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

Jozsa, Monika

Chapter 7

Estimating Kalman representations with given network graphs

The purpose of this chapter is to illustrate how the results of Chapters 2–4 can be applied to estimate linear time-invariant state-space (LTI–SS) representations with a given network graph from data. The notion of network graph refers to a directed graph with some $n \geqslant 2$ nodes that determines the information flow in the LTI–SS representation. More precisely, the LTI–SS representation is decomposed into n subsystems which subsystems correspond to the n nodes of the graph and they communicate with each other according to the directed edges of the network graph.

Applying the results of Chapters 2–4 can help in solving the problem of reverse engineering the network graph of LTI–SS representations. Note that as every process can be represented by infinitely many LTI–SS representations, each of which may have a different network graph, the general problem of reverse engineering the network graph of LTI–SS representations is not well-posed. Instead, by using the results of Chapters 2–4, we know that a well-posed problem is to fix a network graph and to find a so-called Kalman representation of the observed process whose network graph is the designated one, if it exists. Kalman representations are LTI–SS representation with a specific noise process, called the innovation process. This question concerns two tasks, to decide whether a Kalman representation with a given network graph exists and to estimate such representation. In this chapter we primarily focus on the estimation of the representations and secondarily, we touch the topic of testing whether a Kalman representation with a certain network graph exists.

We consider three candidate Kalman representations of the observed processes. First, we work with minimal Kalman representations in the so-called causal block triangular form where the observed process is partitioned into two components and which representations have the two-node network graph with one edge, see Figure 7.1. Second, minimal Kalman representations in the so-called causal coordinated form are considered where the observed process is partitioned into three components and which representations have the three-node star graph as their network graph, see Figure 7.4. At last, we consider minimal Kalman representations with

the so-called causal transitive acyclic directed graph (TADG) zero structure, where the observed process is partitioned into three components and where the network graph is the TADG depicted in Figure 7.5.

From Chapters 2–4, we know that for the existence of the above-mentioned Kalman representations certain Granger non-causality conditions should hold. In practice, these conditions should be tested before reconstructing the network graph of the Kalman representations at hand from the observed data. For this purpose we propose a statistical test to check whether a process Granger causes another process or not. This statistical test is based on a measure, called (time-domain) Geweke–Granger causality (Geweke, 1984), that describes causal relations between processes on a continuous scale from zero to one. Geweke–Granger causality is zero whenever Granger causality is absent and it is positive whenever Granger causality is present. The statistical test accepts a Granger non-causality condition if the corresponding Geweke–Granger causality is small enough.

If the Granger non-causality conditions that are necessary for the existence of a Kalman representation with a specific network graph are satisfied, then the next natural step is to find such Kalman representation. To this end, we first present realization algorithms to calculate the system matrices of these Kalman representations from the observed process and then, relying on the latter realization algorithms, we propose identification algorithms to estimate the system matrices of the same Kalman representation from the observed data. The proposed algorithms are closely related to the realization algorithms in Chapters 2–4.

It is important to note that we do not aim at presenting a complete analysis of the proposed statistical test on Granger non-causality, nor of the proposed algorithms. They rather serve as an idea of how the results of Chapters 2–4 can be put in practice. Much more work is needed to evaluate the performance of the procedures above and to improve them, if necessary.

However, even from the limited study of this chapter, promising results emerge. In addition to showing methods that help in reverse engineering of the network graph of Kalman representations, the proposed procedures open up the possibility of exploiting the network graph of Kalman representations for distributed parameter estimation and for improving the accuracy of system identification algorithms. Distributed parameter estimation can be used for the following reasons: if we know that a process admits a Kalman representation with a certain network graph, then various subsystems of such Kalman representations can be estimated independently of each other by only using data that belongs to some components of the observed process. These subsystems can then be combined to form the whole system. Since the estimation error tends to be higher with the increase of the dimension of the output process, estimating the subsystems in a distributed way may decrease the

overall estimation error.

