Relationship between Granger non-causality and network graph of state-space representations

Jozsa, Monika

Publication date:
2019

Relationship between Granger non-causality and network graphs of state-space representations

Mónika Józsa

This research has been carried out at Johann Bernoulli Institute for Mathematics and Computer Science, Faculty of Mathematics and Natural Sciences, University of Groningen, The Netherlands, and at the Research Unit Informatics and Control Systems (URIA) of IMT-Lille-Douai, France.

The work reported in this dissertation is part of the research program of the Dutch Institute of Systems and Control (DISC). The author has successfully completed the graduate program of DISC.

Printed by: Ridderprint BV, www.ridderprint.nl

ISBN (printed version): 978–94–034–1296–2
ISBN (electronic version): 978–94–034–1295–5

Relationship between Granger non-causality and network graphs of state-space representations

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

25 February 2019 at 9:00 hours

by

Mónika Józsa

born on 4 July 1989
Budapest, Hungary

Supervisors:	Prof. Dr. M. Kanat Camlibel Dr. Mihály Petreczky
Assessment committee:	Prof. Dr. Arjan van der Schaft Prof. Dr. Peter E. Caines Prof. Dr. Ralf L. M. Peeters

Contents

List of Figures

List of Tables

List of Algorithms

List of Acronyms

EEG	Electroencephalogram
fMRI	Functional Magnetic Resonance Imaging
GB–SS	General Bilinear State-Space
LQG	Least Quadratic Gaussian
LTI–SS	Linear Time-Invariant State-Space
MA	Moving-Average
MEG	Magnetoencephalography
STD	Standard Deviation
SVD	Singular Value Decomposition
TADG	Transitive Acyclic Directed Graph
VAR	Vector Autoregressive
VARMA	Vector Autoregressive Moving-Average
ZMSIR	Zero-mean square-integrable with rational spectrum
ZMWSSI	Zero-mean weakly stationary with respect to input

Introduction

Detecting interactions among stochastic processes can be of interest for several applications such as mapping interactions in the brain, predicting economical price movements or understanding social group behaviour. The first step towards these applications is to formulate an appropriate definition of interaction. In this thesis we consider two approaches for defining interaction. The first considers the stochastic processes as outputs of dynamical stochastic systems and concentrates on the information flow between systems that generate the interacting processes. The second approach focuses on statistical properties of the interacting processes. In all cases, the interactions will be one-directional, i.e., where one process influences another but the other does not influence the first. These are particularly interesting as they can be measured more easily compared to bi-directional interactions.

Consider a multidimensional stochastic process that is partitioned into components that interact with each other. Then the first approach is based on the existence of a dynamical system having this process as its output process, such that it is decomposed into subsystems that communicate according to a so-called network graph: Informally, by the network graph of a system we mean a directed graph, whose nodes correspond to subsystems, such that each subsystem generates a component of the output process. There are as many subsystems as there are components of the output process. Regarding the edges of the network graph, there is an edge from one node to another, if the subsystem which corresponds to the source node sends information to the subsystem which corresponds to the target node. By interaction between two subsystems we mean that some processes of one subsystem serve as an input to another subsystem. In case of state-space systems it will mean that there is an edge from one node to another, if the state and noise process of the subsystem corresponding to the source node serve as an input to the subsystem which corresponds to the target node. In case of transfer matrices, we will designate an edge, if the noise process of the subsystem of the source node serves as an input to the subsystem of the target node. As an example, let $\mathbf{y}$ be the output process of a stochastic dynamical system that is partitioned into two components such as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$. Then the network graph of a dynamical system which generates $\mathbf{y}$ has two nodes, the first one corresponds to the subsystem which generates $\mathbf{y}_1$ and the second one corresponds to the subsystem which generates $\mathbf{y}_2$. The edges of the network graph are determined by the information flow between the subsystems: there is a directed edge from one subsystem to another if the first one sends information to the other. If there is a dynamical system generating $\mathbf{y}$, such that in its network graph there is an edge from the first node (corresponding to the subsystem of $\mathbf{y}_1$) to the second node (corresponding to the subsystem of $\mathbf{y}_2$), then we say that $\mathbf{y}_1$ influences $\mathbf{y}_2$. This idea is generalized for processes with several components.

The advantage of the this approach is that it offers an intuitive mechanistic explanation of how one component of the output process influences the other one. The apparent disadvantage, however, is that the same output process can be generated by systems with different network graphs. As a result, the presence or absence of an interaction between two output components depends on the exact dynamical system that we choose to represent the output process.

