a) states that

$$0 = [\mathbf{A}^\alpha]_{ba} = \sum_{\substack{c_1,\dots,c_{\alpha-1} \\ \in\{1,\dots,k\}}} \left([\mathbf{A}]_{bc_1} \cdots [\mathbf{A}]_{c_{\alpha-1}a}\right).$$

That is, in the local causality graph, the indirect influence of component a on b via paths $a \longrightarrow c_1 \longrightarrow \cdots \longrightarrow c_{\alpha-1} \longrightarrow b$ is cancelled out. The purely directed paths $a \longrightarrow \cdots \longrightarrow b$ of equal length offset each other.

Note that paths containing intermediate vertices a or b can be omitted because they contain shorter sub-paths that have already been considered. Formally, a simple calculation shows

$$\begin{aligned} & 0 = \sum_{\substack{c_1,\dots,c_{\alpha-1} \\ \in\{1,\dots,k\}}} \left([\mathbf{A}]_{bc_1} \cdots [\mathbf{A}]_{c_{\alpha-1}a}\right) && \text{for } \alpha = 1,\dots,k \\ \Leftrightarrow\quad & 0 = \sum_{\substack{c_1,\dots,c_{\alpha-1} \\ \in\{1,\dots,k\}\setminus\{a,b\}}} \left([\mathbf{A}]_{bc_1} \cdots [\mathbf{A}]_{c_{\alpha-1}a}\right) && \text{for } \alpha = 1,\dots,k. \end{aligned}$$

Further, observe that if we replace one of the edges (except $a \longrightarrow c_1$) with a directed edge in the other direction or an undirected edge, the resulting path contains a collider $\longrightarrow \cdot \longleftarrow$ or $\longrightarrow \cdot$ ---. So it is not surprising that such paths do not need to be considered.

(b) *Undirected edges:* Suppose a --- $b \notin E_{CG}$. Then Corollary 8.24(b) states that

$$0 = \left[\mathbf{A}^\alpha \Sigma_L (\mathbf{A}^\top)^\beta\right]_{ab} = \sum_{i=1}^{k} \sum_{j=1}^{k} [\mathbf{A}^\alpha]_{ai} [\Sigma_L]_{ij} \left[\mathbf{A}^\beta\right]_{bj}$$

for $\alpha, \beta = 0,\dots,k-1$, which means that paths of the form $a \longleftarrow \cdots \longleftarrow i$ --- $j \longrightarrow \cdots \longrightarrow b$ in the local causality graph compensate each other, i.e., paths with α directed edges in direction a, β directed edges in direction b, and an undirected edge in between. The case $\alpha = \beta = 0$ provides a --- $b \notin E^0_{CG}$. Furthermore, replacing one of the directed edges leads to a path with a collider $\longrightarrow \cdot \longleftarrow$, --- $\cdot \longleftarrow$, $\longrightarrow \cdot$ --- or --- $\cdot$ --- on that path. Again, not surprisingly, such paths do not need to be considered.

Finally, we make some remarks about the *existence* of a given mixed graph as a (local) causality graph. We start with the local causality graph since its existence is obvious.

Proposition 8.28. *Suppose that $G = (V, E)$ is a mixed graph. Then, there exists a k-dimensional Ornstein-Uhlenbeck process and a local causality graph $G^0_{CG} = (V, E^0_{CG})$, such that, for $a, b \in V$ with $a \neq b$,*

$$\begin{aligned} a \longrightarrow b \notin E \quad &\Leftrightarrow \quad [\mathbf{A}]_{ba} = 0 \quad \Leftrightarrow \quad a \longrightarrow b \notin E^0_{CG}, \\ a \text{ --- } b \notin E \quad &\Leftrightarrow \quad [\Sigma_L]_{ab} = 0 \quad \Leftrightarrow \quad a \text{ --- } b \notin E^0_{CG}. \end{aligned}$$

Proof. We define the matrices $\mathbf{A}$ and Σ_L as follows.

$$[\mathbf{A}]_{ab} = \begin{cases} -k, & \text{if } a = b, \\ 1, & \text{if } a \neq b \text{ and } b \longrightarrow a \in E, \\ 0, & \text{if } a \neq b \text{ and } b \longrightarrow a \notin E, \end{cases} \quad [\Sigma_L]_{ab} = \begin{cases} k, & \text{if } a = b, \\ 1, & \text{if } a \neq b \text{ and } a \text{ --- } b \in E, \\ 0, & \text{if } a \neq b \text{ and } a \text{ --- } b \notin E. \end{cases}$$

The deleted absolute row sums of $\mathbf{A}$ satisfy by definition

$$r_a(\mathbf{A}) := \sum_{\substack{b=1\\ b\neq a}}^{k} |[\mathbf{A}]_{ab}| \leq k-1.$$

Then the Gershgorin discs satisfy

$$\begin{aligned}\{z \in \mathbb{C} : |z - [\mathbf{A}]_{aa}| \leq r_a(\mathbf{A})\} = \{z \in \mathbb{C} : |z+k| \leq r_a(\mathbf{A})\} \subseteq \{z \in \mathbb{C} : |z+k| \leq k-1\} \\ \subseteq (-\infty, 0) + i\mathbb{R}.\end{aligned}$$

Since, according to the Gershgorin circle theorem (Horn & Johnson, 2013, Theorem 6.1.1), all eigenvalues of $\mathbf{A}$ lie in the union of the Gershgorin discs, $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ follows. Additionally, the symmetric matrix Σ_L is strictly diagonally dominant, i.e., $\Sigma_L > 0$ (Horn & Johnson, 2013, Theorem 6.1.10). Furthermore, take $L(t) = \Sigma_L^{1/2} B(t)$, where $(B(t))_{t\in\mathbb{R}}$ is a k-dimensional Brownian motion and $\Sigma_L^{1/2}$ is the positive square root. Then $L = (L(t))_{t\in\mathbb{R}}$ is a Levy process with covariance matrix Σ_L (Cont & Tankov, 2003, Theorem 4.1). The resulting k-dimensional Ornstein-Uhlenbeck process $\mathcal{Y}_V$ generates a local causality graph $G^0_{CG} = (V, E^0_{CG})$ based on Corollaries 8.15 and 8.17, which is identical to the undirected graph $G = (V, E)$. ■

For the causality graph, as expected, the existence question is much more difficult, because of the additional conditions on the powers of $\mathbf{A}$. We cannot expect every graph to exist. In fact, not even every 2-dimensional mixed graph can be constructed as a causality graph for an MCAR process. To justify this claim, we give an auxiliary lemma.

Lemma 8.29. *Let $\mathcal{Y}_V$ be an MCAR(p) process with $V = \{1, 2\}$ and $\Sigma_L > 0$. Then*

$$1 \dashrightarrow\!\!\!\!\!\dashleftarrow 2 \notin E_{CG} \quad \Rightarrow \quad 1 \longrightarrow 2 \notin E_{CG} \quad \textit{and} \quad 2 \longrightarrow 1 \notin E_{CG}.$$

Proof. Suppose that $1 \text{ - - - } 2 \notin E_{CG}$. We perform the proof in three steps. First, we show that Σ_L is diagonal. Then, we prove that A_j is diagonal for $j = 1, \dots, p$. Finally, we conclude the assertion.

Step 1: The relation $1 \text{ - - - } 2 \notin E_{CG}$ implies $1 \text{ - - - } 2 \notin E^0_{CG}$ (cf. Remark 8.26), i.e., $[\Sigma_L]_{12} = 0$ and, for symmetry reasons, $[\Sigma_L]_{21} = 0$ also applies. So Σ_L is diagonal.

Step 2: Let $\alpha = p + j$ for $j = 0, \dots, p-1$ and $\beta = p - 1$, and vice versa. Then the assumption $1 \text{ - - - } 2 \notin E_{CG}$, Theorem 8.23(b), and equation (8.13) imply

$$\begin{aligned}0 &= \left[\mathbf{C}\mathbf{A}^{p+j}\mathbf{B}\Sigma_L\mathbf{B}^\top \left(\mathbf{A}^\top\right)^{p-1} \mathbf{C}^\top\right]_{12} = \left[\mathbf{C}\mathbf{A}^{p+j}\mathbf{B}\Sigma_L\right]_{12} \qquad \text{and} \\ 0 &= \left[\mathbf{C}\mathbf{A}^{p-1}\mathbf{B}\Sigma_L\mathbf{B}^\top \left(\mathbf{A}^\top\right)^{p+j} \mathbf{C}^\top\right]_{12} = \left[\Sigma_L\mathbf{B}^\top \left(\mathbf{A}^\top\right)^{p+j} \mathbf{C}^\top\right]_{12}.\end{aligned}$$

Step 1 together with $\Sigma_L > 0$ gives

$$\left[\mathbf{C}\mathbf{A}^{p+j}\mathbf{B}\right]_{12} = 0 \quad \text{and} \quad \left[\mathbf{C}\mathbf{A}^{p+j}\mathbf{B}\right]_{21} = 0 \tag{8.14}$$

for $j = 0, \dots, p-1$. We now show that $[A_j]_{12} = [A_j]_{21} = 0$ holds for $j = 1, \dots, p$ by induction over j. The *base case* holds immediately if we set $j = 0$ in (8.14) and apply (8.13) to obtain

$$0 = [\mathbf{C}\mathbf{A}^p\mathbf{B}]_{12} = [-A_1]_{12} \quad \text{and} \quad 0 = [\mathbf{C}\mathbf{A}^p\mathbf{B}]_{21} = [-A_1]_{21} = 0.$$

Now fix some $j_0 \in \mathbb{N}$ with $1 \leq j_0 \leq p-1$ and assume that $[A_j]_{12} = [A_j]_{21} = 0$ holds for all $1 \leq j \leq j_0$. For the *induction step*, we set $j = j_0$ in (8.14) to obtain, by Lemma A.2 and (8.13),

$$0 = \left[\mathbf{C}\mathbf{A}^{p+j_0}\mathbf{B}\right]_{12} = \left[-\sum_{i=0}^{j_0} A_{j_0-i+1}\mathbf{C}\mathbf{A}^{p+i-1}\mathbf{B}\right]_{12} = \left[-A_{j_0+1} - \sum_{i=1}^{j_0} A_{j_0-i+1}\mathbf{C}\mathbf{A}^{p+i-1}\mathbf{B}\right]_{12}.$$

However, $[A_{j_0-i+1}]_{12} = [A_{j_0-i+1}]_{21} = 0$ and $[\mathbf{C}\mathbf{A}^{p+i-1}\mathbf{B}]_{12} = [\mathbf{C}\mathbf{A}^{p+i-1}\mathbf{B}]_{21} = 0$ for $i = 1, \dots, j_0$ according to the induction hypothesis and equations (8.14). This only leaves $0 = [-A_{j_0+1}]_{12}$. Similarly, we get $0 = [-A_{j_0+1}]_{21}$.

Step 3: Due to Step 2, the matrix $\mathbf{C}\mathbf{A}^\alpha\mathbf{E}_j$ is a sum of products of diagonal matrices. So $\mathbf{C}\mathbf{A}^\alpha\mathbf{E}_j$ itself is diagonal and $[\mathbf{C}\mathbf{A}^\alpha\mathbf{E}_j]_{21} = [\mathbf{C}\mathbf{A}^\alpha\mathbf{E}_j]_{12} = 0$ holds for $\alpha = 1, \dots, kp-1$ and $j = 1, \dots, p$. Then Theorem 8.23(a) gives $1 \longrightarrow 2 \notin E_{CG}$ and $1 \longrightarrow 2 \notin E_{CG}$. ■

Remark 8.30. The previous Lemma 8.29 shows that we cannot generate the graph $G = (V, E)$ with $V = \{1, 2\}$, $1 \longrightarrow 2 \in E$, but $1 \;\text{---}\; 2 \notin E$, as a causality graph for an MCAR process. A directed edge in a 2-dimensional causality graph always implies an undirected edge. We suspect that similar statements hold in higher dimensions. Therefore, the local causality graph G^0_{CG} allows the modelling of more general graphs than the causality graph G_{CG}, which is an advantage of the local causality graph.

CHAPTER 9

(Local) Causality graphs for ICCSS processes

In this chapter, we construct (local) causality graphs for the popular multivariate continuous-time *state space models*, more specifically for their output processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$, $V = \{1,\dots,k\}$ (cf. Chapter 2), in order to visualise the essential dependency structures of such processes. We further derive analytic representations of the edges in both graphs by model parameters.

In the context of MCAR processes, the topic of (local) causality graphs is discussed in Chapter 8. Although MCAR processes are state space processes, the techniques do not apply to, for example, MCARMA(p,q) processes with $q > 0$. The much simpler structure of an MCAR process allows the direct recovery of the input process $\mathcal{X}$ from the output process $\mathcal{Y}_V$. For most state space processes this is not possible. The present chapter can thus be seen as an extension of Chapter 8 to a broader class of models and, of course, we compare our results with those of that chapter.

To the best of our knowledge, the (local) causality graph is the first (mixed) graphical model for this broad class of stochastic processes. Even for discrete-time stochastic processes, the literature on mixed graphical models is limited to AR processes (Eichler, 2007, 2012). Not much is known about mixed graphical models that satisfy certain types of Markov properties for the discrete-time ARMA processes.

Even the *orthogonal projections* of multivariate state space processes and their derivatives required in this chapter have not yet been addressed in the existing literature. Although, as already mentioned, Rozanov (1967), III, 5, is devoted to the topic of prediction for general stationary processes, the representations in that book are based on specific maximal decompositions of the spectral density function, which are generally not expressible as a simple formula. The orthogonal projections of univariate CARMA processes were discussed in the previous paper by Brockwell and Lindner (2015). The authors provide representations for the orthogonal projection of a CARMA process at time t onto the linear space generated by the CARMA process up to time 0, and for the conditional expectation on the σ-algebra generated by the CARMA process up to time 0. A multivariate generalisation of the conditional expectation result using the σ-algebra generated by the MCARMA process up to time 0 can be found in Basse-O'Connor et al. (2019), Corollary 4.11. However, the statement is not consistent with the comprehensible univariate result of Brockwell and Lindner (2015). In any case, in this thesis, we require not only orthogonal projections of the components $Y_a(t+h)$, $a \in V$, $h \geq 0$, on the linear space of the past of the process $\mathcal{Y}_V$ up to time t but also on linear spaces generated by subprocesses $\mathcal{Y}_S$, $S \subset V$. Additionally, we require orthogonal projections of the highest derivatives of the components on these linear spaces.

To calculate the orthogonal projections, we have to overcome the challenge of recovering the input process $\mathcal{X}$ from the output process $\mathcal{Y}_V$, as highlighted in Basse-O'Connor et al. (2019) and

Section 2.4. Therefore, we have to make rather weak assumptions that result in *ICCSS processes*. We briefly summarise these assumptions: First, we assume that the driving Lévy process of a given state space model $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ satisfies the common Assumption 1. Furthermore, we require that the transfer function H of $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ has a coprime right polynomial matrix fraction description $H(z) = Q(z)P(z)^{-1}$ with

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

Then, without loss of generality, we can assume that the state space model is given in the unique controller canonical form $(\mathbf{A}, \mathbf{B}, \overline{\overline{\mathbf{C}}}, L)$ due to Proposition 2.18. Finally, for the controller canonical state space model, we require that $p > q > 0$,

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

In this case, we call the strictly stationary version of the output process $\mathcal{Y}_V$ an ICCSS(p, q) process (cf. Definition 2.27) and summarise the assumptions under this acronym. Of course, due to $q > 0$, the class of MCAR(p) models are excluded in the following considerations. However, $q > 0$ is not an essential limitation, because (local) causality graphs for MCAR processes are treated in Chapter 8.

For ICCSS processes, we are then able to derive *alternative representations* of $Y_a(t + h)$ and its highest derivative. We use these representations to compute the necessary *orthogonal projections*, resulting in the *characterisations of the edges* in the (local) causality graph. We find that these characterisations are interpretatively meaningful and are given by the model parameters of the controller canonical form and the covariance matrix of the driving Lévy process.

It should be noted that (local) causality graphs can be defined for all wide-sense stationary and mean-square continuous state space models $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ that satisfy Assumptions 3 and 4, as we discuss in Remark 9.16. However, in this general context, we are not able to compute the orthogonal projections necessary for the edge characterisations. Since the edge characterisations are the ultimate goal of the present chapter, we restrict our focus to ICCSS processes.

The structure of this chapter is similar to that of Chapter 8 as follows. In Section 9.1, we derive alternative representations of ICCSS processes and their $(p - q - 1)$-th derivatives. Second, we establish orthogonal projections of these processes onto linear spaces. In particular, we discuss the difference quotient of the $(p - q - 1)$-th derivative, its orthogonal projection, and its mean-square limit. We then show that ICCSS processes satisfy the assumptions of the (local) causality graph in Section 9.2, ensuring that both graphical models are well defined and satisfy the preferred Markov properties. Finally, in Section 9.3, we present the main results of the chapter. We characterise the directed and undirected edges using only the model parameters of the ICCSS process. We also provide detailed interpretations of the characterisations and, throughout the chapter, we draw connections to the results for MCAR processes.

9.1. Orthogonal projections of ICCSS processes

In this section, we derive orthogonal projections of ICCSS processes $\mathcal{Y}_V$ and their $(p-q-1)$-th derivative. These projections are not only essential to characterise the directed and undirected edges in the (local) causality graph for ICCSS processes but also to check that the graphical models are well defined. To compute these orthogonal projections, we need *alternative representations* for both $Y_a(t+h)$ and $D^{(p-q-1)}Y_a(t+h)$, $a \in V$. Due to the well-defined integral representation of the truncated state vector in Proposition 2.30, it is obvious that the following alternative representations of $\mathcal{Y}_a(t+h)$ and its $(p-q-1)$-th derivative $D^{(p-q-1)}Y_a(t+h)$ are also well defined. Note that we consider $D^{(p-q-1)}Y_a(t+h)$ since, by Remark 9.6 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.

Theorem 9.1. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$. Then, for $h \geq 0$, $t \in \mathbb{R}$, and $a \in V$, it applies that*

$$\begin{aligned} Y_a(t+h) &= \int_{-\infty}^{t} e_a^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du \\ &\quad + \sum_{m=0}^{p-q-1} e_a^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} D^{(m)} Y_V(t) + e_a^\top \bar{\bar{\varepsilon}}(t,h) \quad \mathbb{P}\text{-}a.s. \quad \textit{and} \\ D^{(p-q-1)} Y_a(t+h) &= \int_{-\infty}^{t} e_a^\top \underline{\underline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du \\ &\quad + \sum_{m=0}^{p-q-1} e_a^\top \underline{\underline{\mathbf{M}}}_m(h) \mathbf{\Theta} D^{(m)} Y_V(t) + e_a^\top \underline{\underline{\varepsilon}}(t,h) \quad \mathbb{P}\text{-}a.s. \end{aligned}$$

We abbreviate

$$\begin{aligned} \overline{\overline{\mathbf{M}}}(h) &= \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \left(\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j \right), & \underline{\underline{\mathbf{M}}}(h) &= \underline{\underline{\mathbf{C}}} e^{\mathbf{A}h} \left(\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j \right), \\ \overline{\overline{\mathbf{M}}}_m(h) &= \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m}, & \underline{\underline{\mathbf{M}}}_m(h) &= \underline{\underline{\mathbf{C}}} e^{\mathbf{A}h} \sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m}, \\ \bar{\bar{\varepsilon}}(t,h) &= \overline{\overline{\mathbf{C}}} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u), & \underline{\underline{\varepsilon}}(t,h) &= \underline{\underline{\mathbf{C}}} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u). \end{aligned}$$

The matrices $\overline{\overline{\mathbf{C}}}$ and $\underline{\underline{\mathbf{C}}}$ are defined in (2.22), $\mathbf{E}_j$, $j = 1, \ldots, p$, *in* (1.1), $\overline{\overline{\mathbf{E}}}$ *and* $\underline{\underline{\mathbf{E}}}$ *in* (1.2).

Proof. Let $t \in \mathbb{R}$, $h \geq 0$, and $a \in V$. First of all, due to Remark 2.26(c), we receive

$$Y_a(t+h) = e_a^\top \overline{\overline{\mathbf{C}}} X(t+h) \quad \text{and} \quad D^{(p-q-1)} Y_a(t+h) = e_a^\top \underline{\underline{\mathbf{C}}} X(t+h).$$

From now on, the proofs of the two representations differ only in the choice of $\overline{\overline{\mathbf{C}}}$ and $\underline{\underline{\mathbf{C}}}$, respectively. Thus, we only continue with the representation of $Y_a(t+h)$. Due to (2.6), we have

$$Y_a(t+h) = e_a^\top \overline{\overline{\mathbf{C}}} \left(e^{\mathbf{A}h} X(t) + \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \right) = e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} X(t) + e_a^\top \bar{\bar{\varepsilon}}(t,h).$$

Furthermore,

$$\begin{aligned} e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} X(t) &= e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \begin{pmatrix} X^q(t) & X^{(q+1)}(t) & \cdots & X^{(p)}(t) \end{pmatrix}^\top \\ &= e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \overline{\overline{\mathbf{E}}} X^q(t) + \sum_{j=1}^{p-q} e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \mathbf{E}_{q+j} X^{(q+j)}(t). \end{aligned}$$

Lemma 2.31 and interchanging the summation order yield

$$\begin{aligned} e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} X(t) &= e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \overline{\overline{\mathbf{E}}} X^q(t) + \sum_{j=1}^{p-q} e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \left(\mathbf{\Lambda}^j X^q(t) + \sum_{m=0}^{j-1} \mathbf{\Lambda}^{j-1-m} \mathbf{\Theta} D^{(m)} Y_V(t) \right) \\ &= e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \left(\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j \right) X^q(t) \\ &\quad + \sum_{m=0}^{p-q-1} \sum_{j=m+1}^{p-q} e_a^\top \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m} \mathbf{\Theta} D^{(m)} Y_V(t) \\ &= e_a^\top \overline{\overline{\mathbf{M}}}(h) X^q(t) + \sum_{m=0}^{p-q-1} e_a^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} D^{(m)} Y_V(t). \end{aligned}$$

Finally, Proposition 2.30 implies

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

Remark 9.2. All abbreviations of deterministic functions and stochastic processes in Theorem 9.1 are marked with two lines at the bottom or top, depending on whether they contain $\overline{\overline{\mathbf{C}}}$ or $\underline{\underline{\mathbf{C}}}$. Furthermore, $\bar{\bar{\varepsilon}}(t,0) = \underline{\underline{\varepsilon}}(t,0) = 0_k \in \mathbb{R}^k$.

Remark 9.3. Let us compare the representations in Theorem 9.1 to those for MCAR processes. For an MCAR process, we establish in Lemma 8.1 that

$$\begin{aligned} Y_a(t+h) &= e_a^\top \mathbf{C} e^{\mathbf{A}h} \sum_{m=0}^{p-1} \mathbf{E}_{m+1} D^{(m)} Y_V(t) + e_a^\top \mathbf{C} \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.} \quad \text{and} \\ D^{(p-1)} Y_a(t+h) &= e_a^\top \mathbf{E}_p^\top e^{\mathbf{A}h} \sum_{m=0}^{p-1} \mathbf{E}_{m+1} D^{(m)} Y_V(t) + e_a^\top \mathbf{E}_p^\top \int_t^{t+h} e^{\mathbf{A}(t+h-u)} \mathbf{B} dL(u) \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

To be able to compare Theorem 9.1 to Lemma 8.1, we have to interpret

$$\begin{aligned} \overline{\overline{\mathbf{M}}}_m(h)\mathbf{\Theta} &\mathrel{\widehat{=}} \mathbf{C} e^{\mathbf{A}h} \mathbf{E}_{m+1}, \quad m = 0, \dots, p-1, \quad \text{and} \quad \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} \mathrel{\widehat{=}} 0_k \in M_k(\mathbb{R}), \\ \underline{\underline{\mathbf{M}}}_m(h)\mathbf{\Theta} &\mathrel{\widehat{=}} \mathbf{E}_p^\top e^{\mathbf{A}h} \mathbf{E}_{m+1}, \quad m = 0, \dots, p-1, \quad \text{and} \quad \underline{\underline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} \mathrel{\widehat{=}} 0_k \in M_k(\mathbb{R}), \end{aligned}$$

in the case $q = 0$. Then the result for MCAR processes can be seen as a special case of the result for ICCSS processes. We heuristically justify that this interpretation is reasonable. First of all, in $\overline{\overline{\mathbf{M}}}_m(h)\mathbf{\Theta}$ the summand $j = m+1$ is mainly relevant. For this summand, $\mathbf{\Lambda}^0 = I_{kq}$ yields

$$\overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \mathbf{E}_{q+m+1} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Theta} = \overline{\overline{\mathbf{C}}} e^{\mathbf{A}h} \mathbf{E}_{q+m+1} C_q^{-1}. \tag{9.1}$$

If $q = 0$ is inserted into $\overline{\overline{\mathbf{M}}}_m(h)\mathbf{\Theta}$, all summands disappear due to the zero dimensionality of $\mathbf{\Lambda}^{j-1-m}$, $j = m+2, \dots, p-q$, except for (9.1). With $C_q = I_k$, it remains as claimed that $\overline{\overline{\mathbf{M}}}_m(h)\mathbf{\Theta} \mathrel{\widehat{=}} \mathbf{C}e^{\mathbf{\Lambda}h}\mathbf{E}_{m+1}$ for $m = 0, \dots, p-1$. The second matrix function $\overline{\overline{\mathbf{M}}}(h)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}$, $u < t$, can be interpreted as a zero matrix for $q = 0$, again due to the zero dimensionality of $\mathbf{\Lambda}$. Although we get a non-zero matrix for $t = u$, this event is a Lebesgue null set. An analogous reasoning is possible for $\underline{\underline{\mathbf{M}}}_m(h)\mathbf{\Theta}$ and $\underline{\underline{\mathbf{M}}}(h)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}$, which we do not repeat.

The representations in Theorem 9.1 suggest that for the orthogonal projection of $Y_a(t+h)$ and $D^{(p-q-1)}Y_a(t+h)$, on the one hand, the past $(Y_V(s))_{s\leq t}$ of all components and, on the other hand, the future of the Lévy process $(L(s)-L(t))_{t\leq s\leq t+h}$ is relevant. However, for a formal proof, we need that all integrals are defined in L^2. Therefore, we show that the integral representation of $\mathcal{X}^q$ in Proposition 2.30 holds in L^2. The proof is based on the ideas of the proof of Theorem 2.8 by Brockwell and Lindner (2015) in the univariate setting.

Proposition 9.4. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$. Then, for $a, v \in V$ and $t \in \mathbb{R}$, the integral*

$$\int_{-\infty}^{t} e_a^\top e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta} e_v Y_v(u)du \in \mathcal{L}_{Y_v}(t)$$

exists as an L^2-limit. In particular,

$$X^q(t) = \int_{-\infty}^{t} e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta} Y_V(u)du$$

exists as an L^2-limit.