The algorithms of this chapter are compared to two other classical methods for estimating the system matrices of minimal Kalman representations, namely to a state-space subspace algorithm (Larimore, 1983) and prediction error method (Ljung, 1999, Chapter 7). For the comparison, we use simulated data from randomly generated minimal Kalman representations with the three different above-mentioned network graphs and apply each method to the simulated data to estimate minimal Kalman representations of the underlying observed process. These estimates are then compared to the original Kalman representation that was used for the data generation. The results show that for the considered data length and parameter settings, the algorithms presented in this chapter can provide better estimates for a minimal Kalman representation of the observed process than the considered classical methods. Moreover, contrary to the other two methods, the proposed algorithms potentially estimate the system matrices in a distributed manner which can be of use in distributed parameter estimation.

The algorithms and analysis were implemented in MATLAB 2017a. Throughout the chapter, we use several functions from the multivariate Granger causality (MVGC) toolbox (Barnett and Seth, 2014) and from the collection of functions provided as supplementary material for the paper (Barnett and Seth, 2015), called state-space Granger causality (SSGC) toolbox.

The chapter is organized as follows: First, we present the realization and identification algorithms for calculating and estimating the system matrices of minimal Kalman representations in causal block triangular form, in causal coordinated form and with causal TADG-zero structure. Second, the statistical test is introduced for checking Granger non-causality. Note that the latter test integrates the proposed algorithm for estimating the system matrices of minimal Kalman representations in causal block triangular form. Then, the proposed algorithms are compared to two classical methods on estimating minimal Kalman representations using simulated data. For this, we first introduce the measures that the evaluation and comparison of the three methods are based on. Finally, we present the results.

7.1 Estimating Kalman representation with specific network graph

In this section, we present realization algorithms for calculating and identification algorithms for estimating the system matrices of minimal Kalman representations in causal block triangular form, in causal coordinated form and with causal TADG-zero structure. The algorithms are closely related to the covariance realization algo-

rithms presented in Chapters 2–4.

The proposed realization algorithms are applied to the covariances of a process and define system matrices of minimal Kalman representations of that process in causal block triangular form, in causal coordinated form or with causal TADG-zero structure, if the representation at hand exists. The proposed identification algorithms are applied to the empirical covariances of a process and they define approximations of the system matrices of the above-mentioned Kalman representations, if those representations exist.

More formally, consider a process $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_k^T]^T \in \mathbb{R}^r$, where $\mathbf{y}_i \in \mathbb{R}^{r_i}$, $i = 1, \ldots, k$ and assume that $\mathbf{y}$ has a minimal Kalman representation $S = (A, K, C, I, \mathbf{e})$ in causal block triangular form, or in causal coordinated form or with causal TADG-zero structure where $\mathbf{e}$ is a Gaussian white noise process. Then, our goal is to estimate matrices $\hat{A}, \hat{K}, \hat{C}$ and $Q^{\hat{\mathbf{e}}}$ such that they determine a minimal Kalman representation $\hat{S} = (\hat{A}, \hat{K}, \hat{C}, I, \hat{\mathbf{e}}, \hat{\mathbf{y}})$ where $\hat{\mathbf{e}}$ is a Gaussian white noise process with variance matrix $Q^{\hat{\mathbf{e}}}$, $\hat{S}$ has the same network graph as S and $\hat{\mathbf{y}}$ is an approximation of the process $\mathbf{y}$, or, in other words, the Kalman representation $\hat{S}$ is an approximation of S.

Intuitively, the algorithms presented in this section try to keep $\hat{S}$ 'close' to S regarding the following measure: Denote a deterministic LTI–SS system

$$\begin{aligned} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t) \end{aligned} \tag{7.1}$$

by (A, B, C, D). Then the measure that determines how close $\hat{S}$ is to S is the relative H_2-error of the deterministic systems $\hat{S}^{\text{det}} = (\hat{A}, \hat{K}, \hat{C}, I)$ and $S^{\text{det}} = (A, K, C, I)$:

$$\varepsilon_{sys}^{H_2}(\hat{S}^{\text{det}}, S^{\text{det}}) = \frac{||\hat{S}^{\text{det}} - S^{\text{det}}||_{H_2}}{||S^{\text{det}}||_{H_2}}, \tag{7.2}$$

where $||.||_{H_2}$ denotes the H_2-norm of a deterministic LTI–SS system (Trentelman et al., 2001, Chapter 11) and $\hat{S}^{\text{det}} - S^{\text{det}}$ is the deterministic system

$$\left(\begin{bmatrix} \hat{A} & 0 \\ 0 & A \end{bmatrix}, \begin{bmatrix} \hat{K} \\ K \end{bmatrix}, \begin{bmatrix} \hat{C} & -C \end{bmatrix}, I \right).$$