Again, consider a multidimensional stochastic process that is an output process of a dynamical system and is partitioned into components. Then the second approach for defining interaction among the components of the process is based on statistical properties of the process. A widely used example of this approach is called Granger causality, a purely statistical concept for defining linear causal relationships between processes: Denote the process in question by $\mathbf{y}$ and let it be partitioned into two components such as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$. Then intuitively, we can say that $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$ if the best linear prediction of $\mathbf{y}_2$ based on the past values of $\mathbf{y}$ is better than the one based on the past values of $\mathbf{y}_2$. For a process $\mathbf{y}$ that is the output process of a linear dynamical system, we define directed interactions in $\mathbf{y}$ by the Granger causalities between $\mathbf{y}_1$ and $\mathbf{y}_2$. This is further generalized for processes with several components by using an extended version of Granger causality, called conditional Granger causality. As an extension of Granger causality, we will also define interaction in a process that is the output process of a so-called bilinear dynamical system.

As the second approach is based on statistical properties of processes, contrary to the first approach, it leads to definitions that do not depend on which dynamical system we use to represent $\mathbf{y}$. However, it does not always offer an explanation of the mechanisms according to which the interaction takes place.

In summary, the first approach focuses on the mechanism inside a dynamical system but is too sensitive to the choice of the system itself. The second approach, solving the issue with the first, is independent from which dynamical system we choose to represent the output process, however, it does not capture the inner mechanism of the interaction in general. It is thus of interest, in order to benefit from the advantages of both approaches, to relate and find equivalence between the network graph of a dynamical system and statistical properties of its output process.

The following dynamical systems are subjects of this thesis: autonomous linear time-invariant state-space (LTI–SS) representations, transfer matrices of autonomous LTI systems and general bilinear state-space (GB–SS) representations. For these system we derive results on the relation between network graphs and causal properties among the components of the output process. For this, we use the so-called innovation form of the representations, where we choose the noise process to be the so-called innovation process. It is then shown that the existence of a dynamical system in innovation form with a certain network graph is equivalent with causal relations among the components of the observed process. The results include algorithms for the construction of the dynamical systems in question.

Contribution

For simplicity, unless stated otherwise, the terms LTI–SS representation, transfer matrix and GB–SS representation mean autonomous stochastic LTI–SS representation, transfer matrix of stochastic autonomous LTI system and stochastic GB–SS representation. To LTI–SS or GB–SS representations and to transfer matrices, we associate a so-called network graph. The network graph is defined based on the view of these systems as an interconnection of several subsystems. In this thesis we will restrict attention to network graphs which belong to one of the following classes of graphs: two node graphs with one edge, star graphs, and transitive acyclic graphs. The objective of this thesis is to associate the existence of systems having one of these network graphs to properties of the observed processes. The properties of the observed processes that help us to impose conditions and show implications of the existence of these systems are the so-called Granger causality, conditional Granger causality and GB–Granger causality. Below, we discuss network graphs and causality in more detail.

Network graph: Let $\mathbf{y}$ be an output process of an LTI–SS representation or a transfer matrix or a GB–SS representation and denote the system that represents $\mathbf{y}$ by $\mathcal{S}$. Assume that $\mathbf{y}$ is partitioned such that $\mathbf{y} = \left[\mathbf{y}_1^T, \ldots, \mathbf{y}_n^T\right]^T$ and consider subsystems $\mathcal{S}_i$, $i = 1, \ldots, n$ of the system $\mathcal{S}$ such that $\mathcal{S}_i$ generates the component $\mathbf{y}_i$. Then, the network graph has nodes $\{1, \ldots, n\}$ and there is edge from node i to

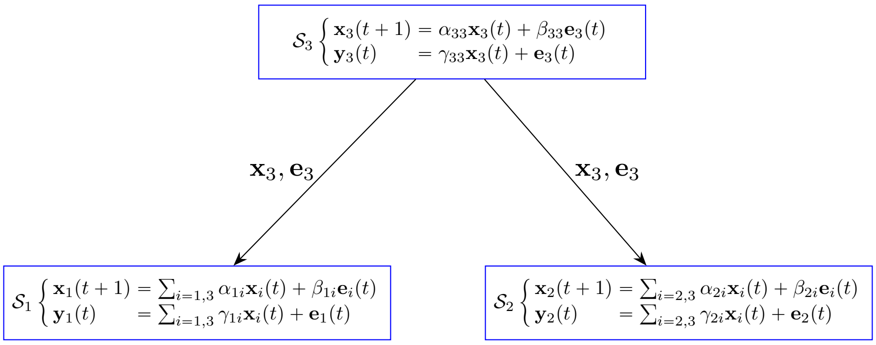

Figure 1: LTI–SS representation of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T]^T$ with the three-node star graph as its network graph: The state and noise process of subsystem $\mathcal{S}_3$ serves as an input to subsystems $\mathcal{S}_1$ and $\mathcal{S}_2$.

