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

Jozsa, Monika

Chapter 6

Causality and network graph in general bilinear state-space representations

In this chapter, we would like to derive similar results to the results presented in Chapter 2 for general bilinear state-space (GB–SS) representations. For background material on GB–SS representation, see Section 1.3. The motivation for this chapter is that in most of the fields, where Granger causality is applied (e.g., econometrics, systems biology, neuroscience), nonlinear models are more desirable due to their ability to describe richer variety of phenomena. Since Granger causality is based on linear relations, it is not suitable when the processes relate to each other in a nonlinear way, e.g., for processes generated by nonlinear dynamical systems. As a first step towards nonlinear systems, a natural choice for the class of nonlinear systems is the class of bilinear systems. This class includes e.g., vector autoregressive moving-average (VARMA), switched linear, and, in case of general bilinear state-space (abbreviated by GB–SS) representations, jump Markov linear models. The reason for choosing bilinear systems is that they can produce richer phenomena than linear systems, yet many of the analytical tools for linear systems are suitable to analyze them. In particular, stochastic realization theory exists for GB–SS representations (Petreczky and René, 2017). This theory serves as a basis for the technicalities of the main results of this chapter.

In order to achieve the objectives of this chapter, we will

1) choose a suitable definition of causality based on the statistical properties of the input-output processes that are represented by bilinear state-space representations;

2) prove an equivalence between the defined causality and properties of the inner structure of bilinear state-space representations.

In order to formalize causality for the outputs of GB–SS representations, we will introduce the concept of GB-Granger causality. GB-Granger causality is an extension of Granger causality and it coincides with Granger causality when applied to outputs of stochastic LTI–SS representations.

In the main result of this chapter, 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 the absence of a GB-Granger causality 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 of which generates $\mathbf{y}_1$ with input $\mathbf{u}$, and another one which generates $\mathbf{y}_2$ with input $\mathbf{u}$, where the former sends no information to the latter. That is, GB-Granger causality, although it is defined purely in terms of the input-output processes, can be equivalently interpreted as a property of the internal structure of a bilinear state-space representation of these processes. The results of this chapter are extensions of results in Chapter 2 on the relationship between Granger causality and internal structure of LTI–SS state-space representations towards GB–SS representations.

The chapter is organizes as follows: To introduce GB-Granger causality and its characterization, we first recall some results from Chapter 2. Then, we define GB-Granger causality and explain its meaning in GB–SS representations. Throughout this chapter, we assume that $\mathbf{y}$ is a ZMWSSI process with respect to an admissible set of processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and $\mathbf{y}$ admits a partitioning $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$, such that $\mathbf{y}_i \in \mathbb{R}^{p_i}$ for $p_i > 0$, $i = 1, 2$.

6.1 Granger causality in LTI–SS representations

Before introducing the concept of GB-Granger causality, we recall the definition of Granger causality from Chapter 2 and its meaning in LTI–SS representations. Informally, $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$, if for all $k \geqslant 0$, the best k-step linear prediction of $\mathbf{y}_2$ based on the past values of $\mathbf{y}_2$ is the same as that of based on the past values of $\mathbf{y}$. Recall that $\mathcal{H}_{t-}^{\mathbf{z}}$ denotes the Hilbert space generated by the elements of the past $\{\mathbf{z}(t-k)\}_{k=1}^{\infty}$ of $\mathbf{z}$. Then Granger causality is defined as follows, see also Definition 2.3:

Definition 6.1 (Granger causality)**.** Consider a zero-mean square-integrable, weakly 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 $t, k \in \mathbb{Z}$, $k \geqslant 0$

$$E_l[\mathbf{y}_2(t+k)|\mathcal{H}_{t-}^{\mathbf{y}_2}] = E_l[\mathbf{y}_2(t+k)|\mathcal{H}_{t-}^{\mathbf{y}}].$$

Otherwise, we say that $\mathbf{y}_1$ *Granger causes* $\mathbf{y}_2$.

In Section 1.3, it was shown that a GB–SS representation defines a linear time-invariant state-space (LTI–SS) representation if $\Sigma = \{1\}$ and $\mathbf{u}_1(t) \equiv 1$. Accordingly, the innovation process of $\mathbf{y}$ in LTI–SS representations is $\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t)|\mathcal{H}_{t-}^{\mathbf{y}}]$

and an innovation LTI–SS representation of $\mathbf{y}$ is in the form of

$$\begin{aligned}\mathbf{x}(t+1) &= A\mathbf{x}(t) + K\mathbf{e}(t)\\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{e}.\end{aligned}$$

Recall that the dimension of an LTI–SS representation is the dimension of the state-process. Furthermore, an LTI–SS representation is minimal if it has minimal dimension among all the LTI–SS representations with the same output process. Granger non-causality among the components of an output of an LTI–SS representation can be characterized by the properties of a minimal innovation LTI–SS representation, see Theorem 2.5. In order to help the understanding and to appreciate the differences and similarities between Theorem 2.5 and Theorem 6.5 (see next section), we present the following statement which is the reformulation of the statement (i) $\iff$ (ii) in Theorem 2.5.

Theorem 6.2. *Consider an LTI–SS representation of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ where $\mathbf{y}_i \in \mathbb{R}^{p_i}$, for some $p_i > 0$, $i = 1, 2$. Then $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ if and only if $\mathbf{y}$ has a minimal innovation LTI–SS representation*

$$\begin{aligned}\begin{bmatrix}\mathbf{x}_1(t+1)\\ \mathbf{x}_2(t+1)\end{bmatrix} &= \begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1(t)\\ \mathbf{x}_2(t)\end{bmatrix} + \begin{bmatrix}K_{11} & K_{12}\\ 0 & K_{22}\end{bmatrix}\begin{bmatrix}\mathbf{e}_1(t)\\ \mathbf{e}_2(t)\end{bmatrix}\\ \begin{bmatrix}\mathbf{y}_1(t)\\ \mathbf{y}_2(t)\end{bmatrix} &= \begin{bmatrix}C_{11} & C_{12}\\ 0 & C_{22}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1(t)\\ \mathbf{x}_2(t)\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1(t)\\ \mathbf{e}_2(t)\end{bmatrix}\end{aligned} \tag{6.1}$$

where $A_{ij} \in \mathbb{R}^{n_i \times n_j}, K_{ij} \in \mathbb{R}^{n_i \times p_j}, C_{ij} \in \mathbb{R}^{p_i \times n_j}$, $i, j = 1, 2$ for some $n_1 \geqslant 0$, $n_2 > 0$ and $(A_{22}, K_{22}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation LTI–SS representation of $\mathbf{y}_2$.