Proof. Let $a, v \in V$ and define $F(t) = e_a^\top e^{\mathbf{\Lambda}t}\mathbf{\Theta}e_v$ for $t \geq 0$. First, for $s, t \in \mathbb{R}$, $s < t$,

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

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

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

follows. This majorant is integrable because

$$\sup_{u\in[0,t-s]}|Y_v(u)| = \sup_{u\in[0,t-s]}\left|e_v^\top\overline{\overline{\mathbf{C}}}X(u)\right| \leq \sup_{u\in[0,t-s]}\left\|e_v^\top\overline{\overline{\mathbf{C}}}\right\|\,\|X(u)\| \leq c\sup_{u\in[0,t-s]}\|X(u)\|$$

is valid for a constant $c \geq 0$. Thus,

$$\mathbb{E}\left[\left(\sup_{u\in[s,t]}|Y_v(u)|\right)^2\right] = \mathbb{E}\left[\left(\sup_{u\in[0,t-s]}|Y_v(u)|\right)^2\right] \leq c^2\,\mathbb{E}\left[\left(\sup_{u\in[0,t-s]}\|X(u)\|\right)^2\right] < \infty \tag{9.2}$$

due to the stationarity of $\mathcal{Y}_V$, Assumption 1, and Lemma A.4 of Brockwell and Schlemm (2013). Furthermore, F is a continuous function, so $\sup_{u\in[0,t-s]}|F(u)| < \infty$. In summary,

$$\int_s^t F(t-u)Y_v(u)du = \underset{n\to\infty}{\text{l.i.m.}}\ \frac{t-s}{n}\sum_{\ell=1}^{n} F\left(t-s-\ell\,\frac{t-s}{n}\right)Y_v\left(s+\ell\,\frac{t-s}{n}\right)$$

and the integral is in $\mathcal{L}_{Y_v}(t)$. For the second step of this proof, we recall that, for $t\in\mathbb{R}$,

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

due to Proposition 2.30. Using the dominated convergence theorem again, we show that the convergence holds in L^2. For $s<t$, substituting $t-u$ for u, and using the triangle inequality, we get

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

To see that this majorant is integrable, we use Fubini, Cauchy-Schwarz inequality, and the stationarity of $\mathcal{Y}_V$, to receive

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

To obtain this finiteness, we finally use relation (9.2) and that $\sum_{n=0}^{\infty}\sup_{u\in[n,n+1]}|F(u)|<\infty$ by definition of F and relation (2.30). In summary, we obtain

$$\int_{-\infty}^t e_a^\top e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}e_vY_v(u)du = \int_{-\infty}^t F(t-u)Y_v(u)du = \underset{s\to-\infty}{\text{l.i.m.}}\int_s^t F(t-u)Y_v(u)du$$

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

Before moving on to orthogonal projections, we introduce one last alternative representation, this time for the difference quotient $(D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t))/h$.

Lemma 9.5. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$. Then, for $h \geq 0$, $t \in \mathbb{R}$, and $a \in V$, we have*

$$\begin{aligned}
&\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h} \\
&= \int_{-\infty}^{t} e_a^\top \underline{\underline{\mathbf{M}}}'(0) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du + \sum_{m=0}^{p-q-1} e_a^\top \underline{\underline{\mathbf{M}}}'_m(0) \mathbf{\Theta} D^{(m)} Y_V(t) \\
&\quad + e_a^\top O(h) R_1 + e_a^\top O(h) R_2 + e_a^\top \frac{\underline{\underline{\varepsilon}}(t,h)}{h},
\end{aligned}$$

where $\underline{\underline{\mathbf{M}}}'(0)$ and $\underline{\underline{\mathbf{M}}}'_m(0)$ denote the first derivatives of $\underline{\underline{\mathbf{M}}}(h)$ and $\underline{\underline{\mathbf{M}}}_m(h)$ in zero, and R_1, R_2 are random vectors in $\mathcal{L}_{Y_V}(t) \subseteq L^2$. The random variable $e_a^\top \underline{\underline{\varepsilon}}(t,h)/h$ is independent of the former addends and

$$\lim_{h \searrow 0} \frac{1}{h} \mathbb{E}\left[\left(e_a^\top \underline{\underline{\varepsilon}}(t,h)\right)^2\right] = e_a^\top \underline{\underline{\mathbf{C}}} \mathbf{B} \Sigma_L \mathbf{B}^\top \underline{\underline{\mathbf{C}}} e_a \neq 0 \quad \text{but} \quad \lim_{h \searrow 0} \frac{1}{h^2} \mathbb{E}\left[\left(e_a^\top \underline{\underline{\varepsilon}}(t,h)\right)^2\right] = \infty.$$

Proof. Recall that due to Theorem 9.1 and $\underline{\underline{\varepsilon}}(t,0) = 0_k \in \mathbb{R}^k$, we have

$$\begin{aligned}
&\frac{D^{(p-q-1)}Y_a(t+h) - D^{(p-q-1)}Y_a(t)}{h} \\
&= \int_{-\infty}^{t} e_a^\top \frac{\underline{\underline{\mathbf{M}}}(h) - \underline{\underline{\mathbf{M}}}(0)}{h} e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du \\
&\quad + \sum_{m=0}^{p-q-1} e_a^\top \frac{\underline{\underline{\mathbf{M}}}_m(h) - \underline{\underline{\mathbf{M}}}_m(0)}{h} \mathbf{\Theta} D^{(m)} Y_V(t) + e_a^\top \frac{\underline{\underline{\varepsilon}}(t,h)}{h} \quad \mathbb{P}\text{-a.s.}
\end{aligned} \tag{9.3}$$

Replacing the matrix exponential with its power series, we obtain

$$\begin{aligned}
\frac{\underline{\underline{\mathbf{M}}}(h) - \underline{\underline{\mathbf{M}}}(0)}{h} &= \underline{\underline{\mathbf{C}}} \frac{e^{\mathbf{A}h} - I_{kp}}{h} \left(\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j \right) \\
&= \underline{\underline{\mathbf{M}}}'(0) + O(h) \left(\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j \right), \\
\frac{\underline{\underline{\mathbf{M}}}_m(h) - \underline{\underline{\mathbf{M}}}_m(0)}{h} &= \underline{\underline{\mathbf{C}}} \frac{e^{\mathbf{A}h} - I_{kp}}{h} \sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m} \\
&= \underline{\underline{\mathbf{M}}}'_m(0) + O(h) \sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m}.
\end{aligned} \tag{9.4}$$

Furthermore, we define the random variables

$$\begin{aligned}
R_1 &= \int_{-\infty}^{t} \left(\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j \right) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du, \\
R_2 &= \sum_{m=0}^{p-q-1} \sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m} \mathbf{\Theta} D^{(m)} Y_V(t).
\end{aligned} \tag{9.5}$$

If we plug equations (9.4) and (9.5) in (9.3), we obtain the stated representation. Moreover, we know from Proposition 9.4 that R_1 is in $\mathcal{L}_{Y_V}(t) \subseteq L^2$ and from Remark 2.26(b) that R_2 is in $\mathcal{L}_{Y_V}(t) \subseteq L^2$. Since $\mathcal{L}_{Y_V(t)}$ and $(L(s)-L(t))_{t\leq s\leq t+h}$ are independent (Marquardt & Stelzer, 2007, Theorem 3.12), we receive that R_1 and R_2 are independent of $\underline{\underline{\varepsilon}}(t,h)$. Finally,

$$\frac{1}{h}\mathbb{E}\left[\left(e_a^\top \underline{\underline{\varepsilon}}(t,h)\right)^2\right] = \frac{1}{h}\int_0^h e_a^\top \underline{\underline{\mathbf{C}}} e^{\mathbf{A}u}\mathbf{B}\Sigma_L\mathbf{B}^\top e^{\mathbf{A}^\top u}\underline{\underline{\mathbf{C}}} e_a du \overset{h\searrow 0}{\longrightarrow} e_a^\top \underline{\underline{\mathbf{C}}}\mathbf{B}\Sigma_L\mathbf{B}^\top \underline{\underline{\mathbf{C}}} e_a.$$

$\underline{\underline{\mathbf{C}}}\mathbf{B}\Sigma_L\mathbf{B}^\top\underline{\underline{\mathbf{C}}}$ is positive definite due to $\Sigma_L > 0$, $\mathbf{B}$ being of full rank by definition, and $\underline{\underline{\mathbf{C}}}$ being of full rank due to the assumptions in (2.23). Therefore, $e_a^\top \underline{\underline{\mathbf{C}}}\mathbf{B}\Sigma_L\mathbf{B}^\top \underline{\underline{\mathbf{C}}} e_a > 0$ and, of course, $\mathbb{E}[(e_a^\top \underline{\underline{\varepsilon}}(t,h))^2]/h^2$ converges to infinity. ■

Remark 9.6. An important consequence of Lemma 9.5 and Remark 2.26 is that the mean-square limit of the difference quotient does not exist. Thus, for all components of the ICCSS process, there are no mean-square derivatives greater than $(p-q-1)$. We must analyse the $(p-q-1)$-derivative for local Granger causality and local contemporaneous correlation. It also becomes clear why we divide by h and not by h^2 in the Definition 5.5 of local contemporaneous correlation.

Finally, we establish the *orthogonal projections* of the two random variables $Y_a(t+h)$ and $D^{(p-q-1)}Y_a(t+h)$ onto the linear spaces $\mathcal{L}_{Y_S}(t)$.

Theorem 9.7. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$. Suppose $S \subseteq V$ and $a \in V$. Then, for $h \geq 0$ and $t \in \mathbb{R}$, we have*

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

and

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

Proof. Based on Theorem 9.1, the proofs of the two orthogonal projections differ only in the choice of $\overline{\overline{\mathbf{M}}}(h)$ or $\underline{\underline{\mathbf{M}}}(h)$, $\overline{\overline{\mathbf{M}}}_m(h)$ or $\underline{\underline{\mathbf{M}}}_m(h)$, and $\bar{\bar{\varepsilon}}(t,h)$ or $\underline{\underline{\varepsilon}}(t,h)$. Thus, we only prove the representation of $P_{\mathcal{L}_{Y_S}(t)}Y_a(t+h)$. Let $h \geq 0$, $t \in \mathbb{R}$, $S \subseteq V$, and $a \in V$. From Theorem 9.1, we recall that

$$Y_a(t+h) = \int_{-\infty}^{t} e_a^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du + \sum_{m=0}^{p-q-1} e_a^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} D^{(m)} Y_V(t) + e_a^\top \bar{\bar{\varepsilon}}(t,h) \quad \mathbb{P}\text{-a.s.}$$

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

$$\begin{aligned} &P_{\mathcal{L}_{Y_S}(t)}\left(\int_{-\infty}^{t} e_a^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} Y_V(u) du\right) \\ &= \sum_{v\in S} \int_{-\infty}^{t} e_a^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) du \\ &\quad + P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{v\in V\setminus S} \int_{-\infty}^{t} e_a^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_v Y_v(u) du\right), \end{aligned}$$

since, because of Proposition 9.4, the integrals are in $\mathcal{L}_{Y_S}(t)$ for $v \in S$. For the second summand, we obtain

$$\begin{aligned} &P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{m=0}^{p-q-1} e_a^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} D^{(m)} Y_V(t)\right) \\ &= \sum_{v\in S}\sum_{m=0}^{p-q-1} e_a^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_v D^{(m)} Y_v(t) + P_{\mathcal{L}_{Y_S}(t)}\left(\sum_{v\in V\setminus S}\sum_{m=0}^{p-q-1} e_a^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_v D^{(m)} Y_v(t)\right), \end{aligned}$$

because, due to Remark 2.26(b), the derivatives of $Y_v(t)$ are in $\mathcal{L}_{Y_S}(t)$ for $v \in S$. For the third summand $e_a^\top \bar{\bar{\varepsilon}}(t,h)$, we note that $(Y_S(s))_{s\leq t}$ and $(L(s)-L(t))_{t\leq s\leq t+h}$ are independent (Marquardt & Stelzer, 2007, Theorem 3.12). Thus, $e_a^\top \bar{\bar{\varepsilon}}(t,h)$ is independent of $\mathcal{L}_{Y_S}(t)$ and we obtain immediately that $P_{\mathcal{L}_{Y_S}(t)} e_a^\top \bar{\bar{\varepsilon}}(t,h) = 0$. If we put all three summands together, we get the assertion. ■

Remark 9.8. When calculating the orthogonal projections, it becomes clear why we require the assumptions in (2.23), which are sufficient assumptions to recover $X(t)$ from $(Y_V(s))_{s\leq t}$. 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 to establish the *limit of the orthogonal projections* of the difference quotients.

Theorem 9.9. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$. Suppose $S \subseteq V$, $a \in V$, and $t \in \mathbb{R}$. Then, for $h \geq 0$,*

$$D^{(p-q-1)}Y_a(t+h) - P_{\mathcal{L}_{Y_V}(t)} D^{(p-q-1)} Y_a(t+h) = e_a^\top \underline{\underline{\varepsilon}}(t,h) \quad \mathbb{P}\text{-}a.s.$$

and

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

Proof. Let $S\subseteq V$, $a\in V$, $h\geq 0$, and $t\in\mathbb{R}$. The first assertion follows directly from Theorems 9.1 and 9.7, giving

$$
\begin{aligned}
&D^{(p-q-1)}Y_a(t+h)-P_{\mathcal{L}_{Y_V}(t)}D^{(p-q-1)}Y_a(t+h)\\
&\quad=\int_{-\infty}^{t} e_a^\top \underline{\underline{\mathbf{M}}}(h)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta} Y_V(u)du+\sum_{m=0}^{p-q-1} e_a^\top \underline{\underline{\mathbf{M}}}_m(h)^\top\mathbf{\Theta} D^{(m)}Y_V(t)+e_a^\top \underline{\underline{\varepsilon}}(t,h)\\
&\qquad-\int_{-\infty}^{t} e_a^\top \underline{\underline{\mathbf{M}}}(h)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta} Y_V(u)du-\sum_{m=0}^{p-q-1} e_a^\top \underline{\underline{\mathbf{M}}}_m(h)\mathbf{\Theta} D^{(m)}Y_V(t)\\
&\quad=e_a^\top\underline{\underline{\varepsilon}}(t,h).
\end{aligned}
$$

The second assertion is also initially based on Theorem 9.7. We obtain

$$
\begin{aligned}
P_{\mathcal{L}_{Y_V}(t)}D^{(p-q-1)}Y_a(t+h)&=\int_{-\infty}^{t} e_a^\top \underline{\underline{\mathbf{M}}}(h)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta} Y_V(u)du+\sum_{m=0}^{p-q-1} e_a^\top \underline{\underline{\mathbf{M}}}_m(h)\mathbf{\Theta} D^{(m)}Y_V(t)\\
&=e_a^\top\underline{\underline{\mathbf{C}}}e^{\mathbf{A}h}X(t).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\lim_{h\to 0}\mathbb{E}\left[\left(P_{\mathcal{L}_{Y_V}(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h)-D^{(p-q-1)}Y_a(t)}{h}\right)-e_a^\top\underline{\underline{\mathbf{C}}}\mathbf{A}X(t)\right)^2\right]\\
&\quad=\lim_{h\to 0}\mathbb{E}\left[\left(e_a^\top\underline{\underline{\mathbf{C}}}\,\frac{e^{\mathbf{A}h}-I_{kp}}{h}X(t)-e_a^\top\underline{\underline{\mathbf{C}}}\mathbf{A}X(t)\right)^2\right]\\
&\quad=\lim_{h\to 0}e_a^\top\underline{\underline{\mathbf{C}}}\left(\frac{e^{\mathbf{A}h}-I_{kp}}{h}-\mathbf{A}\right)c_{XX}(0)\left(\frac{e^{\mathbf{A}h}-I_{kp}}{h}-\mathbf{A}\right)^\top\underline{\underline{\mathbf{C}}}^\top e_a=0.
\end{aligned}
$$

That is,

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

But then, together with Lemma 3.1(a,c), we can conclude that

$$
\begin{aligned}
&\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h)-D^{(p-q-1)}Y_a(t)}{h}\right)\\
&\quad=\underset{h\to 0}{\text{l.i.m.}}\, P_{\mathcal{L}_{Y_S}(t)}P_{\mathcal{L}_{Y_V}(t)}\left(\frac{D^{(p-q-1)}Y_a(t+h)-D^{(p-q-1)}Y_a(t)}{h}\right)\\
&\quad=P_{\mathcal{L}_{Y_S}(t)}\left(e_a^\top\underline{\underline{\mathbf{C}}}\mathbf{A}X(t)\right)\quad \mathbb{P}\text{-a.s.}
\end{aligned}
$$

Again, similar to the proof of Theorem 9.1,

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

We obtain, replacing $\underline{\underline{\mathbf{M}}}(h)$ by $\underline{\underline{\mathbf{M}}}'(0)$ and $\underline{\underline{\mathbf{M}}}_m(h)$ by $\underline{\underline{\mathbf{M}}}'_m(0)$ in the proof of Theorem 9.7,

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

Remark 9.10. Although the derivation of orthogonal projections of MCAR(p) processes differs from that of ICCSS(p,q) processes with $q>0$, the results in Theorem 9.7 are consistent with those in Proposition 8.2 if we interpret $\overline{\overline{\mathbf{M}}}_m(h)\mathbf{\Theta}\mathrel{\widehat{=}}\mathbf{C}e^{\mathbf{A}h}\mathbf{E}_{m+1}$ and $\overline{\overline{\mathbf{M}}}(h)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}\mathrel{\widehat{=}}0_k\in M_k(\mathbb{R})$ as in Remark 9.3. Furthermore, Theorem 9.9 is consistent with Theorem 8.3 if we interpret $\underline{\underline{\mathbf{M}}}'_m(0)\mathbf{\Theta}\mathrel{\widehat{=}}\mathbf{E}_p^\top\mathbf{A}\mathbf{E}_{m+1}$ and $\underline{\underline{\mathbf{M}}}'(0)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta}\mathrel{\widehat{=}}0_k$.

A very important special case of the orthogonal projections in Theorems 9.7 and 9.9 is the case $S=V$, for which certain terms are simplified.

Corollary 9.11. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p>q>0$. Then, for $t\in\mathbb{R}$, $h\geq 0$, and $a\in V$, we have*

(a) $P_{\mathcal{L}_{Y_V}(t)}Y_a(t+h)=e_a^\top\overline{\overline{\mathbf{C}}}e^{\mathbf{A}h}X(t)\quad \mathbb{P}$*-a.s.,*

(b) $P_{\mathcal{L}_{Y_V}(t)}D^{(p-q-1)}Y_a(t+h)=e_a^\top\underline{\underline{\mathbf{C}}}e^{\mathbf{A}h}X(t)\quad \mathbb{P}$*-a.s.,*

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

To conclude the chapter on orthogonal projections, we comment on this result, with particular reference to the literature.

Remark 9.12.

(a) The representations in Corollary 9.11 are, of course, projection-free and analogous to the representations for MCAR processes in Remark 8.4. For $S \subset V$, we again refer to Rozanov (1967), III, 5. Since such explicit representations are not relevant for the derivation of the (local) causality graph, we do explore this topic further.

(b) The orthogonal projections in Corollary 9.11(a) match the orthogonal projections of univariate CARMA processes in Brockwell and Lindner (2015), Theorem 2.8. Basse-O'Connor et al. (2019) derives as well predictions of MCARMA processes, but their results differ from Brockwell and Lindner (2015).

(c) From Corollary 9.11(c), not only the existence of the L^2-limit becomes clear but also

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

The existence of these limits is essential for the well-definedness of local Granger causality and local contemporaneous correlation for ICCSS processes. Furthermore, due to Corollary 9.11(c) and the Definition 4.7 of local Granger non-causality, we realise that $\mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_S$ if and only if

$$P_{\mathcal{L}_{Y_S}(t)}\left(e_b^\top \underline{\underline{\mathbf{C}}}\mathbf{A}X(t)\right) = P_{\mathcal{L}_{Y_{S\setminus\{a\}}}(t)}\left(e_b^\top \underline{\underline{\mathbf{C}}}\mathbf{A}X(t)\right) \quad \mathbb{P}\text{-a.s.}$$

for $t \in \mathbb{R}$. That is, if and only if $e_b^\top \underline{\underline{\mathbf{C}}}\mathbf{A}X(t) \perp \mathcal{L}_{Y_a}(t) \,|\, \mathcal{L}_{Y_{S\setminus\{a\}}}(t)$ for all $t \in \mathbb{R}$ by Proposition 2.4.2 of Lindquist and Picci (2015). In this particular case, we can express local Granger non-causality as a conditional orthogonality relation.

9.2. Establishment of (local) causality graphs for ICCSS processes

In this section, we establish (local) causality graphs for ICCSS processes. To define these graphical models according to Definition 6.1 and for various Markov properties to hold, certain well-definedness requirements must be met by the ICCSS process. The validity of both *wide-sense stationarity* and *continuity in the mean square* is immediately clear from Remarks 2.12 and 2.13. Therefore, we only need to ensure that the ICCSS process satisfies the *Assumptions 3 and 4*.

Theorem 9.13. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. Then $\mathcal{Y}_V$ satisfies Assumptions 3 and 4.*

Proof. The proof of Assumption 3 is elaborate and has already been given in Section 8.2 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 $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Indeed, the assumptions in (2.23) provide that $Q(i\lambda)$ is of full rank and $\mathcal{N}(P) \subseteq (-\infty, 0) + i\mathbb{R}$, so we directly receive that $Q(i\lambda)P(i\lambda)^{-1}$ is also of full rank. Furthermore, we require that $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$, which is also true due to the assumptions in (2.23). Finally, it is a necessity that $\overline{\overline{\mathbf{C}}}\mathbf{B}\Sigma_L\mathbf{B}^\top\overline{\overline{\mathbf{C}}}^\top > 0$. Again, $\Sigma_L > 0$, $\mathbf{B}$ is of full rank by definition, and $\overline{\overline{\mathbf{C}}}$ is of full rank due to the assumptions in (2.23), so $\overline{\overline{\mathbf{C}}}\mathbf{B}\Sigma_L\mathbf{B}^\top\overline{\overline{\mathbf{C}}}^\top > 0$ is satisfied. For the Assumption 4, the proof of Proposition 8.11 applies, requiring $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$. ■

A direct consequence of the results in Chapters 6 and 7 is then the following.

Proposition 9.14. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. If we define $V = \{1, \ldots, k\}$ as the vertices and the edges E_{CG} via*

(a) $a \longrightarrow b \notin E_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V$,

(b) $a \dashrightarrow\!\!\!\!- b \notin E_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V$,

for $a, b \in V$ with $a \neq b$, then the causality graph $G_{CG} = (V, E_{CG})$ for the ICCSS process $\mathcal{Y}_V$ is well defined and satisfies the pairwise, local, block-recursive, global AMP, and global Granger-causal Markov property.

Proposition 9.15. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. If we define $V = \{1, \ldots, k\}$ as the vertices and the edges E^0_{CG} via*

(a) $a \longrightarrow b \notin E^0_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V$,

(b) $a \dashrightarrow\!\!\!\!- b \notin E^0_{CG} \quad \Leftrightarrow \quad \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V$,

for $a, b \in V$ with $a \neq b$, then the local causality graph $G^0_{CG} = (V, E^0_{CG})$ for the ICCSS process $\mathcal{Y}_V$ is well defined and satisfies the pairwise, local, and block-recursive Markov property. Furthermore, the statements of Propositions 7.25 and 7.26 apply.

Remark 9.16. In principle, more general state space models $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*, L)$ also satisfy the Assumptions 3 and 4. The proof of Theorem 9.13 shows that sufficient assumptions are that the driving Lévy process satisfies Assumption 1, $\sigma(\mathbf{A}^*) \subseteq (-\infty, 0) + i\mathbb{R}$, $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$, and $\mathbf{C}^*\mathbf{B}^*\Sigma_L\mathbf{B}^{*\top}\mathbf{C}^{*\top} > 0$. Then the (local) causality graph is also well defined and the various Markov properties are satisfied. However, in this general context, we are not able to calculate the orthogonal projections needed to characterise the edges, which is our main interest.

9.3. Edge characterisations for ICCSS processes

In this section, we finally come to the main part of the chapter, the *characterisation* of the edges in the (local) causality graph based on the model parameters of the ICCSS process. We present several characterisations, provide interpretations, and draw parallels to the results for MCAR processes. The first characterisation of the *directed edges* in the (local) causality graph is developed from the characterisations of (local) Granger non-causality in Theorem 4.4 and Definition 4.7. We further apply the orthogonal projections of ICCSS processes and their derivatives from Theorems 9.7 and 9.9.

Proposition 9.17. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. Let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V \quad \Leftrightarrow \quad e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda} t} \mathbf{\Theta} e_a = 0$ *and* $e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a = 0$ *for all* $m = 0, \ldots, p-q-1$, $0 \leq h \leq 1$, $t \geq 0$,

(b) $\mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \mid \mathcal{Y}_V \quad \Leftrightarrow \quad e_b^\top \underline{\underline{\mathbf{M}}}'(0) e^{\mathbf{\Lambda} t} \mathbf{\Theta} e_a = 0$ *and* $e_b^\top \underline{\underline{\mathbf{M}}}'_m(0) \mathbf{\Theta} e_a = 0$ *for all* $m = 0, \ldots, p-q-1$, $t \geq 0$.

Proof.

(a) Recall that due to Theorem 4.4, we have $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V$ if and only if, for $0 \leq h \leq 1$ and $t \in \mathbb{R}$,

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

From Theorem 9.7, we further obtain

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

for $0 \leq h \leq 1$ and $t \in \mathbb{R}$. We equate the two orthogonal projections and remove the coinciding terms. Then, we receive that $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V$ if and only if

$$\begin{aligned} &\int_{-\infty}^t e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_a Y_a(u) du + \sum_{m=0}^{p-q-1} e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a D^{(m)} Y_a(t) \\ &= P_{\mathcal{L}_{Y_{V \setminus \{a\}}}(t)} \left(\int_{-\infty}^t e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_a Y_a(u) du \right) \\ &\quad + P_{\mathcal{L}_{Y_{V \setminus \{a\}}}(t)} \left(\sum_{m=0}^{p-q-1} e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a D^{(m)} Y_a(t) \right) \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

for $0 \leq h \leq 1$ and $t \in \mathbb{R}$. The expression on the left side of the equation 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 Proposition 3.10, we obtain $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \mid \mathcal{Y}_V$ if and only if

$$\int_{-\infty}^{t} e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_a Y_a(u) du + \sum_{m=0}^{p-q-1} e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a D^{(m)} Y_a(t) = 0 \quad \mathbb{P}\text{-a.s.} \tag{9.6}$$

for $0 \leq h \leq 1$ and $t \in \mathbb{R}$. In the following, we show that this characterisation is equivalent to

$$e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda} t} \mathbf{\Theta} e_a = 0 \quad \text{and} \quad e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a = 0 \tag{9.7}$$

for $m = 0, \dots, p-q-1$, $0 \leq h \leq 1$, and $t \geq 0$.