node j, if the noise, and in case of state-space representations also the state process, of $\mathcal{S}_i$ serve as an input of $\mathcal{S}_j$. In fact, for the systems at hand, an edge (i, j) in the network graph corresponds to non-zero blocks in certain matrix parameters of the system. In parallel, the lack of this edge corresponds to zero blocks in those matrices. Intuitively, an edge in the network graph means that information can flow from the subsystem corresponding to a source node to the subsystem corresponding to the target node, but there is no information flowing the other way around. Figure 1 illustrates the network graph of an LTI–SS representation having the three-node star graph as its network graph.

Causality: For an output process $\mathbf{y} = \left[\mathbf{y}_1^T, \ldots, \mathbf{y}_n^T\right]^T$ of an LTI–SS representation or a transfer matrix, we will study Granger non-causality and conditional Granger non-causality relations among the components $\mathbf{y}_1, \ldots, \mathbf{y}_n$. Granger non-causality (Granger, 1963) can be explained as follows: $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ if the knowledge of the past values of $\mathbf{y}_1$ and $\mathbf{y}_2$ does not yield a more accurate prediction of the future values of $\mathbf{y}_2$ than the knowledge of the past values of only $\mathbf{y}_2$. Conditional Granger non-causality is a general form of Granger non-causality: Informally, $\mathbf{y}_1$ conditionally does not Granger cause $\mathbf{y}_2$ with respect to $\mathbf{y}_3$ if the knowledge of the past values of $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$ does not yield a more accurate prediction of the future values of $\mathbf{y}_2$ than the knowledge of the past values of only $\mathbf{y}_2$ and $\mathbf{y}_3$.

Among the components of the output process of a GB–SS representation, we will formulate an extended definition of Granger causality, called GB–Granger causality. Let $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T\right]^T$ be the output process of a GB–SS representation and $\mathbf{u}$ be its input process. Then GB–Granger non-causality from $\mathbf{y}_1$ to $\mathbf{y}_2$ with respect to $\mathbf{u}$ has the following intuitive meaning: the knowledge of all the products of the past $\mathbf{y}$ and $\mathbf{u}$ values do not give a more accurate prediction of $\mathbf{y}_2$ than the knowledge of all the products of the past $\mathbf{y}_2$ and $\mathbf{u}$ values. In the trivial case when $\mathbf{u}$ is a constant input, GB–Granger non-causality from $\mathbf{y}_1$ to $\mathbf{y}_2$ is equivalent to Granger non-causality from $\mathbf{y}_1$ to $\mathbf{y}_2$.

Results: The presentation of the main results is organized as follows:

In Chapter 2, we show that a process $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T\right]^T$ admits a specific LTI–SS representation in the so-called block triangular form if and only if $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$. Informally, an LTI–SS representation in block triangular form is a system whose network graph has two nodes, corresponding to two subsystems generating $\mathbf{y}_1$ and $\mathbf{y}_2$, and an edge from the node associated with $\mathbf{y}_2$ to the node associated with $\mathbf{y}_1$. For both coercive and non-coercive processes, we give conditions for the minimality of the representations and present algorithms on the construction of the representations.

In a similar manner, it is shown in Chapter 3 that a process $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T \ldots, \mathbf{y}_n^T\right]^T$ admits a specific LTI–SS representation in the so-called coordinated form if and only if $\mathbf{y}_i$ does not Granger cause $\mathbf{y}_n$ and $\mathbf{y}_i$ conditionally does not Granger cause $\mathbf{y}_j$ with respect to $\mathbf{y}_n$ for all $i \neq j$ $i, j = 1, 2 \ldots, n-1$. Informally, an LTI–SS representation in coordinated form is a system whose network graph is a star graph (a tree of depth one which has one root node and all other nodes of which are leaves) such that the root node of the graph corresponds to the subsystem which generates $\mathbf{y}_n$ and the leaves correspond to subsystems which generate $\mathbf{y}_j$, $j = 1, \ldots, n-1$. Chapter 3, together with Chapter 2, is based on the journal paper (Jozsa et al., 2018b).

