Causality and Network Graph in General Bilinear State-Space Representations

Jozsa, Monika; Petreczky, Mihaly; Camlibel, M. Kanat

Published in:
IEEE Transactions on Automatic Control

DOI:
10.1109/TAC.2019.2952033

Publication date:
2020

Causality and Network Graph in General Bilinear State-Space Representations

Mónika Józsa, Mihály Petreczky, and M. Kanat Camlibel

***Abstract*—This article proposes an extension of the well-known concept of Granger causality, called GB–Granger causality. GB–Granger causality is designed to relate the internal structure of bilinear state-space systems and statistical properties of their output processes. That is, if such a system generates two processes, where one does not GB–Granger cause the other, then it can be interpreted as the interconnection of two subsystems: one that sends information to the other, and one which does not send information back.This result is an extension of earlier obtained results on the relationship between Granger causality and the internal structure of linear time-invariant state-space representations.**

***Index Terms*—Interconnected systems, stochastic systems, system realization.**

I. Introduction

Detecting interactions among stochastic processes and relating them to the internal structure of the generating systems can be of interest for several applications, such as mapping interactions in the brain, predicting economical price movements, or understanding social group behavior. The first step toward detecting such interactions is to propose a formal mathematical definition of the concept of interaction. In this article, we propose two formalizations of one directional interactions between two stochastic processes. The stochastic processes are assumed to be outputs of a nonlinear dynamical system. Both formalizations will try to capture causal interactions, i.e., that one process causes the other one. The first formalization concentrates on the information flow between the dynamical systems that generate the processes. The second one focuses on statistical properties of the processes.

Manuscript received June 28, 2019; revised July 6, 2019; accepted October 23, 2019. Date of publication November 7, 2019; date of current version July 28, 2020. This work was supported in part by the CPER Data Project, in part by the European Union, in part by the European Regional Development Fund, in part by the French State and the French Region of Hauts-de-France, and in part by the PEPS Blanc INS2I 2019 project financed by CNRS, France. Recommended by Associate Editor Z. Gao. *(Corresponding author: Mónika Józsa.)*

M. Józsa was with the IMT Lille Douai, University of Lille, Unit de Recherche Informatique Automatique, F-59000 Lille, France, and also with the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence, University of Groningen, 9700 AK Groningen, The Netherlands, He is now with Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: m.jozsa@eng.cam.ac.uk).

M. Petreczky is with the Centre de Recherche en Informatique, Signal et Automatique de Lille, 59650 Villeneuve-d'Ascq, France (e-mail: mihaly.petreczky@ec-lille.fr).

M. K. Camlibel is with the Johann Bernoulli Institute for Mathematics and Computer Science and the Artificial Intelligence, University of Groningen, 9700 AK Groningen, The Netherlands (e-mail: m.k.camlibel@rug.nl).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2019.2952033

More precisely, let $\mathbf{y}$ be a process that is partitioned into two components, such as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$. For the first approach, assume that $\mathbf{y}$ is the output of a dynamical system that can be represented as an interconnection of two subsystems: the first generates $\mathbf{y}_1$, and the second generates $\mathbf{y}_2$ as its output. Furthermore, assume that the first subsystem sends information to the second, but there is no information flowing in the opposite direction. In other words, the *network graph*[1] of this dynamical system has two nodes and one directed edge. Then, according to the first approach, we say that $\mathbf{y}_1$ influences $\mathbf{y}_2$.

This approach offers an intuitive mechanistic explanation of how one component of the output process influences the other. However, the same output process can be generated by systems with different network graphs. As a result, the presence of an interaction between two output components depends on the exact dynamical system representing the output process.

The second approach is based on statistical properties of the joint process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$. A widely used example of this approach is Granger causality [3]. Intuitively, $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$ if the best linear predictions of $\mathbf{y}_2$ based on the past values of $\mathbf{y}$ are better than those only based on the past values of $\mathbf{y}_2$. We then say that $\mathbf{y}_1$ influences $\mathbf{y}_2$, if $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$. Concepts that follow from this second approach lead to definitions that depend only on properties of $\mathbf{y}$ and do not depend on which dynamical system we use to represent $\mathbf{y}$. However, they do 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 solves the issue with the first; however, it generally does not capture the inner mechanism of the interaction. It is thus of interest to relate these approaches to benefit from the advantages of both.

In [1], [3]–[5], Granger causality was formally related to the network graphs of autoregressive (AR), moving-average (MA), and linear-time-invariant state-space (LTI–SS) models. These results show that Granger causality, despite being defined based on statistical properties of a process, can be related to structural properties of linear models of that process. In most of the fields, however, where Granger causality is applied (e.g., econometrics and neuroscience), linear models are insufficient to represent the observed process.

In this article, we extend the result on the relation between Granger causality and linear systems to a more general concept of causality and a class of nonlinear systems. That is, we define a new concept of causality that can describe interaction between processes that relate to each other in a nonlinear way. Compared to other reformulations of

[1]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 is an edge from one node to the other, if the subsystem corresponding to the source node sends information to the subsystem corresponding to the target node (see also [2, Sec. 1.4]).

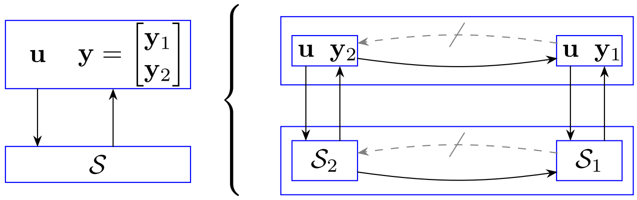

Fig. 1. Illustration of the results: Cascade interconnection structure in a GB–SS representation S with input $\mathbf{u}$ and output $\mathbf{y}$. S is decomposed into subsystems S_1 and S_2 with outputs $\mathbf{y}_1$ and $\mathbf{y}_2$, when $\mathbf{y}_2$ GB–Granger causes $\mathbf{y}_1$ but $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$ with respect to $\mathbf{u}$.

TABLE I
SUMMARY OF NOTATIONS

Σ	$\{1, 2, \ldots, d\}$
Σ^*, Σ^+	the set of all finite sequences of elements of Σ with, and without the unit element ϵ
M_w	the product of matrices $\{M_{\sigma_i}\}_{i=1}^k$ along the elements of the sequence $w = \sigma_1 \cdots \sigma_k \in \Sigma^*$
$\mathbf{r}_w(t)$	product of the processes indexed by the elements of the sequence $w \in \Sigma^*$
$\mathbf{u}_\sigma(t)$	input process indexed by $\sigma \in \Sigma$
p_σ, α_σ	parameters of admissible input processes
$\mathbf{z}_w^{\mathbf{r}}(t)$,$\mathbf{z}_w^{\mathbf{r}+}(t)$	the past and future processes of $\mathbf{r}$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ along the sequence $w \in \Sigma^*$
ZMWSSI, SII	classes of stochastic processes to which the output of GB–SS representations have to belong

Granger causality (see, e.g., [6]), our concept is designed to have a structural interpretation in the chosen class of nonlinear systems.

In order to achieve the objective of this article, we will

1) focus on a specific class of nonlinear dynamical systems;
2) define a new concept of causality as interaction among the components of a process generated by a system chosen in point 1) based on statistical properties of the process at hand;
3) characterize the causality defined in 2) by properties of the internal structure of the system generating the process at hand.

As a first step toward nonlinear systems, a natural choice is to study bilinear systems, which include, e.g., LTI–SS, switched linear, AR MA, and jump Markov linear models. Bilinear systems produce richer phenomena than linear systems, yet many analytical tools for linear systems are suitable to analyze them. In this article, we focus on general bilinear state-space (GB–SS) representations for which stochastic realization theory exists [7]. This theory serves as a basis for the technicalities of the article.

To formalize causality for the outputs of GB–SS representations, we introduce an extension of Granger causality, called GB–Granger causality, that coincides with Granger causality, when applied to outputs of stochastic LTI–SS models.

In the main results, we consider a GB–SS representation with output process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ and input process $\mathbf{u}$. Then, we show that GB–Granger noncausality from $\mathbf{y}_1$ to $\mathbf{y}_2$ with respect to $\mathbf{u}$ is equivalent to the decomposition of the GB–SS representation into the interconnection of two subsystems, one generating $\mathbf{y}_1$, and another one generating $\mathbf{y}_2$, where the former sends no information to the latter (see Fig. 1).

The results of this article are based on realization theory of bilinear systems [7]–[12] and results on Granger causality in linear systems [3], [13]–[17]. We adopt the concept of GB–SS representation from [7] and rely on the realization theory presented there. The advantage of GB–SS representations in [7] is that, contrary to [10] and [12], the input process is not necessarily white, which therefore includes, e.g., jump Markov linear systems. However, contrary to [8]–[10], it does not allow additive input terms in the system. Note that our results depend on realization theory of GB–SS representations. Hence, in order to extend our results to GB–SS representations with additive inputs, realization theory of the latter system class has to be developed. This remains a topic of future research.