If equations (9.7) are satisfied, we immediately obtain that equation (9.6) is valid. Now, suppose equation (9.6) applies. We convert the two summands in (9.6) into their spectral representation. Therefore, note that due Bernstein (2009), Proposition 11.2.2, and $\sigma(\mathbf{\Lambda}) \subseteq (-\infty, 0) + i\mathbb{R}$ the equality

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

holds. The integrator is integrable, since, using Proposition 11.2.2 of Bernstein (2009) again,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \left| e^{-i\lambda s} \mathbb{1}\{s \geq 0\} e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda} s} \mathbf{\Theta} e_a \right| ds \\
&\leq \int_0^\infty \frac{1}{2} \left(e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda} s} \overline{\overline{\mathbf{M}}}(h) e_b + e_a^\top \mathbf{\Theta}^\top e^{\mathbf{\Lambda} s} \mathbf{\Theta} e_a \right) ds \\
&= \frac{1}{2} \left(e_b^\top \overline{\overline{\mathbf{M}}}(h) (-\mathbf{\Lambda})^{-1} \overline{\overline{\mathbf{M}}}(h) e_b + e_a^\top \mathbf{\Theta}^\top (-\mathbf{\Lambda})^{-1} \mathbf{\Theta} e_a \right) < \infty.
\end{aligned} \tag{9.8}$$

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

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

where Φ_a is the random spectral measure from the spectral representation (2.1) of $\mathcal{Y}_a$. For the second summand, we substitute $Y_a(t)$ as well as its derivatives by their spectral representation (cf. Proposition 2.8, Remark 2.9). In summary, we obtain

$$\begin{aligned}
0 &= \int_{-\infty}^{t} e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda}(t-u)} \mathbf{\Theta} e_a Y_a(u) du + \sum_{m=0}^{p-q-1} e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a D^{(m)} Y_a(t) \\
&= \int_{-\infty}^{\infty} e^{i\lambda t} e_b^\top \overline{\overline{\mathbf{M}}}(h) (i\lambda I_{kq} - \mathbf{\Lambda})^{-1} \mathbf{\Theta} e_a \Phi_a(d\lambda) + \sum_{m=0}^{p-q-1} e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a \int_{-\infty}^{\infty} (i\lambda)^m e^{i\lambda t} \Phi_a(d\lambda).
\end{aligned}$$

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

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

Hence, $|\psi(\lambda, h)|^2 f_{Y_a Y_a}(\lambda) = 0$ for (almost) all $\lambda \in \mathbb{R}$. But $f_{Y_a Y_a}(\lambda) > 0$ for $\lambda \in \mathbb{R}$ by Theorem 9.13, which yields $\psi(\lambda, h) = 0$ for $0 \leq h \leq 1$ and (almost) all $\lambda \in \mathbb{R}$. Bernstein (2009) provides in equations (4.4.3) and (4.4.23) that, due to $i\lambda \in \mathbb{C} \setminus \sigma(\mathbf{\Lambda})$,

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

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

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

and multiplication by the characteristic polynomial yields

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

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 powers higher than $kq-1$ in the second summand and their prefactors must be zero. Due to $\gamma_{kq} = 1$, we receive

$$e_b^\top \overline{\overline{\mathbf{M}}}_m(h) \mathbf{\Theta} e_a = 0$$

for $m = 0, \ldots, p-q-1$. Inserting this result into $\psi(\lambda, h) = 0$ yields

$$0 = e_b^\top \overline{\overline{\mathbf{M}}}(h)(i\lambda I_{kq} - \mathbf{\Lambda})^{-1} \mathbf{\Theta} e_a = \int_{-\infty}^{\infty} e^{-i\lambda s} \mathbb{1}\{s \geq 0\} e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda} s} \mathbf{\Theta} e_a ds.$$

Together with the already known integrability (9.8), Pinsky (2009), Corollary 2.2.23, provides

$$e_b^\top \overline{\overline{\mathbf{M}}}(h) e^{\mathbf{\Lambda} t} \mathbf{\Theta} e_a = 0$$

for $t \geq 0$, which finally concludes the proof of (a).

(b) Due to the similarity of the results in Theorems 9.7 and 9.9, we just have to replace $\underline{\underline{\mathbf{M}}}(h)$ by $\underline{\underline{\mathbf{M}}}'(0)$ and $\underline{\underline{\mathbf{M}}}_m(h)$ by $\underline{\underline{\mathbf{M}}}'_m(0)$ in the proof of (a). ■

We introduce a second characterisation of the directed edges in the (local) causality graph, the proof of which is based on Proposition 9.17 and is similar to the proof of Theorem 8.23(a) for MCAR processes.

Theorem 9.18. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. Let $a, b \in V$ with $a \neq b$. Then, we have*

(a) $\mathcal{Y}_a \not\rightarrow \mathcal{Y}_b \,|\, \mathcal{Y}_V \;\Leftrightarrow\;$ $e_b^\top \overline{\overline{\mathbf{C}}} \mathbf{A}^\alpha (\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j) \mathbf{\Lambda}^\beta \mathbf{\Theta} e_a = 0$ *and*
$e_b^\top \overline{\overline{\mathbf{C}}} \mathbf{A}^\alpha (\sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m}) \mathbf{\Theta} e_a = 0$
for all $\alpha = 0, \ldots, kp-1$, $\beta = 0, \ldots, kq-1$, and $m = 0, \ldots, p-q-1$,

(b) $\mathcal{Y}_a \not\rightarrow_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V \;\Leftrightarrow\;$ $e_b^\top \underline{\underline{\mathbf{C}}} \mathbf{A} (\overline{\overline{\mathbf{E}}} + \sum_{j=1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^j) \mathbf{\Lambda}^\beta \mathbf{\Theta} e_a = 0$ *and*
$e_b^\top \underline{\underline{\mathbf{C}}} \mathbf{A} (\sum_{j=m+1}^{p-q} \mathbf{E}_{q+j} \underline{\underline{\mathbf{E}}}^\top \mathbf{\Lambda}^{j-1-m}) \mathbf{\Theta} e_a = 0$
for all $\beta = 0, \ldots, kq-1$ and $m = 0, \ldots, p-q-1$.

Proof.

(a) Based on the characterisation in Proposition 9.17(a), the same steps as in the proof of Theorem 8.23(a) can be carried out. First, we replace the matrix exponential $e^{\mathbf{A}h}$ in Proposition 9.17(a) by powers of the matrix $\mathbf{A}$ and second, we replace $e^{\mathbf{\Lambda}h}$ by powers of $\mathbf{\Lambda}$.

(b) Follows in analogy to (a), using Proposition 9.17(b). ■

Remark 9.19. The characterisations and thus the directed edges in the (local) causality graph do not depend on the chosen Lévy process. Furthermore, the characterisation in Proposition 9.17(a) seems to depend on h. However, this is not the case, as can be seen from Theorem 9.18(a). So it does not matter whether we define directed edges by considering the time span $0 \leq h \leq 1$ or by considering the entire future $h \geq 0$. There is no difference between Granger causality and global Granger causality for ICCSS processes.

Next, we present characterisations of the *undirected edges* in the (local) causality graph. The proofs are again based on orthogonal projections of ICCSS processes and their derivatives. They are similar to the proofs of Proposition 8.16 and Theorem 8.23(b) for MCAR processes. Note that the assumption $\Sigma_L > 0$ is only used for the second characterisation in Proposition 9.20(a).

Proposition 9.20. *Let $\mathcal{Y}_V$ be an ICCSS(p,q) process with $p > q > 0$ and $\Sigma_L > 0$. Let $a, b \in V$ with $a \neq b$. Then, we have*

$$\begin{aligned}
&\text{(a) } \mathcal{Y}_a \nsim \mathcal{Y}_b \,|\, \mathcal{Y}_V && \Leftrightarrow && e_a^\top \int_0^{\min(h,\widetilde{h})} \overline{\overline{\mathbf{C}}} e^{\mathbf{A}(h-s)} \mathbf{B} \Sigma_L \mathbf{B}^\top e^{\mathbf{A}^\top(\widetilde{h}-s)} \overline{\overline{\mathbf{C}}}^\top ds\, e_b = 0 \\
& && && \text{for all } 0 \leq h, \widetilde{h} \leq 1, \\
& && \Leftrightarrow && e_a^\top \overline{\overline{\mathbf{C}}} \mathbf{A}^\alpha \mathbf{B} \Sigma_L \mathbf{B}^\top (\mathbf{A}^\top)^\beta \overline{\overline{\mathbf{C}}}^\top e_b = 0 \\
& && && \text{for all } \alpha, \beta = 0, \ldots, kp-1, \\
&\text{(b) } \mathcal{Y}_a \nsim_0 \mathcal{Y}_b \,|\, \mathcal{Y}_V && \Leftrightarrow && e_a^\top \underline{\underline{\mathbf{C}}} \mathbf{B} \Sigma_L \mathbf{B}^\top \underline{\underline{\mathbf{C}}}^\top e_b = e_a^\top C_q \Sigma_L C_q^\top e_b = 0.
\end{aligned}$$

Proof.

(a) Based on Corollary 9.11(a), the proof of the first characterisation can be carried out in the same way as the proof of Proposition 8.16(a). The second characterisation follows along the lines of the proof of Theorem 8.23(b).

(b) Based on Theorem 9.1 and Corollary 9.11(b), this statement can be proven analogously to Proposition 8.16(b). ■

Remark 9.21. The characterisations and thus the undirected edges in the (local) causality graph depend on the chosen Lévy process only by Σ_L. Furthermore, the second characterisation in Proposition 9.20(a) shows that there is indeed no dependence on the lag h again. It does not matter whether we define undirected edges in the causality graph by considering the time span $0 \leq h, \widetilde{h} \leq 1$ or by considering the entire future $h, \widetilde{h} \geq 0$. There is no difference between contemporaneous correlation and global contemporaneous correlation for ICCSS processes.

We conclude this section with further comments on the characterisations in Propositions 9.17 and 9.20. In particular, we compare the characterisations with each other and with the characterisations for MCAR processes. Additionally, we give interpretations.

Remark 9.22.

(a) The uniqueness of the polynomial matrices $P(z)$ and $Q(z)$ in the decomposition (2.13) of the transfer function (see Proposition 2.18) leads to the uniqueness of the controller canonical state space representation, which in turn leads to the uniqueness of the edges in the (local) causality graph. However, the characterisations of the edges are so complex that it is not reasonable to discuss the existence of a given mixed graph as a causality graph for an ICCSS process.

(b) It can be shown by a simple calculation, which is similar to that in Remark 8.26, that $\overline{\overline{\mathbf{C}}}\mathbf{A}^{p-q} = \underline{\underline{\mathbf{C}}}\mathbf{A}$. If we set $\alpha = p-q$ in Theorem 9.18(a) and compare the result to Theorem 9.18(b), we find that Granger non-causality implies local Granger non-causality, which we know as well from the theory in Proposition 4.17. Similarly, we have $\overline{\overline{\overline{\mathbf{C}}}}\mathbf{A}^{p-q-1} = \underline{\underline{\underline{\mathbf{C}}}}$. If we set $\alpha = \beta = p-q-1$ in Proposition 9.20(a) and compare the result to Proposition 9.20(b), we get that contemporaneous uncorrelatedness implies local contemporaneous uncorrelatedness. This result is again consistent with the theory in Proposition 5.10. In summary, $E_{CG}^{(0)} \subseteq E_{CG}$ and, in general, the sets are not equal, because the opposite implications are usually not true.

Remark 9.23. Once more, we establish relationships between the results for ICCSS processes and the results for MCAR processes.

(a) The *undirected edges* are characterised only by the noise terms $\overline{\overline{\varepsilon}}(t,h)$ and $\underline{\underline{\varepsilon}}(t,h)$, so they do not require the inversion of the process. It is therefore not surprising that the characterisations of the undirected edges for ICCSS(p,q) processes (cf. Proposition 9.20) and the characterisations of the undirected edges for MCAR(p) processes (cf. Proposition 8.16 and Theorem 8.23(b)) are analogous.

(b) Of course, we cannot simply insert $q=0$ in the characterisations of the *directed edges* of the ICCSS(p,q) process, since several matrices become zero-dimensional. However, if we interpret

$$\begin{aligned}
\overline{\overline{\overline{\mathbf{M}}}}_m(h)\mathbf{\Theta} &\mathrel{\widehat{=}} \mathbf{C}e^{\mathbf{A}h}\mathbf{E}_{m+1}, \quad m=0,\dots,p-1, \quad \text{and} \quad \overline{\overline{\overline{\mathbf{M}}}}(h)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta} \mathrel{\widehat{=}} 0_k \in M_k(\mathbb{R}),\\
\underline{\underline{\underline{\mathbf{M}}}}'_m(0)\mathbf{\Theta} &\mathrel{\widehat{=}} \mathbf{E}_p^\top\mathbf{A}\mathbf{E}_{m+1}, \quad m=0,\dots,p-1, \quad \text{and} \quad \underline{\underline{\underline{\mathbf{M}}}}'(0)e^{\mathbf{\Lambda}(t-u)}\mathbf{\Theta} \mathrel{\widehat{=}} 0_k \in M_k(\mathbb{R}),
\end{aligned}$$

as argued in Remark 9.3, the characterisations of the directed edges for MCAR(p) processes (cf. Proposition 8.14 and Theorem 8.23(a)) can be seen as a special case of the characterisations of the directed edges for ICCSS(p,q) processes (cf. Proposition 9.17 and Theorem 9.18).

Interpretation 9.24 (*Causality graph*)**.** To interpret the directed and undirected edges in the causality graph G_{CG}, we recall the representation of the b-th component

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

from Theorem 9.1.

(a) *Directed edges:* A direct application of Proposition 9.17 gives that $a \longrightarrow b \notin E_{CG}$ if and only if neither $Y_a(t)$, $D^{(1)}Y_a(t)$,..., $D^{(p-q-1)}Y_a(t)$ nor the integral over the past of the a-th component have any influence on $Y_b(t+h)$. In the representation of $Y_b(t+h)$, the a-th component always vanishes, because its coefficient functions are zero.

(b) *Undirected edges:* Proposition 9.20 yields

$$a \dashrightarrow\!\!\!\!\!- \, b \notin E_{CG} \quad \Leftrightarrow \quad \mathbb{E}\left[e_a^\top \bar{\bar{\varepsilon}}(t,h) e_b^\top \bar{\bar{\varepsilon}}(t,\widetilde{h})\right] = \mathbb{E}\left[e_a^\top \bar{\bar{\varepsilon}}(0,h) e_b^\top \bar{\bar{\varepsilon}}(0,\widetilde{h})\right] = 0, \quad 0 \leq h, \widetilde{h} \leq 1.$$

This means that the noise terms $e_a^\top \bar{\bar{\varepsilon}}(t,h)$ and $e_b^\top \bar{\bar{\varepsilon}}(t,\widetilde{h})$ of $Y_a(t+h)$ and $Y_b(t+\widetilde{h})$, respectively, are uncorrelated for any $t \geq 0$ and $0 \leq h, \widetilde{h} \leq 1$.

Interpretation 9.25 (*Local causality graph*)**.** The interpretation of the directed and undirected edges in the local causality graph G^0_{CG} is a lot more intricate. The reason for this is that the limit in the mean square of the difference quotient does not exist by definition and Remark 9.6, respectively, but the limit of the projections does. To give an interpretation, we use the following representation of the difference quotient of the b-th component

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

from Lemma 9.5. Then it follows that

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

Although the L^2-limit of $\mathtt{D}_h^{(p-q-1)} Y_b(t,h)$ does not exist, the L^2-limit of $\sqrt{h}\mathtt{D}_h^{(p-q-1)} Y_b(t,h)$ does by Lemma 9.5. The L^2-limits of $P_{\mathcal{L}_{Y_V}(t)} \mathtt{D}_h^{(p-q-1)} Y_b(t,h)$ also exists.

(a) *Directed edges:* By Proposition 9.17(b), we receive that $a \longrightarrow b \notin E^0_{CG}$ if and only if neither $Y_a(t)$, $D^{(1)}Y_a(t)$,..., $D^{(p-q-1)}Y_a(t)$ nor the integral over the past have any influence on $\mathtt{D}_h^{(p-q-1)} Y_b(t,h)$ if h is small. The same holds for $P_{\mathcal{L}_{Y_V}(t)} \mathtt{D}_h^{(p-q-1)} Y_b(t,h)$. *Given* $\mathcal{L}_{Y_V}(t)$, the a-th component does not influence the b-th component in the limit, because the corresponding coefficients are zero.

(b) *Undirected edges:* By Proposition 9.20(b), we receive that $a \dashrightarrow\!\!\!\!\!- \, b \notin E^0_{CG}$ if and only if the limit

$$\begin{aligned} &h\,\mathbb{E}\left[\left(\mathtt{D}_h^{(p-q-1)} Y_a(t,h) - P_{\mathcal{L}_V(t)} \mathtt{D}_h^{(p-q-1)} Y_a(t,h)\right)\left(\mathtt{D}_h^{(p-q-1)} Y_b(t,h) - P_{\mathcal{L}_V(t)} \mathtt{D}_h^{(p-q-1)} Y_b(t,h)\right)\right] \\ &= \frac{1}{h}\mathbb{E}\left[e_a^\top \underline{\underline{\varepsilon}}(t,h) e_b^\top \underline{\underline{\varepsilon}}(t,h)\right] \xrightarrow{h \downarrow 0} e_a^\top \underline{\underline{\mathbf{C}}} \mathbf{B} \Sigma_L \mathbf{B}^\top \underline{\underline{\mathbf{C}}} e_b \end{aligned}$$

is zero. Hence, *given* $\mathcal{L}_{Y_V}(t)$, $\sqrt{h}\mathtt{D}_h^{(p-q-1)}Y_a(t,h)$ and $\sqrt{h}\mathtt{D}_h^{(p-q-1)}Y_b(t,h)$ are uncorrelated in the limit. Equivalently, the noise terms $e_a^\top \underline{\underline{\varepsilon}}(t,h)/\sqrt{h}$ and $e_b^\top \underline{\underline{\varepsilon}}(t,h)/\sqrt{h}$ are uncorrelated in the limit.

To summarise the chapter: We can construct well-defined (local) causality graphs for ICCSS processes. We are also able to derive analytic and interpretatively meaningful edge characterisations through the model parameters of the ICCSS process.

Part II.

Partial correlation graphs

As has been pointed out repeatedly, graphical models for multivariate stochastic processes in continuous time are of great importance and have been little studied. Consequently, in this part of the thesis, we are interested in a third graphical model. In contrast to the previous models, the emphasis is on presenting a graph that is both simple and user-friendly, yet powerful. This graphical model is the undirected *partial correlation graph* for multivariate stochastic processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ in continuous time.

The concept of *partial correlation* is an important and well-studied measure of dependence in statistics. For an $\mathbb{R}^k$-valued random vector $Y_V = (Y_1, \dots, Y_k)^\top$ with positive definite covariance matrix Σ_V, the partial correlation of Y_a and Y_b given $Y_{V\setminus\{a,b\}}$ measures the correlation of the real-valued random variables Y_a and Y_b after removing the linear effects of the remaining random variables $Y_{V\setminus\{a,b\}}$. The partial correlation can be determined as follows: Consider the linear regression problems

$$\beta_\ell = \operatorname{argmin}_{\beta\in\mathbb{R}^{k-2}} \mathbb{E}\left[\left(Y_\ell - \beta^\top Y_{V\setminus\{a,b\}}\right)^2\right] \quad \text{for } \ell \in \{a,b\}. \tag{9.9}$$

These problems have the well-known solution (e.g., Anderson, 1984; Brillinger, 2001; Fujikoshi, Ulyanov, & Shimizu, 2010)

$$\beta_\ell = \left(\Sigma_{V\setminus\{a,b\}V\setminus\{a,b\}}\right)^{-1} \Sigma_{V\setminus\{a,b\}\ell}. \tag{9.10}$$

Furthermore, the residuals $\varepsilon_{\ell|V\setminus\{a,b\}} := Y_\ell - \beta_\ell^\top Y_{V\setminus\{a,b\}}$ satisfy

$$\begin{aligned} c_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}} &:= \mathbb{E}\left[\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}\right] \\ &= \Sigma_{ab} - \Sigma_{aV\setminus\{a,b\}}\left(\Sigma_{V\setminus\{a,b\}V\setminus\{a,b\}}\right)^{-1}\Sigma_{V\setminus\{a,b\}b}, \end{aligned} \tag{9.11}$$

which is the *partial covariance* of Y_a and Y_b given $Y_{V\setminus\{a,b\}}$. Similarly, the correlation of the residuals is called *partial correlation* of Y_a and Y_b given $Y_{V\setminus\{a,b\}}$, also known as coherence, and is given by

$$R_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}} := \frac{c_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}}{\sqrt{c_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{a|V\setminus\{a,b\}}}\, c_{\varepsilon_{b|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}}} = -\frac{\left[\Sigma_V^{-1}\right]_{ab}}{\sqrt{\left[\Sigma_V^{-1}\right]_{aa}\left[\Sigma_V^{-1}\right]_{bb}}}. \tag{9.12}$$

Note that from representation (9.12), we see that the partial correlation is completely determined by the concentration (precision) matrix Σ_V^{-1}. Note also that for a Gaussian random vector, zero partial correlation is even equivalent to Y_a and Y_b being independent given $Y_{V\setminus\{a,b\}}$. Finally, it is important to note that in the linear regression problem (9.9), the random variable $\beta_\ell^\top Y_{V\setminus\{a,b\}}$ is the linear projection of Y_l on the linear space generated by the components of $Y_{V\setminus\{a,b\}}$.

An extension of *partial correlation* to stationary time series $\mathcal{Z}_V = (Z_V(t))_{t\in\mathbb{Z}}$ in *discrete time* has been around for quite some time (Tick, 1963) and is ubiquitous in the analysis of multivariate time series (Brillinger, 2001; Gardner, 1988; Priestley, 1981). For these stationary time series models, the partial covariance function of $\mathcal{Z}_a$ and $\mathcal{Z}_b$ given $\mathcal{Z}_{V\setminus\{a,b\}}$ is zero if and only if the partial spectral density function is zero. Thus, in the frequency domain, the role of the partial correlation function is taken over by the *partial spectral coherence function*, which measures the linear dependence between two components $\mathcal{Z}_a$ and $\mathcal{Z}_b$ after removing the linear effects of the remaining components $\mathcal{Z}_{V\setminus\{a,b\}}$ in the frequency domain. Furthermore, the role of the covariance matrix Σ_V in the partial correlation (9.11) is taken over by the matrix-valued spectral density function $f_{Z_V Z_V}(\lambda)$, $\lambda \in [-\pi, \pi]$, in the partial spectral coherence function. This representation gives a very simple characterisation of zero partial correlation. The applications of the partial spectral coherence function are very broad, especially in signal processing, but the word coherence may have a slightly different meaning in different fields (Gardner, 1992).

However, to the best of our knowledge, a mathematically rigorous theory for the concept of *partial correlation* for *continuous-time* processes $\mathcal{Y}_V$ is missing in the literature, so we include the theory in the first chapter of this part. We formally introduce this concept by orthogonal projections on linear space generated by the process $\mathcal{Y}_{V\setminus\{a,b\}}$, similar to the concept of partial correlation for random vectors in (9.11) and similar to our concept of contemporaneous correlation in Section 5.1. We also relate the partial correlation relation to an optimisation problem resembling the linear regression problem (9.9). This optimisation problem is further the multivariate continuous-time counterpart to the discrete-time optimisation problem of Brillinger (2001) and the continuous-time counterpart to Dahlhaus (2000).

After this introduction to the concept of partial correlation, the main subject of this part is *partial correlation graphs*. The partial correlation graph has its *discrete-time* origins in Brillinger (1996) and Dahlhaus (2000) and is a widely used frequency domain approach for constructing graphs. Dahlhaus (2000) establishes the partial correlation graph for multivariate time series as follows: Each component $\mathcal{Z}_a$, $a \in V = \{1, \dots, k\}$, is represented by a vertex. The undirected edges represent a partial spectral coherence function that is not the zero function, meaning that the component processes are partially correlated given the remaining component process. Dahlhaus (2000) also gives various edge characterisations, in particular, a very simple one using the inverse spectral density function of $\mathcal{Z}_V$, similar to (9.12). He also applies the partial correlation graph to vector autoregressive (VAR) processes.

The partial correlation graph of Dahlhaus (2000) has since been used in a wide variety of *applications*, including air pollution data (Dahlhaus, 2000), vital signs of intensive care patients (Gather, Imhoff, & Fried, 2002), human tremor data (Dahlhaus & Eichler, 2003), financial data (Abdelwahab, Amor, & Abdelwahed, 2008), and neuro-physical signals (Dahlhaus, Eichler, & Sandkühler, 1997; Eichler, Dahlhaus, & Sandkühler, 2003; Medkour, Walden, & Burgess, 2009). The usefulness of partial correlation graphs as a visualisation and analysis tool has thus already been demonstrated in discrete time.

In the second chapter of this part, we rigorously define *partial correlation graphs* for multivariate stochastic processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ in *continuous time* as follows: As usual, each component $\mathcal{Y}_a = (Y_a(t))_{t\in\mathbb{R}}$, $a \in V = \{1, \dots, k\}$, is represented by a vertex. Furthermore, we draw an undirected edge between two vertices a and b if and only if, for all $t \in \mathbb{R}$, $Y_a(t)$ and $Y_b(t)$ are uncorrelated given the linear information provided by the environment $\mathcal{Y}_{V\setminus\{a,b\}}$. That is, if and only if $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are partially correlated given $\mathcal{Y}_{V\setminus\{a,b\}}$.

The proposed partial correlation graph is simple in the sense that there are neither loops from a vertex to itself nor any multiple edges between vertices. We further find that the partial correlation graph satisfies the *Markov properties* desired for undirected graphs, so the graphical model is reasonably defined in this sense. Of course, we also provide *edge characterisations*, in particular using the partial spectral coherence function, and the key edge characterisation using the inverse of the spectral density function of $\mathcal{Y}_V$, similar to (9.12). This important characterisation is very simple and user-friendly, computationally inexpensive, and allows the partial correlation graph to be easily applied to example processes. This is an advantage of the partial correlation graph over the (local) causality graph, where the computation of edges can be quite challenging in certain examples, e.g., for ICCSS processes (cf. Chapter 9). The edge characterisations also allow for a comparison to the (local) causality graph.

Finally, as an example, we apply the partial correlation graph to *multivariate continuous-time autoregressive (MCAR) processes* to provide another graphical representation of dependency structures between the components of this important class of processes. We also give edge characterisations by the model parameters of the process and discuss relations to the causality graph and the local causality graph for MCAR processes.

Throughout the discussion of partial correlation graphs, we consider only wide-sense stationary and mean-square continuous multivariate stochastic processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$, which have a spectral density function $f_{Y_V Y_V}(\lambda)$ for $\lambda \in \mathbb{R}$. For a definition and the main properties of such processes, we refer to Section 2.1. We also emphasise that most of the results for the partial correlation graph were developed before the (local) causality graph, but in this order, the relations between the three graphical models can be better highlighted and the thesis remains more transparent.