The LTI–SS representation (6.1) can be viewed as a cascade interconnection of two subsystems, see Figure 6.1.

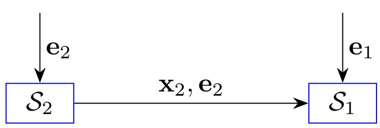

Figure 6.1: Network graph of the LTI–SS representation (6.1).

Consider the system (6.1) and define the dynamical systems $\mathcal{S}_1$ and $\mathcal{S}_2$ below.

$$\mathcal{S}_1 \begin{cases} \mathbf{x}_1(t+1) = \sum_{i=1}^{2}(A_{1i}\mathbf{x}_i(t) + K_{1i}\mathbf{e}_i(t)) \\ \mathbf{y}_1(t) = \sum_{i=1}^{2} C_{1i}\mathbf{x}_i(t)) + \mathbf{e}_2(t) \end{cases}$$

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

Notice that subsystem $\mathcal{S}_2$ sends its state $\mathbf{x}_2$ and noise $\mathbf{e}_2$ to subsystem $\mathcal{S}_1$ as an external input while $\mathcal{S}_1$ does not send information to $\mathcal{S}_2$. Accordingly, the network graph of the representation (6.1) is the two-node star graph with $\mathcal{S}_2$ being the root node and $\mathcal{S}_1$ being the leave. Hence, for this simple case, Theorem 6.2 shows an equivalence between the network graph and the statistical properties of the observed process. In the next section, we extend this result to GB–SS representations and GB-Granger causality.

6.2 GB–Granger causality in GB–SS representations

As we mentioned earlier, LTI–SS representations form a special subclass of GB–SS representations. Therefore, one could naturally ask what Granger causality means in GB–SS representations. However, Granger causality is based on approximating the output process by the linear combination of its own past values. Note that innovation LTI–SS representations give rise to a linear operator from the past values of the innovation process (and hence of the past outputs) to future outputs. Hence, LTI–SS representations can be related to the best linear prediction of future outputs based on past outputs, which allow us to relate Ganger causality with properties of LTI–SS representations, as it is stated in Theorem 6.2.

Unfortunately, for GB–SS representations the approach above no longer works. In fact, an innovation GB–SS representation defines a relationship between the elements of the Hilbert space generated by the products of the past values of $\mathbf{y}$ and $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$, to the elements of the Hilbert space generated by the products of future values of $\mathbf{y}$ and $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. More precisely, consider an innovation GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$. Then for all $v \in \Sigma^*$, $E[\mathbf{z}_v^{\mathbf{y}+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = p_v CA_v\mathbf{x}(t)$, where $\mathbf{x}(t) \in \mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}$ (see Lemma 6.9 in Appendix 6.A). That is, a GB–SS representation says very little about the best linear prediction of the future outputs based on the past outputs. For this reason, there is little hope of deriving counterparts of Theorem 6.2 for GB–SS representations, while using the classical definition of Granger causality. However, the discussion above shows us a way out of this problem. Namely, it follows that a GB–SS representation says something about the best linear prediction of the future of the output with respect to the inputs, denoted by $\mathbf{z}_v^{\mathbf{y}+}(t)$ in Definition 1.14, based on the past of the output with respect to the inputs, denoted by $\mathbf{z}_w^{\mathbf{y}}(t)$ in Definition 1.13. In fact, through the state process, it reveals a linear relation between them. Moreover, if $\Sigma = 1$ and $\mathbf{u}_1 = 1$, then $\mathbf{z}_v^{\mathbf{y}+}(t) = \mathbf{y}(t+|v|)$ represents future outputs, and $\mathbf{z}_w^{\mathbf{y}}(t) = \mathbf{y}(t-|w|)$ represents past outputs. This then opens up the possibility of extending Granger causality by using the process $\mathbf{z}_v^{\mathbf{y}+}(t)$ rather than $\mathbf{y}(t+|v|)$ and $\mathbf{z}_v^{\mathbf{y}-}(t)$ rather than $\mathbf{y}(t-|v|)$. We define the following extension of Granger causality.

Definition 6.3 (GB–Granger causality)**.** Consider the processes $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ where $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is admissible and $\mathbf{y}$ is ZMWSSI with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and is decomposed such that $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$. We say that $\mathbf{y}_1$ *does not GB–Granger cause* $\mathbf{y}_2$ with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ if for all $v \in \Sigma^*$ and $t \in \mathbb{Z}$

$$E_l[\mathbf{z}_v^{\mathbf{y}_2+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = E_l[\mathbf{z}_v^{\mathbf{y}_2+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}}]. \tag{6.2}$$

Otherwise, we say that $\mathbf{y}_1$ *GB–Granger causes* $\mathbf{y}_2$ *with respect to* $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$.

Notice that the Hilbert space $\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}$ is generated by the past $\{\mathbf{z}_w^{\mathbf{y}}\}_{w\in\Sigma^+}$ of $\mathbf{y}$ with respect to the admissible set of processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. Thus, the projections in (6.2) are based on the past of $\mathbf{y}$ and $\mathbf{y}_2$ with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. Then, informally we can say that $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$, if the best linear prediction of the future of $\mathbf{y}_2$ with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ based on the past of $\mathbf{y}$ with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ is the same as that of based on the past of $\mathbf{y}_2$ with respect to $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$.