Granger causality between stochastic processes was studied for AR, MA models [3], transfer functions [4], [5], [18], and for stochastic LTI–SS representations [1], [13]–[16], [19]. To extend the concept of Granger causality in GB–SS representations, we rely on similar methodology as in [1], [14], [15], and [17]. However, in contrast to [1], [14], [15], and [17], which consider LTI–SS representations, in this article, we consider the more general class of GB–SS representations.

The structure of this article is as follows, first, we introduce the terminology in Section II, which is followed by a brief summary on realization theory of GB–SS representations in Section III. Then, in Section IV, the main results on GB–Granger causality and GB–SS representations are presented. Finally, the proofs of the results can be found in the Appendix.

II. Preliminaries and Notations

The terminology adopted from [7] that GB–SS representations rely on is presented next; see Table 1 for the summary of notations.

We consider discrete-time, square-integrable, multivariate, wide-sense stationary stochastic processes with real entries. Throughout the article, we fix a probability space $(\Omega, \mathcal{F}, P)$ and all the random variables and stochastic processes are understood with respect to $(\Omega, \mathcal{F}, P)$. The random variable of a process $\mathbf{z}$ at time t is denoted by $\mathbf{z}(t)$, where t is from the discrete-time axis of integers $\mathbb{Z}$. Using standard notation, the expected value of a random variable $\mathbf{z}(t)$ is written as $E[\mathbf{z}(t)]$ and the covariance matrix between two random variables $\mathbf{z}_1(t)$ and $\mathbf{z}_2(t)$ is denoted by $E[(\mathbf{z}_1(t) - E[\mathbf{z}_1(t)])(\mathbf{z}_2(t) - E[\mathbf{z}_2(t)])^T]$. Note that if the processes $\mathbf{z}_1$ and $\mathbf{z}_2$ are zero-mean, the latter simplifies to $E[\mathbf{z}_1(t)\mathbf{z}_2^T(t)]$. The conditional expectation of a random variable $\mathbf{z}$ to a σ-algebra $\mathcal{F}$ is denoted by $E[\mathbf{z}|\mathcal{F}]$. When a process $\mathbf{z}$ or a random variable $\mathbf{z}(t)$ takes its values from $\mathbb{R}^n$, then we write $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{z}(t) \in \mathbb{R}^n$. Consider a process $\mathbf{z}$ and a present time $t \in \mathbb{Z}$. The σ-algebras generated by the random variables in the present, past, and future of $\mathbf{z}$ are denoted by $\mathcal{F}_t^{\mathbf{z}} = \sigma(\mathbf{z}(t))$, $\mathcal{F}_{t-}^{\mathbf{z}} = \sigma(\{\mathbf{z}(k)\}_{k=-\infty}^{t-1})$, and $\mathcal{F}_{t+}^{\mathbf{z}} = \sigma(\{\mathbf{z}(k)\}_{k=t}^{\infty})$, respectively, where for a set Z of random variables, $\sigma(Z)$ denotes the smallest σ-algebra, which contains each σ-algebra generated by an element of Z.

In the rest of this section, we will introduce tools that will help us to define GB–SS representations in Section III.

Throughout this article, we denote the finite set $\{1, 2, \ldots, d\}$ by Σ, where d is a positive integer.

Let Σ^+ be the set of finite sequences of elements of Σ, i.e., an element of Σ^+ is a sequence of the form $w = \sigma_1 \cdots \sigma_k$, where $\sigma_1, \ldots, \sigma_k \in \Sigma$, $k > 0$. We define the concatenation operation on Σ^+: if $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$ and $v = \hat{\sigma}_1 \cdots \hat{\sigma}_l \in \Sigma^+$, then the concatenation of w and v, denoted by wv, is defined by $wv = \sigma_1 \cdots \sigma_k \hat{\sigma}_1 \cdots \hat{\sigma}_l$. It will be convenient to extend Σ^+ by a formal unit element $\epsilon \notin \Sigma^+$. We denote this set by $\Sigma^* = \Sigma^+ \cup \{\epsilon\}$. The concatenation operation is extended to Σ^* as follows: $\epsilon\epsilon = \epsilon$, and for any $w \in \Sigma^+$, $\epsilon w = w\epsilon = w$. We define the length of a sequence $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$ by $|w| = k$ and the length of ϵ by $|\epsilon| = 0$. Consider a set of matrices $\{M_\sigma\}_{\sigma\in\Sigma}$, where

$M_\sigma \in \mathbb{R}^{n\times n}$, $n \geq 1$ for all $\sigma \in \Sigma$ and let $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$. Then, we denote the matrix $M_{\sigma_k} \cdots M_{\sigma_1}$ by M_w and we define $M_\epsilon = I$. In addition, for a set of processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and for $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$, we denote the process $\mathbf{u}_{\sigma_k}(t) \cdots \mathbf{u}_{\sigma_1}(t - |w| + 1)$ by $\mathbf{u}_w(t)$ and define $\mathbf{u}_\epsilon(t) \equiv 1$.

In order to describe the behavior of the processes of GB–SS representations, we introduce the following processes.

Definition 1: Consider a process $\mathbf{r}$ and a set of processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. Let $\sigma \in \Sigma$ and $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$. Then, we define the process $\mathbf{z}^{\mathbf{r}}_w(t) = \mathbf{r}(t - |w|)\mathbf{u}_w(t-1)$ which we call the *past* of $\mathbf{r}$ with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ along w, and we define the process $\mathbf{z}^{\mathbf{r}+}_w(t) = \mathbf{r}(t + |w|)\mathbf{u}_w(t + |w| - 1)$ which we call the *future* of $\mathbf{r}$ with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ along w.[2]

Note that for $w = \epsilon$, both the past $\mathbf{z}^{\mathbf{r}}_\epsilon(t)$ and the future $\mathbf{z}^{\mathbf{r}+}_\epsilon(t)$ of $\mathbf{r}$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ equals $\mathbf{r}(t)$.

Next, we define *admissible sets of processes*, see [7, Definition 1], which will help us to formulate a Markovianlike property of the input processes of GB–SS representations.

Definition 2 (admissible set of processes): A set of processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is called *admissible* if the following holds.

1) $[\mathbf{u}_v^T, \mathbf{u}_w^T]^T$ is wide-sense stationary for all $v, w \in \Sigma^*$.
2) There exist $\{\alpha_\sigma\}_{\sigma\in\Sigma} \in \mathbb{R}$ such that $\sum_{\sigma\in\Sigma} \alpha_\sigma \mathbf{u}_\sigma(t) \equiv 1$.
3) There exist (strictly) positive numbers $\{p_\sigma\}_{\sigma\in\Sigma}$, such that $E[\mathbf{u}_{v_1\sigma_1}(t)\mathbf{u}_{v_2\sigma_2}(t)| \bigvee_{\sigma\in\Sigma} \mathcal{F}^{\mathbf{u}_\sigma}_{t-}]$
$$= \begin{cases} p_{\sigma_1}\mathbf{u}_{v_1}(t-1)\mathbf{u}_{v_2}(t-1) & \sigma_1 = \sigma_2 \text{ and } v_1v_2 \in \Sigma^+ \\ 0 & \sigma_1 \neq \sigma_2 \end{cases}$$
for any $\sigma_1, \sigma_2 \in \Sigma$ and $v_1, v_2 \in \Sigma^*$, where $\bigvee_{\sigma\in\Sigma} \mathcal{F}^{\mathbf{u}_\sigma}_{t-}$ is the smallest σ-algebra, s.t. $\bigvee_{\sigma\in\Sigma} \mathcal{F}^{\mathbf{u}_\sigma}_{t-} \supseteq \mathcal{F}^{\mathbf{u}_\sigma}_{t-}$ for all $\sigma \in \Sigma$.

The next definition is based on [7, Definitions 1and 5] and it introduces a class of processes that the output, state, and noise processes of GB–SS representations belong to. The definition involves the concept of conditionally independent σ-algebras [20]: Two σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ are conditionally independent w.r.t. a third one $\mathcal{F}_3$, if for every $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$, $P(A_1 \cap A_2|\mathcal{F}_3) = P(A_1|\mathcal{F}_3)P(A_2|\mathcal{F}_3)$.

Definition 3 (ZMWSSI and ZMWSSI–SII processes): A stochastic process $\mathbf{r}$ is called *zero-mean wide-sense stationary* w.r.t. an admissible set of processes $\{\mathbf{u}\}_{\sigma\in\Sigma}$ (ZMWSSI) if $\mathcal{F}^{\mathbf{r}}_{(t+1)-}$ and $\mathcal{F}^{\mathbf{u}}_{t+}$ are conditionally independent w.r.t. $\mathcal{F}^{\mathbf{u}}_{t-}$, and $[\mathbf{r}^T, (\mathbf{z}^{\mathbf{r}}_v)^T, (\mathbf{z}^{\mathbf{r}}_w)^T]^T$ is zero-mean wide-sense stationary for all $v, w \in \Sigma^+$. Furthermore, a ZMWSSI process $\mathbf{r}$ is said to be *square integrable with respect to* $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ (ZMWSSI–SII), if for all $w \in \Sigma^+$, the process $\mathbf{z}^{\mathbf{r}+}_w$ is square integrable.