Note that (7.2) is equal to the following

$$\varepsilon_{sys}^{H_2}(\hat{S}^{\text{det}}, S^{\text{det}}) = \frac{\sqrt{\sum_{k=0}^{\infty} \|\hat{C}\hat{A}^k\hat{K} - CA^kK\|_2^2}}{\sqrt{\sum_{k=0}^{\infty} \|CA^kK\|_2^2}}.$$

It is easy to see that $\varepsilon_{sys}^{H_2}(\hat{S}^{\text{det}}, S^{\text{det}}) = 0$ when $(\hat{A}, \hat{K}, \hat{C}, I)$ and (A, K, C, I) are

isomorphic (see Definition 1.11). In fact, $\varepsilon_{sys}^{H_2}(\hat{S}^{\text{det}}, S^{\text{det}})$ measures the accuracy of the Markov parameters $\{\hat{C}\hat{A}^k\hat{K}\}_{k=0}^{\infty}$ of the deterministic system $(\hat{A}, \hat{K}, \hat{C}, I)$ as estimates for the Markov parameters $\{CA^kK\}_{k=0}^{\infty}$ of the deterministic system (A, K, C, I). Note that the variance matrix $E[\mathbf{e}(t)\mathbf{e}^T(t)]$ will be estimated based on classical methods and thus its accuracy will be the same for the proposed methods on estimating system matrices of Kalman representation and the considered classical methods. The algorithms which are to be presented in this sections can be reformulated with any other error measure on estimates of Kalman representations.

7.1.1 Estimating system matrices of minimal Kalman representations in causal block triangular form

We start by presenting a realization algorithm (Algorithm 14) which calculates system matrices of a minimal Kalman representation of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ in causal block triangular form if $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$. Note that if $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$ then the output of Algorithm 14 does not define a Kalman representation of $\mathbf{y}$. However, if $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ then Algorithm 14 is equivalent to Algorithm 5 in Chapter 2, see Remark 7.1. Algorithm 14 serves as a basis of an identification algorithm presented in Algorithm 15. In fact, the latter algorithm boils down to applying Algorithm 14 to empirical covariances of the output process.

Similar to Algorithm 14, Algorithm 5 also uses output covariances to calculate the system matrices of a minimal Kalman representation in causal block triangular form if the corresponding Granger causality condition holds. There are two reasons why we use Algorithm 14 and not Algorithm 5 in this chapter: First, if the inputs of Algorithm 5 are not the exact covariances of $\mathbf{y}$ (e.g., we use empirical covariances) then the output matrix K of Algorithm 5 is, in general, not in block triangular form. Second, Algorithm 5 does not exploit the fact that, if $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ then the matrices A, K and C of a minimal Kalman representation $(A, K, C, I, \mathbf{e}, \mathbf{y})$ in causal block triangular form can be calculated in a distributer manner.

Algorithm 14 below goes as follows: First, it follows the steps of Algorithm 5, however, it changes appropriate blocks of its output matrices to zero blocks in order to guarantee that they are in block triangular form. Note that if $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ then this change is unnecessary, the matrices are automatically in block triangular form. Second, it calculates another set of estimates of the system matrices by using a distributed method. To decide which estimates to choose as final estimates, we take reference matrices A^*, K^*, C^* that define a minimal Kalman representation $(A^*, K^*, C^*, I, \mathbf{e}, \mathbf{y})$ and compare the two sets of estimates to the reference matrices using the relative H_2-error defined in (7.2). This latter step helps to reformulate Algorithm 14 as an identification algorithm, and it sets the final estimates in a way that

they correspond to the output matrices of Algorithm 5 if the Granger non-causality condition from $\mathbf{y}_1$ to $\mathbf{y}_2$ holds.