Structure of the part

The structure of this part is as follows. In Chapter 10, we establish and interpret the concept of partial correlation. We further provide characterisations of this relation and compare it to its discrete-time counterpart. This preliminary work then leads to the definition of the partial correlation graph in Chapter 11. In this chapter, we also discuss edge characterisations and Markov properties, and provide a comparison to the causality graph. As an example, in Chapter 12, we apply the partial correlation graph to MCAR processes and provide edge characterisation by model parameters. We also address the relationship to the causality graph and the local causality graph in this special case.

CHAPTER 10

PRELIMINARIES

In this chapter, we define and elaborate on the concept of partial correlation for continuous-time processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$. Roughly speaking, two subprocesses $\mathcal{Y}_A$ and $\mathcal{Y}_B$ are partially uncorrelated given $\mathcal{Y}_C$ if and only if, for all $t \in \mathbb{R}$, $Y_A(t)$ and $Y_B(t)$ are uncorrelated after removing the linear information provided by $\mathcal{Y}_C$. Here and throughout this chapter, A, B, C are non-empty subsets of V with cardinality $\#A = \alpha$, $\#B = \beta$, and $\#C = \gamma$.

We proceed as follows. In Section 10.1, we define and interpret the partial correlation relation and compute the orthogonal projections therein. Additionally, we study properties of $Y_A(t)$ given the linear information provided by $\mathcal{Y}_C$, i.e., the noise process resulting from the partial correlation relation. Section 10.2 is then devoted to characterisations of the partial correlation relation. We provide characterisations in terms of the spectral density function and the spectral coherence function of the noise processes. Importantly, we present the main characterisation involving the inverse of the spectral density function of the underlying process $\mathcal{Y}_{A\cup B\cup C}$. We conclude the section with the key result that the partial correlation relation satisfies the graphoid properties.

10.1. PARTIAL CORRELATION RELATION

Let us introduce the concept of partial correlation and comment on this definition.

Definition 10.1. Two subprocesses $\mathcal{Y}_A$ and $\mathcal{Y}_B$ of $\mathcal{Y}_V$ are *partially uncorrelated* given another subprocess $\mathcal{Y}_C$ if and only if

$$\mathbb{E}\left[\left(Y_a(t) - P_{\mathcal{L}_{Y_C}} Y_a(t)\right)\overline{\left(Y_b(t) - P_{\mathcal{L}_{Y_C}} Y_b(t)\right)}\right] = 0 \quad \forall\, a \in A,\ b \in B,\ t \in \mathbb{R}.$$

In short, we write $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \,|\, \mathcal{Y}_C$.

Remark 10.2.

(a) In Definition 10.1, it is important to recall that $P_{\mathcal{L}_{Y_C}}$ denotes the orthogonal projection on

$$\mathcal{L}_{Y_C} := \overline{\left\{\sum_{i=1}^{\ell}\sum_{c\in C}\gamma_{c,i}Y_c(t_i) : \gamma_{c,i} \in \mathbb{C},\ -\infty < t_1 \leq \ldots \leq t_\ell < \infty,\ \ell \in \mathbb{N}\right\}}.$$

This closed linear space describes the linear information provided by the process $\mathcal{Y}_C$ over the entire time span. The relation $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \,|\, \mathcal{Y}_C$ thus states, as desired, that, for all $t \in \mathbb{R}$, $Y_A(t)$ and $Y_B(t)$ are uncorrelated given the linear information provided by $\mathcal{Y}_C$. Since for random vectors, as noted above, $\beta_a^\top Y_{V\setminus\{a,b\}}$ and $\beta_b^\top Y_{V\setminus\{a,b\}}$ are the projections of Y_a and Y_b, respectively, onto the linear space generated by $Y_{V\setminus\{a,b\}}$, Definition 10.1 can be seen as an extension of zero partial covariance (9.11) and zero partial correlation (9.12) for random vectors to stochastic processes.

(b) In terms of the conditional orthogonality relation $\perp$ (cf. Definition 3.2), we obtain

$$\begin{aligned}\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \,|\, \mathcal{Y}_C \quad &\Leftrightarrow \quad \mathbb{E}\left[\left(Y^A - P_{\mathcal{L}_{Y_C}} Y^A\right)\left(Y^B - P_{\mathcal{L}_{Y_C}} Y^B\right)\right] = 0\\ &\qquad\quad \forall\, Y^A \in L_{Y_A}(t),\ Y^B \in L_{Y_B}(t),\ t \in \mathbb{R},\\ &\Leftrightarrow \quad L_{Y_A}(t) \perp L_{Y_B}(t) \,|\, \mathcal{L}_{Y_C} \quad \forall\, t \in \mathbb{R}.\end{aligned}$$

(c) Certainly, the partial correlation relation is symmetric and we have

$$\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \,|\, \mathcal{Y}_C \quad \Leftrightarrow \quad \mathcal{Y}_a \perp\!\!\!\perp \mathcal{Y}_b \,|\, \mathcal{Y}_C \quad \forall\, a \in A, b \in B,$$

which is useful for verifying zero partial correlation. Furthermore, statements can be made for $A = \{a\}$ and $B = \{b\}$, and the corresponding multivariate results follow immediately.

To work with the partial correlation relation and to give characterisations, it is crucial to have a frequency representation of the linear space $\mathcal{L}_{Y_C}$.

Lemma 10.3. *Suppose that*

$$\mathcal{L}^*_{Y_C} := \left\{\int_{-\infty}^{\infty} e^{i\lambda t}\varphi(\lambda)\Phi_C(d\lambda) : \varphi \in L^2\left(f_{Y_C Y_C}\right),\ t \in \mathbb{R}\right\},$$

where Φ_C is the random spectral measure from the spectral representation (2.1) *of $\mathcal{Y}_C$ and*

$$L^2\left(f_{Y_C Y_C}\right) := \left\{\varphi^\top : \mathbb{R} \to \mathbb{C}^\gamma : \varphi \text{ measurable, } \int_{-\infty}^{\infty} \left|\varphi(\lambda) f_{Y_C Y_C}(\lambda)\overline{\varphi(\lambda)}^\top\right| d\lambda < \infty\right\}.$$

Then the equality $\mathcal{L}_{Y_C} = \mathcal{L}^*_{Y_C}$ applies.

Proof. The relation $\mathcal{L}_{Y_C} \subseteq \mathcal{L}^*_{Y_C}$ is obvious, since $Y_c(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\Phi_c(d\lambda) \in \mathcal{L}^*_{Y_C}$ for all $c \in C$ and $t \in \mathbb{R}$, and $\mathcal{L}^*_{Y_C}$ is closed. The relation $\supseteq$ is established by Rozanov (1967) on p. 34. ■

We also compute the *orthogonal projections* contained in Definition 10.1 in the next proposition and obtain, as might be expected, the frequency domain analogue of $\beta_\ell^\top Y_{V\setminus\{a,b\}}$ with β_ℓ as in (9.10). We also give an optimisation problem in the frequency domain, resembling the linear regression problem (9.9) for random vectors. Note that the requirement for the existence of a partially positive definite spectral density function allows an explicit representation of the orthogonal projections.

Proposition 10.4. *Suppose that $f_{Y_C Y_C}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Then, for $t \in \mathbb{R}$ and $a \in A$,*

$$P_{\mathcal{L}_C} Y_a(t) = \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}\Phi_C(d\lambda),$$

where Φ_C is the random spectral measure from the spectral representation (2.1) *of $\mathcal{Y}_C$. Furthermore, $P_{\mathcal{L}_C} Y_a(t)$ is the solution to the optimisation problem*

$$\min_{\varphi_{a|C} \in L^2(f_{Y_C Y_C})} \mathbb{E}\left[\left|Y_a(t) - \int_{-\infty}^{\infty} e^{i\lambda t} \varphi_{a|C}(\lambda) \Phi_C(d\lambda)\right|^2\right]. \tag{10.1}$$

Finally, $P_{\mathcal{L}_C} Y_A(t) = (P_{\mathcal{L}_C} Y_a(t))_{a \in A}$ can be calculated component-wise.

Proof. Let $t \in \mathbb{R}$, $a \in A$, and assume that $\{a\} \cap C = \emptyset$, since the statements apply trivially for $a \in C$. To simplify the notation, we abbreviate

$$\widehat{Y}_{a|C}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1} \Phi_C(d\lambda).$$

The proof is divided into three steps. In the first step, we derive that $\widehat{Y}_{a|C}(t) \in \mathcal{L}_{Y_C}$. In the second step, we show that $Y_a(t) - \widehat{Y}_{a|C}(t) \in \mathcal{L}_{Y_C}^{\perp}$. Both together then give the assertion $\widehat{Y}_{a|C}(t) = P_{\mathcal{L}_C} Y_a(t)$. Then, in a third step, we conclude that $f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}$, $\lambda \in \mathbb{R}$, is the solution function to the optimisation problem (10.1).

Step 1: Given that $\mathcal{L}_{Y_C} = \mathcal{L}_{Y_C}^*$ due to Lemma 10.3, we can establish the measurability and integrability of the function $f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}$, $\lambda \in \mathbb{R}$. For the measurability, we first find that the functions $f_{Y_a Y_C}$ and $f_{Y_C Y_C}$ are measurable as derivatives. Furthermore, sums and products of measurable functions are measurable. If we set $\lambda/0 := 0$ for $\lambda \in \mathbb{R}$, then their quotients are also measurable (Klenke, 2020, Theorem 1.91). Now we compute $f_{Y_C Y_C}(\lambda)^{-1}$ by Gaussian elimination and find that $f_{Y_C Y_C}(\lambda)^{-1}$, $\lambda \in \mathbb{R}$, is measurable. Thus, $f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}$, $\lambda \in \mathbb{R}$, is also measurable.

For the integrability, we first note that $f_{Y_{\{a\}\cup C} Y_{\{a\}\cup C}}(\lambda) \geq 0$ due to Lemma 2.4(c). Furthermore, $f_{Y_C Y_C}(\lambda) > 0$ by assumption, so Proposition 8.2.4 of Bernstein (2009) gives

$$f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1} f_{Y_C Y_a}(\lambda) \leq f_{Y_a Y_a}(\lambda).$$

Since further $f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1} f_{Y_C Y_a}(\lambda) \geq 0$ and the integral is monotonous, we obtain

$$\begin{aligned}
&\int_{-\infty}^{\infty} \left| f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1} f_{Y_C Y_C}(\lambda) \overline{f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}}^{\top} \right| d\lambda \\
&= \int_{-\infty}^{\infty} f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1} f_{Y_C Y_a}(\lambda) d\lambda \\
&\leq \int_{-\infty}^{\infty} f_{Y_a Y_a}(\lambda) d\lambda < \infty,
\end{aligned}$$

where the finiteness follows from Lemma 2.4(a). In summary, $\widehat{Y}_{a|C}(t) \in \mathcal{L}_{Y_C}$.

Step 2: Due to Rozanov (1967), I, (7.2), any $Y^C \in \mathcal{L}_{Y_C}$ has a spectral representation

$$Y^C = \int_{-\infty}^{\infty} \varphi(\lambda) \Phi_C(d\lambda) \quad \mathbb{P}\text{-a.s.},$$

where $\varphi \in L^2\left(f_{Y_C Y_C}\right)$. Now, writing $Y_a(t)$ in its spectral representation (2.1), it applies that

$$\begin{aligned}
&\mathbb{E}\Big[\Big(Y_a(t) - \widehat{Y}_{a|C}(t)\Big)\overline{Y^C}\Big] \\
&= \mathbb{E}\left[\left(\int_{-\infty}^{\infty} e^{i\lambda t}\Phi_a(d\lambda) - \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}\Phi_C(d\lambda)\right)\overline{\int_{-\infty}^{\infty}\varphi(\lambda)\Phi_C(d\lambda)}\right] \\
&= \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_a Y_C}(\lambda)\overline{\varphi(\lambda)}^{\top} d\lambda - \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1} f_{Y_C Y_C}(\lambda)\overline{\varphi(\lambda)}^{\top} d\lambda = 0.
\end{aligned}$$

Thus, $Y_a(t) - \widehat{Y}_{a|C}(t) \in \mathcal{L}_{Y_C}^{\perp}$ for $t \in \mathbb{R}$.

Step 3: Since $\mathcal{L}_{Y_C} = \mathcal{L}_{Y_C}^*$ due to Lemma 10.3, the optimisation problem (10.1) is equivalent to

$$\min_{Y^C \in \mathcal{L}_{Y_C}} \mathbb{E}\left[\left|Y_a(t) - Y^C\right|^2\right].$$

From the minimality property of the orthogonal projection, we obtain that $Y^C = P_{\mathcal{L}_C} Y_a(t)$ is the optimal solution to this optimisation problem. Due to Step 1 and Step 2, the function $\varphi_{a|C}(\lambda) = f_{Y_a Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}$, $\lambda \in \mathbb{R}$, is then the optimal function in (10.1). ■

Remark 10.5. For time series in discrete time, the partial correlation relation is also motivated by an *optimisation problem* (Brillinger, 2001, Theorem 8.3.1; Dahlhaus, 2000, relation (2.1) and Definition 2.1) similar to (9.9) for random vectors. However, this optimisation problem is in the time domain instead of the frequency domain. To see the correspondence to (10.1), suppose that the function $\varphi_{a|C}(\lambda)$, $\lambda \in \mathbb{R}$, in the optimisation problem (10.1) is the Fourier transform of an *integrable* function $d_{a|C}(t)$, $t \in \mathbb{R}$. Then, for $t \in \mathbb{R}$, Rozanov (1967), I, Example 8.3, provides

$$\int_{-\infty}^{\infty} e^{i\lambda t}\varphi_{a|C}(\lambda)\Phi_C(d\lambda) = \int_{-\infty}^{\infty} d_{a|C}(t-s) Y_C(s) ds.$$

With this integral representation, we have the similarity of our optimisation problem (10.1) to the discrete-time optimisation problem

$$\min_{d_{a|C}} \mathbb{E}\left[\left|Z_a(t) - \sum_{u=-\infty}^{\infty} d_{a|C}(t-u) Z_C(u)\right|^2\right].$$

Given this parallelism, similarities with Dahlhaus (2000) are to be expected in various sections of this part. However, the advantage of our frequency domain approach is that we require weaker assumptions.

Finally, for $t \in \mathbb{R}$, we define the multivariate *noise process* of the partial correlation relation

$$\varepsilon_{A|C}(t) := Y_A(t) - P_{\mathcal{L}_C} Y_A(t) = Y_A(t) - \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_A Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1}\Phi_C(d\lambda).$$

Below, we present *properties* of these noise processes, which are crucial for the characterisations of the partial correlation relation.

Lemma 10.6. *Suppose that $f_{Y_CY_C}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Then the noise processes $(\varepsilon_{A|C}(t))_{t\in\mathbb{R}}$ and $(\varepsilon_{B|C}(t))_{t\in\mathbb{R}}$ are wide-sense stationary and stationary correlated with (cross-) spectral density function*

$$f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) = f_{Y_AY_B}(\lambda) - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}f_{Y_CY_B}(\lambda) \quad \textit{for almost all } \lambda \in \mathbb{R},$$

and (cross-) covariance function

$$c_{\varepsilon_{A|C}\varepsilon_{B|C}}(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\left(f_{Y_AY_B}(\lambda) - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}f_{Y_CY_B}(\lambda)\right)d\lambda \quad \forall\, t \in \mathbb{R}.$$

Proof. First of all, we can write

$$\begin{aligned}\varepsilon_{A|C}(t) &= \int_{-\infty}^{\infty} e^{i\lambda t}\Phi_A(d\lambda) - \int_{-\infty}^{\infty} e^{i\lambda t} f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}\Phi_C(d\lambda)\\ &= \int_{-\infty}^{\infty} e^{i\lambda t}\left(E_A^\top - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}E_C^\top\right)\Phi_V(d\lambda),\end{aligned}$$

where $E_A \in M_{k\times\alpha}(\mathbb{R})$ (and analogously $E_C \in M_{k\times\gamma}(\mathbb{R})$) is the matrix defined by its entries

$$[E_A]_{ij} = \begin{cases} 1, & i = j,\ i,j \in A,\\ 0, & else. \end{cases}$$

Therefore, $(\varepsilon_{A|C}(t))_{t\in\mathbb{R}}$ is a linear transformation of the wide-sense stationary process $\mathcal{Y}_V$ with spectral characteristic $E_A^\top - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}E_C^\top$, $\lambda \in \mathbb{R}$. Due to Rozanov (1967), I, (8.2), the noise process $(\varepsilon_{A|C}(t))_{t\in\mathbb{R}}$ is also wide-sense stationary. Furthermore, Rozanov (1967), I, (8.13), provides that this linear transformation has a spectral density function, which is, for $\lambda \in \mathbb{R}$,

$$\begin{aligned}f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) &= \left(E_A^\top - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}E_C^\top\right) f_{Y_VY_V}(\lambda)\overline{\left(E_A^\top - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}E_C^\top\right)}^\top\\ &= f_{Y_AY_A}(\lambda) - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}f_{Y_CY_A}(\lambda).\end{aligned}$$

Then, Lemma 2.4(b) yields

$$c_{\varepsilon_{A|C}\varepsilon_{A|C}}(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\left(f_{Y_AY_A}(\lambda) - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}f_{Y_CY_A}(\lambda)\right)d\lambda$$

for $t \in \mathbb{R}$. In particular, the spectral density function of $(\varepsilon_{A\cup B|C}(t))_{t\in\mathbb{R}}$ is given by

$$f_{\varepsilon_{A\cup B|C}\varepsilon_{A\cup B|C}}(\lambda) = f_{Y_{A\cup B}Y_{A\cup B}}(\lambda) - f_{Y_{A\cup B}Y_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}f_{Y_CY_{A\cup B}}(\lambda)$$

for almost all $\lambda \in \mathbb{R}$. Thus, the cross-spectral density function is

$$\begin{aligned}f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) &= E_A^\top\left(f_{Y_{A\cup B}Y_{A\cup B}}(\lambda) - f_{Y_{A\cup B}Y_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}f_{Y_CY_{A\cup B}}(\lambda)\right)E_B\\ &= f_{Y_AY_B}(\lambda) - f_{Y_AY_C}(\lambda)f_{Y_CY_C}(\lambda)^{-1}f_{Y_CY_B}(\lambda)\end{aligned}$$

for almost all $\lambda \in \mathbb{R}$. Finally $(\varepsilon_{A|C}(t))_{t\in\mathbb{R}}$ and $(\varepsilon_{B|C}(t))_{t\in\mathbb{R}}$ are stationary correlated with

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

for all $t \in \mathbb{R}$. ■

Remark 10.7. For the function $f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda)$ we use the terms (cross-) spectral density function of $(\varepsilon_{A|C}(t))_{t\in\mathbb{R}}$ and $(\varepsilon_{B|C}(t))_{t\in\mathbb{R}}$ as well as partial spectral density function of $\mathcal{Y}_A$ and $\mathcal{Y}_B$ given $\mathcal{Y}_C$ interchangeably. The terms (cross-) covariance function of $(\varepsilon_{A|C}(t))_{t\in\mathbb{R}}$ and $(\varepsilon_{B|C}(t))_{t\in\mathbb{R}}$ as well as partial covariance function of $\mathcal{Y}_A$ and $\mathcal{Y}_B$ given $\mathcal{Y}_C$ are also interchangeable.

10.2. Characterisations and properties of partial correlation

In this section, we study several characterisations of the partial correlation relation. We start with simple characterisations in terms of the (cross-) covariance function, the (cross-) spectral density function, and the spectral coherence function of the noise processes $(\varepsilon_{A|C}(t))_{t\in\mathbb{R}}$ and $(\varepsilon_{B|C}(t))_{t\in\mathbb{R}}$, analogous to the discrete-time results in Remark 2.3 of Dahlhaus (2000). The spectral coherence function of the noise processes is also called *partial spectral coherence function* of $\mathcal{Y}_A$ and $\mathcal{Y}_B$ given $\mathcal{Y}_C$ in analogy to the partial correlation for random vectors (9.12).

Proposition 10.8. *Suppose that $f_{Y_C Y_C}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Then, we have*

$$\begin{aligned} \mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_C \quad &\Leftrightarrow \quad c_{\varepsilon_{A|C}\varepsilon_{B|C}}(t) = 0_{\alpha\times\beta} \quad \text{for all } t \in \mathbb{R}, \\ &\Leftrightarrow \quad f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) = 0_{\alpha\times\beta} \quad \text{for almost all } \lambda \in \mathbb{R}. \end{aligned}$$

In particular, these conditions imply that, for almost all $\lambda \in \mathbb{R}$, the spectral coherence function satisfies $R_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) = 0_{\alpha\times\beta}$. If $f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) > 0$ and $f_{\varepsilon_{B|C}\varepsilon_{B|C}}(\lambda) > 0$ for almost all $\lambda \in \mathbb{R}$, then the converse holds as well.

Proof. Suppose that $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B | \mathcal{Y}_C$. By definition of the partial correlation relation, we obtain the first characterisation $c_{\varepsilon_{A|C}\varepsilon_{B|C}}(t) = 0_{\alpha\times\beta}$ for all $t \in \mathbb{R}$. Second, suppose that $c_{\varepsilon_{A|C}\varepsilon_{B|C}}(t) = 0_{\alpha\times\beta}$ for all $t \in \mathbb{R}$. Then the Fourier inversion formula (Pinsky, 2009, Proposition 2.2.37) yields $f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) = 0_{\alpha\times\beta}$ for almost all $\lambda \in \mathbb{R}$. The converse implication applies by Lemma 2.4(b). For the third characterisation, suppose that $f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) = 0_{\alpha\times\beta}$ for almost all $\lambda \in \mathbb{R}$. Then Definition 2.6 implies that $R_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) = 0_{\alpha\times\beta}$ for almost all $\lambda \in \mathbb{R}$. If we additionally assume that $f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) > 0$ and $f_{\varepsilon_{B|C}\varepsilon_{B|C}}(\lambda) > 0$ for almost all $\lambda \in \mathbb{R}$, then Definition 2.6 gives the converse implication. ■

Remark 10.9.

(a) The assumption that $f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) > 0$ for almost all $\lambda \in \mathbb{R}$ excludes the case where $\varepsilon_{A|C}(t) = 0_\alpha$ $\mathbb{P}$-a.s. for $t \in \mathbb{R}$, i.e., the case where $Y_a(t) \in \mathcal{L}_{Y_C}$ for $a \in A$. This can be explained as follows. If $\varepsilon_{A|C}(t) = 0_\alpha \in \mathbb{R}^\alpha$ $\mathbb{P}$-a.s. for $t \in \mathbb{R}$, then $c_{\varepsilon_{A|C}\varepsilon_{A|C}}(t) = 0_\alpha \in M_\alpha(\mathbb{R})$ for $t \in \mathbb{R}$. As a consequence, $f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) = 0_\alpha$ for almost all $\lambda \in \mathbb{R}$. This matrix is certainly not positive definite. From now on, we assume that $A \cap C = \emptyset$.

(b) If $A \cap C = \emptyset$, Bernstein (2009) provides in Proposition 8.2.4 that $f_{Y_{A\cup C}Y_{A\cup C}}(\lambda) > 0$ if and only if $f_{Y_C Y_C}(\lambda) > 0$ and $f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) > 0$. Furthermore, $f_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda) > 0$ is sufficient for $f_{Y_{A\cup C}Y_{A\cup C}}(\lambda) > 0$ and $f_{Y_{B\cup C}Y_{B\cup C}}(\lambda) > 0$.

(c) If $A \cap C = \emptyset$ and $f_{Y_{A\cup C}Y_{A\cup C}}(\lambda) > 0$ for $\lambda \in \mathbb{R}$, then $f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) > 0$ and Proposition 10.8 results in $\mathcal{Y}_A \not\perp\!\!\!\perp \mathcal{Y}_A \mid \mathcal{Y}_C$. In the following, we always assume a sufficient condition for $f_{Y_{A\cup C}Y_{A\cup C}}(\lambda) > 0$, so we can also exclude the case $A \cap B \neq \emptyset$ from our analysis and assume throughout the remaining chapter that $A, B, C \subseteq V$ are disjoint.

We further present a very simple characterisation of the partial correlation relation in terms of an *inverse spectral density function*, which we denote, for $A \subseteq V$ and $\lambda \in \mathbb{R}$, by

$$g_{Y_A Y_A}(\lambda) := f_{Y_A Y_A}(\lambda)^{-1}.$$

The corresponding discrete-time result is given in Theorem 2.4 of Dahlhaus (2000). The proof is not immediately obvious due to the use of the representation of the inverse of a matrix by block matrices twice. Additionally, we could not find an elaborated proof in the literature, so we provide it for the sake of completeness.

Proposition 10.10. *Suppose that $A, B, C \subseteq V$ are disjoint and $f_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Then we have*

$$\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_C \quad \Leftrightarrow \quad [g_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda)]_{AB} = 0_{\alpha\times\beta} \quad \textit{for almost all } \lambda \in \mathbb{R}.$$

Proof. Let $\lambda \in \mathbb{R}$. For notational convenience, we assume without loss of generality that $A := \{1, 2, \ldots, \alpha\}$, $B := \{\alpha+1, \ldots, \alpha+\beta\}$, and $C := \{\alpha+\beta+1, \ldots, \alpha+\beta+\gamma\}$. Then, we can decompose

$$f_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda) = \begin{pmatrix} f_{Y_{A\cup B}Y_{A\cup B}}(\lambda) & f_{Y_{A\cup B}Y_C}(\lambda) \\ f_{Y_C Y_{A\cup B}}(\lambda) & f_{Y_C Y_C}(\lambda) \end{pmatrix}.$$

As $f_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda) > 0$ and $f_{Y_C Y_C}(\lambda) > 0$ by assumption, e.g., Theorem 2.1 of Lu and Shiou (2002) gives

$$\begin{aligned} [g_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda)]_{A\cup B A\cup B} &= \left(f_{Y_{A\cup B}Y_{A\cup B}}(\lambda) - f_{Y_{A\cup B}Y_C}(\lambda) f_{Y_C Y_C}(\lambda)^{-1} f_{Y_C Y_{A\cup B}}(\lambda)\right)^{-1} \\ &= f_{\varepsilon_{A\cup B|C}\varepsilon_{A\cup B|C}}(\lambda)^{-1}. \end{aligned}$$

The matrix $f_{\varepsilon_{A\cup B|C}\varepsilon_{A\cup B|C}}(\lambda)$ itself has the decomposition

$$f_{\varepsilon_{A\cup B|C}\varepsilon_{A\cup B|C}}(\lambda) = \begin{pmatrix} f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) & f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) \\ f_{\varepsilon_{B|C}\varepsilon_{A|C}}(\lambda) & f_{\varepsilon_{B|C}\varepsilon_{B|C}}(\lambda) \end{pmatrix}.$$

Note that $f_{\varepsilon_{A\cup B|C}\varepsilon_{A\cup B|C}}(\lambda) > 0$ and $f_{\varepsilon_{B|C}\varepsilon_{B|C}}(\lambda) > 0$ due to $f_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda) > 0$ and Remark 10.9(b). Thus, we can make use of the matrix inversion formula again and Theorem 2.1 of Lu and Shiou (2002) yields

$$\begin{aligned}[g_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda)]_{AB} &= \left[f_{\varepsilon_{A\cup B|C}\varepsilon_{A\cup B|C}}(\lambda)^{-1}\right]_{AB} \\ &= -\left(f_{\varepsilon_{A|C}\varepsilon_{A|C}}(\lambda) - f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda)f_{\varepsilon_{B|C}\varepsilon_{B|C}}(\lambda)^{-1}f_{\varepsilon_{B|C}\varepsilon_{A|C}}(\lambda)\right)^{-1} \\ &\quad\cdot f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda)f_{\varepsilon_{B|C}\varepsilon_{B|C}}(\lambda)^{-1}.\end{aligned}$$

This representation shows that $[g_{Y_{A\cup B\cup C}Y_{A\cup B\cup C}}(\lambda)]_{AB} = 0_{\alpha\times\beta}$ for almost all $\lambda \in \mathbb{R}$ if and only if $f_{\varepsilon_{A|C}\varepsilon_{B|C}}(\lambda) = 0_{\alpha\times\beta}$ for almost all $\lambda \in \mathbb{R}$. Proposition 10.8 concludes the proof. ■

The following lemma applies Proposition 10.10 to introduce a relationship between the inverse of the spectral density function of a full process $\mathcal{Y}_V$ and the inverse of the spectral density function of a process $\mathcal{Y}_V$ reduced by a *confounder process* $\mathcal{Y}_C$. This result is the continuous-time counterpart to Dahlhaus (2000), Remark 2.5. Again, since we could not find an elaborate proof in the literature, we include it for completeness.