III. GB–SS Representations

This section introduces GB–SS representations and innovation GB–SS representations, see also [7]. To begin with, we define GB–SS representations.

Definition 4 (GB–SS representation): A system of the form

$$\begin{aligned} \mathbf{x}(t+1) &= \sum_{\sigma\in\Sigma} (A_\sigma \mathbf{x}(t) + K_\sigma \mathbf{v}(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{v}(t) \end{aligned} \tag{1}$$

where $A_\sigma \in \mathbb{R}^{n\times n}$, $K_\sigma \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{k\times n}$, $D \in \mathbb{R}^{k\times m}$, $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{v}(t) \in \mathbb{R}^m$, $\mathbf{y}(t) \in \mathbb{R}^k$, and $\mathbf{u}_\sigma(t) \in \mathbb{R}$, $\sigma \in \Sigma$ are called GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ if the following holds.

[2]We can obtain the processes in Definition 1 by multiplying the parallel processes used in [7] with a scalar, see, e.g., [7, eq. (6)].

1) $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is admissible.
2) $[\mathbf{x}^T, \mathbf{v}^T]$ is ZMWSSI with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$.
3) For $w \in \Sigma^+$, $E[\mathbf{z}^{\mathbf{v}}_w(t)\mathbf{v}^T(t)] = 0$ and $E[\mathbf{z}^{\mathbf{x}}_w(t)\mathbf{v}^T(t)] = 0$.
4) For $\hat{\sigma}, \sigma \in \Sigma$, $E[\mathbf{z}^{\mathbf{x}}_{\hat{\sigma}}(t)(\mathbf{z}^{\mathbf{v}}_\sigma(t))^T] = 0$.
5) $\sum_{\sigma\in\Sigma} p_\sigma A_\sigma \otimes A_\sigma$ is stable, i.e., all its eigenvalues are inside the open unit disk.

We refer to a GB–SS representation (1) as GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, D, \mathbf{v}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ or as GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, D, \mathbf{v})$ of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$, where note that $\{A_\sigma, K_\sigma, \mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and $\mathbf{v}$ determine the state process. Furthermore, notice that $\mathbf{y}$ is the linear combination of $\mathbf{x}$ and $\mathbf{v}$ and, thus, it is also ZMWSSI w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$.

Depending on the choice of the input processes, the behavior of a GB–SS representation can significantly vary. The constraint on the input, formulated in Definition 2, gives scope to choosing $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$, for example, in the following ways.

1) $\Sigma = 1$ and $\mathbf{u}_1(t) \equiv 1$, then $\mathbf{u}_1$ is admissible and the GB–SS representation defines an LTI–SS representation.
2) $\mathbf{u}_\sigma(t)$ is zero-mean, square-integrable, independent identically distributed (i.i.d.) process for all $\sigma \in \Sigma$ and $\mathbf{u}_{\sigma_1}(t)$ and $\mathbf{u}_{\sigma_2}(t)$ are independent for all $\sigma_1, \sigma_2 \in \Sigma, \sigma_1 \neq \sigma_2$, then $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is admissible [7, Example 1].
3) $\mathbf{u}_\sigma(t) = \chi(\Theta(t) = \sigma)$, where Θ is an i.i.d. process taking values in Σ, then $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is admissible [7, Example 2].

More examples can be found in [7]. Note that Definition 2 gives a stricter definition of admissible set of processes than [7, Definition 1].[3] The results of this article remain valid with the definition of admissible set of processes in [7]; however, we use Definition 2 in order to avoid technicalities.

A. Innovation GB–SS Representations

Below, we define innovation processes and innovation GB–SS representations. The latter class of representations plays a key role throughout the rest of this article.

To this end, we recall from [7] the following notations. The real valued zero-mean square integrable random variables form a Hilbert space $\mathcal{H}$ with the covariance as the inner product (see [21] for details). Let $\mathbf{r}$ be a ZMWSSI process w.r.t. a set of admissible processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. Then, the one-dimensional components of $\mathbf{r}(t)$ and $\mathbf{z}^{\mathbf{r}}_w(t)$ (see Definition 1) belong to $\mathcal{H}$ for all $t \in \mathbb{Z}$. We denote the Hilbert spaces generated by the one-dimensional components of $\mathbf{r}(t)$ and of $\{\mathbf{z}^{\mathbf{r}}_w(t)\}_{w\in\Sigma^+}$ by $\mathcal{H}^{\mathbf{z}}_t$ and $\mathcal{H}^{\mathbf{z}^{\mathbf{r}}_w}_{t,w\in\Sigma^+}$, respectively. The (orthogonal) linear projection of $\mathbf{r}(t)$ onto a closed subspace $\mathcal{M}$ of $\mathcal{H}$ is meant elementwise and it is denoted by $E_l[\mathbf{r}(t)|\mathcal{M}]$. If all the components of $\mathbf{r}(t)$ are in $\mathcal{M} \subset \mathcal{H}$, then we write $\mathbf{r}(t) \in \mathcal{M}$.

Definition 5 (GB–innovation process): The GB–innovation process of a ZMWSSI process $\mathbf{y}$ w.r.t. the processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is defined by $\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}]$.

Definition 6 (innovation GB–SS representation): A GB–SS representation (1) is called innovation GB–SS representation if the noise process $\mathbf{v}$ is the GB–innovation process $\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}]$ of $\mathbf{y}$ with respect to the input $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and the matrix D of (1) is the identity matrix.

In the specific case, when $\Sigma = \{1\}$ and $\mathbf{u}_1(t) \equiv 1$, innovation GB–SS representations define innovation LTI–SS representations (called Kalman representation in [1] and [2]).

Finally, we make a technical assumption that requires the definition of full rank processes.

[3]The set of admissible words used in [7] is here the trivial Σ^+ set.

Definition 7: An output process $\mathbf{y}$ of a GB–SS representation is called *full rank* if for all $\sigma \in \Sigma$ and $t \in \mathbb{Z}$, the matrix $E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}_\sigma^2(t)]$ is strictly positive definite, where $\mathbf{e}$ is the GB–innovation process of $\mathbf{y}$ w.r.t. the input $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$.

Definition 7 is equivalent to that the random variable $\mathbf{z}_\sigma^{\mathbf{e}}(t)$ has positive definite variance matrix for all $\sigma \in \Sigma$ and $t \in \mathbb{Z}$. The next assumption will be in force in the rest of this article.

Assumption 1: The output process $\mathbf{y}$ is ZMWSSI–SII with respect to the input $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and it is full rank.

Among innovation GB–SS representations, we will focus on the so-called minimal ones: We define the *dimension* of a GB–SS representation as the dimension of its state process. Then, a GB–SS representation is called *minimal* if it has minimal dimension among all GB–SS representations of the same input–output processes. Minimal innovation GB–SS representations have several advantageous properties as described in Remark 1 below.

Remark 1 (Realization theory): According to [7, Th. 3, 5, 6], if Assumption 1 holds, then $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ has a minimal innovation GB–SS representation. The latter GB–SS representation can be calculated from any GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ using [7, Algorithm 1], or from suitable high-order moments of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ using [7, Algorithm 2]. Finally, any two minimal innovation GB–SS representations of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ are isomorphic (see [7, Sec. III.B]) for the formal definition of isomorphism. That is, without loss of generality, we can restrict attention to minimal innovation GB–SS representations.

IV. GB–Granger Causality in GB–SS Representations

In this section, we present the main results of this article on an extended form of Granger causality, called GB–Granger causality and properties of GB–SS representations. First, we introduce GB–Granger causality, and then, we present its characterization by properties of GB–SS representations. Throughout the rest of this article, $\mathbf{y}$ is partitioned as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$, where $\mathbf{y}_i \in \mathbb{R}^{k_i}$ for some $k_i > 0$, $i = 1, 2$.

A. Extending Granger Causality

Informally, $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$, if the best linear predictions of $\mathbf{y}_2$ based on the past values of $\mathbf{y}$ are the same as those based only on the past values of $\mathbf{y}_2$. Recall that $\mathcal{H}_{t-}^{\mathbf{z}}$ denotes the Hilbert space generated by the past $\{\mathbf{z}(t-k)\}_{k=1}^{\infty}$ of $\mathbf{z}$. Then, Granger causality is defined as follows.

Definition 8 (Granger causality): [1, Definition 5] Consider a zero-mean square integrable, wide-sense stationary process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$. We say that $\mathbf{y}_1$ *does not Granger cause* $\mathbf{y}_2$ if for all $t, k \in \mathbb{Z}$, $k \geq 0 E_l[\mathbf{y}_2(t+k)|\mathcal{H}_{t-}^{\mathbf{y}}] = E_l[\mathbf{y}_2(t+k)|\mathcal{H}_{t-}^{\mathbf{y}_2}]$. Otherwise, we say that $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$.