Remark 7.1. Let the input of Algorithm 14 be as follows: $\{A^*, K^*, C^*\}$ are system matrices of a minimal Kalman representation of $\mathbf{y}$, $\{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2N}$ is the covariance sequence of $\mathbf{y}$ and $\{r_1, r_2, n, n_2\}$ are such that $\mathbf{y}_1 \in \mathbb{R}^{r_1}$, $\mathbf{y}_2 \in \mathbb{R}^{r_2}$ and n, n_2 are the dimensions of minimal Kalman representations of $\mathbf{y}$ and $\mathbf{y}_2$, respectively. If $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ then the output matrices of Algorithms 14 and 5 define the same output matrices with the nuance that the matrix $\hat{T}$ in Step 2 of Algorithm 14 is a particular choice for the matrix T in Step 2 of Algorithm 5.

Algorithm 14 Minimal Kalman representation in causal block triangular form

Input $\{A^*, K^*, C^*\}, \{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$ and $\{r_1, r_2, n, n_2\}$: System matrices of a minimal Kalman representation of $\mathbf{y}$, covariance sequence of $\mathbf{y}$ and dimensions of the components of $\mathbf{y}$ and minimal Kalman representations of $\mathbf{y}$ and $\mathbf{y}_2$, respectively

Output $\{A, K, C\}$: System matrices of a minimal Kalman representation of $\mathbf{y}$ in causal block triangular form

Step 1 Calculate the system matrices $\{\hat{A}, \hat{K}, \hat{C}\}$ of a minimal Kalman representation $(\hat{A}, \hat{K}, \hat{C}, I, \mathbf{e}, \mathbf{y})$, e.g., by applying Algorithm 1 with input $\{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$.

Step 2 Let $\hat{C} = \left[\hat{C}_1^T \hat{C}_2^T\right]^T$ be such that $\hat{C}_i \in \mathbb{R}^{r_i \times n}$. Denote the observability matrix of (A, C_2) up to n by $\hat{\mathcal{O}}_n$ and take its SVD as $\hat{\mathcal{O}}_n = USV^T$, where $U \in \mathbb{R}^{nr \times nr}$, $S \in \mathbb{R}^{nr \times n}$ and $V \in \mathbb{R}^{n \times n}$. Define $\hat{T} = \left[V(:, n_2 + 1 : n)\ V(:, 1 : n_2)\right]$, where $V(:, i : i+j)$ denotes the matrix in $\mathbb{R}^{n \times j+1}$ whose $1, \ldots, j+1$ columns are the $i, i+1, \ldots, i+j$ columns of V, respectively.

Step 3 Define $n_1 = n - n_2$ and partition the matrices

$$\hat{T}\hat{A}\hat{T}^+ = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}, \ \hat{T}\hat{K} = \begin{bmatrix} \hat{K}_{11} & \hat{L}_{12} \\ \hat{K}_{21} & \hat{K}_{22} \end{bmatrix}, \ \hat{C}\hat{T}^+ = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} \\ \hat{C}_{21} & \hat{C}_{22} \end{bmatrix}, \tag{7.3}$$

such that $\hat{A}_{ij} \in \mathbb{R}^{r_i \times r_j}$, $\hat{K}_{ij} \in \mathbb{R}^{n_i \times r_j}$ and $\hat{C}_{ij} \in \mathbb{R}^{r_i \times n_j}$ for $i, j = 1, 2$. Let

$$A_{v_1} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ 0 & \hat{A}_{22} \end{bmatrix}, \ C_{v_1} = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} \\ 0 & \hat{C}_{22} \end{bmatrix}, \ K_{v_1} = \begin{bmatrix} \hat{K}_{11} & \hat{K}_{12} \\ 0 & \hat{K}_{22} \end{bmatrix}.$$

Step 4 Calculate the system matrices $\{A_{22}, K_{22}, C_{22}\}$ of a minimal Kalman representation $(A_{22}, K_{22}, C_{22}, I, \mathbf{e}, \mathbf{y})$, e.g., by applying Algorithm 1 with input $\{\Lambda_k^{\mathbf{y}_2}\}_{k=0}^{2n}$ (the $r_2 \times r_2$ right bottom sub-matrices of $\{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$).