In Chapter 4, the existence of an LTI–SS representation with the so-called transitive acyclic directed graph (TADG) zero structure is characterized by the series of conditional Granger non-causality conditions: Let $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T \ldots, \mathbf{y}_n^T\right]^T$ be the output process of an LTI–SS representation that has zero structure with a TADG $G = (V, E)$, where the set of nodes is $V = \{1, 2 \ldots, n\}$. Informally, an LTI–SS representation has a TADG zero structure with the graph G if its network graph is the graph G. Then we associate each component $\mathbf{y}_i$ of $\mathbf{y}$ with a node i of G. The conditional Granger non-causality conditions can be explained as follows: If (j, i) is not an edge of G then $\mathbf{y}_i$ does not conditionally Granger cause $\mathbf{y}_j$ with respect to the collection of the components of $\mathbf{y}$ that correspond to the parent nodes of j. Chapter 4 is partially based on the conference paper (Jozsa et al., 2017a).

As a counterpart of Chapter 4, Chapter 5 studies transfer matrices of LTI–SS representations with TADG network graphs. It deals with transfer matrices with TADG zero structure and shows that transfer matrices with TADG zero structure can be characterized by the same series of conditional Granger non-causality conditions as LTI–SS representations with TADG zero structure in Chapter 4. The results of

Chapter 5 have been reported in the conference paper (Jozsa et al., 2017b).

In Chapter 6 the main result of Chapter 2 are extended to general bilinear state-space (GB–SS) representations and GB–Granger causality. It deals with GB–SS representations where a multiplicative input is present in the system and thus the relations between the processes of the system are nonlinear. It introduces an extended form of Granger causality to a more general statistical property of the input and output processes, called GB–Granger causality. It then shows that a process $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T\right]^T$ together with an input process admit a specific GB–SS representation in block triangular form if and only if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$. The block triangular form of the GB–SS representation defines a network graph that has two nodes, corresponding to two subsystems generating $\mathbf{y}_1$ and $\mathbf{y}_2$, and an edge from the node associated with $\mathbf{y}_2$ to the node associated with $\mathbf{y}_1$. Chapter 6 is based on the submitted journal paper (Jozsa et al., 2018a).

Finally, in Chapter 7, we illustrate how the results and algorithms of Chapters 3–5 can be applied to simulated data. The results show that with the help of the results in Chapters 3–5 we can identify the LTI–SS representations with different network graphs with precision comparable to classical identification methods.

Motivation

The contributions of this thesis are useful for reverse engineering of the network graph of stochastic dynamical systems. In addition, they can be relevant for distributed estimation/control and for structure preserving model reduction.

Reverse engineering of the network graph: By reverse engineering of the network graph we mean finding out the network graph of a system based on the observed output of the system. This problem arises in several domains such as systems biology (Nordling and Jacobsen, 2011; Julius et al., 2009; Kang et al., 2015; Valdes-Sosa et al., 2011; Westra et al., 2007), neuroscience (Roebroeck et al., 2011c), smart grids (Bolognani et al., 2013; Zhang et al., 2017), etc. To solve this problem, we first need to understand when the observed behavior can be realized by a system with a specific network graph.

An emerging application in neuroscience (Roebroeck et al., 2011c; Valdes-Sosa et al., 2011; Friston et al., 2003; Goebel et al., 2003) is to detect and model interactions between brain regions using e.g., fMRI, EEG, MEG data. For this purpose, both Granger causality based methods (Goebel et al., 2003) and state-space based methods (Friston et al., 2003) were used. In the former case, the presence of an interaction was identified with the presence of Granger causality between the outputs associated with various brain regions. In the latter case, the presence of an interaction was interpreted as the presence of an edge in the network graph of a state-space representation, whose parameters were identified from data. However, the formal relationship between these methods was not always clear. This has lead to a lively debate regarding the advantages/disadvantages of both methods (Valdes-Sosa et al., 2011; David, 2011; Roebroeck et al., 2011b). The results of Chapters 2–4 imply that the network graph of a specific LTI–SS representation defines causal relations in the output process and the causal properties of the output process restricts the network graph of a potential LTI–SS representation. In fact, considering conditional Granger non-causality conditions and LTI–SS representations in innovation form having a certain network graph, the two approaches are formally equivalent and produce the same outcome. Therefore, these results serve as an answer to the debate on the relationship between state-space methods and Granger causality in LTI–SS representations. Besides, it provides important knowledge for reverse engineering of the network graphs of LTI–SS representations.