Recall that a GB–SS representation defines an LTI–SS representation if $\Sigma = \{1\}$ and $\mathbf{u}_1(t) \equiv 1$. Accordingly, the innovation process of an output $\mathbf{y}$ in an innovation LTI–SS representations is $\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t)|\mathcal{H}_{t-}^{\mathbf{y}}]$. It is easy to see that in an innovation LTI–SS representation, the output process can be expressed by the linear combination of its own past values. However, it is no longer true in the more general class of GB–SS representations. In fact, an innovation GB–SS representation defines a linear relationship between the future of its output w.r.t. the inputs, denoted by $\mathbf{z}_v^{\mathbf{y}+}(t)$ and the past of its output w.r.t. the inputs, denoted by $\mathbf{z}_w^{\mathbf{y}}(t)$, see Definition 1. This motivates our extension of Granger causality, where we use the process $\mathbf{z}_v^{\mathbf{y}+}(t)$ rather than $\mathbf{y}(t+|v|)$ and $\mathbf{z}_v^{\mathbf{y}-}(t)$ rather than $\mathbf{y}(t-|v|)$.

Definition 9 (GB–Granger causality): Consider the processes $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$, where $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is admissible and $\mathbf{y}$ is ZMWSSI w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. We say that $\mathbf{y}_1$ *does not GB–Granger cause* $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ if for all $v \in \Sigma^*$ and $t \in \mathbb{Z}$

$$E_l[\mathbf{z}_v^{\mathbf{y}_2+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = E_l[\mathbf{z}_v^{\mathbf{y}_2+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}}]. \tag{2}$$

Otherwise, $\mathbf{y}_1$ GB–Granger causes $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$.

Informally, $\mathbf{y}_1$ *does not GB–Granger cause* $\mathbf{y}_2$, if the best linear predictions of the future of $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ along v is the same based on the past of $\mathbf{y}$ or based on the past of $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ along $\{w\}_{w\in\Sigma^+}$.

Remark 2: If $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$, then it implies that $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$. Moreover, in the specific case, when $\Sigma = \{1\}$ and $\mathbf{u}_1(t) \equiv 1$, the processes $\mathbf{z}_v^{\mathbf{y}+}(t)$ and $\mathbf{z}_w^{\mathbf{y}}(t)$ are $\mathbf{y}(t+|v|)$ and $\mathbf{y}(t-|w|)$, respectively, and thus Definitions 8 and 9 coincide. The relationship between GB–Granger causality and other concepts of causality, such as conditional independence [6], seems to be more involved and remains a topic of future research.

Notice that Granger causality is defined purely by statistical properties of a stochastic process. However, if this process is the output of an LTI–SS representation, Granger causality can also be related to the internal structure of the representation ([1, Th. 1]). In the next section, we derive an extension of the latter results, on the relationship between GB–Granger causality and the internal structure of GB–SS representations.

B. Main Results

Next, we present the main results of this article on the relationship between GB–Granger causality and network graphs of GB–SS representations. The representations in question are minimal innovation GB–SS representations that can be constructed algorithmically (see Remark 1 and Algorithm 1 later on in this section).

Theorem 1: With Assumption 1, consider a GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$ and let $\mathbf{e} = [\mathbf{e}_1^T, \mathbf{e}_2^T]^T$ be the GB–innovation process of $\mathbf{y}$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$, where $\mathbf{e}_i \in \mathbb{R}^{k_i}$, $i = 1, 2$. Then, $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ if and only if there exists a minimal innovation GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ such that for all $\sigma \in \Sigma$

$$A_\sigma = \begin{bmatrix} A_{\sigma,11} & A_{\sigma,12} \\ 0 & A_{\sigma,22} \end{bmatrix}$$

$$K_\sigma = \begin{bmatrix} K_{\sigma,11} & K_{\sigma,12} \\ 0 & K_{\sigma,22} \end{bmatrix} \quad C = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix} \tag{3}$$

where for some $n_1, k_1 \geq 0$, $n_2, k_2 > 0$ $A_{\sigma,ij} \in \mathbb{R}^{n_i \times n_j}$, $K_{\sigma,ij} \in \mathbb{R}^{n_i \times k_j}$, $C_{ij} \in \mathbb{R}^{k_i, n_j}$, $i, j = 1, 2$, and $(\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$.

The proof can be found in the Appendix. If $\Sigma = \{1\}$ and $\mathbf{u}_1(t) \equiv 1$, the GB–SS representation reduces to an LTI–SS representation and Definitions 8 and 9 coincide, see Remark 2. As a result, Theorem 3 reduces to earlier results on LTI–SS representations and Granger causality (see [1, Th. 1]).

An innovation GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ that satisfies (3) can be viewed as a cascade interconnection of two subsystems. Define the subsystems

$$\mathcal{S}_1 \begin{cases} \mathbf{x}_1(t+1) = \sum_{\sigma\in\Sigma}(A_{\sigma,11}\mathbf{x}_1(t) + K_{\sigma,11}\mathbf{e}_1(t))\mathbf{u}_\sigma(t) \\ \qquad\qquad + \sum_{\sigma\in\Sigma}(A_{\sigma,12}\mathbf{x}_2(t) + K_{\sigma,12}\mathbf{e}_2(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}_1(t) = \sum_{i=1}^{2} C_{1i}\mathbf{x}_i(t)) + \mathbf{e}_2(t) \end{cases}$$

$$\mathcal{S}_2 \begin{cases} \mathbf{x}_2(t+1) = (A_{\sigma,22}\mathbf{x}_2(t) + K_{\sigma,22}\mathbf{e}_2(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}_2(t) = C_{22}\mathbf{x}_2(t) + \mathbf{e}_2(t) \end{cases}.$$

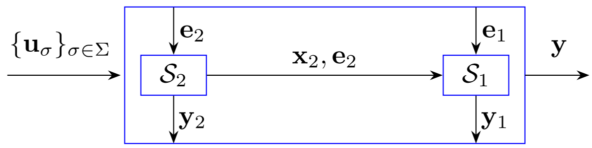

Fig. 2. Cascade interconnection in a GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ with system matrices as in (3).

Algorithm 1: Block Triangular Minimal Innovation GB–SS Representation.

Input $\tilde{A}_\sigma \in \mathbb{R}^{n\times n}, \tilde{K}_\sigma \in \mathbb{R}^{n\times m}, \sigma \in \Sigma, \tilde{C} \in \mathbb{R}^{k\times n}$

Output $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C)$

Step 1 Define the submatrix consisting of the last k_2 rows of $\tilde{C}$ by $\tilde{C}_2 \in \mathbb{R}^{k_2\times n}$ and take the observability matrix $\tilde{O}_{M(n)}$ of $(\{\tilde{A}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}_2)$ up to n. If $\tilde{O}_{M(n)}$ is not of full column rank then define the non-singular matrix $T^{-1} = \begin{bmatrix} T_1 & T_2 \end{bmatrix}$ such that $T_1 \in \mathbb{R}^{n\times n_1}$ spans the kernel of $\tilde{O}_{M(n)}$. If $\tilde{O}_{M(n)}$ is of full column rank, then set $T = I$.

Step 2 Define the matrices $A_\sigma = T\tilde{A}_\sigma T^{-1}$, $K_\sigma = T\tilde{K}_\sigma$ for $\sigma \in \Sigma$ and $C = \tilde{C}T^{-1}$.

Notice that $\mathcal{S}_2$ sends its state $\mathbf{x}_2$ and noise $\mathbf{e}_2$ to $\mathcal{S}_1$ as an external input, whereas $\mathcal{S}_1$ does not send information to $\mathcal{S}_2$. The corresponding network graph is illustrated in Fig. 2.

The necessity part of the proof of Theorem 1 is constructive, and it is based on calculating an innovation GB–SS representation described in Theorem 1. For this calculation, we present Algorithm 1 below, along with the statement of its correctness.

Before presenting Algorithm 1, we define a (complete) *lexicographic ordering* $(\prec)$ *on* Σ^*: $v \prec w$ if either $|v| < |w|$ or if $v = \nu_1 \ldots \nu_k$, $w = \sigma_1 \ldots \sigma_k$, and $\exists\ l \in \{1, \ldots, k\}$ such that $\nu_i = \sigma_i$, $i < l$ and $\nu_l < \sigma_l$. Let the ordered elements of Σ^* be $v_1 = \epsilon, v_2 = \sigma_1, \ldots$ and define $M(j)$ as the number of words of length at most j. We then define the observability matrix $\mathscr{O}_l$ up to l of $(\{A_\sigma\}_{\sigma\in\Sigma}, C)$ as $\mathscr{O}_l = [(CA_{v_1})^T \cdots (CA_{v_l})^T]^T$, where $\{A_\sigma\}_{\sigma\in\Sigma} \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{k\times n}$.