Step 5 Let T be such that $T = \hat{\mathcal{O}}_{n_2}^{+}\mathcal{O}_{n_2}$, where $\hat{\mathcal{O}}_{n_2}$ and $\mathcal{O}_{n_2}$ are observability matrices of $(\hat{A}_{22}, \hat{C}_{22})$ and (A_{22}, C_{22}) up to n_2, respectively.
Step 6 Let

$$A_{v_2} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12}T \\ 0 & A_{22} \end{bmatrix},\ C_{v_2} = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12}T \\ 0 & C_{22} \end{bmatrix},\ K_{v_2} = \begin{bmatrix} \hat{K}_{11} & \hat{K}_{12} \\ 0 & K_{22} \end{bmatrix}.$$

Step 7 Define the deterministic systems

$$\hat{S} = (A_{v_1}, K_{v_1}, C_{v_1}, I),\ S = (A_{v_2}, K_{v_2}, C_{v_2}, I),\ S^* = (A^*, K^*, C^*, I).$$

If $\varepsilon_{sys}^{H_2}(\hat{S}, S^*) \leqslant \epsilon_{sys}^{H_2}(S, S^*)$ (see (7.2)) then define $A = A_{v_1}$, $K = K_{v_1}$ and $C = C_{v_1}$. Otherwise, define $A = A_{v_2}$, $K = K_{v_2}$ and $C = C_{v_2}$.

In Algorithm 15 below, we show how to apply Algorithm 14 to empirical covariances and estimated system matrices of a minimal Kalman representation. Algorithm 14 estimates the system matrices of a minimal Kalman representation of $\mathbf{y}$ in causal block triangular form with its input being the dimensions $\{r_1, r_2, n, n_2\}$ and a finite sample $\{y(i)\}_{i=1}^{N}$ of $\mathbf{y}$, i.e., a sample of $\mathbf{y}(t), \ldots, \mathbf{y}(t+N-1)$ for a $t \in \mathbb{Z}$.

Algorithm 15 Estimating system matrices of minimal Kalman representations in causal block triangular form

Input $\{y(i)\}_{i=1}^{N}$ and $\{r_1, r_2, n, n_2\}$: Finite sample of $\mathbf{y}$ and dimensions of the components of $\mathbf{y}$ and a minimal Kalman representation of $\mathbf{y}$ and $\mathbf{y}_2$, respectively

Output $\{A, K, C\}$: Estimate for system matrices of a minimal Kalman representation of $\mathbf{y}$ in causal block triangular form

Step 1 Calculate the empirical covariances $\{\Lambda_k^y\}_{k=0}^{2n}$ by using the MATLAB function `cov`, see also see Remark 7.2.
Step 2 Apply the MATLAB function `s4sid_CCA` from the SSGC toolbox (see (Barnett and Seth, 2015) for details on SSGC) with input $\{\Lambda^y\}_{k=0}^{2n}$ and denote its output by A^*, K^*, C^* and $\Lambda_0^{\hat{\mathbf{e}}}$, where $(A^*, K^*, C^*, I, \hat{\mathbf{e}})$ is an estimated minimal Kalman representation for some white noise process $\hat{\mathbf{e}}$ such that $\Lambda_0^{\hat{\mathbf{e}}} = E[\hat{\mathbf{e}}(t)\hat{\mathbf{e}}^T(t)]$.
Step 3 Apply Algorithm 14 with input A^*, K^*, C^*, $\{\Lambda_k^y\}_{k=0}^{2n}$ and $\{r_1, r_2, n, n_2\}$ where in Steps 1 and 4, estimate the system matrices of minimal Kalman representations by using `s4sid_CCA` with the corresponding empirical covariances as its input.

Note that in Algorithm 15, `s4sid_CCA` could be changed to any other identification algorithm on minimal Kalman representations. We will rely on Algorithm 15 in the rest of the chapter to estimate system matrices of a minimal Kalman representa-

tion in causal block triangular form.

Remark 7.2. We calculate empirical covariances by using the MATLAB function `cov`, i.e., the kth empirical covariance corresponding to $\Lambda_k^{\mathbf{y}} = E[\mathbf{y}(t+k)\mathbf{y}^T(t)]$ is

$$\Lambda_k^y = \frac{1}{N-k-1}\sum_{i=1}^{N-k}(y_{i+k} - \frac{1}{N}\sum_{j=1}^{N} y_j)(y_i - \frac{1}{N}\sum_{j=1}^{N} y_j)^T. \tag{7.4}$$

7.1.2 Estimating system matrices of minimal Kalman representations in causal coordinated form

In this section, we present a realization algorithm (Algorithm 16) which calculates system matrices of a minimal Kalman representation of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T]^T$ in causal coordinated form if the latter representation exists (see conditions (i)–(ii) in Theorem 3.5). Note that if the representation above does not exist then the output of Algorithm 16 does not define a Kalman representation of $\mathbf{y}$. However, if the Kalman representation exists then Algorithm 16 is equivalent to Algorithm 7 in Chapter 3, see Remark 7.3. Algorithm 16 serves as a basis of the identification algorithm, Algorithm 17 in a way that if the Kalman representation exists then Algorithm 17 boils down to applying Algorithm 16 to empirical covariances of the output process.