Lemma 10.11. *Suppose that $A, B, C \subseteq V$ are disjoint and $f_{Y_VY_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Then, for $\lambda \in \mathbb{R}$, it applies that*

$$[g_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)]_{AB} = [g_{Y_VY_V}(\lambda)]_{AB} - [g_{Y_VY_V}(\lambda)]_{AC}\left([g_{Y_VY_V}(\lambda)]_{CC}\right)^{-1}[g_{Y_VY_V}(\lambda)]_{CB}\,.$$

Proof. Without loss of generality, we assume that $C = \{k-\gamma+1, k-\gamma+2, \ldots, k\}$. We further note that $f_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda) > 0$ and $[g_{Y_VY_V}(\lambda)]_{CC} > 0$ for $\lambda \in \mathbb{R}$. Then, for $a, b \in V \setminus C$ and $\lambda \in \mathbb{R}$, we first show the identity

$$\begin{aligned}\left[h_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)\right]_{ab} &:= [g_{Y_VY_V}(\lambda)]_{ab} - [g_{Y_VY_V}(\lambda)]_{aC}\left([g_{Y_VY_V}(\lambda)]_{CC}\right)^{-1}[g_{Y_VY_V}(\lambda)]_{Cb} \\ &= \left[g_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)\right]_{ab}.\end{aligned}$$

The main idea of the proof is to show that $h_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)$ is the inverse of $f_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)$ and thus $h_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda) = g_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)$. In doing so, we make use of the fact that $g_{Y_VY_V}(\lambda)$ is the inverse of $f_{Y_VY_V}(\lambda)$ and hence,

$$\sum_{j=1}^{k}[g_{Y_VY_V}(\lambda)]_{aj}\,[f_{Y_VY_V}(\lambda)]_{jb} = \delta_{a=b}, \tag{10.2}$$

where $\delta_{a=b}$ is the Kronecker Delta. For $a, b \in V \setminus C$ and $\lambda \in \mathbb{R}$, it applies that

$$\begin{aligned}&\left[h_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)f_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)\right]_{ab} \\ &= \sum_{j=1}^{k-\gamma}\left([g_{Y_VY_V}(\lambda)]_{aj} - [g_{Y_VY_V}(\lambda)]_{aC}\left([g_{Y_VY_V}(\lambda)]_{CC}\right)^{-1}[g_{Y_VY_V}(\lambda)]_{Cj}\right)[f_{Y_VY_V}(\lambda)]_{jb} \\ &= \sum_{j=1}^{k-\gamma}[g_{Y_VY_V}(\lambda)]_{aj}\,[f_{Y_VY_V}(\lambda)]_{jb} \\ &\quad - [g_{Y_VY_V}(\lambda)]_{aC}\left([g_{Y_VY_V}(\lambda)]_{CC}\right)^{-1}\begin{pmatrix}\sum_{j=1}^{k-\gamma}[g_{Y_VY_V}(\lambda)]_{k-\gamma+1j}\,[f_{Y_VY_V}(\lambda)]_{jb} \\ \vdots \\ \sum_{j=1}^{k-\gamma}[g_{Y_VY_V}(\lambda)]_{kj}\,[f_{Y_VY_V}(\lambda)]_{jb}\end{pmatrix}.\end{aligned}$$

Multiple use of equation (10.2) and $\delta_{k-\gamma+1=b} = \ldots = \delta_{k=b} = 0$ yields

$$\begin{aligned}
&\left[h_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda) f_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)\right]_{ab} \\
&\quad = \delta_{a=b} - \sum_{j=k-\gamma+1}^{k} \left[g_{Y_V Y_V}(\lambda)\right]_{aj} \left[f_{Y_V Y_V}(\lambda)\right]_{jb} \\
&\qquad - \left[g_{Y_V Y_V}(\lambda)\right]_{aC} \left(\left[g_{Y_V Y_V}(\lambda)\right]_{CC}\right)^{-1} \begin{pmatrix} -\sum_{j=k-\gamma+1}^{k} \left[g_{Y_V Y_V}(\lambda)\right]_{k-\gamma+1j} \left[f_{Y_V Y_V}(\lambda)\right]_{jb} \\ \vdots \\ -\sum_{j=k-\gamma+1}^{k} \left[g_{Y_V Y_V}(\lambda)\right]_{kj} \left[f_{Y_V Y_V}(\lambda)\right]_{jb} \end{pmatrix} \\
&\quad = \delta_{a=b} - \left[g_{Y_V Y_V}(\lambda)\right]_{aC} \left[f_{Y_V Y_V}(\lambda)\right]_{Cb} \\
&\qquad + \left[g_{Y_V Y_V}(\lambda)\right]_{aC} \left(\left[g_{Y_V Y_V}(\lambda)\right]_{CC}\right)^{-1} \left[g_{Y_V Y_V}(\lambda)\right]_{CC} \left[f_{Y_V Y_V}(\lambda)\right]_{Cb} \\
&\quad = \delta_{a=b}.
\end{aligned}$$

The function $h_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)$ is thus the left inverse of $f_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)$. The right inverse property can be shown analogously. Finally, let $A, B \subseteq V$. Then, for $\lambda \in \mathbb{R}$,

$$\begin{aligned}
\left[g_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)\right]_{AB} &= \left(\left[g_{Y_{V\setminus C}Y_{V\setminus C}}(\lambda)\right]_{ab}\right)_{a\in A, b\in B} \\
&= \left(\left[g_{Y_V Y_V}(\lambda)\right]_{ab} - \left[g_{Y_V Y_V}(\lambda)\right]_{aC} \left(\left[g_{Y_V Y_V}(\lambda)\right]_{CC}\right)^{-1} \left[g_{Y_V Y_V}(\lambda)\right]_{Cb}\right)_{a\in A, b\in B} \\
&= \left[g_{Y_V Y_V}(\lambda)\right]_{AB} - \left[g_{Y_V Y_V}(\lambda)\right]_{AC} \left(\left[g_{Y_V Y_V}(\lambda)\right]_{CC}\right)^{-1} \left[g_{Y_V Y_V}(\lambda)\right]_{CB}. \qquad \blacksquare
\end{aligned}$$

Interpretation 10.12. For an interpretation of Lemma 10.11 (cf. Remark 2.5 of Dahlhaus, 2000), we analyse the case

$$\left[g_{Y_{V\setminus c}Y_{V\setminus c}}(\lambda)\right]_{ab} = \left[g_{Y_V Y_V}(\lambda)\right]_{ab} - \left[g_{Y_V Y_V}(\lambda)\right]_{ac} \left(\left[g_{Y_V Y_V}(\lambda)\right]_{cc}\right)^{-1} \left[g_{Y_V Y_V}(\lambda)\right]_{cb},$$

$\lambda \in \mathbb{R}$, as an example. This equation explains the relation between the partial correlation of the full process $\mathcal{Y}_V$ and the partial correlation of the reduced process $\mathcal{Y}_{V\setminus\{c\}}$: Suppose $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are partially uncorrelated given $\mathcal{Y}_{V\setminus\{a,b\}}$ ($[g_{Y_V Y_V}(\lambda)]_{ab} = 0$ for almost all $\lambda \in \mathbb{R}$), but there is a partial correlation between $\mathcal{Y}_a$ and $\mathcal{Y}_c$ given $\mathcal{Y}_{V\setminus\{a,c\}}$ and between $\mathcal{Y}_b$ and $\mathcal{Y}_c$ given $\mathcal{Y}_{V\setminus\{b,c\}}$ with $[g_{Y_V Y_V}(\lambda)]_{ac} \neq 0$ and $[g_{Y_V Y_V}(\lambda)]_{cb} \neq 0$ on some non-zero set. This causes a partial correlation between $\mathcal{Y}_a$ and $\mathcal{Y}_b$ given $\mathcal{Y}_{V\setminus\{a,b,c\}}$ ($[g_{Y_{V\setminus c}Y_{V\setminus c}}(\lambda)]_{ab} \neq 0$) in the reduced process $\mathcal{Y}_{V\setminus\{c\}}$.

Finally, we establish that the partial correlation relation satisfies the *graphoid properties* – the main result of this section.

Theorem 10.13. *Suppose that $A, B, C, D \subseteq V$ are disjoint and $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. Then the partial correlation relation defines a graphoid, i.e., it satisfies the following properties:*

(P1) Symmetry: $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_C \Rightarrow \mathcal{Y}_B \perp\!\!\!\perp \mathcal{Y}_A \mid \mathcal{Y}_C$.

(P2) Decomposition: $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_{B\cup C} \mid \mathcal{Y}_D \Rightarrow \mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_D$.

(P3) Weak union: $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_{B\cup C} \mid \mathcal{Y}_D \Rightarrow \mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_{C\cup D}$.

(P4) Contraction: $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_D$ *and* $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_C \mid \mathcal{Y}_{B\cup D} \Rightarrow \mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_{B\cup C} \mid \mathcal{Y}_D$.

(P5) Intersection: $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_{C\cup D}$ *and* $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_C \mid \mathcal{Y}_{B\cup D} \Rightarrow \mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_{B\cup C} \mid \mathcal{Y}_D$.

Proof.
(P1) is immediately clear, since the partial correlation relation is symmetric by definition.

(P2, P3, P5) were established by Dahlhaus (2000) in discrete time in his Lemma 3.1. Since the proof is based only on zero functions in inverse spectral density functions and therefore does not depend on the setting of discrete or continuous time, it is directly applicable.

(P4) follows from Proposition 10.10 and Lemma 10.11. First, the relations $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \mid \mathcal{Y}_D$, $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_C \mid \mathcal{Y}_{B\cup D}$, and Proposition 10.10 result in

$$[g_{Y_{A\cup B\cup D}Y_{A\cup B\cup D}}(\lambda)]_{AB} = 0_{\alpha\times\beta} \quad \text{and} \quad [g_{Y_{A\cup B\cup C\cup D}Y_{A\cup B\cup C\cup D}}(\lambda)]_{AC} = 0_{\alpha\times\gamma} \tag{10.3}$$

for almost all $\lambda \in \mathbb{R}$. Along with Lemma 10.11, we obtain

$$\begin{aligned} 0_{\alpha\times\beta} &= [g_{Y_{A\cup B\cup D}Y_{A\cup B\cup D}}(\lambda)]_{AB} \\ &= [g_{Y_{A\cup B\cup C\cup D}Y_{A\cup B\cup C\cup D}}(\lambda)]_{AB} - [g_{Y_{A\cup B\cup C\cup D}Y_{A\cup B\cup C\cup D}}(\lambda)]_{AC} \\ &\quad \left([g_{Y_{A\cup B\cup C\cup D}Y_{A\cup B\cup C\cup D}}(\lambda)]_{CC}\right)^{-1} [g_{Y_{A\cup B\cup C\cup D}Y_{A\cup B\cup C\cup D}}(\lambda)]_{CB} \\ &= [g_{Y_{A\cup B\cup C\cup D}Y_{A\cup B\cup C\cup D}}(\lambda)]_{AB} \end{aligned} \tag{10.4}$$

for almost all $\lambda \in \mathbb{R}$. In summary, equations (10.3) and (10.4) imply

$$[g_{Y_{A\cup B\cup C\cup D}Y_{A\cup B\cup C\cup D}}(\lambda)]_{A(B\cup C)} = 0_{\alpha\times(\beta+\gamma)}$$

for almost all $\lambda \in \mathbb{R}$, and Proposition 10.10 yields $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_{B\cup C} \mid \mathcal{Y}_D$. ■

Remark 10.14. The partial correlation relation can be characterised by a conditional orthogonality relation (cf. Remark 10.2), which satisfies the graphoid property (cf. Lemma 3.3). However, this result does not directly prove the graphoid property of the partial correlation relation. The reason is as follows: The relation $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_{B\cup C} \mid \mathcal{Y}_D$ is equivalent to $L_{Y_A}(t) \perp L_{Y_{B\cup C}}(t) \mid \mathcal{L}_{Y_D}$, $t \in \mathbb{R}$. Thus, the weak union property of the conditional orthogonality relation gives $L_{Y_A}(t) \perp L_{Y_B}(t) \mid \overline{\mathcal{L}_{Y_D} + L_{Y_C}(t)}$ for all $t \in \mathbb{R}$. This relation is not the same as $L_{Y_A}(t) \perp L_{Y_B}(t) | \mathcal{L}_{Y_{C\cup D}}$, $t \in \mathbb{R}$, i.e., $\mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B | \mathcal{Y}_{C\cup D}$. Similar problems arise for (P4) and (P5).

The peculiarity of the partial correlation relation in continuous time is that it defines a graphoid under *minimal assumptions*. We only require wide-sense stationarity, mean-square continuity, and a positive definite spectral density function. For many graphoids, the intersection property (P5) is quite difficult to verify and, unlike for the proofs of (P1)–(P4), additional, possibly strict, assumptions are required. For example, the conditional orthogonality relation for linear spaces satisfies the intersection property only under the additional assumption of conditional linear separation of underlying linear spaces (Eichler, 2007, Proposition A.1). Thus, graphical models for stochastic processes using conditional orthogonality have additional assumptions on the spectral density, see Eichler (2007), equation (2.1), for processes in discrete time and our Assumption 3 for processes in continuous time, which guarantee that conditional linear separation holds. Similarly, graphical models based on conditional independence also require additional assumptions (cf. Lauritzen, 2004, Proposition 3.1; Eichler, 2011, Assumption S).

CHAPTER 11

Partial correlation graphs

In this chapter, we visualise the partial correlation structure between the components of a multivariate continuous-time process $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ in a graphical model, called the *partial correlation graph* $G_{PC} = (V, E_{PC})$, as follows: Each component $\mathcal{Y}_a$, $a \in V = \{1, \dots, k\}$, is represented by a vertex. Furthermore, $a \,—\, b \notin E_{PC}$ if and only if the components $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are partially uncorrelated given the environment $\mathcal{Y}_{V\setminus\{a,b\}}$. As the relation $\mathcal{Y}_a \perp\!\!\!\perp \mathcal{Y}_b \,|\, \mathcal{Y}_{V\setminus\{a,b\}}$ is symmetric, we use undirected edges in G_{PC}.

We proceed as follows. In Section 11.1, we define the partial correlation graph and provide edge characterisations based on the preliminary work of Section 10.2. In particular, the key characterisation by the inverse spectral density function of $\mathcal{Y}_V$ is given. We also show that the partial correlation graph satisfies several Markov properties. Then, in Section 11.2, we study similarities and differences between the partial correlation graph and the causality graph. We find that, in general, there are no direct relationships between the edges of the two graphical models. However, if we construct the undirected augmented causality graph, the edges of the partial correlation graph are a subset of the edges of the augmented causality graph. Note that the comparison to the local causality graph is discussed in Section 12.2.

11.1. Definition, edge characterisations, and Markov properties

The motivation from the previous introduction leads to the following definition.

Definition 11.1. Suppose that $\mathcal{Y}_V$ is wide-sense stationary, mean-square continuous, and has a spectral density function with $f_{Y_VY_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. If we define $V = \{1, \dots, k\}$ as the vertices and the edges E_{PC} via

$$a \,—\, b \notin E_{PC} \quad \Leftrightarrow \quad \mathcal{Y}_a \perp\!\!\!\perp \mathcal{Y}_b \,|\, \mathcal{Y}_{V\setminus\{a,b\}},$$

for $a, b \in V$ with $a \neq b$, then $G_{PC} = (V, E_{PC})$ is called *partial correlation graph* for $\mathcal{Y}_V$.

Remark 11.2.

(a) The name partial correlation graph is clearly based on the partial correlation relation.

(b) For the definition of G_{PC} it is not necessary to require that $f_{Y_VY_V}(\lambda) > 0$, but sufficient that $f_{Y_{V\setminus\{a,b\}}Y_{V\setminus\{a,b\}}}(\lambda) > 0$ for all $a, b \in V$ with $a \neq b$. However, $f_{Y_VY_V}(\lambda) > 0$ is essential for the graphoid properties and thus for the Markov properties for the graph. Observe that $f_{Y_VY_V}(\lambda) \geq 0$ applies generally (cf. Lemma 2.4), so $f_{Y_VY_V}(\lambda) > 0$ is only a mild assumption.

(c) A direct consequence of Remark 10.9(c) is that we would always have $a \text{---} a \in E_{PC}$ for $a \in V$. Since such self-loops do not help to visualise the partial correlation structure and do not change the properties of the graph, we omit them for the sake of simplicity.

Below, we provide *edge characterisations* for the partial correlation graph. First, Proposition 10.8 directly gives the following characterisations.

Lemma 11.3. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph for $\mathcal{Y}_V$. Then, for $a, b \in V$ with $a \neq b$, the following equivalences apply.*

$$\begin{aligned} a \text{---} b \notin E_{PC} \quad &\Leftrightarrow \quad c_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(t) = 0 \quad \textit{for all } t \in \mathbb{R}, \\ &\Leftrightarrow \quad f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) = 0 \quad \textit{for almost all } \lambda \in \mathbb{R}, \\ &\Leftrightarrow \quad R_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) = 0 \quad \textit{for almost all } \lambda \in \mathbb{R}. \end{aligned}$$

Note that the characterisation by the partial spectral coherence function $R_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda)$, $\lambda \in \mathbb{R}$, is valid, since $f_{Y_V Y_V}(\lambda) > 0$ by assumption and thus $f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{a|V\setminus\{a,b\}}}(\lambda) > 0$ and $f_{\varepsilon_{b|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) > 0$ (cf. Remark 10.9), which results in a non-vanishing denominator. Note also that the partial spectral coherence function provides a measure of the strength of the dependence between $\mathcal{Y}_a$ and $\mathcal{Y}_b$ given $\mathcal{Y}_{V\setminus\{a,b\}}$, an extension of the partial correlation (coherence) (9.12) for random vectors to stochastic processes.

In addition, Proposition 10.10 gives the key representation using the *inverse of the spectral density function*, which is also in analogy to the alternative representation of the partial correlation (coherence) in (9.12). The corresponding edge characterisation for time series in discrete time is established by Dahlhaus (2000) in Theorem 2. We include the proof for completeness.

Proposition 11.4. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph for $\mathcal{Y}_V$. Then, for $a, b \in V$ with $a \neq b$, it applies that*

$$R_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) = -\frac{[g_{Y_V Y_V}(\lambda)]_{ab}}{\left([g_{Y_V Y_V}(\lambda)]_{aa}\,[g_{Y_V Y_V}(\lambda)]_{bb}\right)^{1/2}} \quad \forall\, \lambda \in \mathbb{R}. \tag{11.1}$$

Furthermore,

$$a \text{---} b \notin E_{PC} \quad \Leftrightarrow \quad [g_{Y_V Y_V}(\lambda)]_{ab} = 0 \quad \textit{for almost all } \lambda \in \mathbb{R}.$$

Proof. Let $\lambda \in \mathbb{R}$ and, without loss of generality, set $a = 1$ and $b = 2$. On the one hand,

$$f_{\varepsilon_{\{a,b\}|V\setminus\{a,b\}}\varepsilon_{\{a,b\}|V\setminus\{a,b\}}}(\lambda) = \begin{pmatrix} f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{a|V\setminus\{a,b\}}}(\lambda) & f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) \\ f_{\varepsilon_{b|V\setminus\{a,b\}}\varepsilon_{a|V\setminus\{a,b\}}}(\lambda) & f_{\varepsilon_{b|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) \end{pmatrix},$$

$$f_{\varepsilon_{\{a,b\}|V\setminus\{a,b\}}\varepsilon_{\{a,b\}|V\setminus\{a,b\}}}(\lambda)^{-1} = \frac{1}{D(\lambda)} \begin{pmatrix} f_{\varepsilon_{b|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) & -f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda) \\ -f_{\varepsilon_{b|V\setminus\{a,b\}}\varepsilon_{a|V\setminus\{a,b\}}}(\lambda) & f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{a|V\setminus\{a,b\}}}(\lambda) \end{pmatrix},$$

where $D(\lambda) \coloneqq \det(f_{\varepsilon_{\{a,b\}|V\setminus\{a,b\}}\varepsilon_{\{a,b\}|V\setminus\{a,b\}}}(\lambda))$. On the other hand, $f_{\varepsilon_{\{a,b\}|V\setminus\{a,b\}}\varepsilon_{\{a,b\}|V\setminus\{a,b\}}}(\lambda)^{-1}$ corresponds to the upper left (2×2)-dimensional block of $g_{Y_V Y_V}(\lambda)$, as in the proof of Proposition 10.10. Comparing $f_{\varepsilon_{\{a,b\}|V\setminus\{a,b\}}\varepsilon_{\{a,b\}|V\setminus\{a,b\}}}(\lambda)^{-1}$ and $[g_{Y_V Y_V}(\lambda)]_{\{a,b\}\{a,b\}}$ therefore yields

$$\begin{aligned}\frac{[g_{Y_V Y_V}(\lambda)]_{ab}}{\left([g_{Y_V Y_V}(\lambda)]_{aa}\,[g_{Y_V Y_V}(\lambda)]_{bb}\right)^{1/2}} &= -\frac{f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda)}{\left(f_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{a|V\setminus\{a,b\}}}(\lambda) f_{\varepsilon_{b|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda)\right)^{1/2}}\\ &= -R_{\varepsilon_{a|V\setminus\{a,b\}}\varepsilon_{b|V\setminus\{a,b\}}}(\lambda).\end{aligned}$$

The second part of the assertion follows from Proposition 10.10. ■

Remark 11.5. A significant advantage of Proposition 11.4 over other edge characterisations is that the criterion is computationally inexpensive. One only needs to know the spectral density function and then perform a singular matrix inversion to obtain all the edges in the graph simultaneously. Furthermore, relation (11.1) even equips us with a simple measure of the strength of the connection between the components.

Remark 11.6. Dahlhaus (2000) states that the partial correlation graph can be compared to the *concentration graph* for random vectors. The concentration graph for a random vector $Z \in \mathbb{R}^k$ with $\mathbb{E}\|Z\|^2 < \infty$, $\mathbb{E}(Z) = 0_k$, and $\Sigma_Z := \mathbb{E}[ZZ^\top] > 0$ is defined as follows. Let the vertices be defined as $V = \{1, \dots, k\}$ and the edges E_{CO} via

$$\begin{aligned} a \longrightarrow b \notin E_{CO} \quad &\Leftrightarrow \quad \left[\Sigma_Z^{-1}\right]_{ab} = 0\\ &\Leftrightarrow \quad [\Sigma_Z]_{ab} - [\Sigma_Z]_{aV\setminus\{a,b\}}\left([\Sigma_Z]_{V\setminus\{a,b\}V\setminus\{a,b\}}\right)^{-1}[\Sigma_Z]_{V\setminus\{a,b\}b} = 0,\end{aligned}$$

for $a, b \in V$ with $a \neq b$. Then $G_{CO} = (V, E_{CO})$ is called concentration graph for Z.

The concentration graph G_{CO} visualises the sparsity pattern of the inverse covariance matrix of Z, also called the concentration matrix, hence the name concentration graph. However, because of equations (9.11) and (9.12), the name partial correlation graph would also be appropriate.

The definition of the concentration graph illustrates why the partial correlation graph G_{PC} for continuous-time stochastic processes is the counterpart to the concentration graph G_{CO} for random vectors: A missing edge $a \longrightarrow b \notin E_{CO}$ in the concentration graph for Z reflects that Z_a and Z_b are partially uncorrelated given $Z_{V\setminus\{a,b\}}$. Similarly, $a \longrightarrow b \notin E_{PC}$ in the partial correlation graph for $\mathcal{Y}_V$ means that the stochastic processes $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are partially uncorrelated given $\mathcal{Y}_{V\setminus\{a,b\}}$. Finally, the edges in the partial correlation graph are characterised by the inverse of a spectral density function, which is the frequency domain counterpart to the inverse of a covariance matrix. For an independent and identically distributed sequence of random vectors with distribution Z, the spectral density is even equal to $(2\pi)^{-1}\Sigma_Z$.

Note that the concentration graph is usually defined only for multivariate Gaussian random vectors (Maathuis et al., 2019, p. 218), but this definition is a natural generalisation. For Gaussian random vectors, however, missing edges even correspond to conditional independence relations (Maathuis et al., 2019, Corollary 9.1.2).

To conclude this section, we establish *Markov properties* for G_{PC}. To do this, we first recall some crucial terminology from Part I for the reader's convenience.

In an undirected graph $G = (V, E)$, we denote by $\mathrm{ne}(a) = \{v \in V | a —— v \in E\}$ the set of *neighbours* of $a \in V$. A *path* of length n from a vertex a to a vertex b is a sequence $\alpha_0 = a, \ldots, \alpha_n = b$ of vertices, such that $\alpha_{i-1} —— \alpha_i \in E$ for $i = 1, \ldots, n$. For $A, B, C \subseteq V$, we say that C *separates* A and B if every path from a vertex in A to a vertex in B contains at least one vertex from the separating set C, and we write $A \bowtie B \,|\, C$ for short.