Lemma 1: Denote the GB–innovation process of $\mathbf{y}$ by $\mathbf{e}$. Assume that $(\{\tilde{A}_\sigma, \tilde{K}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}, I, \mathbf{e})$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ of dimension n. Let $\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}$ and C denote the matrices returned by Algorithm 1. Then, $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$, and the matrices $\{A_\sigma\}_{\sigma\in\Sigma}$ and C are in the form of

$$A_\sigma = \begin{bmatrix} A_{\sigma,11} & A_{\sigma,12} \\ 0 & A_{\sigma,22} \end{bmatrix} \quad C = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix} \tag{4}$$

where $A_{\sigma,ij} \in \mathbb{R}^{n_i,n_j}$ and $C_{ij} \in \mathbb{R}^{k_i,n_j}$, $i, j = 1, 2$ for some $n_1, k_1 \geq 0$, $n_2, k_2 > 0$. In addition, if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$, then the matrices $\{K_\sigma\}_{\sigma\in\Sigma}$ are in the form of

$$K_\sigma = \begin{bmatrix} K_{\sigma,11} & K_{\sigma,12} \\ 0 & K_{\sigma,22} \end{bmatrix} \tag{5}$$

where $K_{\sigma,ij} \in \mathbb{R}^{n_i\times k_j}, i, j \in \{1, 2\}$ and $(\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$.

The proof of Lemma 1 can be found in the Appendix.

Remark 3: From Lemma 1, it follows that if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$, then Algorithm 1 calculates the system matrices of the GB–SS representation described in Theorem 1. A minimal innovation GB–SS representation can be calculated from any GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ using [7, Algorithm 1], see Remark 1. Having a minimal innovation GB–SS representation as the input, Algorithm 1 provides a constructive proof of the necessity part of Theorem 1 by calculating a minimal innovation GB–SS representation in the form of (3) that characterizes GB–Granger noncausality.

Remark 4 (Checking GB–Granger causality): Algorithm 1 can be used for checking GB–Granger causality as follows. Apply Algorithm 1 and check if the matrices $\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}$ and C returned by Algorithm 1 satisfy (4) and (5), and if $\mathcal{S}_2 = (\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$. By Lemma 1 and Theorem 1, both tests are positive if and only if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$. We check whether $\mathcal{S}_2$ is a minimal innovation GB–SS representation as follows. We use [7, Algorithm 1] to compute a minimal innovation GB–SS representation $\bar{\mathcal{S}}_2$ of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$ and the covariances $\bar{Q}_\sigma = E[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}_\sigma^2(t)], \sigma \in \Sigma$ of the innovation process $\mathbf{v}$ of $\mathbf{y}_2$. Then, $\mathcal{S}_2$ is a minimal innovation GB–SS representation, if and only if $\mathcal{S}_2$ and $\bar{\mathcal{S}}_2$ have the same dimension and the same noise process, i.e., $\mathbf{v} = \mathbf{e}_2$. To check the latter, we can use the following lemma.

Lemma 2: $\mathbf{v}(t) = \mathbf{e}_2(t)$ if and only if for all $i = 1, \ldots, k_2$, $\sum_{\sigma\in\Sigma} \alpha_\sigma^2 \bar{Q}_{\sigma,ii} = \sum_{\sigma\in\Sigma} \alpha_\sigma^2 Q_{\sigma,(k_1+i)(k_1+i)}$, where $\{\alpha_\sigma\}_{\sigma\in\Sigma}$ are as in Definition 2, and $Q_{\sigma,rl}$ and $\bar{Q}_{\sigma,kl}$ denote the $(k,l)th$ entry of the matrices Q_σ and $\bar{Q}_\sigma$, respectively.

The proof of Lemma 2 is presented in the Appendix. Since a minimal innovation GB–SS representation can be calculated from suitable high-order moments of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ using [7, Algorithm 2], and the latter moments can be estimated from sampled data, the procedure above could be a starting point of a statistical test for checking GB–Granger causality, similar to the one of Granger causality in [2]. This remains a topic of future research .

Example 1: Consider a GB–SS representation $(\{\bar{A}_\sigma, \bar{K}_\sigma\}_{\sigma\in\Sigma}, \bar{C}, \bar{D}, \bar{\mathbf{v}}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$, where $\Sigma = \{1, 2\}$, $\mathbf{u}_\sigma(t) = \chi(\theta(t) = \sigma)$ with χ being the characteristic function and $\theta(t) \in \{1, 2\}$ being an i.i.d. process; $\mathbf{v}(t)$ is a normalized Gaussian white noise process, s.t. the σ-algebras generated by $\{\mathbf{v}(t)\}_{t\in\mathbb{Z}}$ and $\{\theta(t)\}_{t\in\mathbb{Z}}$ are independent. The system matrices are given by

$$\bar{A}_1 = \begin{bmatrix} 0.8 & 0.9 & -0.8 & 0.3 \\ 1.9 & 0.4 & -1.4 & 1.5 \\ 2.9 & 1.7 & -2.3 & 0.9 \\ 0.9 & 0.4 & -0.6 & 0 \end{bmatrix} \quad \bar{K}_1 = \begin{bmatrix} 1.1 & 1.5 \\ 1.1 & 0.9 \\ 2.3 & 3 \\ 0.6 & 0.7 \end{bmatrix}$$

$$\bar{A}_2 = \begin{bmatrix} -1.38 & -0.42 & 1.24 & -1.64 \\ -0.66 & -0.58 & 0.68 & -0.52 \\ -2.76 & -1.08 & 2.48 & -2.84 \\ -0.68 & -0.32 & 0.6 & -0.56 \end{bmatrix} \quad \bar{K}_2 = \begin{bmatrix} 0.2 & 1.24 \\ 0.2 & 0.84 \\ 0.44 & 2.52 \\ 0.12 & 0.56 \end{bmatrix}$$

$$\bar{C} = \begin{bmatrix} 8.5 & 5.5 & -8 & 11 \\ 3.5 & -1.5 & -2 & 4 \end{bmatrix} \quad \bar{D} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

By [7, Example 2], $\{\mathbf{u}_\sigma(t)\}_{\sigma=1}^2$ satisfies Definition 2 with $\alpha_\sigma = 1$, $p_\sigma = P(\theta(t) = \sigma)$, $\sigma \in \Sigma$. Notice that $E[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}_\sigma^2(t)] = p_\sigma I_2$, $\sigma \in \Sigma$. We assume that $p_1 = 0.3$ and $p_2 = 0.7$. We transform this GB–SS representation to a minimal innovation GB–SS representation using [7, Algorithm 1] and then we apply Algorithm 1 with the partitioning of the output $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$, $\mathbf{y}_i \in \mathbb{R}$. The output matrices of Algorithm 1 define an innovation GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$, where

$Q_\sigma = E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}_\sigma^2(t)] = p_\sigma I_2$, and the matrices are

$$A_1 = \begin{bmatrix} -1.72 & -2.64 & -2.75 & -0.68 \\ 0.86 & 1.42 & 2.45 & 0.76 \\ 0 & 0 & -0.62 & -0.03 \\ 0 & 0 & 1.11 & -0.18 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.98 & 1.53 & 4.23 & 1.32 \\ -0.56 & -0.9 & -2.47 & -0.89 \\ 0 & 0 & -0.26 & -0.02 \\ 0 & 0 & 1.04 & 0.14 \end{bmatrix}$$

$$K_1 = \begin{bmatrix} 2.33 & 3 \\ -1.63 & -1.83 \\ 0 & -0.09 \\ 0 & 0.45 \end{bmatrix} \quad K_2 = \begin{bmatrix} 0.45 & 2.5 \\ -0.3 & -1.59 \\ 0 & -0.02 \\ 0 & 0.32 \end{bmatrix}$$

$$C = \begin{bmatrix} -2.24 & -5.39 & -14.41 & -6.76 \\ 0 & 0 & -5.86 & 0.43 \end{bmatrix}.$$

Hence, they satisfy (4) and (5) with $n_1 = n_2 = 2$. Following Remark 4, we can check that $(\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$ is a minimal innovation GB–SS representation: the condition of Lemma 2 is satisfied, minimal innovation GB–SS representations of $\mathbf{y}_2$ have dimension 2, and the variance of their noise process is 1. From this, we can conclude that $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$.

V. Conclusion

In this article, we proposed a new concept, called GB–Granger causality for defining causality in a statistical manner between processes that are outputs of GB–SS representations. We showed that GB–Granger causality can be characterized by structural properties of GB–SS representations, namely, we showed that the absence of GB–Granger causality is equivalent to the existence of a GB–SS representation, which is a cascade interconnection of two subsystems. Moreover, we proposed an algorithm for calculating such a GB–SS representation. When applied to LTI–SS representations, these results boil down to the known correspondence between Granger causality and structural properties of LTI–SS representations [1], [14], [17].

The results could be used for developing statistical hypothesis testing for GB–Granger causality, in a similar manner as it was done for linear systems and Granger causality [2]. This extension, which would have potential applications in, e.g., neuroscience and econometrics, remains a future work.

Appendix
Proof

Proof of Lemma 1: In order to prove Lemma 1, we use the following result.