Definition 6.3 is a generalization of Granger causality in the sense that if $\Sigma = \{1\}$ and $\mathbf{u}_1(t) \equiv 1$ then Definition 6.3 coincides with Definition 2.3. Furthermore, if $|v| = k$ then, using that there exist $\{\alpha_\sigma\}_{\sigma\in\Sigma}$ such that $\sum_{\sigma\in\Sigma} \alpha_\sigma \mathbf{u}_\sigma(t) \equiv 1$ (see Definition 1.15), (6.2) implies that

$$E_l[\mathbf{y}_2(t+k)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = E_l[\mathbf{y}_2(t+k)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}}]. \tag{6.3}$$

Although, (6.3) is more intuitive as an extension of Granger causality, we use (6.2) in Definition 6.3 for technical reasons.

Next, we will characterize the relationship between GB–Granger causality and the structure of GB–SS representations in a similar manner as it was done in Theorem 6.2 for Granger causality and LTI–SS representations. That is, we will show that GB–Granger non-causality is equivalent to the existence of a minimal innovation GB–SS representation with block triangular system matrices. For this, we first define the class of GB–SS representations in question.

Definition 6.4. An innovation GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ of the processes $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$ where $\mathbf{e}_i \in \mathbb{R}^{r_i}$, $i = 1,2$ is called *innovation*

GB–SS representation in block triangular form if for all $\sigma \in \Sigma$

$$A_\sigma = \begin{bmatrix} A_{\sigma,11} & A_{\sigma,12} \\ 0 & A_{\sigma,22} \end{bmatrix}, \quad K_\sigma = \begin{bmatrix} K_{\sigma,11} & K_{\sigma,12} \\ 0 & K_{\sigma,22} \end{bmatrix}, \quad C = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix} \tag{6.4}$$

where $A_{\sigma,ij} \in \mathbb{R}^{n_i \times n_j}$, $K_{\sigma,ij} \in \mathbb{R}^{n_i \times p_j}$, $C_{ij} \in \mathbb{R}^{p_i, n_j}$ for some $n_1 \geqslant 0$, $n_2 > 0$. If, in addition for all $\sigma \in \Sigma$, $(\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$ then $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ is called *an innovation GB–SS representation of* $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ *in causal block triangular form.*

Now we are ready to state the main results of the chapter.

Theorem 6.5. *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* $\mathbf{y}$ *with respect to* $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ *where* $\mathbf{e}_i \in \mathbb{R}^{p_i}, i = 1, 2$. *Then,* $\mathbf{y}_1$ *does not GB–Granger cause* $\mathbf{y}_2$ *with respect to* $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ *if and only if there exists a minimal innovation GB–SS representation of* $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ *in causal block triangular form.*

The proof can be found in Appendix 6.A.

An innovation GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ of the processes $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ in causal block triangular form can be viewed as a cascade interconnection of two subsystems in a similar manner as the LTI–SS representation (2.1) was viewed in the previous section, see Figure 2.1. Define the subsystems

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

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

Notice that subsystem $\mathcal{S}_2$ sends its state $\mathbf{x}_2$ and noise $\mathbf{e}_2$ to subsystem $\mathcal{S}_1$ as an external input while $\mathcal{S}_1$ does not send information to $\mathcal{S}_2$. Accordingly, the network graph of the GB–SS representation $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ is as in Figure 6.2. Theorem 6.5 shows an equivalence between the network graph of an innovation GB–SS representation and statistical properties of the observed processes $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and $\mathbf{y}$.

The necessity part of the proof of Theorem 6.5 is constructive and it is based on an algorithm which calculates an innovation GB–SS representation described in Theorem 6.5. We present this algorithm in Algorithm 13 below, along with the statement of its correctness. Algorithm 13 is an extended form of Algorithm 3 in Chapter 1.

Next, we present a number of lemmas that show that Algorithm 13 calculates the GB–SS representation in Theorem 6.5. Assume that the processes $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} =$

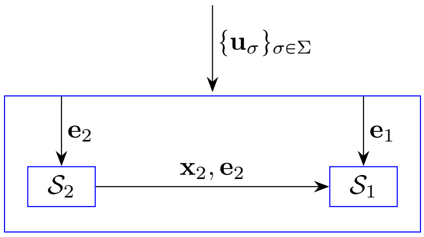

Figure 6.2: Network graph of a GB–SS representation $(\{A_{\sigma,}, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$ in block triangular form.

Algorithm 13 Minimal innovation GB–SS representation in causal block triangular form

Input $\{\Psi_w^{\mathbf{y}}\}_{\{w\in\Sigma*,|w|\leqslant N\}}$ and $\{E[\mathbf{z}_\sigma^{\mathbf{y}}(t)(\mathbf{z}_\sigma^{\mathbf{y}}(t))^T]\}_{\sigma\in\Sigma}$: Covariance sequence of $\mathbf{y}$ and its past and variances of $\mathbf{z}_\sigma^{\mathbf{y}}$

Output $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C)$: System matrices of (6.4)

Step 1 Apply Algorithm 3 with input $\{\Psi_w^{\mathbf{y}}\}_{\{w\in\Sigma*,|w|\leqslant N\}}$, $\{E[\mathbf{z}_\sigma^{\mathbf{y}}(t)(\mathbf{z}_\sigma^{\mathbf{y}}(t))^T]\}_{\sigma\in\Sigma}$ and denote its output by $(\{\tilde{A}_\sigma, \tilde{K}_\sigma, Q_\sigma\}_{\sigma\in\Sigma}, \tilde{C})$.
Step 2 Define the sub-matrix consisting of the last p_2 rows of $\tilde{C}$ by $\tilde{C}_2 \in \mathbb{R}^{p_2\times n}$ 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 3 Define the matrices $A_\sigma = T\tilde{A}_\sigma T^{-1}$, $K_\sigma = T\tilde{K}_\sigma$ for $\sigma \in \Sigma$ and $C = \tilde{C}T^{-1}$.

$[\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$ have a GB–SS representation with dimension n and that $N \geqslant n$. Then we have the following statement on the output $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C)$ of Algorithm 13 with input $\{\Psi_w^{\mathbf{y}}\}_{\{w\in\Sigma*,|w|\leqslant N\}}$ and $\{E[\mathbf{z}_\sigma^{\mathbf{y}}(t)(\mathbf{z}_\sigma^{\mathbf{y}}(t))^T]\}_{\sigma\in\Sigma}$:

Lemma 6.6. *Let the GB–innovation process of* $\mathbf{y}$ *be* $\mathbf{e}$. *Then, the tuple* $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ *is a minimal innovation GB–SS representation of* $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$.