Proposition 11.7. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph for $\mathcal{Y}_V$. Then $\mathcal{Y}_V$ satisfies*

(P) the pairwise Markov property with respect to G_{PC}, i.e., for $a, b \in V$,

$$a \text{ — } b \notin E_{PC} \quad \Rightarrow \quad \mathcal{Y}_a \perp\!\!\!\perp \mathcal{Y}_b \,|\, \mathcal{Y}_{V \setminus \{a,b\}},$$

(L) the local Markov property with respect to G_{PC}, i.e., for $a \in V$,

$$\mathcal{Y}_{V \setminus (ne(a) \cup \{a\})} \perp\!\!\!\perp \mathcal{Y}_a \,|\, \mathcal{Y}_{ne(a)},$$

(G) the global Markov property with respect to G_{PC}, i.e., for disjoint $A, B, C \subseteq V$,

$$A \bowtie B \,|\, C \quad \Rightarrow \quad \mathcal{Y}_A \perp\!\!\!\perp \mathcal{Y}_B \,|\, \mathcal{Y}_C.$$

Proof. The pairwise Markov property is satisfied by definition. Furthermore, the partial correlation relation defines a graphoid by Theorem 10.13. Thus, Lauritzen (2004) states in Theorem 3.7 that the pairwise, local, and global Markov properties are equivalent, so the local and global Markov properties are also valid. ■

As in the (local) causality graph, the pairwise and local Markov properties are properties that are intuitively expected to apply in any reasonably defined graphical model. The global Markov property is especially important, because it provides a graphical criterion for deciding whether two subprocesses $\mathcal{Y}_A$ and $\mathcal{Y}_B$ are partially uncorrelated given a third subprocess $\mathcal{Y}_C$. Although the graph itself is defined only by pairwise partial correlation relations, we obtain partial correlation relations between multivariate subprocesses given any subprocesses through path analysis. For visualisation purposes, we provide an example whose existence is guaranteed by Remark 12.4.

Example 11.8. Let $\mathcal{Y}_V$ be a 5-dimensional continuous-time process, with an associated partial correlation graph given in Figure 11.1.

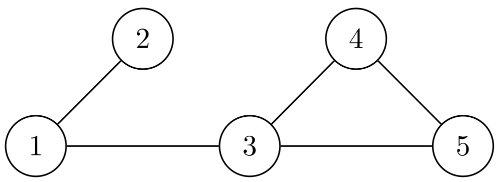


Figure 11.1.: Partial correlation graph for Example 11.8

Since $1 —— 5 \notin E_{PC}$, the pairwise Markov property implies $\mathcal{Y}_1 \perp\!\!\!\perp \mathcal{Y}_5 \,|\, \mathcal{Y}_{\{2,3,4\}}$. Moreover, using the local Markov property, it follows that $\mathcal{Y}_{\{4,5\}} \perp\!\!\!\perp \mathcal{Y}_1 \,|\, \mathcal{Y}_{\{2,3\}}$, since the vertex 1 has the vertices 2 and 3 as neighbours. Finally, the global Markov property even provides that $\mathcal{Y}_{\{4,5\}} \perp\!\!\!\perp \mathcal{Y}_1 \,|\, \mathcal{Y}_3$, since $\{4, 5\} \bowtie \{1\}|\{3\}$.

11.2. Comparison of partial correlation graphs and causality graphs

In this section, we study relations between the partial correlation graph and the causality graph. For the reader's convenience, we restate a possible definition of the causality graph, using specifically the edge characterisations from Remark 4.3 and Lemma 5.2.

Proposition 11.9. *Suppose that $\mathcal{Y}_V$ is wide-sense stationary, mean-square continuous, and satisfies Assumptions 3 and 4. If we define $V = \{1, \ldots, k\}$ as the vertices and the edges E_{CG} via*

(a) $a \longrightarrow b \notin E_{CG} \quad \Leftrightarrow \quad L_{Y_b}(t+h) \perp \mathcal{L}_{Y_a}(t) \mid \mathcal{L}_{Y_{V\setminus\{a\}}}(t) \quad \forall\, 0 \le h \le 1,\ t \in \mathbb{R},$

(b) $a \dashrightarrow\!\!\!\!\! - \, b \notin E_{CG} \quad \Leftrightarrow \quad L_{Y_a}(t+h) \perp L_{Y_b}(t+h') \mid \mathcal{L}_{Y_V}(t) \quad \forall\, 0 \le h, h' \le 1,\ t \in \mathbb{R},$

for $a, b \in V$ with $a \neq b$, then $G_{CG} = (V, E_{CG})$ is the causality graph for $\mathcal{Y}_V$.

Note that Assumption 3 includes the positive definiteness of the spectral density function, so the partial correlation graph is well defined given this assumption.

Remark 11.10. To highlight the differences between the undirected edges in the causality graph and the undirected edges in the partial correlation graph, we recall from Remark 10.2 that in the partial correlation graph, we have

$$a \;\text{---}\; b \notin E_{PC} \quad \Leftrightarrow \quad L_{Y_a}(t) \perp L_{Y_b}(t) \mid \mathcal{L}_{Y_{V\setminus\{a,b\}}} \quad \forall\, t \in \mathbb{R},$$

for $a, b \in V$ with $a \neq b$. The concept of contemporaneous uncorrelatedness in Proposition 11.9(b) thus differs from zero partial correlation in two ways. First, for zero partial correlation, we always project on the linear space of the whole process $\mathcal{Y}_{V\setminus\{a,b\}} = (Y_{V\setminus\{a,b\}}(t))_{t\in\mathbb{R}}$, whereas, for contemporaneous uncorrelatedness, we project on the past $(Y_V(s))_{s\le t}$. Second, in the case of contemporaneous uncorrelatedness, the correlation has to be considered not only at identical time points but also at mixed time points one time step into the future.

Despite the differences between the two concepts (which is also confirmed by the analysis of MCAR processes in Example 12.8), there are relationships between *edges* in the partial correlation graph and *paths* in the mixed causality graph. To show such a relation, we apply the concept of m-separation and the global AMP Markov property for the causality graph.

Lemma 11.11. *Suppose $G_{PC} = (V, E_{PC})$ is the partial correlation graph and $G_{CG} = (V, E_{CG})$ is the causality graph for $\mathcal{Y}_V$. Then, for $a, b \in V$ with $a \neq b$, we have*

$$\{a\} \bowtie_m \{b\} \mid V \setminus \{a,b\} \;\; [G_{CG}] \quad \Rightarrow \quad a \;\text{---}\; b \notin E_{PC}.$$

Proof. The global AMP Markov property, Theorem 7.8, provides that

$$\{a\} \bowtie_m \{b\} \mid V \setminus \{a,b\} \;\; [G_{CG}] \quad \Rightarrow \quad \mathcal{L}_{Y_a} \perp \mathcal{L}_{Y_b} \mid \mathcal{L}_{Y_{V\setminus\{a,b\}}}.$$

This conditional orthogonality relation immediately implies $L_{Y_a}(t) \perp L_{Y_b}(t) \mid \mathcal{L}_{Y_{V\setminus\{a,b\}}}$ for all $t \in \mathbb{R}$ by subset arguments. This relation in turn implies $a \;\text{---}\; b \notin E_{PC}$ due to Remark 10.2(b). ■

The advantage of this result is that the concept of m-separation has several different characterisations in the literature, leading to more sufficient criteria for $a — b \notin E_{PC}$. One approach is, as in Section 7.2, to construct the undirected augmented causality graph (cf. Definition 7.16) and to relate it to the undirected partial correlation graph.

Lemma 11.12. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph, $G_{CG} = (V, E_{CG})$ is the causality graph, and $G^a_{CG} = (V, E^a_{CG})$ is the augmented causality graph for $\mathcal{Y}_V$. Then, for $a, b \in V$ with $a \neq b$, we have*

$$a — b \notin E^a_{CG} \quad \Leftrightarrow \quad dis\,(a \cup ch(a)) \cap dis\,(b \cup ch(b)) = \emptyset \;\; \textit{in } G_{CG}, \tag{11.2}$$

$$\Leftrightarrow \quad \{a\} \bowtie \{b\} \mid V \setminus \{a, b\} \;\; [G^a_{CG}]. \tag{11.3}$$

In particular, each of these conditions implies that $a — b \notin E_{PC}$. Furthermore, $E_{PC} \subseteq E^a_{CG}$.

Proof. By definition, we have

$$a — b \notin E^a_{CG} \quad \Leftrightarrow \quad a \text{ and } b \text{ are not collider connected in } G_{CG}.$$

Furthermore, Eichler (2011) provides in Lemma 3.2 that

$$a \text{ and } b \text{ are not collider connected in } G_{CG} \quad \Leftrightarrow \quad \mathrm{dis}\,(a \cup \mathrm{ch}(a)) \cap \mathrm{dis}\,(b \cup \mathrm{ch}(b)) = \emptyset \text{ in } G_{CG},$$

$$\Leftrightarrow \quad \{a\} \bowtie_m \{b\} \mid V \setminus \{a, b\} \;\; [G_{CG}].$$

In addition, Eichler (2011) states in Theorem 3.1 that

$$\{a\} \bowtie_m \{b\} \mid V \setminus \{a, b\} \;\; [G_{CG}] \quad \Leftrightarrow \quad \{a\} \bowtie \{b\} \mid V \setminus \{a, b\} \;\; [G^a_{CG}].$$

These statements are then, of course, all sufficient for $a — b \notin E_{PC}$ due to the previous Lemma 11.11. In particular, $E_{PC} \subseteq E^a_{CG}$ is valid. ■

Remark 11.13.

(a) This result gives several possibilities to make inferences about the partial correlation graph from the causality graph. On the one hand, the criterion (11.2) is particularly useful, since we can work with the original mixed graph and it is easy to implement algorithmically. On the other hand, the characterisation (11.3) is of particular interest, as it uses the classical separation in an undirected graph, which is a common way to define global Markov properties for mixed graphs. Finally, the inclusion property $E_{PC} \subseteq E^a_{CG}$ gives a simple connection between the edges in both graphs.

(b) The causality graph $G_{CG} = (V, E_{CG})$ and the local causality graph $G^0_{CG} = (V, E^0_{CG})$ satisfy the relation $E^0_{CG} \subseteq E_{CG}$ (cf. Proposition 4.17 and 5.10). Thus, in the augmented (local) causality graph, $E^{a,0}_{CG} \subseteq E^a_{CG}$ is satisfied. We expect that $E_{PC} \subseteq E^{a,0}_{CG} \subseteq E^a_{CG}$ applies. In general, however, we cannot prove the statement $E_{PC} \subseteq E^{a,0}_{CG}$, since we do not have a global AMP Markov property for the local causality graph. We derive this subset relation in Section 12.2, where we restrict ourselves to MCAR(p) processes.

CHAPTER 12

Partial correlation graphs for MCAR processes

In this chapter, we apply the partial correlation graph to causal multivariate continuous-time autoregressive processes of order p, or MCAR(p) processes for short, to illustrate the partial correlation structure between the components of this important and versatile class of processes. In addition to the (local) causality graph, the partial correlation graph provides another visualisation of dependency structures within MCAR processes. We emphasise that partial correlation graphs for Ornstein-Uhlenbeck processes can be obtained as a special case of this chapter. Furthermore, Gaussian MCAR processes and Gaussian Ornstein-Uhlenbeck processes, where the driving Lévy process is a Brownian motion, are also valid special cases.

Throughout this chapter, we use the following matrix notation. Let $A_1, A_2, \ldots, A_p \in M_k(\mathbb{R})$, $p \geq 1$, and denote

$$\mathbf{A} = \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -A_p & -A_{p-1} & \cdots & \cdots & -A_1 \end{pmatrix} \in M_{kp}(\mathbb{R}), \quad \mathbf{B} = \begin{pmatrix} 0_k \\ \vdots \\ 0_k \\ I_k \end{pmatrix} \in M_{kp\times k}(\mathbb{R}),$$
$$\mathbf{C} = \begin{pmatrix} I_k & 0_k & \cdots & 0_k \end{pmatrix} \in M_{k\times kp}(\mathbb{R}).$$

Finally, suppose $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$. The unique causal, strictly stationary output processes $\mathcal{Y}_V = (Y_V(t))_{t\in\mathbb{R}}$ of the corresponding state space models $(\mathbf{A}, \mathbf{B}, \mathbf{C}, L)$ are the causal MCAR(p) processes. For a more detailed introduction to state space models, we refer to Section 2.2.

The structure of the chapter is as follows. In Section 12.1, we ensure that the partial correlation graph for causal MCAR processes is well defined and we establish this graphical model. We also provide edge characterisations by model parameters, along with comparisons to the literature. Moving on to Section 12.2, we study relations between the partial correlation graph and the (local) causality graph. We confirm that, in general, there are no simple direct relations between the edges in the causality graph and the edges in the partial correlation graph, which is a continuation of Section 11.2. We do the same for the local causality graph and the partial correlation graph. For MCAR processes, we additionally show that the edges of the partial correlation graph are a subset of the edges of the augmented local causality graph.

12.1. ESTABLISHMENT AND EDGE CHARACTERISATIONS

To define the partial correlation graph for causal MCAR processes, we must first ensure that the graphical model is *well defined*. However, the assumptions of the partial correlation graph were already studied in Section 2.2, so we only give a brief summary: The MCAR process $\mathcal{Y}_V$ is wide-sense stationary and mean-square continuous, see Remarks 2.12 and 2.13. Furthermore, in equation (2.18), we state that its spectral density function is

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

where, in the notation of the present chapter, the AR polynomial P is

$$P(z) = I_k z^p + A_1 z^{p-1} + \cdots + A_p, \quad z \in \mathbb{C}. \tag{12.1}$$

Finally, we need to ensure that $f_{Y_V Y_V}(\lambda) > 0$ for $\lambda \in \mathbb{R}$. But this condition is already met when $\Sigma_L > 0$ and $\sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ (cf. Lemma 8.5). In particular, we obtain the representation of the inverse spectral density function

$$g_{Y_V Y_V}(\lambda) = 2\pi P(-i\lambda)^\top \Sigma_L^{-1} P(i\lambda), \quad \lambda \in \mathbb{R}.$$

By Definition 11.1, Proposition 11.4, and Proposition 11.7, we then obtain the following result.

Proposition 12.1. *Suppose that $\mathcal{Y}_V$ is a causal MCAR(p) process with $\Sigma_L > 0$. If we define $V = \{1, \dots, k\}$ as the vertices and the edges E_{PC} via*

$$a \text{ --- } b \notin E_{PC} \quad \Leftrightarrow \quad \mathcal{Y}_a \perp\!\!\!\perp \mathcal{Y}_b \mid \mathcal{Y}_{V \setminus \{a,b\}} \quad \Leftrightarrow \quad \Big[P(-i\lambda)^\top \Sigma_L^{-1} P(i\lambda)\Big]_{ab} = 0 \quad \forall\, \lambda \in \mathbb{R},$$

for $a, b \in V$ with $a \neq b$, then the partial correlation graph $G_{PC} = (V, E_{PC})$ for the MCAR process $\mathcal{Y}_V$ is well defined and satisfies the pairwise, local, and global Markov property.

Note that partial correlation graphs can, of course, be defined for more general state space models, but we find that MCAR processes are sufficient for our illustrative purposes. Note also that for MCAR processes, $g_{Y_V Y_V}(\lambda)$, $\lambda \in \mathbb{R}$, has a very simple representation, it is a polynomial matrix. As a result, we can give the following *edge characterisations* based on the coefficient matrices $A_1, A_2, \dots, A_p$ of the AR polynomial, and the covariance matrix Σ_L of the driving Lévy process.

Proposition 12.2. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph for the causal MCAR(p) process $\mathcal{Y}_V$ with AR polynomial P given by* (12.1)*, where we define $A_0 := I_k$. Then, for $a, b \in V$ with $a \neq b$, we obtain the edge characterisation*

$$a \text{ --- } b \notin E_{PC} \quad \Leftrightarrow \quad \left[\sum_{m=0 \vee n-p}^{n \wedge p} (-1)^m A_{p-m}^\top \Sigma_L^{-1} A_{p-n+m}\right]_{ab} = 0 \quad \text{for } n = 0, \dots, 2p.$$

This characterisation is reduced in the following cases.

(a) Suppose $\Sigma_L = \sigma^2 I_k > 0$. *Then*

$$a \text{ --- } b \notin E_{PC} \quad \Leftrightarrow \quad \left[\sum_{m=0\vee n-p}^{n\wedge p} (-1)^m A_{p-m}^\top A_{p-n+m}\right]_{ab} = 0 \quad \textit{for } n = 0, \dots, 2p.$$

(b) Suppose A_m *is a diagonal matrix for all* $m = 1, \dots, p$. *Then*

$$a \text{ --- } b \notin E_{PC} \quad \Leftrightarrow \quad \left[\Sigma_L^{-1}\right]_{ab} = 0.$$

Proof. First of all, we insert the AR polynomial P in $g_{Y_V Y_V}(\lambda)$, $\lambda \in \mathbb{R}$, to get

$$\begin{aligned} g_{Y_V Y_V}(\lambda) &= 2\pi \left(\sum_{m=0}^{p} A_{p-m}^\top (-i\lambda)^m\right) \Sigma_L^{-1} \left(\sum_{\ell=0}^{p} A_{p-\ell}(i\lambda)^\ell\right) \\ &= 2\pi \sum_{n=0}^{2p} \sum_{m=0\vee n-p}^{n\wedge p} (-1)^m A_{p-m}^\top \Sigma_L^{-1} A_{p-n+m}(i\lambda)^n, \end{aligned}$$

where we rearrange the addends according to the degree of λ and substitute $n = \ell + m$, where $n = 0, \dots, 2p$. Since $0 \le \ell = n - m \le p$ and $0 \le m \le p$, we obtain the boundary $0 \vee n - p \le m \le n \wedge p$. Since the components of $g_{Y_V Y_V}(\lambda)$ are polynomials, the components are the zero function if and only if the corresponding coefficients are zero. Then, by Proposition 12.1,

$$a \text{ --- } b \notin E_{PC} \quad \Leftrightarrow \quad \left[\sum_{m=0\vee n-p}^{n\wedge p} (-1)^m A_{p-m}^\top \Sigma_L^{-1} A_{p-n+m}\right]_{ab} = 0 \quad \text{for } n = 0, \dots, 2p. \qquad (12.2)$$

(a) Assume that $\Sigma_L = \sigma^2 I_k$. Then $\sigma^2 > 0$ and $\Sigma_L^{-1} = 1/\sigma^2 I_k$ provides that the relation (12.2) is equivalent to Proposition 12.2(a).

(b) Assume that A_m, $m = 1, \dots, p$, are diagonal matrices. Then the AR polynomial P is a diagonal polynomial matrix and $a \text{ --- } b \notin E_{PC}$ is equivalent to

$$0 = \left[P(-i\lambda)\Sigma_L^{-1} P(i\lambda)\right]_{ab} = \left[P(-i\lambda)\right]_{aa} \left[\Sigma_L^{-1}\right]_{ab} \left[P(i\lambda)\right]_{bb},$$

for all $\lambda \in \mathbb{R}$. Due to $\mathcal{N}(P) = \sigma(\mathbf{A}) \subseteq (-\infty, 0) + i\mathbb{R}$ and because P is diagonal, the diagonal elements of $P(i\lambda)$ are not zero and $a \text{ --- } b \notin E_{PC}$ is equivalent to $[\Sigma_L^{-1}]_{ab} = 0$. ∎

Remark 12.3. The edge characterisations in Proposition 12.2 for MCAR(p) processes are, as might be expected, similar to those for VAR(p) processes in Example 2.2 of Dahlhaus and Eichler (2003). The authors state that in the discrete-time partial correlation graph $G_{PC}^d = (V, E_{PC}^d)$,

$$a \text{ --- } b \notin E_{PC}^d \quad \Leftrightarrow \quad \left[\sum_{m=0\vee n-p}^{n\wedge p} \Phi_m^\top \Sigma_\varepsilon^{-1} \Phi_{m-n+p}\right]_{ab} = 0 \quad \text{for } n = 0, \dots, 2p,$$

with the AR coefficient matrices $\Phi_m \in M_k(\mathbb{R})$, $m = 1, \dots, p$, $\Phi_0 = -I_k$, and the covariance matrix $0 < \Sigma_\varepsilon \in M_k(\mathbb{R})$ of the white noise process. The two characterisations of the continuous-time and the discrete-time multivariate AR process match exactly if we neglect the factor $(-1)^m$.

This small difference is due to the fact that in the continuous-time model, the spectral density function is defined by an AR polynomial at $\pm i\lambda$, whereas in the discrete-time setting the spectral density function is defined by an AR polynomial at $e^{\pm i\lambda}$. The similarity of the characterisations again supports that our partial correlation graph is the continuous-time counterpart to the discrete-time partial correlation graph.

Remark 12.4. A consequence of Proposition 12.2(b) is that for any undirected graph $G = (V, E)$ and any $p \in \mathbb{N}$, there exists an MCAR(p) process with $G_{PC} = G$. Indeed, we can define

$$A_m = \binom{p}{m} I_k \quad \text{for } m = 0, \dots, p, \quad \text{and} \quad \left[\Sigma_L^{-1}\right]_{ab} = \begin{cases} k, & \text{if } a = b, \\ 1, & \text{if } a \neq b \text{ and } a \,\text{---}\, b \in E, \\ 0, & \text{if } a \neq b \text{ and } a \,\text{---}\, b \notin E. \end{cases}$$

Consequently, Marquardt and Stelzer (2007), Corollary 3.8, and the binomial identity provide

$$\begin{aligned} \sigma(\mathbf{A}) &= \mathcal{N}(P) = \{z \in \mathbb{C} : \det(P(z)) = 0\} \\ &= \left\{ z \in \mathbb{C} : 0 = \det\left(\sum_{m=0}^{p} \binom{p}{m} I_k z^{p-m} \right) = \left(\sum_{m=0}^{p} \binom{p}{m} z^{p-m} \right)^k = (1+z)^{kp} \right\} \\ &= \{-1\} \subseteq (-\infty, 0) + i\mathbb{R}. \end{aligned}$$

Furthermore, Σ_L^{-1} is strictly diagonally dominant, i.e., positive definite (Horn & Johnson, 2013, Theorem 6.1.10). Σ_L is also positive definite and there exists a Lévy process with this covariance matrix (Cont & Tankov, 2003, Theorem 4.1). Due to Proposition 12.2(b), the resulting k-dimensional MCAR(p) process $\mathcal{Y}_V$ generates a partial correlation graph $G_{PC} = (V, E_{PC})$ which is identical to the undirected graph $G = (V, E)$. This fact is a major advantage of the partial correlation graph over the causality graph, where it is not clear if a given mixed graph can be constructed as a causality graph of a continuous-time process.

Furthermore, the following sufficient condition for an edge between a and b in the partial correlation graph can be obtained by setting $n = 2p$ in Proposition 12.2.

Lemma 12.5. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph for the causal MCAR(p) process $\mathcal{Y}_V$. Then, for $a, b \in V$ with $a \neq b$, we have*

$$a \,\text{---}\, b \notin E_{PC} \quad \Rightarrow \quad \left[\Sigma_L^{-1}\right]_{ab} = 0.$$

Remark 12.6. The matrix Σ_L^{-1} is the concentration matrix of the random vector $L(1)$, so it defines the concentration graph $G_{CO} = (V, E_{CO})$ of $L(1)$. Lemma 12.5 thus gives the relation $E_{CO} \subseteq E_{PC}$. In other words, the partial correlation of the random variables $L_a(1)$ and $L_b(1)$ given $L_{V \setminus \{a,b\}}(1)$ implies an edge in the partial correlation graph for the continuous-time process $\mathcal{Y}_V$, i.e., the stochastic processes $\mathcal{Y}_a$ and $\mathcal{Y}_b$ are partially correlated given $\mathcal{Y}_{V \setminus \{a,b\}}$. If we additionally assume that A_m, $m = 1, \dots, p$, are diagonal, then Proposition 12.2(b) yields $E_{CO} = E_{PC}$.

Finally, for a visualisation of the previous edge characterisations in Propositions 12.1 and 12.2, we give an example.

Example 12.7. Suppose that $\mathcal{Y}_V$ is a 4-dimensional Ornstein-Uhlenbeck process with $\Sigma_L = I_4$ and

$$\mathbf{A} = \begin{pmatrix} -2 & 0 & 1 & 1 \\ 0 & -2 & -1 & -1 \\ -1 & -1 & -2 & -1 \\ 1 & -1 & -1 & -2 \end{pmatrix}.$$

Then a straightforward calculation yields $\sigma(\mathbf{A}) = \{-1, -1, -2, -4\} \subseteq (-\infty, 0) + i\mathbb{R}$. Furthermore, for an Ornstein-Uhlenbeck process $\mathcal{Y}_V$, the inverse spectral density function is simplified to $g_{Y_V Y_V}(\lambda) = 2\pi(-i\lambda I_k - \mathbf{A}^\top)\Sigma_L^{-1}(i\lambda I_k - \mathbf{A})$ for $\lambda \in \mathbb{R}$. We obtain

$$g_{Y_V Y_V}(\lambda) = 2\pi \begin{pmatrix} \lambda^2 + 6 & 0 & 2i\lambda - 1 & -3 \\ 0 & \lambda^2 + 6 & 5 & 5 \\ -2i\lambda - 1 & 5 & \lambda^2 + 7 & 6 \\ -3 & 5 & 6 & \lambda^2 + 7 \end{pmatrix}.$$

The corresponding partial correlation graph $G_{PC} = (V, E_{PC})$ is then given in Figure 12.1.

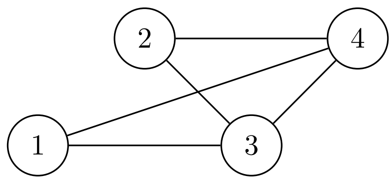


Figure 12.1.: Partial correlation graph for Example 12.7

Furthermore, for an Ornstein-Uhlenbeck process with $\Sigma_L = I_k$, the edge characterisation in Proposition 12.2(a) is simplified to

$$a - b \notin E_{PC} \quad \Leftrightarrow \quad [\mathbf{A}]_{ba} - [\mathbf{A}]_{ab} = 0, \ \left[\mathbf{A}^\top \mathbf{A}\right]_{ab} = 0. \tag{12.3}$$

Of course, relation (12.3) also provides the edges in Figure 12.1. For example, $1 - 3 \in E_{PC}$, as

$$[\mathbf{A}]_{31} - [\mathbf{A}]_{13} = -1 - 1 \neq 0.$$

Furthermore, $1 - 2 \notin E_{PC}$, since

$$[\mathbf{A}]_{21} - [\mathbf{A}]_{12} = 0 - 0 = 0 \quad \text{and}$$
$$\left[\mathbf{A}^\top \mathbf{A}\right]_{12} = (-2) \cdot 0 + 0 \cdot (-2) + (-1) \cdot (-1) + 1 \cdot (-1) = 0.$$

In summary, Example 12.7 once again highlights the advantage of edge characterisation in Proposition 12.1, which is the ability to obtain all edges simultaneously through the inverse of the spectral density function.

12.2. Comparison of partial correlation graphs and (local) causality graphs

In this section, we relate the partial correlation graph to the causality graph and the local causality graph of Chapter 6, in the special case of MCAR processes. This discussion can be seen as a continuation of Section 11.2.

We start with relations between the partial correlation graph and the *causality graph*. In the comparison in Section 11.2, we suspected that, in general, there is no direct relation between the edges in the causality graph and the partial correlation graph, although $E_{PC} \subseteq E_{CG}^{a}$. We now confirm this presumption with two counterexamples.