Lemma 3: Consider an innovation GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ with state process $\mathbf{x}$. Then, $E_l[\mathbf{z}_v^{\mathbf{y}+}(t) | \mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = CA_v\mathbf{x}(t)$, for all $v \in \Sigma^+$.

Proof: Recall that $\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}$ is the Hilbert space generated by the past $\{\mathbf{z}_w^{\mathbf{y}}\}_{w\in\Sigma^+}$ of $\mathbf{y}$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. From [7, eq. (38)], we know that $E[\mathbf{z}_v^{\mathbf{y}+}(t)(\mathbf{z}_{\sigma w}^{\mathbf{y}}(t))^T] = E[\mathbf{y}(t)(\mathbf{z}_{\sigma wv}^{\mathbf{y}}(t))^T] = CA_vA_wG_\sigma$ for all $\sigma\in\Sigma, v\in\Sigma^+, w\in\Sigma^*$, where for $\sigma\in\Sigma$ $G_\sigma = A_\sigma P_\sigma C^T + K_\sigma Q_\sigma$ and $P_\sigma = E[\mathbf{x}(t)(\mathbf{x}(t))^T\mathbf{u}_\sigma^2(t)]$. Also, from [7, Lemma 12], we know that $E[\mathbf{x}(t)(\mathbf{z}_{\sigma w}^{\mathbf{y}}(t))^T] = A_wG_\sigma$ for all $\sigma\in\Sigma, w\in\Sigma^*$. Hence, $E[\mathbf{z}_v^{\mathbf{y}+}(t)(\mathbf{z}_{\sigma w}^{\mathbf{y}}(t))^T] = CA_vE[\mathbf{x}(t)(\mathbf{z}_{\sigma w}^{\mathbf{y}}(t))^T]$ for all $v, \sigma w \in \Sigma^+$. Considering that $\mathbf{x}(t) \in \mathcal{H}_{t,w}^{\mathbf{z}_w^{\mathbf{e}}}$, and that $\mathcal{H}_{t,w}^{\mathbf{z}_w^{\mathbf{e}}} \subseteq \mathcal{H}_{t,w}^{\mathbf{z}_w^{\mathbf{y}}}$, we obtain that $E_l[\mathbf{z}_v^{\mathbf{y}+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = CA_v\mathbf{x}(t)$. ■

Cont. proof of Lemma 1: The following statements should be proven:

1) C is of the form (4);
2) A_σ is of the form (4);
3) if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$, then first, K_σ is of the form (5);
4) second, $(\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$.

Below, we prove statements 1)–4) one by one. Throughout the proof, $T = [T_1 \quad T_2]^{-1}$ denotes the matrix defined in Step 1 of Algorithm 1.

1) Since the first k_2 rows of $\tilde{\mathscr{O}}_{M(n)}$ equal C_2 and the columns of T_1 span $\ker \tilde{\mathscr{O}}_{M(n)}$, we have that $C_2\tilde{\ }T^{-1} = [0 \quad C_{22}]$ with some $C_{22} \in \mathbb{R}^{n_2\times n_2}$, $0 < n_2 \leq n$.

2) We first show that $\ker \tilde{O}_{M(n)} = \ker \tilde{O}_{M(n+1)}$. Define $X_k = \ker \tilde{O}_{M(k)}$ for $k = 0, \ldots\ldots n+1$, where $\tilde{O}_{M(k)}$ is the observability matrix of $(\{\tilde{A}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}_2)$ up to k. Then, either $\tilde{C}_2 = 0$, in which case $\ker \tilde{O}_{M(n)} = \ker \tilde{O}_{M(n+1)}$ trivially holds, or $\dim(X_0) = \dim(\ker \tilde{C}_2) < n$. Notice that $X_{k-1} \supseteq X_k$ for $k = 1, \ldots\ldots n+1$, which together with that $\dim(X_0) < n$ implies that there exists an $l \in \{1, \ldots, n\}$, such that for all $k = l, \ldots, n$ $\dim(X_k) = \dim(X_{k+1})$ and $X_k = X_{k+1}$. By using that $X_n = X_{n+1}$ and that the rows of $\tilde{O}_{M(n)}$ and $\tilde{O}_{M(n)}\tilde{A}_\sigma$ are rows of $\tilde{O}_{M(n+1)}$, we obtain that X_n is A_σ-invariant for all $\sigma \in \Sigma$. Hence, considering that the columns of T_1 span X_n, we obtain that $\tilde{A}_\sigma T_1 = T_1 N \in X_n$ for a suitable matrix $N \in \mathbb{R}^{n_1\times n_1}$. Let

$$A_\sigma = T\tilde{A}_\sigma T^{-1} = \begin{bmatrix} A_{\sigma,11} & A_{\sigma,12} \\ A_{\sigma,21} & A_{\sigma,22} \end{bmatrix}$$

where $A_{\sigma,ij} \in \mathbb{R}^{n_i\times n_j}$ and notice that

$$T\tilde{A}_\sigma T^{-1} = \begin{bmatrix} T\tilde{A}_\sigma T_1 & \tilde{A}_\sigma T_2 \end{bmatrix} = \begin{bmatrix} TT_1N & \tilde{A}_\sigma T_2 \end{bmatrix}.$$

Then, $(TT_1N)^T = [N0]$ implies that $A_{\sigma,21} = 0$.

3) In order to see that the matrices $\{K_\sigma\}_{\sigma\in\Sigma}$ are as in (5), we prove a sequence of statements 3a–3b–3c–3d–3e and 3f, in the following, where 3f states that $\{K_\sigma\}_{\sigma\in\Sigma}$ satisfy (5).

a) $\mathbf{x}_2(t) \in \mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}}$.
b) $E[\mathbf{z}_w^{\mathbf{y}}(t)(\mathbf{z}_v^{\mathbf{e}}(t))^T] = 0$ for all $|v| < |w|, w, v \in \Sigma^+$.
c)
$$\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}} = \bigoplus_{\sigma\in\Sigma}\left(\mathcal{H}_{t+1,w\in\Sigma^+}^{\mathbf{z}_{w\sigma}^{\mathbf{y}_2}} \oplus \mathcal{H}_{t+1}^{\mathbf{z}_\sigma^{\mathbf{e}_2}}\right)$$
where $\oplus$ denotes the direct sum of orthogonal closed subspaces and $\mathcal{H}_{t+1,w\in\Sigma^+}^{\mathbf{z}_{w\sigma}^{\mathbf{y}_2}}$ denotes the Hilbert space generated by $\{\mathbf{z}_{w\sigma}^{\mathbf{y}_2}(t+1)\}_{w\in\Sigma^+}$, see also 3d in the following.
d) There exist $\{N_\sigma\}_{\sigma\in\Sigma} \in \mathbb{R}^{n_2\times k_2}$, $\mathbf{r} \in \bigoplus_{\sigma\in\Sigma}\mathcal{H}_{t+1,w\in\Sigma^+}^{\mathbf{z}_{w\sigma}^{\mathbf{y}_2}}$, such that $\mathbf{x}_2(t+1) = \mathbf{r} + \sum_{\sigma\in\Sigma} N_\sigma \mathbf{z}_\sigma^{\mathbf{e}_2}(t+1)$.
e) Let $K_\sigma = [K_{\sigma,21}K_{\sigma,22}]$, such that $K_{\sigma,21} \in \mathbb{R}^{n_2\times k_1}$, $K_{\sigma,22} \in \mathbb{R}^{n_2\times k_2}$. Then, for $\sigma\in\Sigma$
$$\begin{aligned}[K_{\sigma,21}K_{\sigma,22}]&E[\mathbf{z}_\sigma^{\mathbf{e}}(t+1)(\mathbf{z}_\sigma^{\mathbf{e}}(t+1))^T] \\ &= N_\sigma E[\mathbf{z}_\sigma^{\mathbf{e}_2}(t+1)(\mathbf{z}_\sigma^{\mathbf{e}}(t+1))^T].\end{aligned}$$
f) $K_{\sigma,21} = 0$ for all $\sigma \in \Sigma$.

Next, we prove 3a–3f, one-by-one.