The cited applications (Nordling and Jacobsen, 2011; Julius et al., 2009; Kang et al., 2015; Roebroeck et al., 2011c; Valdes-Sosa et al., 2011; Friston et al., 2003) and the related papers such as (Yuan et al., 2015; Yuan et al., 2011) use nonlinear state-space representations with inputs. For those state-space representations, there exist no simple methods for checking Granger non-causality. However, we show in Chapter 6 that by extending the concept of Granger causality to the so-called GB–Granger causality, GB–Granger non-causality can be translated to the existence of a GB–SS representation in innovation form having the two-node graph with one edge as its network graph. This can be a first step for further research on relating causality in the observed processes of nonlinear state-space systems to intrinsic properties, or more specifically to the network graph of those systems.

Distributed estimation/control: For the design of interconnected systems, choosing alternative network graphs realizing the same functionality can be beneficial. For example, for deterministic coordinated LTI–SS systems with inputs (Kempker et al., 2014b; Kempker, 2012), several control problems such as stabilization can be solved in a distributed manner: in order to stabilize the coordinator, no knowledge of the state of the agents is required, and in order to stabilize each agent, only the state of this agent and of the coordinator are needed.

It is therefore of interest to know which observed behaviors can be represented by LTI–SS representations with certain network graphs and how to convert an LTI–SS representation into an LTI–SS representation with certain network graphs while preserving its observed behavior. Such a transformation would allow the design of distributed controllers for systems which do not initially have the given network graph. Chapters 2–4 provide such conditions and transformations for autonomous

LTI–SS representations with transitive acyclic directed network graphs. Furthermore, Chapter 6 provides similar results on GB–SS representations with the network graph that has two nodes and one directed edge. Although we do not study control design in this thesis, the results/analysis we provide are necessary first steps towards solving more general cases. In addition, these systems are already useful for distributed state estimation, as for such systems we know the following: the state of each subsystem that corresponds to a node in the network graph can be estimated only by using its own output and the output of the subsystems corresponding to parent nodes in the network graph. Hence, the state of the LTI–SS or GB–SS representations that are the subject of this thesis can be estimated in a distributed manner.

Moreover, the proposed algorithms for constructing the representations in Chapters 2–4 and 6, open up the possibility of distributed parameter estimation; for calculating a subsystem that is represented by a node in the corresponding network graph, only the observed process of the subsystem and the observed processes of the subsystems represented by the parent nodes are needed. That is, the results of Chapters 2–4 and 6 provide representations that are suitable for distributed parameter estimation which serves as basis for distributed control.

Structure preserving model reduction: The results of the thesis could also be of interest for structure preserving model reduction, where the goal is to replace an interconnected model of the system by another, smaller dimensional (in terms of the dimension of states) interconnected model which has the same or similar network graph as the original model, see (van der Schaft, 2015; Sandberg and Murray, 2009) and (Monshizadeh et al., 2014). To model an observed behaviour by interconnected systems or systems with a certain network graphs, we first need to understand what property of an observed process allows for representing them by such models. Restricting ourselves to interconnected models that are subjects of this thesis, by our results, one can analyze a process to decide whether or not an interconnected model of the generating system exists. Furthermore, if it exists then by using the methods of the thesis, the interconnected model can be constructed. The constructed model has the following useful property: the reduction of the order of a subsystem corresponding to a node j of the network graph has local effect, meaning that the subsystems that correspond to any other node i of the network graph such that there is no directed path from i to j or from j to i remain unchanged. Therefore, if the model reduction of a subsystem keeps the interconnection structure of the subsystem locally then the global interconnection structure remains unchanged. If it does not then still, even though the local interconnection structure changes, the global interconnection structure is partially preserved.

Related work

Besides the concept of network graph introduced in this thesis, there are several other notions for describing the structure of a system or the network of subsystems in a system. Examples of such notions are: feedback systems (Caines and Chan, 1975; Caines, 1976; Gevers and Anderson, 1982; Hsiao, 1982), dynamical structure function (Gonçalves et al., 2007; Howes et al., 2008; Yuan et al., 2015), dynamic network modeling (Weerts, 2018; Dankers, 2014; Weerts et al., 2018; Van den Hof et al., 2013), dynamic causal modeling (Friston et al., 2003; Havlicek et al., 2015; Penny et al., 2004), and causality graphs (Eichler, 2007; Eichler, 2012). Also, besides Granger and GB–Granger causality, there are several examples for statistical notions that have essential role to understand the relation between stochastic processes. We can mention here conditional orthogonality, transfer entropy (Barnett et al., 2009) and directional mutual information (L. Massey, 1990; Kramer, 1998).