Furthermore, we have the following statements about the matrices $\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}$ and C:

Lemma 6.7. *The matrices* $\{A_\sigma\}_{\sigma\in\Sigma}$ *and* C *are in the form*

$$A_\sigma = \begin{bmatrix} A_{\sigma,11} & A_{\sigma,12} \\ 0 & A_{\sigma,22} \end{bmatrix} \quad C = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix} \tag{6.5}$$

where $A_{\sigma,ij} \in \mathbb{R}^{n_i,n_j}$, $C_{ij} \in \mathbb{R}^{p_i,n_j}$, $i, j = 1, 2$ *for some* $n_1 \geqslant 0$, $n_2 > 0$. *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*

$$K_\sigma = \begin{bmatrix} K_{\sigma,11} & K_{\sigma,12} \\ 0 & K_{\sigma,22} \end{bmatrix}, \tag{6.6}$$

where $K_{\sigma,ij} \in \mathbb{R}^{n_i \times p_j}, i, j \in \{1,2\}$ *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 proofs of Lemmas 6.6 and 6.7 can be found in Appendix 6.A.

From Lemmas 6.6 and 6.7, it follows that if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$ then Algorithm 13 calculates the system matrices of the GB–SS representation described in Theorem 6.5. Hence, Algorithm 13 enables the calculation of a minimal innovation GB–SS representation in causal block triangular form that characterizes GB–Granger non-causality. It also provides a constructive proof of the necessity part of Theorem 6.5.

Remark 6.8. From Lemmas 1.26 and 6.7, it follows that the output matrices of Algorithms 3 and 13 define isomorphic GB–SS representations. By Remark 1.24, it also follows that Algorithm 13 can be modified to calculate a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ in causal block triangular form from any GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$, provided that $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$.

6.3 Conclusions

The results of this chapter show that GB–Granger causality among the components of processes that are outputs of GB–SS representations can be characterized by structural properties of GB–SS representations. More precisely, it is shown that GB–Granger non-causality among the components of an output process is equivalent to the existence of a GB–SS representation in causal block triangular form. Notice that GB–Granger causality is an extension of the classical Granger causality and innovation GB–SS representations in causal block triangular form are extensions of Kalman representations in causal block triangular form. Hence, the results of this chapter extend the correspondence between structural properties of LTI-SS representations and Granger causality of their outputs to GB–SS representations.

6.A Proofs

Proof of Lemma 6.6. To prove the statement, we show that the output matrices of Algorithm 3 and Algorithm 13 with input $\{\Psi^{\mathbf{y}}_w\}_{\{w\in\Sigma*,|w|\leqslant N\}}$ and $\{E[\mathbf{z}^{\mathbf{y}}_\sigma(t)(\mathbf{z}^{\mathbf{y}}_\sigma(t))^T]\}_{\sigma\in\Sigma}$ define system matrices of isomorphic GB–SS representations. Denote the output matrices of Algorithm 13 with input $\{\Psi^{\mathbf{y}}_w\}_{\{w\in\Sigma*,|w|\leqslant N\}}$ and $\{E[\mathbf{z}^{\mathbf{y}}_\sigma(t)(\mathbf{z}^{\mathbf{y}}_\sigma(t))^T]\}_{\sigma\in\Sigma}$ by $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C)$. Likewise, denote the output matrices of Algorithm 3 with input $\{\Psi^{\mathbf{y}}_w\}_{\{w\in\Sigma*,|w|\leqslant N\}}$ and $\{E[\mathbf{z}^{\mathbf{y}}_\sigma(t)(\mathbf{z}^{\mathbf{y}}_\sigma(t))^T]\}_{\sigma\in\Sigma}$ by $(\{\tilde{A}_\sigma, \tilde{K}_\sigma, \tilde{Q}_\sigma\}_{\sigma\in\Sigma}, \tilde{C})$. From (Petreczky and René, 2017, Theorem 3) we know that $(\{\tilde{A}_\sigma, \tilde{K}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}, I, \mathbf{e})$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$ where $\mathbf{e}$ denotes the GB–innovation process of $\mathbf{y}$ w.r.t. the input $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. By Step 3 of Algorithm 13, we also know that $A_\sigma = T\tilde{A}_\sigma T^{-1}$, $K_\sigma = T\tilde{K}_\sigma$ for $\sigma \in \Sigma$ and $C = \tilde{C}T^{-1}$ with a non-singular T matrix. Notice that T defines a linear transformation between the tuple $(\{\tilde{A}_\sigma, \tilde{K}_\sigma\}_{\sigma\in\Sigma}, \tilde{C})$ and $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C)$ that does not depend on the input or output processes. Denote the state process of $(\{\tilde{A}_\sigma, \tilde{K}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}, I, \mathbf{e})$ by $\tilde{\mathbf{x}}$. Since $(\{\tilde{A}_\sigma, \tilde{K}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}, I, \mathbf{e})$ is a minimal innovation GB–SS representation, $(\{T\tilde{A}_\sigma T^{-1}, T\tilde{K}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}T^{-1}, I, \mathbf{e})$ also defines an innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ with state process $T\tilde{\mathbf{x}}$. Since T is non-singular, it implies that the Kalman representation

$$(\{T\tilde{A}_\sigma T^{-1}, T\tilde{K}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}T^{-1}, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}),$$

or equivalently $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e}, \{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$, is also minimal, which completes the proof. □

We need the following auxiliary result in order to prove Lemma 6.7.