Example 12.8. Recall that for the Ornstein-Uhlenbeck process with $\Sigma_L = I_k$, due to Proposition 12.2 with $p = 1$, it applies that

$$a \text{ --- } b \notin E_{PC} \quad \Leftrightarrow \quad [\mathbf{A}]_{ba} - [\mathbf{A}]_{ab} = 0, \quad \left[\mathbf{A}^\top \mathbf{A}\right]_{ab} = 0,$$

in the partial correlation graph. In addition, by Corollary 8.24, we obtain the edge characterisations

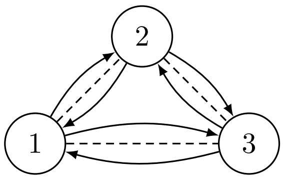
$$\begin{aligned} a \longrightarrow b \notin E_{CG} \quad &\Leftrightarrow \quad [\mathbf{A}^\alpha]_{ba} = 0, && \alpha = 1, \ldots, k-1, \\ a \dashrightarrow b \notin E_{CG} \quad &\Leftrightarrow \quad \left[\mathbf{A}^\alpha \left(\mathbf{A}^\top\right)^\beta\right]_{ab} = 0, && \alpha, \beta = 0, \ldots, k-1, \end{aligned}$$

for the causality graph.

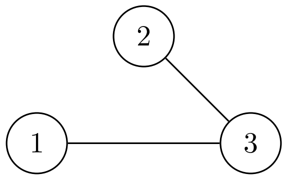
(a) Suppose that $\mathcal{Y}_V$ is a 3-dimensional Ornstein-Uhlenbeck process with $\Sigma_L = I_3$ and

$$\mathbf{A} = \begin{pmatrix} -3 & 1 & 1 \\ 1 & -3 & 1 \\ 6 & 1 & -8 \end{pmatrix},$$

$\sigma(\mathbf{A}) = \{-1, -4, -9\} \subseteq (-\infty, 0) + i\mathbb{R}$. Then $[\mathbf{A}]_{21} - [\mathbf{A}]_{12} = 0$ and $[\mathbf{A}^\top \mathbf{A}]_{12} = 0$, so $1 \text{ --- } 2 \notin E_{PC}$. However, $1 \longrightarrow 2 \in E_{CG}$, $2 \longrightarrow 1 \in E_{CG}$, and $1 \dashrightarrow 2 \in E_{CG}$, since $[\mathbf{A}]_{21} \neq 0$ and $[\mathbf{A}]_{12} \neq 0$. While there exist all three edges between 1 and 2 in the causality graph, there is no edge between 1 and 2 in the partial correlation graph. The corresponding graphical models are given in Figure 12.2.



(a) Causality graph



(b) Partial correlation graph

Figure 12.2.: Graphical models for Example 12.8(a)

(b) Suppose that $\mathcal{Y}_V$ is a 3-dimensional Ornstein-Uhlenbeck process with $\Sigma_L = I_3$ and

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & -2 \end{pmatrix},$$

$\sigma(\mathbf{A}) = \{-1, -1, -2\} \subseteq (-\infty, 0) + i\mathbb{R}$. Then a simple calculation shows that

$$[\mathbf{A}^\alpha]_{21} = [\mathbf{A}^\alpha]_{12} = 0, \quad \alpha = 1, 2, \quad \left[\mathbf{A}^\alpha \left(\mathbf{A}^\top\right)^\beta\right]_{12} = 0, \quad \alpha, \beta = 0, 1, 2.$$

Therefore, $1 \longrightarrow 2 \notin E_{CG}$, $2 \longrightarrow 1 \notin E_{CG}$, and $1 \dashrightarrow 2 \notin E_{CG}$. However, $1 — 2 \in E_{PC}$, since $[\mathbf{A}^\top \mathbf{A}]_{12} = 1$. While there is no edge between the vertices 1 and 2 in the causality graph, there is an edge between 1 and 2 in the partial correlation graph. The corresponding graphical models are given in Figure 12.3.

Figure 12.3.: Graphical models for Example 12.8(b)

In summary, even in the special case $\Sigma_L = I_k$, there are no direct relations between the edges. In the partial correlation graph, the orthogonality of the columns in $\mathbf{A}$ is characteristic, whereas in the causality graph, the orthogonality of the rows is relevant for the undirected edges, and the orthogonality of the rows and columns is relevant for the directed edges. Of course, in some special cases, there are simple relations between the edges in the partial correlation graph and the edges in the causality graph. Because of the reasoning about orthogonality, an obvious special case is a symmetric matrix $\mathbf{A}$.

Lemma 12.9. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph and $G_{CG} = (V, E_{CG})$ is the causality graph for the causal Ornstein-Uhlenbeck process $\mathcal{Y}_V$, where $\mathbf{A} = \mathbf{A}^\top$ and $\Sigma_L = I_k$. Then, for $a, b \in V$ with $a \neq b$, it applies that*

$$a \dashrightarrow b \notin E_{CG} \quad \Rightarrow \quad a — b \notin E_{PC}.$$

Proof. The assumptions that $\Sigma_L = I_k$, $\mathbf{A}$ is symmetric, and Corollary 8.24 imply

$$[\mathbf{A}]_{ba} - [\mathbf{A}]_{ab} = 0, \quad \left[\mathbf{A}^\top \mathbf{A}\right]_{ab} = \left[\mathbf{A}\mathbf{A}^\top\right]_{ab} = 0.$$

Thus, Proposition 12.2 yields $a — b \notin E_{PC}$. ■

Next, we provide a comparison of the partial correlation graph to the *local causality graph.* To avoid going too deep into the intricate definition of the local causality graph in its generality, but to improve readability, we restate the local causality graph only for MCAR processes via the characterisations from Propositions 8.14 and 8.16.

Proposition 12.10. *Suppose that $\mathcal{Y}_V$ is a causal MCAR(p) process with $\Sigma_L > 0$. If we define $V = \{1, \dots, k\}$ as the vertices and the edges E^0_{CG} via*

(a) $a \longrightarrow b \notin E^0_{CG} \quad \Leftrightarrow \quad [A_j]_{ba} = 0 \quad$ *for* $j = 1, \dots, p$,

(b) $a \dashrightarrow\!\!\!\!\!\!- b \notin E^0_{CG} \quad \Leftrightarrow \quad [\Sigma_L]_{ab} = 0$,

for $a, b \in V$ with $a \neq b$, then $G^0_{CG} = (V, E^0_{CG})$ is the local causality graph for $\mathcal{Y}_V$.

Remark 12.11. We emphasise that the undirected edges in the local causality graph are characterised by Σ_L and not Σ_L^{-1} as in the partial correlation graph, and these matrices generally do not match. The local causality graph considers the direct correlation of $L_a(1)$ and $L_b(1)$, while the partial correlation graph considers the correlation of $L_a(1)$ and $L_b(1)$ given the environment $L_{V\setminus\{a,b\}}(1)$. Of course, in the case of no environment $k = 2$, we obtain that $[\Sigma_L]_{ab} = 0$ if and only if $[\Sigma_L^{-1}]_{ab} = 0$, and $a \longrightarrow\!\!\!\!\!\!- b \notin E_{PC}$ implies $a \dashrightarrow\!\!\!\!\!\!- b \notin E^0_{CG}$.

Again, because of the different definitions, there are generally no direct relations between the edges in the partial correlation graph and the edges in the local causality graph, not even in the special case of $\Sigma_L = I_k$. To illustrate this, we continue with Example 12.8.

Example 12.12. First, we observe that in Example 12.8(a,b), we assume that $\Sigma_L = I_k$. Thus, $a \dashrightarrow\!\!\!\!\!\!- b \notin E^0_{CG}$ is always true. Continuing Example 12.8(a), we still get $1 \longrightarrow\!\!\!\!\!\!- 2 \notin E_{PC}$, but $1 \longrightarrow 2 \in E^0_{CG}$ and $2 \longrightarrow 1 \in E^0_{CG}$, since $[\mathbf{A}]_{21} = [\mathbf{A}]_{12} \neq 0$. While there are both directed edges between the vertices 1 and 2 in the local causality graph, there is no edge between 1 and 2 in the partial correlation graph. Furthermore, in Example 12.8(b), we still get $1 \longrightarrow\!\!\!\!\!\!- 2 \in E_{PC}$, but $1 \longrightarrow 2 \notin E^0_{CG}$ and $2 \longrightarrow 1 \notin E^0_{CG}$, since $[\mathbf{A}]_{21} = [\mathbf{A}]_{12} = 0$. While there are no edges between the vertices 1 and 2 in the local causality graph, there is an edge between 1 and 2 in the partial correlation graph. The corresponding local causality graphs are given in Figure 12.4.

(a) Local causality graph for Example 12.8(a) (b) Local causality graph for Example 12.8(b)

Figure 12.4.: Graphical models for Example 12.8 and Example 12.12

However, as for the causality graph, we can establish relations between *edges* in the partial correlation graph and *paths* in the local causality graph via the concept of m-separation and augmentation separation, although no global AMP Markov property could be shown for the local causality graph.

Lemma 12.13. *Suppose that $G_{PC} = (V, E_{PC})$ is the partial correlation graph, $G^0_{CG} = (V, E^0_{CG})$ is the local causality graph, and $G^{0,a}_{CG} = (V, E^{0,a}_{CG})$ is the augmented local causality graph for the causal MCAR(p) process $\mathcal{Y}_V$. Then, for $a, b \in V$ with $a \neq b$, we have*

$$\begin{aligned} a \text{---} b \notin E^{0,a}_{CG} \quad &\Leftrightarrow \quad \{a\} \bowtie_m \{b\} \mid V \setminus \{a,b\} \quad [G^0_{CG}], \\ &\Leftrightarrow \quad \mathit{dis}\,(a \cup ch(a)) \cap \mathit{dis}\,(b \cup ch(b)) = \emptyset \text{ in } G^0_{CG}, \\ &\Leftrightarrow \quad \{a\} \bowtie \{b\} \mid V \setminus \{a,b\} \quad [G^{0,a}_{CG}]. \end{aligned}$$

In particular, each of these conditions implies that $a \text{---} b \notin E_{PC}$. Furthermore, $E_{PC} \subseteq E^{0,a}_{CG}$.

Proof. In Lemma 11.12, we already established the equivalences

$$\begin{aligned} a \text{---} b \notin E^{0,a}_{CG} \quad &\Leftrightarrow \quad a \text{ and } b \text{ are not collider connected in } G^0_{CG}, \\ &\Leftrightarrow \quad \mathrm{dis}\,(a \cup \mathrm{ch}(a)) \cap \mathrm{dis}\,(b \cup \mathrm{ch}(b)) = \emptyset \text{ in } G^0_{CG}, \\ &\Leftrightarrow \quad \{a\} \bowtie \{b\} \mid V \setminus \{a,b\} \quad [G^{0,a}_{CG}], \\ &\Leftrightarrow \quad \{a\} \bowtie_m \{b\} \mid V \setminus \{a,b\} \quad [G^0_{CG}], \end{aligned}$$

regardless of the specific definition of the graph. Thus, we only need to prove $a \text{---} b \notin E_{PC}$, which we do by contradiction. So suppose that $a \text{---} b \in E_{PC}$. Then there exists a $\lambda \in \mathbb{R}$, such that

$$0 \neq \left[P(-i\lambda)^\top \Sigma_L^{-1} P(i\lambda)\right]_{ab} = \sum_{c \in V} \sum_{d \in V} [P(-i\lambda)]_{ca} \left[\Sigma_L^{-1}\right]_{cd} [P(i\lambda)]_{db}.$$

Consequently, there exist vertices $c, d \in V$, such that

$$[P(-i\lambda)]_{ca} \neq 0, \quad \left[\Sigma_L^{-1}\right]_{cd} \neq 0, \quad [P(i\lambda)]_{db} \neq 0.$$

The statements about the AR polynomial yield that there exist directed edges $a \longrightarrow c$ and $b \longrightarrow d$ in G^0_{CG}. The statement $[\Sigma_L^{-1}]_{cd} \neq 0$ yields that there exists a path π between c and d of only undirected edges in the local causality graph G^0_{CG} (Eichler, 2007, p. 341).

Indeed, consider an Ornstein-Uhlenbeck process $\widetilde{\mathcal{Y}}_V$ with $\mathbf{A} = -I_k$, that is driven by the same Lévy process as the MCAR process $\mathcal{Y}_V$. Then $\widetilde{\mathcal{Y}}_V$ generates a partial correlation graph $\widetilde{G}_{PC} = (V, \widetilde{E}_{PC})$ and a causality graph $\widetilde{G}_{CG} = (V, \widetilde{E}_{CG})$. For these graphical models, we have $c \text{---} d \notin \widetilde{E}_{PC}$ if and only if $[\Sigma_L^{-1}]_{cd} = 0$ and $c \dashrightarrow d \notin \widetilde{E}_{CG}$ if and only if $[\Sigma_L]_{cd} = 0$. Additionally, there are no directed edges in the causality graph. Then, a consequence of Lemma 11.12 is that $c \text{---} d \in \widetilde{E}_{PC}$ ($[\Sigma_L^{-1}]_{cd} \neq 0$) implies $c \text{---} d \in \widetilde{E}^a_{CG}$ and $\mathrm{dis}(c) \cap \mathrm{dis}(d) \neq \emptyset$ in $\widetilde{E}_{CG}$. Thus, there exists a path $\widetilde{\pi}$ of only undirected edges between c and d in the causality graph $\widetilde{G}_{CG}$, i.e., for some $c = \alpha_1, \ldots, \alpha_\ell = d \in V$, we have $[\Sigma_L]_{\alpha_i \alpha_{i+1}} \neq 0$ for $i = 1, \ldots, \ell - 1$. This result implies a path $\widetilde{\pi}$ of only undirected edges between c and d in the local causality graph G^0_{CG}.

Finally, if $\widetilde{\pi}$ does not already provide a path between a and b, we complete $\widetilde{\pi}$ with one or both directed edges from above, to get a path π between a and b on which every intermediate vertex is a collider. This path contradicts the premise and the statement $a \text{---} b \notin E_{PC}$ is valid. In particular, $E_{PC} \subseteq E^{0,a}_{CG}$ is satisfied. ■

As discussed in Remark 11.13, Lemma 12.13 provides several ways to make statements about the partial correlation graph from the local causality graph. Again, we support the result with an example and continue with Examples 12.8 and 12.12.

Example 12.14. In both Example 12.8(a) and (b), we obtain the augmented local causality graph given in Figure 12.5, emphasising the relation $E_{PC} \subseteq E_{CG}^{0,a}$. In particular, in Example 12.8(a) we have the true subset relation $E_{PC} \subset E_{CG}^{0,a}$, whereas in Example 12.8(b) the equality $E_{PC} = E_{CG}^{0,a}$ applies.

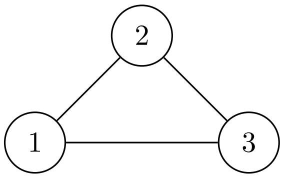


Figure 12.5.: Augmented local causality graph for Example 12.8(a,b) and Example 12.14

CHAPTER 13

Conclusion and outlook

In this thesis, we establish three graphical models that visualise different types of dependencies between the components of stationary multivariate stochastic processes in continuous time. In the following, we summarise the main results of this work and give an outlook on resulting research questions.

The first two graphical models, established in Part I, are the *causality graph* and the *local causality graph*. In both graphical models, the vertices stand for the respective components $\mathcal{Y}_1, \ldots, \mathcal{Y}_k$ of an underlying multivariate process $\mathcal{Y}_V$. The directed edges represent (local) Granger causality, while the undirected edges represent (local) contemporaneous correlation between the components. We now highlight some important aspects of both mixed graphs.

First, the edge definitions in the (local) causality graph are inspired by approaches from the literature on dependency structures in discrete-time and continuous-time processes. We take these concepts and adapt them to stationary multivariate processes in continuous time, to obtain appropriate and meaningful definitions of dependency structures between the components of $\mathcal{Y}_V$. Second, the (local) causality graph satisfies various levels of Markov properties. Although the graphical models are defined only by pairwise relationships between the components, these properties can be used together with path analysis to identify multivariate Granger non-causality relations and information about multivariate subprocesses being contemporaneously uncorrelated, even with respect to partial information. In doing so, only few assumptions about the underlying process are required. That is, wide-sense stationarity, mean-square continuity, purely non-determinism, and the existence of a spectral density function that must satisfy a mild regularity assumption. However, there is an important difference between the two types of graphical models. The advantage of the causality graph is that it satisfies the global AMP and the global Granger-causal Markov property. The local causality graph probably does not satisfy global Markov properties with the same generality, so we provide additional but easily verifiable assumptions. Finally, due to the universal applicability of both graphical models, the (local) causality graph applies to very general output processes of state space models. For MCAR processes and ICCSS processes, we are even able to establish edge characterisations by the model parameters. For this purpose, we first prove essential properties of the processes and, in particular, determine orthogonal projections of these processes on linear spaces. The resulting edge characterisations by model parameters are meaningful in terms of interpretation and comparison to the literature, emphasising once again that the edges, respectively dependency structures, are sensibly defined. Furthermore, as expected, the characterisations for the edges in the local causality graph are less restrictive than those for the edges in the causality graph. The local causality graph has fewer edges than the causality graph and, in general, the graphs are not identical. The advantage of the local causality graph over the causality graph is that it allows more general graphs to be generated. In fact, it is guaranteed that any mixed graph can be obtained as a local causality graph, e.g., the one in Figure 13.1a.

(a) A mixed graph

(b) An undirected graph

Figure 13.1.: Two graphical models

In Part II, we establish a third graphical model, the *partial correlation graph*, which is an undirected graph. As in the (local) causality graph, the vertices stand for the respective components $\mathcal{Y}_1, \dots, \mathcal{Y}_k$ of the underlying multivariate process $\mathcal{Y}_V$. However, the undirected edges in the graph represent partial correlation relations. Again, we highlight some key aspects of this graphical model.

First, in defining this graph, we take the concept of partial correlation for random vectors and time series in discrete time and adapt it to continuous-time processes. In this way, we again obtain a sensible definition of dependency structures between the components of the process. However, unlike the discrete-time partial correlation relation, our continuous-time concept is based on the insights that have already been gained from Part I, i.e., on orthogonal projections, rather than on an optimisation problem. Nevertheless, an alternative approach via an optimisation problem is available, which emphasises the similarity of the discrete-time and continuous-time partial correlation relation.

Second, an advantage of the partial correlation graph is that the assumptions about the process can be relaxed compared to the (local) causality graph. We only require that the underlying process is wide-sense stationary, mean-square continuous, and has a positive definite spectral density function. Still, the partial correlation graph satisfies the usual Markov properties for undirected graphs, in particular the global Markov property. Again, this means that even if the edges are defined by pairwise partial correlation relations, multivariate relations can be derived by path analysis, including those with respect to partial information.

However, the advantage of the partial correlation graph is not only its generality but also its simplicity. The edges have a very simple characterisation in terms of zero entry functions in the inverse of the spectral density function. This characterisation allows all edges in the graph to be determined simultaneously in a very simple and computationally inexpensive manner. At the same time, the inverse of the spectral density function determines the partial spectral coherence function, allowing for statements about the strength of the dependence.

Due to its generality, the partial correlation graph can, of course, also be applied to state space processes, such as MCAR processes, for which the inverse of the spectral density function can be used to derive edge characterisations in terms of the model parameters of the process. The edge characterisations for MCAR processes are interpretatively meaningful and allow to ensure the existence of all undirected graphs as a partial correlation graph, e.g., the one in Figure 13.1b.

Finally, we note that the partial correlation graph can be related to the (local) causality graph. Although there are no simple direct relations between the edges, the partial correlation graph is related to the augmented (local) causality graph. More specifically, the edges in the partial correlation graph are a subset of the edges in the augmented causality graph. At least for MCAR processes, the edges in the partial correlation graph are even a subset of the edges in the augmented local causality graph.

Despite the comprehensive answers provided in this thesis, there arise some research questions worthy of further investigation. Addressing these questions could lead to further developments and applications of the models presented and to new models that go beyond them. We divide the research questions into the following areas: open questions for the (local) causality graph, open questions for the partial correlation graph, and suggestions for alternative graphical models.

First, we present important open questions for the (local) causality graph, both for the general model and the application to MCAR and ICCSS processes.

(1) In Section 7.3, we address global Markov properties for the local causality graph, making additional assumptions. Thus, we do not derive global Markov properties for the local causality graph with the same generality as for the causality graph. This motivates the question of whether the assumptions can be relaxed.

(2) In the case of MCAR and ICCSS processes, it is sufficient for our purposes to specify the orthogonal projections onto linear subspaces only semi-explicitly, as commented, e.g., in Remark 8.4. The question arises to what extent these projections can be specified. This, in turn, leads to the study of whether and how the maximal decomposition of a spectral density function can be explicitly specified.

(3) For the causality graph, we cannot generate every mixed graph as a causality graph using MCAR processes, see Remark 8.30. Thus, we wonder which graphs can be generated and which cannot.

(4) One of the assumptions for our state space processes, contained in the acronym ICCSS processes, is the existence of a coprime right polynomial matrix fraction description (2.13) of the transfer function with polynomials as in (2.14). The existence of this decomposition is claimed in the literature without proof in the context of state space models and thus remains an open question.

(5) From a statistical point of view, a natural question is how to test for the presence of edges in our (local) causality graph. Based on our edge characterisations by model parameters, it seems feasible to derive tests for the directed and undirected edges in the (local) causality graph, e.g., for MCAR processes. For the model parameters, which are unknown in practice, estimators from the literature can be used. These are the quasi-maximum likelihood estimator (Schlemm & Stelzer, 2012b) or the Whittle estimator (Fasen-Hartmann & Mayer, 2022) for state space models.

Secondly, we present open questions for the partial correlation graph, where, based on our Proposition 11.4, the edges are characterised by zero entry functions in the inverse of the spectral density function. Again, it is of interest to test for edges in the partial correlation graph, respectively for zero entry functions in the the inverse spectral density function. To do this, it is natural to apply estimators for the spectral density function from the literature.

(6) In the low-frequency sampling scheme, we can use the aforementioned parametric approach, e.g., for MCAR processes. The model parameters can be estimated by quasi-maximum likelihood estimation (Schlemm & Stelzer, 2012b) or Whittle estimation (Fasen-Hartmann & Mayer, 2022). From these estimators, an estimator for the inverse spectral density function can be constructed.

(7) In the high-frequency sampling scheme, it seems straightforward to generalise the smoothed normalised periodogram for CARMA processes (Fasen & Fuchs, 2013a, 2013b) to the multivariate setting. Alternatively, the lag-window spectral density estimator for stationary processes in a Hilbert space (Kartsioukas, Stoev, & Hsing, 2023) has potential. Due to the generality of the statements made in this paper, a common but unwieldy cumulant condition is required for non-Gaussian processes. However, by exploiting the structure of state space processes, it should be possible to provide alternative proofs that do not rely on this assumption.

Finally, we present ideas for alternative graphical models that could further help to understand, visualise, and communicate dependency structures in stochastic processes.

(8) In our graphical models, we restrict ourselves to linear relationships between components of multivariate processes. Relationships between components that are non-linear in nature cannot be fully captured. An obvious approach would therefore be to define a mixed graph using the conditional expected values given sigma fields instead of orthogonal projections onto linear spaces. The properties of such models can then be explored further, where the proof of the Markov properties should be transferable from Chapter 7 and Eichler (2012). However, we expect more technical assumptions than in our (local) causality graph.

(9) Another conceivable modification would be to adapt the definition of contemporaneous correlation by adding conditioning on the future of the environment, analogous to Eichler (2012), as commented in Remark 7.9. The Markov properties of the resulting graphical model should again carry over. Of particular interest, however, are the edge characterisations. We wonder whether the characterisations of the undirected edges, e.g., for Ornstein-Uhlenbeck processes, now contain the inverse of the covariance matrix of the driving Lévy process rather than the covariance matrix itself. This outcome would be expected given the results of Eichler (2012) and the edge characterisations in our partial correlation graph.

With this summary of the key findings of this thesis and the initiation of further research, we conclude our exploration.

Bibliography

Aalen, O. (1987). Dynamic modelling and causality. *Scand. Actuar. J.*, *1987*(3–4), 177–190. https://doi.org/10.1080/03461238.1987.10413826

Abdelwahab, A., Amor, O., & Abdelwahed, T. (2008). The analysis of the interdependence structure in international financial markets by graphical models. *Int. J. Finance Econ.*, *15*, 291–306.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: John Wiley & Sons.

Andresen, A., Benth, F. E., Koekebakker, S., & Zakamulin, V. (2014). The CARMA interest rate model. *IJTAF*, *17*(02), 1–27. https://doi.org/10.1142/S0219024914500083

Applebaum, D. (2011). *Lévy processes and stochastic calculus* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809781

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *J. R. Stat. Soc. (Supplement)*, *8*(1), 27–41. https://doi.org/10.2307/2983611

Basse-O'Connor, A., Nielsen, M. S., Pedersen, J., & Rohde, V. (2019). Multivariate stochastic delay differential equations and CAR representations of CARMA processes. *Stoch. Process. their Appl.*, *129*(10), 4119–4143. https://doi.org/10.1016/j.spa.2018.11.011

Benth, F. E., Klüppelberg, C., Müller, G., & Vos, L. (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Econ.*, *44*, 392–406. https://doi.org/10.1016/j.eneco.2014.03.020

Benth, F. E., & Saltyte Benth, J. (2009). Dynamic pricing of wind futures. *Energy Econ.*, *31*(1), 16–24. https://doi.org/10.1016/j.eneco.2008.09.009

Bergmann, T., & Hartwigsen, G. (2021). Inferring causality from noninvasive brain stimulation in cognitive neuroscience. *J. Cogn. Neurosci.*, *33*(2), 195–224. https://doi.org/10.1162/jocn_a_01591

Bergstrom, A. R. (1997). Gaussian estimation of mixed-order continuous-time dynamic models with unobservable stochastic trends from mixed stock and flow data. *Econom. Theory*, *13*(4), 467–505. https://doi.org/10.1017/S0266466600005971

Bernstein, D. S. (2009). *Matrix mathematics: Theory, facts, and formulas* (2nd ed.). Princeton: Princeton University Press. https://doi.org/10.1515/9781400833344

Bhatia, R. (1997). *Matrix analysis*. New York: Springer. https://doi.org/10.1007/978-1-4612-0653-8

Brillinger, D. R. (1996). Remarks concerning graphical models for time series and point processes. *Braz. Rev. Econom.*, *16*(1), 1–23. https://doi.org/10.12660/bre.v16n11996.2878

Brillinger, D. R. (2001). *Time series: Data analysis and theory*. Philadelphia: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898719246

Brockwell, P. (2014). Recent results in the theory and applications of CARMA processes. *Ann. Inst. Stat. Math.*, *66*, 647–685. https://doi.org/10.1007/s10463-014-0468-7

Brockwell, P., & Davis, R. (1991). *Time series: Theory and methods* (2nd ed.). New York: Springer. https://doi.org/10.1007/978-1-4419-0320-4

Brockwell, P., & Lindner, A. (2015). Prediction of Lévy-driven CARMA processes. *J. Econom.*, *189*(2), 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. *J. Multivar. Anal.*, *115*, 217–251. https://doi.org/10.1016/j.jmva.2012.09.004

Bühler, T., & Salamon, D. A. (2018). *Functional analysis.* Providence: American Mathematical Society. https://doi.org/10.1090/gsm/191

Chamberlain, G. (1982). The general equivalence of Granger and Sims causality. *Econometrica*, *50*(3), 569–581. https://doi.org/10.2307/1912601

Commenges, D., & Gégout-Petit, A. (2009). A general dynamical statistical model with causal interpretation. *J. R. Stat. Soc. Ser. B.*, *71*(3), 719–736. https://doi.org/10.1111/j.1467-9868.2009.00703.x

Comte, F., & Renault, E. (1996). Noncausality in continuous time models. *Econom. Theory*, *12*(2), 215–256. https://doi.org/10.1017/S0266466600006575

Cont, R., & Tankov, P. (2003). *Financial modelling with jump processes.* New York: Chapman and Hall. https://doi.org/10.1201/9780203485217

Courgeau, V., & Veraart, A. E. D. (2022). Likelihood theory for the graph Ornstein-Uhlenbeck process. *Stat. Inference Stoch. Process.*, *25*(2), 227–260. https://doi.org/10.1007/s11203-021-09257-1

Cox, L., & Popken, D. (2015). Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? *Ann. Epidemiol.*, *25*(3), 162–73. https://doi.org/10.1016/j.annepidem.2014.11.006

Cramér, H. (1940). On the theory of stationary random processes. *Ann. Math.*, *41*(1), 215–230. https://doi.org/10.2307/1968827

Cramér, H. (1961). On the structure of purely non-deterministic stochastic processes. *Ark. Mat.*, *4*, 249–266. https://doi.org/10.1007/BF02592012

Cramér, H. (1971). *Structural and statistical problems for a class of stochastic processes: The first Samuel Stanley Wilks lecture at Princeton University, March 7, 1970.* Princeton University Press.