Proof of 3a: By using (4), for any $v \in \Sigma^+$

$$CA_v = \begin{bmatrix} C_{11}(A_v)_{11} & N \\ 0 & C_{22}(A_v)_{22} \end{bmatrix} \quad (6)$$

where $(A_v)_{11} \in \mathbb{R}^{n_1 \times n_1}$ and $(A_v)_{22} \in \mathbb{R}^{n_2 \times n_2}$ are the upper and lower block diagonal submatrices of A_v, and $N \in \mathbb{R}^{k_1 \times n_2}$ is an appropriate matrix. From this, it is easy to see that by choosing an appropriate permutation matrix P, we have that

$$P\mathscr{O}_{M(n)} = \begin{bmatrix} N_1 & N_2 \\ 0 & O_{M(n)} \end{bmatrix}$$

where $O_{M(n)}$ is the observability matrix of $(\{A_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22})$ up to n and N_1, N_2 are appropriate matrices. Notice now that $\mathbf{x}(t) = \mathscr{O}^+_{M(n)} E_l[Z^{\mathbf{y}}_n(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w}]$, where $Z^{\mathbf{y}}_n(t) = [(\mathbf{z}^{\mathbf{y}+}_{v_1}(t))^T, \dots, (\mathbf{z}^{\mathbf{y}+}_{v_{M(n-1)}}(t))^T]^T$ is a vector of the future of $\mathbf{y}(t)$ w.r.t. the input (see Definition 1). Since $P^TP = I$, and hence, $(P\mathscr{O}_{M(n)})^+ = \mathscr{O}^+_{M(n)}P^T$, it follows that $\mathbf{x}(t) = (P\mathscr{O}^+_{M(n)})E_l[PZ^{\mathbf{y}}_n(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w}]$. Note that $PZ^{\mathbf{y}}_n(t) = [(Z^{\mathbf{y}_1}_n(t))^T (Z^{\mathbf{y}_2}_n(t))^T]^T$ where $Z^{\mathbf{y}_i}_n(t) = [(\mathbf{z}^{\mathbf{y}_i+}_{v_1}(t))^T, \dots, (\mathbf{z}^{\mathbf{y}_i+}_{v_{M(n-1)}}(t))^T]^T$, $i = 1, 2$, and thus, for $\mathbf{x}_2$, we have that $\mathbf{x}_2(t) = O^+_{M(n)}E_l[Z^{\mathbf{y}_2}_n(t)| \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}]$. Then, the GB–Granger noncausality condition $E_l[Z^{\mathbf{y}_2}_n(t)| \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}] = E_l[Z^{\mathbf{y}_2}_n(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}]$ implies that $\mathbf{x}_2(t) \in \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}$.

Proof of 3b: From [7, Lemma 14], it follows that $[\mathbf{y}^T, \mathbf{e}^T]^T$ is ZMWSSI. Hence, we can apply [7, Lemma 7] for $[\mathbf{y}^T, \mathbf{e}^T]^T$: Let $w = w_1 \dots w_k \in \Sigma^*$ and $v = v_1 \dots v_l \in \Sigma^*$, s.t. $|v| < |w|$. If $w_{k-i} \neq v_{l-i}$ for some $i = 0, \dots, l-1$, then [7, Lemma 7] implies that the covariance $E[\mathbf{z}^{\mathbf{y}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = 0$. If $w_{k-i} = v_{l-i}$ for all $i = 0, \dots, l-1$, then $E[\mathbf{z}^{\mathbf{y}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = p_{v_2\dots v_l}E[\mathbf{z}^{\mathbf{y}}_{w_1\dots w_{k-l-1}}(t)(\mathbf{z}^{\mathbf{e}}_{v_1}(t))^T] = p_vE[\mathbf{z}^{\mathbf{y}}_{w_1\dots w_{k-l-1}}(t)\mathbf{e}^T(t)] = 0$ where for the last equation, we used that from Definition 4, $E[\mathbf{z}^{\mathbf{y}}_{w_1\dots w_{k-l-1}}(t)\mathbf{e}^T(t)] = 0$.

Proof of 3c: Consider an innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$ and note that by the GB–Granger noncausality condition, the GB–innovation process of $\mathbf{y}_2$ is $\mathbf{e}_2$. Then, by [7, Lemma 16], we can decompose $\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}$ as in 3c.

Proof of 3d: From 3a, we have that $\mathbf{x}_2(t+1) \in \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t+1,w\in\Sigma^+}$. Then, by using 3c, $\mathbf{x}_2(t+1) = \mathbf{r} + \sum_{\sigma\in\Sigma} N_\sigma \mathbf{z}^{\mathbf{e}_2}_\sigma(t+1)$ for some $\mathbf{r} \in \bigoplus_{\sigma\in\Sigma} \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma}}_{t+1,w\in\Sigma^+}$ and $\{N_\sigma\}_{\sigma\in\Sigma} \in \mathbb{R}^{n_2\times k_2}$.

Proof of 3e: To shorten the expressions, define $k = t+1$. Notice that by using the block triangular form of $\{A_\sigma\}_{\sigma\in\Sigma}$, we obtain that $\mathbf{x}_2(k) = \sum_{\sigma\in\Sigma}\sum A_{\sigma,22}\mathbf{z}^{\mathbf{x}_2}_\sigma(k) + [K_{\sigma,21} K_{\sigma,22}]\mathbf{z}^{\mathbf{e}}_\sigma(k)$. From [7, Lemma 14], it follows that $[\mathbf{e}^T, \mathbf{y}^T, \mathbf{x}^T]^T$ is ZMWSSI, and hence, $[\mathbf{e}^T, \mathbf{x}_2^T]^T$ is also ZMWSSI w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. By applying [7, Lemma 7] for $[\mathbf{e}^T, \mathbf{x}_2^T]^T$, we have that if $\sigma \neq \sigma^*$, then $E[\mathbf{z}^{\mathbf{e}}_\sigma(k)(\mathbf{z}^{\mathbf{x}_2}_{\sigma^*}(k))^T] = E[\mathbf{z}^{\mathbf{e}}_\sigma(k)(\mathbf{z}^{\mathbf{e}}_{\sigma^*}(k))^T] = 0$. Also, by Definition 4, $E[\mathbf{z}^{\mathbf{e}}_{\sigma^*}(k)(\mathbf{z}^{\mathbf{x}}_\sigma(k))^T] = 0$ for $\sigma = \sigma^*$, and since for any $\sigma \in \Sigma$, $\mathbf{z}^{\mathbf{x}_2}_\sigma$ is formed by a component of $\mathbf{z}^{\mathbf{x}}_\sigma$, we have that for $\sigma = \sigma^*$, $E[\mathbf{z}^{\mathbf{e}}_{\sigma^*}(k)(\mathbf{z}^{\mathbf{x}_2}_\sigma(k))^T] = 0$. Hence, $E[\mathbf{x}_2(k)(\mathbf{z}^{\mathbf{e}}_\sigma(k))^T] = [K_{\sigma,21} K_{\sigma,22}]Q_\sigma$, where $Q_\sigma = E[\mathbf{z}^{\mathbf{e}}_\sigma(k)(\mathbf{z}^{\mathbf{e}}_\sigma(k))^T]$. By using 3d, we also obtain that $E[\mathbf{x}_2(k)(\mathbf{z}^{\mathbf{e}}_{\sigma^*}(k))^T] = E[\mathbf{r}\mathbf{z}^{\mathbf{e}}_{\sigma^*}(k))^T] + \sum_{\sigma\in\Sigma} N_\sigma E[\mathbf{z}^{\mathbf{e}_2}_\sigma(\mathbf{z}^{\mathbf{e}}_{\sigma^*}(k))^T]$. Notice that from 3b and from $\mathbf{r} \in \bigoplus_{\sigma\in\Sigma} \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma}}_{k,w\in\Sigma^+}$, we know that $E[\mathbf{r}\mathbf{z}^{\mathbf{e}}_{\sigma^*}(k))^T] = 0$. Hence, $E[\mathbf{x}_2(k) (\mathbf{z}^{\mathbf{e}}_\sigma(k))^T] = N_\sigma E[\mathbf{z}^{\mathbf{e}_2}_\sigma(k)(\mathbf{z}^{\mathbf{e}}_\sigma(k))^T]$, which equals $[K_{\sigma,21} K_{\sigma,22}]Q_\sigma$. Substituting $k = t+1$, we obtain 3e.

Proof of 3f: Since $\mathbf{e}_2$ is formed by the last k_2 components of $\mathbf{e}$, we have that $N_\sigma E[\mathbf{z}^{\mathbf{e}_2}_\sigma(t+1)(\mathbf{z}^{\mathbf{e}}_\sigma(t+1))^T] = [0 \quad N_\sigma]Q_\sigma$, and hence, $[0 \quad N_\sigma]Q_\sigma = [K_{\sigma,21} \quad K_{\sigma,22}]Q_\sigma$. By Assumption 1, Q_σ is positive definite, which implies that $[0 \quad N_\sigma] = [K_{\sigma,21} \quad K_{\sigma,22}]$, hence $K_{\sigma,21} = 0$.

4) Denote the state process of the minimal innovation GB–SS representation $\mathcal{G}$ of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$, that the output matrices of Algorithm 1 define, by $[\mathbf{x}_1^T, \mathbf{x}_2^T]^T$, where $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{n_2}$. To see that $\mathcal{G}_2 = (\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$, with state process $\mathbf{x}_2$, defines a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$, notice that from the GB–Granger noncausality condition, and from Definition 4, it follows that $\mathcal{G}_2$ is an innovation GB–SS representation. Assume indirectly that $\mathcal{G}_2$ is not minimal, i.e., that there exists a minimal innovation GB–SS representation $\tilde{\mathcal{G}}_2 = (\{\tilde{A}_{\sigma,22}, \tilde{K}_{\sigma,22}\}_{\sigma\in\Sigma}, \tilde{C}_{22}, I, \mathbf{e}_2)$ of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$ with state $\tilde{\mathbf{x}}_2 \in \mathbb{R}^{\tilde{n}_2}$, where $\tilde{n}_2 < n_2$.