Lemma 6.9. *Let $(\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ be an innovation GB–SS representation of the processes $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ with state process $\mathbf{x}$. Then the equation $E_l[\mathbf{z}^{\mathbf{y}+}_v(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}] = p_v C A_v \mathbf{x}(t)$ holds.*

Proof. Recall that $\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}$ is the Hilbert space generated by the past $\{\mathbf{z}^{\mathbf{y}}_w\}_{w\in\Sigma^+}$ of $\mathbf{y}$ with respect to the input $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. From equation (38) in (Petreczky and René, 2017) we know that for $\sigma \in \Sigma, v \in \Sigma^+, w \in \Sigma^*$

$$E[\mathbf{z}^{\mathbf{y}+}_v(t)(\mathbf{z}^{\mathbf{y}}_{\sigma w}(t))^T] = E[\mathbf{y}(t)(\mathbf{z}^{\mathbf{y}}_{\sigma wv}(t))^T] = p_{wv} C A_v A_w G_\sigma,$$

with $G_\sigma = A_\sigma P_\sigma C^T + K_\sigma Q_\sigma$ for $\sigma \in \Sigma$ where $P_\sigma = E[\mathbf{x}(t)(\mathbf{x}(t))^T \mathbf{u}^2_\sigma(t)]$. In addition, from (Petreczky and René, 2017, Lemma 12) we know that $E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{y}}_{\sigma w}(t))^T] = p_w A_w G_\sigma$ for all $\sigma \in \Sigma, w \in \Sigma^*$. Hence, $E[\mathbf{z}^{\mathbf{y}+}_v(t)(\mathbf{z}^{\mathbf{y}}_{\sigma w}(t))^T] = p_v C A_v E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{y}}_{\sigma w}(t))^T]$ for any $v, \sigma w \in \Sigma^+$. Considering that $\mathbf{x}(t) \in \mathcal{H}^{\mathbf{z}^{\mathbf{e}}_w}_{t,w}$, see (1.8), and that $\mathcal{H}^{\mathbf{z}^{\mathbf{e}}_w}_{t,w} \subseteq \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w}$, see Definition 1.18, implies that $E_l[\mathbf{z}_v^{\mathbf{y}+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = p_v C A_v \mathbf{x}(t)$. □

Proof of Lemma 6.7. To help the reader, we recall the Steps of Algorithm 13: Step 1 of Algorithm 13 applies Algorithm 3 and denotes its output by $(\{\tilde{A}_\sigma, \tilde{K}_\sigma, Q_\sigma\}_{\sigma\in\Sigma}, \tilde{C})$.

Step 2 of Algorithm 13 goes as follows: denote the sub-matrix consisting of the last p_2 rows of $\tilde{C}$ by $\tilde{C}_2 \in \mathbb{R}^{p_2\times n}$ and take the observability matrix $\tilde{\mathscr{O}}_{M(n)}$ of $(\{\tilde{A}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}_2)$ up to n. If $\tilde{\mathscr{O}}_{M(n)}$ is full column rank then define the matrix $T = I$. If $\tilde{\mathscr{O}}_{M(n)}$ is not full column rank, then denote its rank by n_2 and let $n_1 = n - n_2$. Furthermore, 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 $\tilde{\mathscr{O}}_{M(n)}$.

Step 3 of Algorithm 13 goes as follows: Define the matrices $A_\sigma := T\tilde{A}_\sigma T^{-1}$, $K_\sigma := T\tilde{K}_\sigma$ for $\sigma \in \Sigma$ and $C := \tilde{C}T^{-1}$.

The following statements should be proven:

1) C is of the form (6.5),

2) A_σ is of the form (6.5),

3) if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$, then K_σ is of the form (6.6), and

4) if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$ then $(\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22}, I, \mathbf{e}_2)$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$.

Below, we prove the statements 1–4 one by one.

1) If $T = I$ then with $n_1 = 0$ and $n_2 = n$ the matrices $(\{\tilde{A}_\sigma\}_{\sigma\in\Sigma}, \tilde{C})$ are in the form of (6.5). Since the first p_2 rows of $\tilde{\mathscr{O}}_{M(n)}$ equal C_2 and T_1 spans the kernel of $\tilde{\mathscr{O}}_{M(n)}$, we have that $C_2T^{-1} = \begin{bmatrix} 0 & C_{22} \end{bmatrix}$ with some $C_{22} \in \mathbb{R}^{n_2\times n_2}$ full column rank matrix.

2) For $k = 0, \ldots, n+1$, let us denote the observability matrix of $(\{\tilde{A}_\sigma\}_{\sigma\in\Sigma}, \tilde{C}_2)$ up to k by $\tilde{O}_{M(k)}$. We will first show that $\ker \tilde{O}_{M(n)} = \ker \tilde{O}_{M(n+1)}$. Define $X_k := \ker \tilde{O}_{M(k)}$ for $k = 0, \ldots, n+1$. Then either $\tilde{C}_2 = 0$ or $\dim(X_0) = \dim(\ker \tilde{C}_2) < n$. If $\ker \tilde{C}_2 = 0$, then for any $k = 0, \ldots, n+1$, all the entries of $\tilde{O}_{M(k)} = 0$ are zero, and hence $\ker \tilde{O}_{M(n)} = \ker \tilde{O}_{M(n+1)}$ trivially holds. 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 of $\tilde{O}_{M(n)}\tilde{A}_\sigma$ 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 matrix T_1 spans 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 now

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

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

$$T\tilde{A}_\sigma T^{-1} = \begin{bmatrix} T\tilde{A}_\sigma T_1 & \tilde{A}_\sigma T_2 \end{bmatrix} = \begin{bmatrix} TT_1 N & \tilde{A}_\sigma T_2 \end{bmatrix}.$$

Then,

$$TT_1 = \begin{bmatrix} I_{n_1} \\ 0_{n_2 \times n_1} \end{bmatrix}, \quad TT_1 N = \begin{bmatrix} N \\ 0_{n_2 \times n_1} \end{bmatrix}$$

implies that $A_{\sigma,21} = 0$.

3) Next, we show that if $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$ then the output matrices $\{K_\sigma\}_{\sigma\in\Sigma}$ are also in block triangular form as in (6.6). In order to see this, we will need some technical results. In fact, we will prove the statements (i)–(vi) below: each statement uses the proceeding ones and the final one is equivalent to K_σ satisfying eq. (6.6) for all $\sigma \in \Sigma$.