The above-mentioned concepts are strongly related to each other, however, discussing this in detail is outside the scope of this thesis. Since Granger causality (Granger, 1963; Granger, 1988; Wiener, 1956) is a key concept in this thesis and in several related papers, to give more insight about the above-mentioned concepts, we will focus on relating them to Granger causality. Note that in this thesis Granger causality is defined on covariance-stationary or in other words, weakly-stationary processes, however, it is possible to extend it to non-stationary, e.g., to co-integrated processes (Engle and Granger, 1987; Papana et al., 2014). Also, in this thesis Granger causality is defined as a logical value, i.e., it is either present or absent (see (Geweke, 1984) for measures on Granger causality) and in such a way that it does not change in time (see (Lütkepohl, 1993; Dufour and Renault, 1998; Triacca, 2000) for Granger causality in short and long time horizons).

Below, we first present previous research on model dependent formalization of interaction among processes and its model free counterpart. Then, we discuss existing results on the relationship between these two approaches.

Model dependent formalization of interaction among processes

Feedback: Even though the idea behind Granger causality and feedback are different, in particular Granger causality is a model free concept in contrast to feedback, the two concepts are strongly related (Granger, 1963). The definition of feedback between processes is primarily based on feedback systems, i.e., models where a process serves as an input in the model of another process. It can, however, be defined in many ways. Note that this thesis studies the lack of Granger causality (Granger non-causality) and thus we now focus on the lack of feedback, or in other words, on the feedback free property of processes. In (Caines and Chan, 1975; Gevers and Anderson, 1982) and in (Caines, 1988, Chapter 10), the feedback free property is defined based on network graphs of moving-average (MA) models. In (Caines, 1976) the definition of the feedback free property in (Caines and Chan, 1975) is renamed to weakly feedback free property and a stronger notion, called strongly feedback free property, is introduced. Both the strongly and weakly feedback free properties are then characterized by the model-free probabilistic concept, called conditional orthogonality (Caines, 1976). Moreover, it is pointed out that the lack of causality defined in (Granger, 1963) is equivalent to the definition of weakly feedback free property, although also the strongly feedback free property is discussed in (Granger, 1963, Section VII), see (Caines, 1988, Chapter 10) for more details. It is important to mention that in this thesis, Granger non-causality is equivalent to weakly-feedback free property of processes ((Caines, 1988, Definition 2.1, Chapter 10)), which is in several applications more realistic than the strongly feedback free property. We note that (Caines, 1988, Chapter 10) studies a more general class of processes than this thesis, namely there is no restriction to processes with rational spectral density matrix, or equivalently, to processes that have LTI–SS representations.

The results in (Granger, 1963; Caines and Chan, 1975; Caines, 1976; Gevers and Anderson, 1982) are fundamental for our work since they relate Granger causality to network graph of MA and vector autoregressive (VAR) models. The results of Chapter 2 can be viewed as a counterpart of the results in the cited papers for LTI–SS representations. In addition, the results of Chapter 5 can be viewed as an extension of the results on Granger causality and MA models, on processes that have LTI–SS representations, where a collection of Granger causalities is characterized in terms of the network graph of MA models, or equivalently, using the terminology of the thesis, transfer matrices.

Dynamical structure function: Dynamical structure function (Gonçalves et al., 2007; Howes et al., 2008; Yuan et al., 2015; Yuan et al., 2011; Gonçalves and Warnick, 2008) describes structural properties of systems. It was first defined on deterministic linear systems such as deterministic LTI–SS and VAR systems. In (Yue et al., 2015) it was further extended to stochastic VAR systems and related to Granger causality. Dynamical structure function is a similar concept to what we call in this thesis network graph, however, it has not been applied to stochastic LTI–SS representations.

Dynamic network modeling: A dynamic network model in the sense of (Weerts, 2018; Dankers, 2014; Weerts et al., 2018; Van den Hof et al., 2013) is another approach for defining a network of interacting linear systems. The class of systems defined in (Weerts, 2018; Dankers, 2014; Weerts et al., 2018; Van den Hof et al., 2013) are different from the one considered in this thesis. In the cited papers, there are no general results on the relationship between Granger causality and the network graph of dynamic network models, however, examples were discussed in (Dankers, 2014).