From Lemma 3, it follows that $E_l[Z^{\mathbf{y}_2}_{n_2}(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}] = \tilde{O}_{M(n_2)}\tilde{\mathbf{x}}_2(t)$, where $\tilde{O}_{M(n_2)}$ is the observability matrix of $(\{\tilde{A}_{\sigma,22}\}_{\sigma\in\Sigma}, \tilde{C}_{22})$ up to n_2. Then, by defining $L = O^+_{M(n_2)}\tilde{O}_{M(n_2)}$, where $O_{M(n_2)}$ is the observability matrix of $(\{A_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22})$ up to n_2, we have that $\mathbf{x}_2 = L\tilde{\mathbf{x}}_2$.

By using L, we can transform $\mathcal{G}$ into an innovation GB–SS representation $\tilde{\mathcal{G}}$ of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ with state process $[\mathbf{x}_1^T, \tilde{\mathbf{x}}_2^T]^T$.

However, $\tilde{\mathcal{G}}$ has dimension $n_1 + \tilde{n}_2 < n_1 + n_2 = n$, which is a contradiction since n is the dimension of a minimal innovation GB–SS representation. ■

Proof of Theorem 1: The sufficiency part of the proof follows Lemma 1.

To prove the necessity part, let $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ be a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$, such that (3) holds and that $\mathcal{G}_2 = (\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$. To prove that $\mathbf{y}_1$ does not GB–Granger causes $\mathbf{y}_2$, we need to see that

$$E_l[\mathbf{z}^{\mathbf{y}_2+}_v(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}] = E_l[\mathbf{z}^{\mathbf{y}_2+}_v(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}] \quad (7)$$

for all $v \in \Sigma^*$. For $v = \epsilon$, (7) directly follows form that $\mathbf{e}_2$ is the GB–innovation process of $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$.

By (3), the matrices $\{A_\sigma\}_{\sigma\in\Sigma}$ and C are block triangular, hence CA_v is as in (6). It then follows from Lemma 3 that

$$E_l[\mathbf{z}^{\mathbf{y}_2+}_v(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}] = C_{22}(A_v)_{22}\mathbf{x}_2(t). \quad (8)$$

By projecting both side of (8) onto $\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}$, and by using that $\mathbf{x}_2(t) \in \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}$ (see [7, Th. 5]), we get that $E_l[\mathbf{z}^{\mathbf{y}_2+}_v(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}] = C_{22}(A_v)_{22}\mathbf{x}_2(t)$. By considering (8), the latter implies (7), i.e., that there is no GB–Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$. ■

Proof of Lemma 2: Notice that $\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+} \subseteq \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}$, $\mathbf{v}(t) = \mathbf{y}_2(t) - E_l[\mathbf{y}_2(t) \mid \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}]$, and $\mathbf{e}_2(t) = \mathbf{y}_2(t) - E_l[\mathbf{y}_2(t) \mid \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}]$, hence by the minimal distance property of orthogonal projections, $\mathbf{e}_2(t) = \mathbf{v}(t)$ if and only if $E[(\mathbf{v}(t))_i^2] = E[(\mathbf{e}_2(t))_i^2]$, $i = 1, \dots, k_2$, where $(\mathbf{v}(t))_i$ and $(\mathbf{e}_2(t))_i$ are the ith entry of $\mathbf{v}(t)$ and $\mathbf{e}_2(t)$, respectively. Note that $\mathbf{e}_2(t) = \sum_{\sigma\in\Sigma}\alpha_\sigma\mathbf{e}_2(t)\mathbf{u}_\sigma(t)$, $\mathbf{v}(t) = \sum_{\sigma\in\Sigma}\alpha_\sigma\mathbf{v}(t)\mathbf{u}_\sigma(t)$. As $\mathbf{e}_2, \mathbf{v}$ are ZMWSSI processes, $E[\mathbf{e}_2(t)\mathbf{e}_2^T(t)\mathbf{u}_\sigma(t)\mathbf{u}_{\sigma'}(t)] = 0$, $E[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}_\sigma(t)\mathbf{u}_{\sigma'}(t)] = 0$ for all $\sigma \neq \sigma', \sigma, \sigma' \in \Sigma$. Hence, $E[(\mathbf{v}(t))_i^2] = \sum_{\sigma\in\Sigma}\alpha^2_\sigma(t)(\bar{Q}_\sigma)_{i,i}$, $E[(\mathbf{e}_2(t))_i^2] = \sum_{\sigma\in\Sigma}\alpha^2_\sigma(t)(Q_\sigma)_{k_1+i,k_1+i}$. ■

References

[1] M. Jozsa, M. Petreczky, and M. K. Camlibel, “Relationship between Granger noncausality and network graph of state-space representations,” *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 912–927, Mar. 2019.

[2] M. Jozsa, “Relationship between Granger non-causality and network graphs of state-space representations,” Ph.D. dissertation, Dept. Math. Natural Sci., Univ. Groningen, Groningen, The Netherlands, 2019.

[3] C. W. J. Granger, "Economic processes involving feedback," *Inf. Control*, vol. 6, no. 1, pp. 28–48, 1963.

[4] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[5] P. E. Caines and C. Chan, "Feedback between stationary stochastic processes," *IEEE Trans. Autom. Control*, vol. AC-20, no. 4, pp. 498–508, Aug. 1975.

[6] P.-O. Amblard and O. J. Michel, "On directed information theory and Granger causality graphs," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 7–16, 2011.

[7] M. Petreczky and V. René, "Realization theory for a class of stochastic bilinear systems," *IEEE Trans. Autom. Control*, vol. 63, no. 1, pp. 69–84, Jan. 2017.

[8] H. Chen and J. M. Maciejowski, "A new subspace identification method for bilinear systems," Univ. Cambridge, Cambridge, U.K., Tech. Rep. CUED/FINFENG/TR.337, 2001.

[9] P. D'Alessandro, A. Isidori, and A. Ruberti, "Realization and structure theory of bilinear dynamical systems," *SIAM J. Control*, vol. 12, no. 3, pp. 517–535, 1974.

[10] W. Favoreel, B. D. Moor, and P. V. Overschee, "Subspace identification of bilinear systems subject to white inputs," *IEEE Trans. Autom. Control*, vol. 44, no. 6, pp. 1157–1165, Jun. 1999.

[11] A. Isidori, "Direct construction of minimal bilinear realizations from nonlinear input-output maps," *IEEE Trans. Autom. Control*, vol. AC-18, no. 6, pp. 626–631, Dec. 1973.

[12] U. Desai, "Realization of bilinear stochastic systems," *IEEE Trans. Autom. Control*, vol. AC-31, no. 2, pp. 189–192, Feb. 1986.

[13] M. Jozsa, M. Petreczky, and M. K. Camlibel, "Towards realization theory of interconnected linear stochastic systems," in *Proc. 22nd Int. Symp. Math. Theory Netw. Syst.*, 2016, pp. 120–122.

[14] P. E. Caines, R. Deardon, and H. P. Wynn, "Bayes nets of time series: Stochastic realizations and projections," in *Proc. Optimal Des. Related Areas Optim. Statist.* (Springer Optimization and Its Applications), L. Pronzato and A. Zhigljavsky, Eds. New York, N, USA: Springer, 2009, vol. 28, pp. 155–166.

[15] M. Jozsa, M. Petreczky, and M. K. Camlibel, "Causality based graph structure of stochastic linear state-space representations," in *Proc. 56th IEEE Conf. Decis. Control*, Melbourne, VIC, USA, 2017, pp. 2442–2447.

[16] M. Jozsa, M. Petreczky, and M. K. Camlibel, "Relationship between causality of stochastic processes and zero blocks of their joint innovation transfer matrices," in *Proc. 20th IFAC World Congr.*, 2017, pp. 4954–4959.

[17] P. E. Caines and H. P. Wynn, "An algebraic framework for Bayes nets of time series," in *Modeling, Estimation and Control* (Lecture Notes in Control and Information Sciences), A. Chiuso, S. Pinzoni, and A. Ferrante, Eds. Berlin, Germany: Springer, 2007, vol. 364, pp. 45–57.

[18] P. E. Caines, "Weak and strong feedback free processes," *IEEE Trans. Autom. Control*, vol. AC-21, no. 5, pp. 737–739, Oct. 1976.

[19] L. Barnett and A. K. Seth, "Granger causality for state space models," *Phys. Rev. E*, vol. 91, no. 4, pp. 737–739, 2015.

[20] J. Pearl, *Causality: Models, Reasoning and Inference*, 1st ed. Cambridge, MA, USA: Cambridge Univ. Press, 2000.

[21] I. Gikhman and A. Skorokhod, *The Theory of Stochastic Processes II* (Classics in Mathematics). Berlin, Germany: Springer , 2004, vol. 1.