Cramér, H., & Leadbetter, M. R. (1967). *Stationary and related stochastic processes: Sample function properties and their applications.* New York: Wiley.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, *51*, 157–172. https://doi.org/10.1007/s001840000055

Dahlhaus, R., & Eichler, M. (2003). Causality and graphical models in time series analysis. In P. J. Green, N. L. Hjort, & S. Richardson (Eds.), *Highly structured stochastic systems* (pp. 115–144). Oxford: Oxford University Press.

Dahlhaus, R., Eichler, M., & Sandkühler, J. (1997). Identification of synaptic connections in neural ensembles by graphical models. *J. Neurosci. Methods*, *77*(1), 93–107. https://doi.org/10.1016/S0165-0270(97)00100-3

Deck, T. (2006). *Der Itô-Kalkül: Einführung und Anwendungen.* Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-29265-9

Didelez, V. (2006). Asymmetric separation for local independence graphs. In R. Dechter & T. Richardson (Eds.), *Proceedings of the twenty-second conference on uncertainty in artificial intelligence* (pp. 130–137). Arlington: AUAI Press.

Didelez, V. (2007). Graphical models for composable finite Markov processes. *Scand. J. Stat.*, *34*(1), 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. *J. R. Stat. Soc. Ser. B.*, *70*(1), 245–264. https://doi.org/10.1111/j.1467-9868.2007.00634.x

Doob, J. L. (1944). The elementary Gaussian processes. *Ann. Math. Stat.*, *15*(3), 229–282. https://doi.org/10.1214/aoms/1177731234

Doob, J. L. (1953). *Stochastic processes* (3rd ed.). New York: Wiley.

Dufour, J.-M., & Renault, E. (1998). Short run and long run causality in time series: Theory. *Econometrica*, *66*(5), 1099–1126. https://doi.org/10.2307/2999631

Edwards, D. (2000). *Introduction to graphical modelling* (2nd ed.). New York: Springer. https://doi.org/10.1007/978-1-4612-0493-0

Eichler, M. (2001). *Markov properties for graphical time series models.* (Tech. Rep.). Heidelberg: University of Heidelberg.

Eichler, M. (2006). Graphical modeling of dynamic relationships in multivariate time series. In B. Schelter, M. Winterhalder, & J. Timmer (Eds.), *Handbook of time series analysis: Recent theoretical developments and applications* (pp. 335–372). Berlin: Wiley. https://doi.org/10.1002/9783527609970.ch14

Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econom.*, *137*(2), 334–353. https://doi.org/10.1016/j.jeconom.2005.06.032

Eichler, M. (2009). *Granger-causality graphs for multivariate time series.* (Tech. Rep.). Heidelberg: University of Heidelberg.

Eichler, M. (2011). *A note on global Markov properties for mixed graphs.* https://doi.org/10.48550/arXiv.1107.3036

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

Eichler, M. (2013). Causal inference with multiple time series: Principles and problems. *Philos. Trans. Royal Soc. A*, *371*(1997). https://doi.org/10.1098/rsta.2011.0613

Eichler, M., Dahlhaus, R., & Dueck, J. (2017). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *J. Time Series Anal.*, *38*(2), 225–242. https://doi.org/10.1111/jtsa.12213

Eichler, M., Dahlhaus, R., & Sandkühler, J. (2003). Partial correlation analysis for identification of synaptic connections. *Biol. Cybern.*, *89*, 289–302. https://doi.org/10.1007/s00422-003-0400-3

Fasen, V., & Fuchs, F. (2013a). On the limit behavior of the periodogram of high-frequency sampled stable CARMA processes. *Stoch. Process. their Appl.*, *123*(1), 229–273. https://doi.org/10.1016/j.spa.2012.08.003

Fasen, V., & Fuchs, F. (2013b). Spectral estimates for high-frequency sampled continuous-time autoregressive moving average processes. *J. Time Ser. Anal.*, *34*(5), 532–551. `https://doi.org/10.1111/jtsa.12029`

Fasen-Hartmann, V., & Mayer, C. (2022). Whittle estimation for continuous-time stationary state space models with finite second moments. *Ann. Inst. Stat. Math.*, *74*(2), 233–270. `https://doi.org/10.1007/s10463-021-00802-6`

Fasen-Hartmann, V., & Schenk, L. (2023a). *Mixed causality graphs for continuous-time state space models and orthogonal projections.* `https://doi.org/10.48550/arXiv.2311.04478`

Fasen-Hartmann, V., & Schenk, L. (2023b). *Mixed causality graphs for continuous-time stationary processes.* `https://doi.org/10.48550/arXiv.2308.08890`

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

Feshchenko, I. (2012). On closeness of the sum of n subspaces of a Hilbert space. *Ukr. Math. J.*, *63*, 1566–1622. `https://doi.org/10.1007/s11253-012-0601-9`

Florens, J.-P., & Fougère, D. (1996). Noncausality in continuous time. *Econometrica*, *64*(5), 1195–1212. `https://doi.org/10.2307/2171962`

Florens, J.-P., & Mouchart, M. (1982). A note on noncausality. *Econometrica*, *50*(3), 583–591. `https://doi.org/10.2307/1912602`

Florens, J.-P., & Mouchart, M. (1985). A linear theory for noncausality. *Econometrica*, *53*(1), 157–175. `https://doi.org/10.2307/1911729`

Fujikoshi, Y., Ulyanov, V. V., & Shimizu, R. (2010). *Multivariate statistics: High-dimensional and large-sample approximations.* New York: John Wiley & Sons. `https://doi.org/10.1002/9780470539873`

Gardner, W. A. (1988). *Statistical spectral analysis: A nonprobabilistic theory.* New Jersey: Prentice-Hall.

Gardner, W. A. (1992). A unifying view of coherence in signal processing. *Signal Process.*, *29*(2), 113–140. `https://doi.org/10.1016/0165-1684(92)90015-O`

Gather, U., Imhoff, M., & Fried, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Stat. Med.*, *21*(18), 2685–2701. `https://doi.org/10.1002/sim.1209`

Gibbs, J. W. (1902). *Elementary principles in statistical mechanics: Developed with especial reference to the rational foundation of thermodynamics.* Cambridge: Cambridge University Press. `https://doi.org/10.1017/CBO9780511686948`

Gladyshev, E. G. (1958). On multi-dimensional stationary random processes. *Theory Probab. Appl.*, *3*(4), 425–428. `https://doi.org/10.1137/1103035`

Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution. *Ann. Math. Stat.*, *34*(1), 152–177. `https://doi.org/10.1214/aoms/1177704250`

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438. `https://doi.org/10.2307/1912791`

Halmos, P. R. (1957). *Introduction to Hilbert space and the theory of spectral multiplicity* (2nd ed.). New York: Chelsea Publishing.

Hannan, E. J., & Deistler, M. (2012). *The statistical theory of linear systems.* Philadelphia: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611972191

Harvey, A. C., & Stock, J. H. (1985a). Continuous time autoregressive models with common stochastic trends. *J. Econ. Dyn. Control*, *12*(2–3), 365–384. https://doi.org/10.1016/0165-1889(88)90046-2

Harvey, A. C., & Stock, J. H. (1985b). The estimation of higher-order continuous time autoregressive models. *Econom. Theory*, *1*(1), 97–112. https://doi.org/10.1017/S0266466600011026

Harvey, A. C., & Stock, J. H. (1989). Estimating integrated higher-order continuous time autoregressions with an application to money-income causality. *J. Econom.*, *42*(3), 319–336. https://doi.org/10.1016/0304-4076(89)90056-0

Heerah, S., Molinari, R., Guerrier, S., & Marshall-Colon, A. (2021). Granger-causal testing for irregularly sampled time series with application to nitrogen signalling in Arabidopsis. *Bioinformatics*, *37*(16), 2450–2460. https://doi.org/10.1093/bioinformatics/btab126

Hongke, D. C. D. (2008). A new characterization of the closedness of the sum of two subspaces. *Acta Math. Sci.*, *28*(1), 17–23. https://doi.org/10.1016/S0252-9602(08)60003-X

Horn, R. A., & Johnson, C. R. (2013). *Matrix analysis* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139020411

Höfer, T., Przyrembel, H., & Verleger, S. (2004). New evidence for the theory of the stork. *Paediatr. Perinat. Epidemiol.*, *18*(1), 88–92. https://doi.org/10.1111/j.1365-3016.2003.00534.x

Imbens, G. W. (2022). Causality in econometrics: Choice vs. chance. *Econometrica*, *90*(6), 2541–2566. https://doi.org/10.3982/ECTA21204

Jung, C. (1969). *Synchronicity: An acausal connecting principle.* Princeton: Princeton University Press.

Kailath, T. (1980). *Linear systems.* Englewood Cliffs: Prentice-Hall.

Kartsioukas, R., Stoev, S., & Hsing, T. (2023). *Spectral density estimation of function-valued spatial processes.* https://doi.org/10.48550/arXiv.2302.02247

Khintchine, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.*, *109*, 604–615. https://doi.org/10.1007/BF01449156

Klenke, A. (2020). *Wahrscheinlichkeitstheorie* (4th ed.). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-62089-2

Koopmans, L. H. (1964a). On the coefficient of coherence for weakly stationary stochastic processes. *Ann. Math. Stat.*, *35*(2), 532–549. https://doi.org/10.1214/aoms/1177703553

Koopmans, L. H. (1964b). On the multivariate analysis of weakly stationary stochastic processes. *Ann. Math. Stat.*, *35*(4), 1765–1780. https://doi.org/10.1214/aoms/1177700398

Koster, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Stat.*, *26*(3), 413–431. https://doi.org/10.1111/1467-9469.00157

Kuersteiner, G. M. (2010). Granger-Sims causality. In S. N. Durlauf & L. E. Blume (Eds.), *Macroeconometrics and time series analysis* (pp. 5413–5425). Palgrave Macmillan. https://doi.org/10.1057/9780230280830_14

Kuzma, R., Cruickshank, I. J., & Carley, K. M. (2022). Influencing the influencers: Evaluating person-to-person influence on social networks using Granger causality. In R. M. Benito, C. Cherifi, H. Cherifi, E. Moro, L. M. Rocha, & M. Sales-Pardo (Eds.), *Complex networks & their applications X* (pp. 89–99). Cham: Springer. https://doi.org/10.1007/978-3-030-93413-2_8

Lauritzen, S. L. (2004). *Graphical models* (2nd ed.). Oxford: Clarendon Press.

Lax, P. (2002). *Functional analysis.* New York: Wiley.

Lemmens, J., Valkenburg, P., & Peter, J. (2011). The effects of pathological gaming on aggressive behavior. *J. Youth Adolescence*, *40*, 38–47. https://doi.org/10.1007/s10964-010-9558-x

Levitz, M., Perlman, M. D., & Madigan, D. (2001). Separation and completeness properties for AMP chain graph Markov models. *Ann. Statist.*, *29*(6), 1751–1784. https://doi.org/10.1214/aos/1015345961

Lindquist, A., & Picci, G. (2015). *Linear stochastic systems: A geometric approach to modeling, estimation and identification.* Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-45750-4

Lu, T.-T., & Shiou, S.-H. (2002). Inverses of 2×2 block matrices. *Comput. Math. with Appl.*, *43*(1–2), 119–129. https://doi.org/10.1016/S0898-1221(01)00278-4

Maathuis, M., Drton, M., Lauritzen, S., & Wainwright, M. (2019). *Handbook of graphical models.* Boca Raton: CRC Press. https://doi.org/10.1201/9780429463976

Mahnke, R., Kaupužs, J., & Lubaševskij, I. (2009). *Physics of stochastic processes: How randomness acts in time.* Weinheim: Wiley. https://doi.org/10.1002/9783527626090

Marquardt, T., & Stelzer, R. (2007). Multivariate CARMA processes. *Stoch. Process. their Appl.*, *117*(1), 96–120. https://doi.org/10.1016/j.spa.2006.05.014

Matveev, R. F. (1961). On multi-dimensional regular stationary processes. *Theory Probab. Appl.*, *6*(2), 149–165. https://doi.org/10.1137/1106020

Medkour, T., Walden, A., & Burgess, A. (2009). Graphical modelling for brain connectivity via partial coherence. *J. Neurosci. Methods*, *180*(2), 374–383. https://doi.org/10.1016/j.jneumeth.2009.04.003

Mogensen, S. W., & Hansen, N. R. (2020). Markov equivalence of marginalized local independence graphs. *Ann. Math. Stat.*, *48*(1), 539–559. https://doi.org/10.1214/19-AOS1821

Mogensen, S. W., & Hansen, N. R. (2022). Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli*, *28*(4), 3023–3050. https://doi.org/10.3150/21-bej1446

Pearl, J. (1994). *Probabilistic reasoning in intelligent systems: Networks of plausible inference* (3rd ed.). San Francisco: Morgan Kaufmann. https://doi.org/10.1016/C2009-0-27609-4

Petrovic, L., & Dimitrijevic, S. (2012). Causality with finite horizon of the past in continuous time. *Stat. Probab. Lett.*, *82*(7), 1219–1223. https://doi.org/10.1016/j.spl.2012.03.032

Phillips, A. W. (1959). The estimation of parameters in systems of stochastic differential equations. *Biometrika*, *46*(1/2), 67–76. https://doi.org/10.2307/2332809

Pinsky, M. A. (2009). *Introduction to Fourier analysis and wavelets.* Providence: American Mathematical Society.

Priestley, M. B. (1981). *Spectral analysis and time series, Volume II.* London: Academic Press.

Protter, P. E. (2005). *Stochastic integration and differential equations* (2nd ed.). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-10061-5

Renault, E., & Szafarz, A. (1991). *True versus spurious instantaneous causality.* (Tech. Rep.). Bruxelles: Universite Libre de Bruxelles.

Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.*, *30*(1), 145–157. https://doi.org/10.1111/1467-9469.00323

Robertson, J. B. (1968). Orthogonal decompositions of multivariate weakly stationary stochastic processes. *Can. J. Math.*, *20*, 368–383. https://doi.org/10.4153/CJM-1968-033-x

Rozanov, J. A. (1967). *Stationary random processes.* San Francisco: Holden-Day.

Rugh, W. J. (1996). *Linear system theory* (2nd ed.). Upper Saddle River: Prentice Hall.

Sato, K. (2007). *Lévy processes and infinitely divisible distributions* (2nd ed.). Cambridge: Cambridge University Press.

Sato, K., & Yamazato, M. (1983). Stationary processes of Ornstein-Uhlenbeck type. In K. Itô & J. Prokhorov (Eds.), *Probability theory and mathematical statistics* (pp. 541–551). Berlin: Springer. https://doi.org/10.1007/BFb0072949

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*(1), 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. *Electron. J. Stat.*, *6*, 2185–2234. https://doi.org/10.1214/12-EJS743

Schochetman, I., Smith, R., & Tsui, S.-K. (2001). On the closure of the sum of closed subspaces. *Int. J. Math. Math. Sci.*, *26*, 257–267. https://doi.org/10.1155/S0161171201005324

Schweder, T. (1970). Composable Markov processes. *J. Appl. Probability*, *7*(2), 400–410. https://doi.org/10.2307/3211973

Shojaie, A., & Fox, E. B. (2022). Granger causality: A review and recent advances. *Annu. Rev. Stat. Appl.*, *9*(1), 289–319. https://doi.org/10.1146/annurev-statistics-040120-010930

Sims, C. A. (1972). Money, income, and causality. *Am. Econ. Rev.*, *62*(4), 540–552.

Sinoquet, C., & Mourad, R. (2014). *Probabilistic graphical models for genetics, genomics, and postgenomics.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709022.001.0001

Spirtes, P., Richardson, T., Meek, C., Scheines, R., & Glymour, C. (1998). Using path diagrams as a structural equation modeling tool. *Sociol. Methods Res.*, *27*(2), 182–225. https://doi.org/10.1177/0049124198027002003

Stelzer, R. (2011). *CARMA processes driven by non-Gaussian noise.* https://doi.org/10.48550/arXiv.1201.0155

Sucar, L. E. (2020). *Probabilistic graphical models: Principles and applications.* Cham: Springer. https://doi.org/10.1007/978-3-030-61943-5

Tick, L. J. (1963). Conditional spectra, linear systems, and coherency. In *Proceedings of the symposium on time series analysis (Brown University, 1962)* (pp. 197–203). New York: Wiley.

Weidmann, J. (1980). *Linear operators in Hilbert spaces.* New York: Springer. https://doi.org/10.1007/978-1-4612-6027-1

Whittaker, J. (2008). *Graphical models in applied multivariate statistics* (2nd ed.). Chichester: Wiley.

Wright, S. (1921). Correlation and causation. *J. Agric. Res.*, *20*(3), 557–585.

Wright, S. (1934). The method of path coefficients. *Ann. Math. Statist.*, *5*(3), 161–215. https://doi.org/10.1214/aoms/1177732676

Yaglom, A. M. (1987). *Correlation theory of stationary and related random functions.* New York: Springer.

Zadnik, K., Jones, L., Irvin, B., Kleinstein, R. N., Manny, R. E., Shin, J. A., & Mutti, D. O. (2000). Myopia and ambient night-time lighting. *Nature*, *404*, 143–144. https://doi.org/10.1038/35004661

APPENDIX A

Auxiliary technical results

In this chapter, we present two results that have been moved from the main body of the thesis to improve readability. The corresponding proofs are simple mathematical inductions.

A.1. Discrete time difference operator

Lemma A.1. *Let $f : \mathbb{R} \to \mathbb{R}^k$ and define the discrete-time difference operator $D^{(j)}$ iteratively by*

$$\mathcal{D}^{(1)} f(t) = f(t) - f(t-1), \quad \mathcal{D}^{(j)} f(t) = \mathcal{D}^{(j-1)} \left(f(t) - f(t-1) \right), \quad j \in \mathbb{N},\ j \geq 2,$$

for $t \in \mathbb{R}$, where we set $\mathcal{D}^{(0)} f(t) = f(t)$. Further, define $\binom{0}{0} := 1$. Then

$$f(t-n) = \sum_{j=1}^{n+1} \binom{n}{j-1} (-1)^{j-1} \mathcal{D}^{(j-1)} f(t), \quad n \in \mathbb{N}_0,\ t \in \mathbb{R}.$$

Proof. Let $t \in \mathbb{R}$. We show the statement by induction over n. The *base case* $n = 0$ holds immediately because of

$$f(t) = \sum_{j=1}^{1} \binom{0}{j-1} (-1)^{j-1} \mathcal{D}^{(j-1)} f(t).$$

Now assume the *induction hypothesis*

$$f(t-(n-1)) = \sum_{j=1}^{n} \binom{n-1}{j-1} (-1)^{j-1} \mathcal{D}^{(j-1)} f(t) \tag{A.1}$$

for some $n \in \mathbb{N}_0$. For the *induction step*, the operator definition and (A.1) yield

$$\begin{aligned} f(t-n) &= f(t-(n-1)) - \mathcal{D}^{(1)} f(t-(n-1)) \\ &= \sum_{j=1}^{n} \binom{n-1}{j-1} (-1)^{j-1} \mathcal{D}^{(j-1)} f(t) - \sum_{j=1}^{n} \binom{n-1}{j-1} (-1)^{j-1} \mathcal{D}^{(j)} f(t). \end{aligned}$$

An index shift and Pascal's rule for binomial coefficients (e.g., Bernstein, 2009, Fact 1.7.1) give

$$\begin{aligned} &f(t-n) \\ &= f(t) + \sum_{j=2}^{n} \binom{n-1}{j-1} (-1)^{j-1} \mathcal{D}^{(j-1)} f(t) - \sum_{j=1}^{n-1} \binom{n-1}{j-1} (-1)^{j-1} \mathcal{D}^{(j)} f(t) - (-1)^{n-1} \mathcal{D}^{(n)} f(t) \end{aligned}$$

$$= f(t) + \sum_{j=2}^{n} \left(\binom{n-1}{j-1} + \binom{n-1}{j-2} \right) (-1)^{j-1} \mathcal{D}^{(j-1)} f(t) + (-1)^n \mathcal{D}^{(n)} f(t)$$
$$= f(t) + \sum_{j=2}^{n} \binom{n}{j-1} (-1)^{j-1} \mathcal{D}^{(j-1)} f(t) + (-1)^n \mathcal{D}^{(n)} f(t)$$
$$= \sum_{j=1}^{n+1} \binom{n}{j-1} (-1)^{j-1} \mathcal{D}^{(j-1)} f(t). \qquad \blacksquare$$

A.2. Properties of matrices defining MCAR processes

Lemma A.2. *Define*

$$\mathbf{A} = \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -A_p & -A_{p-1} & \cdots & \cdots & -A_1 \end{pmatrix} \in M_{kp}(\mathbb{R}), \quad \mathbf{B} = \begin{pmatrix} 0_k \\ \vdots \\ 0_k \\ I_k \end{pmatrix} \in M_{kp\times k}(\mathbb{R}),$$
$$\mathbf{C} = \begin{pmatrix} I_k & 0_k & \cdots & 0_k \end{pmatrix} \in M_{k\times kp}(\mathbb{R}),$$

where $k, p \in \mathbb{N}$. Then, for $j = 0, \ldots, p-1$, it applies that

$$\mathbf{A}^j\mathbf{B} = \begin{pmatrix} 0_{k(p-j-1)\times k} \\ \mathbf{C}\mathbf{A}^{p-1}\mathbf{B} \\ \vdots \\ \mathbf{C}\mathbf{A}^{p+j-1}\mathbf{B} \end{pmatrix} \quad \textit{and} \quad \mathbf{C}\mathbf{A}^{p+j}\mathbf{B} = -\sum_{i=0}^{j} A_{j-i+1}\mathbf{C}\mathbf{A}^{p+i-1}\mathbf{B}.$$

Proof. We first derive the representation of $\mathbf{A}^j\mathbf{B}$ by induction over j. The *base case* $j = 0$ holds immediately due to the relation (8.13). We obtain

$$\mathbf{A}^0\mathbf{B} = \mathbf{B} = \begin{pmatrix} 0_{k(p-1)\times k} \\ I_k \end{pmatrix} = \begin{pmatrix} 0_{k(p-1)\times k} \\ \mathbf{C}\mathbf{A}^{p-1}\mathbf{B} \end{pmatrix}.$$

Let us now assume the *induction hypothesis*

$$\mathbf{A}^j\mathbf{B} = \begin{pmatrix} 0_{k(p-j-1)\times k} \\ \mathbf{C}\mathbf{A}^{p-1}\mathbf{B} \\ \vdots \\ \mathbf{C}\mathbf{A}^{p+j-1}\mathbf{B} \end{pmatrix} \tag{A.2}$$

for some $j \in \mathbb{N}_0$ with $0 \le j < p-1$. Furthermore, note that, for $0 \le j < p-1$, the induction hypothesis (A.2) and the relation (8.13) yield

$$\mathbf{CA}^{p+j}\mathbf{B} = \mathbf{CA}^{p}\mathbf{A}^{j}\mathbf{B} = \begin{pmatrix} -A_p & \cdots & -A_1 \end{pmatrix} \begin{pmatrix} 0_{k(p-j-1)\times k} \\ \mathbf{CA}^{p-1}\mathbf{B} \\ \vdots \\ \mathbf{CA}^{p+j-1}\mathbf{B} \end{pmatrix} = -\sum_{i=0}^{j} A_{j-i+1}\mathbf{CA}^{p+i-1}\mathbf{B}. \quad \text{(A.3)}$$

Now, because of the induction hypothesis (A.2) and equation (A.3), it follows the *induction step*

$$\begin{aligned}\mathbf{A}^{j+1}\mathbf{B} = \mathbf{A}\mathbf{A}^{j}\mathbf{B} &= \begin{pmatrix} 0_k & I_k & 0_k & \cdots & 0_k \\ 0_k & 0_k & I_k & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_k \\ 0_k & \cdots & \cdots & 0_k & I_k \\ -A_p & -A_{p-1} & \cdots & \cdots & -A_1 \end{pmatrix} \begin{pmatrix} 0_{k(p-j-1)\times k} \\ \mathbf{CA}^{p-1}\mathbf{B} \\ \vdots \\ \mathbf{CA}^{p+j-1}\mathbf{B} \end{pmatrix} \\ &= \begin{pmatrix} 0_{k(p-j-2)\times k} \\ \mathbf{CA}^{p-1}\mathbf{B} \\ \vdots \\ \mathbf{CA}^{p+j-1}\mathbf{B} \\ -\sum_{i=0}^{j} A_{j-i+1}\mathbf{CA}^{p+i-1}\mathbf{B} \end{pmatrix} = \begin{pmatrix} 0_{k(p-j-2)\times k} \\ \mathbf{CA}^{p-1}\mathbf{B} \\ \vdots \\ \mathbf{CA}^{p+j-1}\mathbf{B} \\ \mathbf{CA}^{p+j}\mathbf{B} \end{pmatrix}.\end{aligned}$$

Finally, the representation of $\mathbf{CA}^{p+j}\mathbf{B}$ for all $j = 0, \ldots, p-1$ follows analogously to the calculation (A.3) from the previous induction result. ■