(i) $\mathbf{x}_2(t) \in \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}$.

(ii) $E[\mathbf{z}^{\mathbf{y}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = 0$ for all $|v| < |w|$, $w, v \in \Sigma^+$.

(iii) $\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+} = \bigoplus_{\sigma_1\in\Sigma} \left(\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}}_{t+1,w\in\Sigma^+} \oplus \mathcal{H}^{\mathbf{z}^{\mathbf{e}_2}_{\sigma_1}}_{t+1}\right)$, where $\oplus$ denotes the direct sum of orthogonal closed subspaces and $\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}}_{t+1,w\in\Sigma^+}$ denotes the Hilbert space generated by $\{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}(t+1)\}_{w\in\Sigma^+}$.

(iv) There exist matrices $\{N_{\sigma_1}\}_{\sigma_1\in\Sigma} \in \mathbb{R}^{n_2\times p_2}$ and a process $\mathbf{r} \in \bigoplus_{\sigma_1\in\Sigma} \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}}_{t+1,w\in\Sigma^+}$, such that $\mathbf{x}_2(t+1) = \mathbf{r} + \sum_{\sigma_1\in\Sigma} N_{\sigma_1}\mathbf{z}^{\mathbf{e}_2}_{\sigma_1}(t+1)$, where $\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}}_{t+1,w\in\Sigma^+}$ is the Hilbert-space generated by the components of the set of random variables $\{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}(t+1)\}_{w\in\Sigma^+}$.

(v) For all $\sigma_1 \in \Sigma$

$$[K_{\sigma_1,21} K_{\sigma_1,22}]E[\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1))^T] = N_{\sigma_1}E[\mathbf{z}^{\mathbf{e}_2}_{\sigma_1}(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1))^T],$$

where $K_\sigma = \begin{bmatrix} K_{\sigma_1,21} & K_{\sigma_1,22} \end{bmatrix}$ and $K_{\sigma_1,21} \in \mathbb{R}^{n_2\times p_1}$, $K_{\sigma_1,22} \in \mathbb{R}^{n_2\times p_2}$.

(vi) $K_{\sigma_1,21} = 0$ for all $\sigma_1 \in \Sigma$.

Next, we will prove (i)–(vi).

(i): By using (6.5), we obtain that

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

for any $v \in \Sigma^+$ where $(A_v)_{11} \in \mathbb{R}^{n_1 \times n_1}$ is the upper block diagonal sub-matrix of A_v, $(A_v)_{22} \in \mathbb{R}^{n_2 \times n_2}$ is the lower block diagonal sub-matrix of A_v and $N \in \mathbb{R}^{p_1 \times n_2}$ is an appropriate matrix. From this, it is easy to see that we can rearrange the rows of $\mathscr{O}_{M(n)}$ in such a way that after rearranging the rows, $\mathscr{O}_{M(n)}$ takes the form of

$$\begin{bmatrix} N_1 & N_2 \\ 0 & O_{M(n)} \end{bmatrix},$$

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. More specifically, by choosing an appropriate permutation matrix P, we have that

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

Notice now that (see (1.11))

$$\mathbf{x}(t) = \begin{bmatrix} p_{v_1} I_p & 0 & \cdots & 0 \\ 0 & p_{v_2} I_p & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & p_{v_{M(n)}} I_p \end{bmatrix}^{-1} \mathscr{O}^+_{M(n)} E_l[Z^{\mathbf{y}}_n(t) | \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w}],$$

where I_p is the $p \times p$ identity matrix and $Z^{\mathbf{y}}_n(t) = [(\mathbf{z}^{\mathbf{y}+}_{v_1}(t))^T, \ldots, (\mathbf{z}^{\mathbf{y}+}_{v_{M(n-1)}}(t))^T]^T$ is a vector of the future of $\mathbf{y}(t)$ w.r.t. the input, see Definition 1.14. Denote the matrix

$$L(M(n), p) = \begin{bmatrix} p_{v_1} I_p & 0 & \cdots & 0 \\ 0 & p_{v_2} I_p & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & p_{v_{M(n)}} I_p \end{bmatrix} \tag{6.7}$$

Then, since P is a permutation matrix and hence $P^T P = I$, we know that $(P\mathscr{O}_{M(n)})^+ = \mathscr{O}^+_{M(n)} P^T$. It then follows that

$$\mathbf{x}(t) = (PL^{-1}(M(n), p)\mathscr{O}^+_{M(n)}) E_l[PZ^{\mathbf{y}}_n(t) | \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w}].$$

Note that

$$PZ^{\mathbf{y}}_n(t) = \left[(Z^{\mathbf{y}_1}_n(t))^T \ (Z^{\mathbf{y}_2}_n(t))^T\right]^T,$$

where $Z^{\mathbf{y}_i}_n(t) = [(\mathbf{z}^{\mathbf{y}_i+}_{v_1}(t))^T, \ldots, (\mathbf{z}^{\mathbf{y}_i+}_{v_{M(n-1)}}(t))^T]^T$ is a vector of the future of $\mathbf{y}_i(t)$,

$i = 1, 2$ w.r.t. the input and thus for $\mathbf{x}_2$ we have that

$$\mathbf{x}_2(t) = L^{-1}(M(n), p_2) O^{+}_{M(n)} E_l[Z^{\mathbf{y}_2}_n(t) | \mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t, w \in \Sigma^+}],$$

where recall that $O_{M(n)}$ is the observability matrix of $(\{A_{\sigma,22}\}_{\sigma \in \Sigma}, C_{22})$. Then, the GB–Granger non-causality 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 $\mathbf{x}_2(t) \in \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t, w \in \Sigma^+}$ which proves (i).