Dynamical causal modeling: Dynamical causal modeling (Friston et al., 2003; Havlicek et al., 2015; Penny et al., 2004) is defined on both linear and nonlinear deterministic models. It is compared to Granger causality in (Roebroeck et al., 2011a), explaining that dynamical causal modeling captures mechanistic inference among variables of a system whereas Granger causality only shows a statistical connection between them. Hence, (Roebroeck et al., 2011a) relates Granger causality to dynamical causal modeling in a similar way to how we relate Granger causality to network graphs of stochastic representations, however, it does not aim to show equivalence between the two concepts.

Coordinated systems: The work in (Kempker, 2012; Kempker et al., 2014a; Ran and van Schuppen, 2014), where deterministic LTI–SS representations in coordinated form were introduced, gave strong motivation for Chapter 3. In (Kempker, 2012) and (Kempker et al., 2014a), a general method was presented to transform a system into coordinated form. In (Kempker, 2012) and (Pambakian, 2011), Gaussian coordinated systems and their LQG control were studied. Note that the cited papers did not relate the coordinated system structure to properties of the observed process.

Model free formalization of interaction among processes

Causality graph: In (Eichler, 2005; Eichler, 2007; Eichler, 2012) interactions in systems are defined with the help of causality graphs. Causality graphs are introduced using the combination of Granger causality and instantaneous coupling, see also (Yue et al., 2015). Note that in this thesis we do not study the notion of instantaneous coupling that is defined between the observed processes of the systems that represent them. From this perspective, the cited papers aim to study more complex statistical properties of stochastic processes. However, in the cited papers causality graphs are defined on processes that are outputs of stochastic VAR models and are related to parameters of VAR models, not the more general class of systems, LTI–SS representations. In fact, the relation between causality graphs and parameters of VAR models in the cited paper are similar to the relation between the collection of causalities and LTI–SS representations showed in this thesis. Therefore, the results of this thesis can help in extending causality graphs defined in (Eichler, 2005; Eichler, 2007; Eichler, 2012) to processes that are outputs of LTI–SS representations.

Information theoretic concepts: A branch of information theory, called directed information theory (L. Massey, 1990; Kramer, 1998) studies directional relation between stochastic processes. For the purpose of Granger causality, an important notion from this field is the so-called conditional directional mutual information or simply directed information (L. Massey, 1990) that is based on the notions of conditional transfer entropy and mutual information. In fact, conditional directional mutual information can be formulated using the probabilistic notion of conditional independence which, in Gaussian case coincides with conditional orthogonality. Since Granger causality (as well as conditional Granger causality) can be formalized by conditional orthogonality, it is not surprising that under certain conditions, conditional directional mutual information can also provide an equivalent form of Granger causality, see (Barnett et al., 2009; Amblard and Michel, 2013).

To sum up the paragraphs above, we discussed different notions for describing statistical properties of stochastic processes and structures of dynamical systems. These were compared to the concepts considered in this thesis, in particular to Granger causality. In contrast to the definitions that were discussed above, we study network graphs of LTI–SS representations, LTI transfer matrices and GB–SS representations. The network graphs of the LTI systems are then related to conditional and unconditional Granger causalities among the components of the output process and the network graphs of the GB–SS representations are related to the extended form of Granger causality, called GB–Granger causality.

Relationship between model free and model dependent approaches

State-space representation: The first results on Granger causality in terms of LTI–SS representations were presented in (Barnett and Seth, 2015; Solo, 2016). The cited papers characterize Granger causality in the properties of LTI–SS representation by using transfer matrix approach. The idea behind Chapter 2 and the cited papers are similar, however, we provide a different characterization of Granger causality in terms of properties of LTI–SS representations and, contrary to (Barnett and Seth, 2015; Solo, 2016), we give a construction for LTI–SS representations whose network graph characterizes Granger causality. Note that constructing such an LTI–SS representation is interesting since it provides mechanistic explanation for Granger causality in its observed process.

The results in (Caines et al., 2003; Caines et al., 2009) are the closest ones to the results in Chapter 3. The cited papers provide necessary and sufficient conditions for the existence of LTI–SS representations in the so-called conditional orthogonal form. Conditionally orthogonal LTI–SS representations form a specific subclass of LTI–SS representations in coordinated form discussed in Chapter 3 with additional assumptions on the covariance matrix of the noise process.