As Algorithm 16, Algorithm 7 also calculates the system matrices of a minimal Kalman representation in causal coordinated form, if the latter representation exists. Here we use Algorithm 16 instead of Algorithm 7 because it exploits the possibility of using distributed parameter estimation and it returns matrices that are guaranteed to have zero blocks such that they define a minimal Kalman representations in causal coordinated form.

The idea of Algorithm 16 is as follows: First, it applies Algorithm 14 to calculate system matrices of minimal Kalman representations of $[\mathbf{y}_1^T, \mathbf{y}_3^T]^T$ and of $[\mathbf{y}_2^T, \mathbf{y}_3^T]^T$ in causal block triangular form, if those representations exist. Then, it combines the matrices obtained in the previous step into matrices that are system matrices of a Kalman representation of $\mathbf{y}$ in causal coordinated form, if such representation exists. The latter step is carried out in two variations: the system matrices of the coordinator system, i.e., of $\mathbf{y}_3$ are from the system matrices of the Kalman representation of $[\mathbf{y}_1^T, \mathbf{y}_3^T]^T$ in the first case and in the second case, they are from the system matrices of the Kalman representation of $[\mathbf{y}_2^T, \mathbf{y}_3^T]^T$. To decide which matrices to choose as system matrices of the combined system, we compare them to so-called reference matrices A^*, K^*, C^* that define a minimal Kalman representation $(A^*, K^*, C^*, I, \mathbf{e}, \mathbf{y})$. For the comparison, we use the relative H_2-error defined in (7.2). This comparison is designed to set the output matrices in a way that they

correspond to the output matrices of Algorithm 7. Algorithm 16 serves as a basis for the identification algorithm, Algorithm 17 below.

Algorithm 16 Minimal Kalman representation in causal coordinated form

Input $\{A^*, K^*, C^*\}, \{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$ and $\{r_i, n_i\}_{i=1}^{3}$: System matrices of a minimal Kalman representation of $\mathbf{y}$, covariances of $\mathbf{y}$, dimensions of $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ and the state components of a minimal Kalman representation of $\mathbf{y}$ in causal coordinated form

Output $\{A, K, C\}$: System matrices of a minimal Kalman representation of $\mathbf{y}$ in causal coordinated form

for $i = 1:2$

Step 1 Calculate the system matrices $\{\hat{A}_{i3}, \hat{K}_{i3}, \hat{C}_{i3}\}$ of a minimal Kalman representation $(\hat{A}_{i3}, \hat{K}_{i3}, \hat{C}_{i3}, I, [\mathbf{e}_i^T, \mathbf{e}_3^T], [\mathbf{y}_i^T, \mathbf{y}_3^T])$, e.g., by applying Algorithm 1 with input $\{\Lambda_k^{\mathbf{y}_{i,3}}\}_{k=0}^{2(n_i+n_3)}$, where $\Lambda_k^{\mathbf{y}_{i,3}} = E[[\mathbf{y}_i^T(t+k), \mathbf{y}_3^T(t+k)]^T[\mathbf{y}_i^T(t), \mathbf{y}_3^T(t)]]$ is obtained from $\Lambda_k^{\mathbf{y}}$.