(ii): From (Petreczky and René, 2017, Lemma 14) it follows that $[\mathbf{y}^T, \mathbf{e}^T]^T$ is ZMWSSI. Therefore, we can apply (Petreczky and René, 2017, Lemma 7) for $[\mathbf{y}^T, \mathbf{e}^T]^T$: Consider the covariance $E[\mathbf{z}^{\mathbf{y}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T]$ for $w = w_1 \dots w_k \in \Sigma^*$ and $v = v_1 \dots v_l \in \Sigma^*$, such that $|v| < |w|$. Then (Petreczky and René, 2017, Lemma 7) implies that $E[\mathbf{z}^{\mathbf{y}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = 0$ whenever $w_{k-i} \neq v_{l-i}$ for some $i = 0, \dots, l-1$. On the other hand, if $w_{k-i} = v_{l-i}$ for all $i = 0, \dots, l-1$ then

$$E[\mathbf{z}^{\mathbf{y}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = p_{v_2 \dots v_l} E[\mathbf{z}^{\mathbf{y}}_{w_1 \dots w_{k-l-1}}(t)(\mathbf{z}^{\mathbf{e}}_{v_1}(t))^T] = p_v E[\mathbf{z}^{\mathbf{y}}_{w_1 \dots w_{k-l-1}}(t)\mathbf{e}^T(t)] = 0,$$

where for the last equation we used that $E[\mathbf{z}^{\mathbf{y}}_{w_1 \dots w_{k-l-1}}(t)\mathbf{e}^T(t)] = 0$, see Definition 1.17.

(iii): Consider an innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma \in \Sigma}, \mathbf{y}_2)$ and note that the GB–innovation process of $\mathbf{y}_2$ is $\mathbf{e}_2$ due to the condition that $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$. Then, by (Petreczky and René, 2017, Lemma 16) we can decompose the space $\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t, w \in \Sigma^+}$ as in (iii).

(iv): From (i) 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 (iii), $\mathbf{x}_2(t+1)$ can be written as

$$\mathbf{x}_2(t+1) = \mathbf{r} + \sum_{\sigma_1 \in \Sigma} N_{\sigma_1} \mathbf{z}^{\mathbf{e}_2}_{\sigma_1}(t+1)$$

for some random variable $\mathbf{r} \in \bigoplus_{\sigma_1 \in \Sigma} \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}}_{t+1, w \in \Sigma^+}$ and matrices $\{N_{\sigma_1}\}_{\sigma_1 \in \Sigma} \in \mathbb{R}^{n_2 \times p_2}$.

(v): Notice that by using the block triangular form of the matrices $\{A_\sigma\}_{\sigma \in \Sigma}$, the process $\mathbf{x}_2(t+1)$ can be written as

$$\mathbf{x}_2(t+1) = \sum_{\sigma_1 \in \Sigma} A_{\sigma_1,22} \mathbf{z}^{\mathbf{x}_2}_{\sigma_1}(t+1) + [K_{\sigma_1,21} K_{\sigma_1,22}] \mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1).$$

From (Petreczky and René, 2017, 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 ZMWSSI too. By applying (Petreczky and René, 2017, Lemma 7) for $[\mathbf{e}^T, \mathbf{x}_2^T]^T$, we have that if $\sigma_1 \neq \sigma_2$ then

$$E[\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1)(\mathbf{z}^{\mathbf{x}_2}_{\sigma_2}(t+1))^T] = 0, \quad E[\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma_2}(t+1))^T] = 0.$$

Moreover, by Definition 1.17 it is also true that $E[\mathbf{z}^{\mathbf{e}}_{\sigma_2}(t+1)(\mathbf{z}^{\mathbf{x}}_{\sigma_1}(t+1))^T] = 0$ for $\sigma_1 = \sigma_2$, 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 obtain that $E[\mathbf{z}^{\mathbf{e}}_{\sigma_2}(t+1)(\mathbf{z}^{\mathbf{x}_2}_{\sigma_1}(t+1))^T] = 0$ for $\sigma_1 = \sigma_2$. Hence,

$$E[\mathbf{x}_2(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1))^T] = [K_{\sigma,21} K_{\sigma,22}]Q_{\sigma},$$

where $Q_{\sigma} = E[\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1))^T]$. If we use (iv), and take the covariance of both sides of the equation with $(\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1)$ then we obtain that

$$E[\mathbf{x}_2(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1))^T] = E[\mathbf{r}\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1))^T] + \sum_{\sigma_1 \in \Sigma} N_{\sigma_1} E[\mathbf{z}^{\mathbf{e}_2}_{\sigma_1}(\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1))^T]. \quad (6.8)$$

Notice that by (ii) and since $\mathbf{r} \in \bigoplus_{\sigma_1 \in \Sigma} \mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_{w\sigma_1}}_{t+1, w \in \Sigma^+}$ we know that $E[\mathbf{r}\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1))^T] = 0$. Hence,

$$E[\mathbf{x}_2(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1))^T] = N_{\sigma_1} E[\mathbf{z}^{\mathbf{e}_2}_{\sigma_1}(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1))^T]. \quad (6.9)$$

Combining (6.8) and (6.9), we obtain (v).

(vi): Since $\mathbf{e}_2$ is formed by the last p_2 components of $\mathbf{e}$, we have that

$$N_{\sigma_1} E[\mathbf{z}^{\mathbf{e}_2}_{\sigma_1}(t+1)(\mathbf{z}^{\mathbf{e}}_{\sigma_1}(t+1))^T] = \begin{bmatrix} 0 & N_{\sigma_1} \end{bmatrix} Q_{\sigma_1}$$

and hence $\begin{bmatrix} 0 & N_{\sigma_1} \end{bmatrix} Q_{\sigma_1} = \begin{bmatrix} K_{\sigma_1,21} & K_{\sigma_1,22} \end{bmatrix} Q_{\sigma_1}$. Note that by Assumption 1.21, Q_{σ_1} is positive definite which implies that $\begin{bmatrix} 0 & N_{\sigma_1} \end{bmatrix} = \begin{bmatrix} K_{\sigma_1,21} & K_{\sigma_1,22} \end{bmatrix}$, hence $K_{\sigma_1,21} = 0$.