The conditions of (Caines et al., 2003; Caines et al., 2009) for the existence of such systems are much stronger than the conditions proposed in Chapter 4. The paper (Caines and Wynn, 2007) is the closest one to the results in Chapters 4 and 5. The cited paper studies LTI–SS representations and their transfer matrices of Gaussian processes in a form that is a subclass of the LTI–SS representations with transitive acyclic directed graph (TADG) zero structure discussed in Chapter 4 with additional assumptions on the covariance matrix of the noise process. This additional assumption is closely related to the notion of strongly feedback free property of processes. Recall that we study Granger non-causality that corresponds to the weakly feedback free property of processes.

In (Caines and Wynn, 2007), also transfer matrices of the class of LTI–SS representations are studied in terms of conditional orthogonality. The existence of these systems are characterized by stronger conditional orthogonal conditions than the conditional orthogonality condition that are counterparts of the Granger causality conditions proposed in Chapters 4 and 5. Furthermore, the class of output processes that are modeled are more restrictive. Regarding the proofs of the statements of (Caines et al., 2003; Caines and Wynn, 2007; Caines et al., 2009), only the proof of existence of the LTI–SS representation in conditional orthogonal form can be found in (Caines et al., 2003) and it essentially differs from the proofs of the statements presented in Chapters 4 and 5. Note that (Caines et al., 2003; Caines and Wynn, 2007; Caines et al., 2009) did not provide algorithms to calculate the representations.

Causality in bilinear systems: The results of Chapter 6 relate the network graph of bilinear systems to an extended notion of Granger causality, called GB–Granger causality. To the best of our knowledge, the approach of extending Granger causality to capture nonlinear causal relation among processes based on the properties of the nonlinear system that generates the processes, is new. We adopt the GB–SS representations from (Petreczky and René, 2017). The advantage of the adopted GB–SS representations is that, contrary to (Favoreel et al., 1999; Desai, 1986), the input process is not necessarily white noise. The disadvantage of the GB–SS representations considered in (Petreczky and René, 2017) and in Chapter 6 is that, contrary to (Chen and Maciejowski, 2001; D'Alessandro et al., 1974; Favoreel et al., 1999), it does not allow additional input term in the system, only multiplicative one.

Outline

Below, we briefly address the outline of the thesis:

For providing background materials, Chapter 1 introduces linear time-invariant state-space (LTI–SS) representations, general bilinear state-space (GB–SS) representations and presents results on realization theory of these systems. For background material on LTI–SS representations and their realization theory, we refer to (Lindquist and Picci, 2015; Katayama, 2005; Ljung, 1999; Hannan and Deistler, 1988; Van Overschee and De Moor, 1996). The background material on GB–SS representations and their realization theory can be found in (Petreczky and René, 2017) and the references therein.

Chapter 2 introduces Granger causality between two stochastic processes and presents results on characterizing it by the existence of the so-called Kalman representations in block triangular form.

The results of Chapter 2 are generalized in Chapter 3, introducing conditional Granger causality. A collection of conditional Granger causality and Granger causality conditions is then characterized by the existence of so-called Kalman representations in coordinated form. Chapters 2 and 3 are based on the journal paper (Jozsa et al., 2018b).

As a collection of conditional and unconditional Granger causality, Chapter 4 introduces transitive acyclic directed graph (TADG) causality structure of ZMSIR processes. By using the results of Chapters 2 and 3, Chapter 4 presents results on characterizing TADG-causality structure with the existence of Kalman representations with the so-called TADG-zero structure. Chapter 4 includes the results from the conference papers (Jozsa et al., 2017a), however, several additional statements are presented here that were not included in the cited paper.

The implications of the results presented in Chapter 4 on transfer matrices are formulated in Chapter 5. The results of Chapter 5 are presented independently from the results of the previous chapters. The results of Chapter 5 are based on the conference paper (Jozsa et al., 2017b).

Moving away from linear state-space models towards nonlinear state-space models, Chapter 6 deals with GB–SS representations. More precisely, we study GB–SS representations in the so-called innovation form and with a certain network graph. The main result shows that the existence of GB–SS representations in innovation form with a certain network graph is equivalent to an extended form of Granger causality, called GB–Granger causality.

Chapter 7 illustrates the main results of Chapters 2–4 in practice, by applying the algorithms from Chapters 2–4 on simulated data.