4) It remains to show that

$$\mathcal{G}_2 = (\{A_{\sigma,22}, K_{\sigma,22}\}_{\sigma \in \Sigma}, C_{22}, I, \mathbf{e}_2)$$

defines a minimal innovation GB–SS representation of $(\{\mathbf{u}_{\sigma}\}_{\sigma \in \Sigma}, \mathbf{y}_2)$. For this, we will use Lemma 6.9. First, notice that due to Definition 1.17, $\mathcal{G}_2$ is a GB–SS representation. Second, using the GB–Granger non-causality condition, it is easy to see that $\mathbf{e}_2(t)$ is the GB–innovation process of $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_{\sigma}\}_{\sigma \in \Sigma}$, thus $\mathcal{G}_2$ 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 process $\tilde{\mathbf{x}}_2$ such that $\tilde{\mathbf{x}}_2 \in \mathbb{R}^{\tilde{n}_2}$ where $\tilde{n}_2 < n_2$.

From Lemma 6.9, it follows that $E_l[Z^{\mathbf{y}_2}_{n_2}(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}] = L(M(n_2), p_2)\tilde{O}_{M(n_2)}\tilde{\mathbf{x}}_2(t)$, where $L(M(n_2), p_2)$ is defined in (6.7) and $\tilde{O}_{M(n_2)}$ is the observability matrix (up to n_2) of $(\{\tilde{A}_{\sigma,22}\}_{\sigma\in\Sigma}, \tilde{C}_{22})$. Define the matrix $T = O^+_{M(n_2)}\tilde{O}_{M(n_2)}$, where $O_{M(n_2)}$ is the observability matrix of $(\{A_{\sigma,22}\}_{\sigma\in\Sigma}, C_{22})$ up to n_2, and notice that $\mathbf{x}_2 = T\tilde{\mathbf{x}}_2$. Then, define a system $\tilde{\mathcal{G}}$ as below

$$\begin{bmatrix}\mathbf{x}_1(t+1)\\ \tilde{\mathbf{x}}_2(t+1)\end{bmatrix} = \sum_{\sigma\in\Sigma}\left(\begin{bmatrix}A_{\sigma,11} & A_{\sigma,21}T\\ 0 & \tilde{A}_{\sigma,22}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1(t)\\ \tilde{\mathbf{x}}_2(t)\end{bmatrix} + \begin{bmatrix}K_{\sigma,11} & K_{\sigma,21}\\ 0 & \tilde{K}_{\sigma,22}\end{bmatrix}\begin{bmatrix}\mathbf{e}_1(t)\\ \mathbf{e}_2(t)\end{bmatrix}\right)\mathbf{u}_\sigma(t)$$
$$\begin{bmatrix}\mathbf{y}_1(t)\\ \mathbf{y}_2(t)\end{bmatrix} = \begin{bmatrix}C_{11} & C_{21}T\\ 0 & \tilde{C}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1(t)\\ \tilde{\mathbf{x}}_2(t)\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1(t)\\ \mathbf{e}_2(t)\end{bmatrix}.$$

We obtain that $\tilde{\mathcal{G}}$ is an innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$ with 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 of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y})$. As a result, $\mathcal{G}_2$ is minimal and it completes our proof. □

Proof of Theorem 6.5. The sufficiency part of the proof follows from Lemmas 6.6 and 6.7.

To prove the necessity part, let $\mathcal{G} = (\{A_\sigma, K_\sigma\}_{\sigma\in\Sigma}, C, I, \mathbf{e})$ be a minimal innovation GB–SS representation of the input-output processes $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T)$ where $\mathbf{y}_i \in \mathbb{R}^{p_i}$ for some $p_i > 0$, $i = 1, 2$ in causal block triangular form such that (6.4) 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 $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$. We will prove that the existence of such system implies that $\mathbf{y}_1$ does not GB–Granger cause $\mathbf{y}_2$ w.r.t. the input $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. Since $\mathbf{e}$ is the GB–innovation process of $\mathbf{y}$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ and $\mathbf{e}_2$ is the GB–innovation process of $\mathbf{y}_2$ w.r.t. $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$, we obtain that $\mathbf{e}_2(t)$ equals

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

which implies that

$$E_l[\mathbf{y}_2(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}] = E_l[\mathbf{y}_2(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}]. \tag{6.10}$$

For GB–Granger non-causality from $\mathbf{y}_1$ to $\mathbf{y}_2$ we need to see that for all $v \in \Sigma^*$

$$E_l[\mathbf{z}^{\mathbf{y}_2+}_v(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}}_w}_{t,w\in\Sigma^+}] = E_l[\mathbf{z}^{\mathbf{y}_2+}_v(t)|\mathcal{H}^{\mathbf{z}^{\mathbf{y}_2}_w}_{t,w\in\Sigma^+}]. \tag{6.11}$$

Note that for $v = v_0 = \epsilon$, (6.11) reduces to (6.10). To prove that (6.11) holds for a general $v \in \Sigma^*$, first note that since (6.4) holds, the matrices $\{A_\sigma\}_{\sigma\in\Sigma}$ and C are block triangular. Therefore,

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

for any $v \in \Sigma^+$ where $(A_v)_{11} \in \mathbb{R}^{n_1 \times n_1}$ is the upper block diagonal sub-matrix of A_v, $(A_v)_{22} \in \mathbb{R}^{n_2 \times n_2}$ is the lower block diagonal sub-matrix of A_v and $N \in \mathbb{R}^{p_1 \times n_2}$ is an appropriate matrix. It then follows from Lemma 6.9 that

$$E_l[\mathbf{z}_v^{\mathbf{y}_2+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}}}] = p_v C_{22}(A_v)_{22}\mathbf{x}_2(t). \tag{6.12}$$

Using that $\mathcal{G}_2$ is a minimal innovation GB–SS representation of $(\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}, \mathbf{y}_2)$ with $\mathbf{x}_2$ as its state process, we also know that $\mathbf{x}_2(t) \in \mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}}$ (see (1.11)). Hence, projecting both side of (6.12) onto $\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}}$, we get that $E_l[\mathbf{z}_v^{\mathbf{y}_2+}(t)|\mathcal{H}_{t,w\in\Sigma^+}^{\mathbf{z}_w^{\mathbf{y}_2}}] = p_v C_{22}(A_v)_{22}\mathbf{x}_2(t)$, which, considering (6.12), implies (6.11) i.e., that there is no GB–Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$. □