Step 2 Apply Algorithm 14 with input $\{\hat{A}_{i3}, \hat{K}_{i3}, \hat{C}_{i3}\}, \{\Lambda_k^{\mathbf{y}_{i,3}}\}_{k=0}^{2n}$ and $\{r_i, r_3, n_i + n_3, n_3\}$ and denote its output by

$$\begin{bmatrix} A_{11}^{v_i} & A_{12}^{v_i} \\ 0 & A_{22}^{v_i} \end{bmatrix}, \begin{bmatrix} C_{11}^{v_i} & C_{12}^{v_i} \\ 0 & C_{22}^{v_i} \end{bmatrix}, \begin{bmatrix} K_{11}^{v_i} & K_{12}^{v_i} \\ 0 & K_{22}^{v_i} \end{bmatrix}$$

end for

Step 3 Define $T^{v_1} = (\mathcal{O}_{n_3}^{v_2})^+ \mathcal{O}_{n_3}^{v_1}$ and $T^{v_2} = (\mathcal{O}_{n_3}^{v_1})^+ \mathcal{O}_{n_3}^{v_2}$, where $\mathcal{O}_{n_2}^{v_1}$ and $\mathcal{O}_{n_2}^{v_2}$ are observability matrices of $(A_{22}^{v_1}, C_{22}^{v_1})$ and $(A_{22}^{v_2}, C_{22}^{v_2})$, respectively up to n_3.

Step 4 Define the matrices

$$A_{v_1} = \begin{bmatrix} A_{11}^{v_1} & 0 & A_{12}^{v_1} \\ 0 & A_{11}^{v_2} & A_{12}^{v_2} T^{v_1} \\ 0 & 0 & A_{22}^{v_1} \end{bmatrix} \quad K_{v_1} = \begin{bmatrix} K_{11}^{v_1} & 0 & K_{12}^{v_1} \\ 0 & K_{11}^{v_2} & K_{12}^{v_2} \\ 0 & 0 & K_{22}^{v_1} \end{bmatrix} \quad C_{v_1} = \begin{bmatrix} C_{11}^{v_1} & 0 & C_{12}^{v_1} \\ 0 & C_{11}^{v_2} & C_{12}^{v_2} T^{v_1} \\ 0 & 0 & C_{22}^{v_1} \end{bmatrix}$$

$$A_{v_2} = \begin{bmatrix} A_{11}^{v_1} & 0 & A_{12}^{v_1} T^{v_2} \\ 0 & A_{11}^{v_2} & A_{12}^{v_2} \\ 0 & 0 & A_{22}^{v_2} \end{bmatrix} \quad K_{v_2} = \begin{bmatrix} K_{11}^{v_1} & 0 & K_{12}^{v_1} \\ 0 & K_{11}^{v_2} & K_{12}^{v_2} \\ 0 & 0 & K_{22}^{v_2} \end{bmatrix} \quad C_{v_2} = \begin{bmatrix} C_{11}^{v_1} & 0 & C_{12}^{v_1} T^{v_2} \\ 0 & C_{11}^{v_2} & C_{12}^{v_2} \\ 0 & 0 & C_{22}^{v_2} \end{bmatrix}.$$

Step 5 Define the deterministic systems

$$\hat{S} = (A_{v_1}, K_{v_1}, C_{v_1}, I), \; S = (A_{v_2}, K_{v_2}, C_{v_2}, I), \; S^* = (A^*, K^*, C^*, I).$$

If $\varepsilon_{sys}^{H_2}(\hat{S}, S^*) \leqslant \epsilon_{sys}^{H_2}(S, S^*)$ (see (7.2)) then define $A = A_{v_1}$, $K = K_{v_1}$ and $C = C_{v_1}$. Otherwise, define $A = A_{v_2}$, $K = K_{v_2}$ and $C = C_{v_2}$.

Remark 7.3. Let the input of Algorithm 16 be as follows: $\{A^*, K^*, C^*\}$ are system

matrices of a minimal Kalman representation of $\mathbf{y}$, $\{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2N}$ is the covariance sequence of $\mathbf{y}$ and n_1, n_2, n_3 are the dimensions of the state components of a minimal Kalman representation of $\mathbf{y}$ in causal coordinated form. If (i) and (ii) in Theorem 3.5 hold, then the output matrices of Algorithm 16 define system matrices of a minimal Kalman representation in causal coordinated form. In this case Algorithms 16 and 7 essentially define the same output matrices, see also Remark 7.1.

Next, we present Algorithm 17 on estimating system matrices of a minimal Kalman representation of $\mathbf{y}$ in causal coordinated form from a finite sample $\{y(i)\}_{i=1}^{N}$ of $\mathbf{y}$ and the dimensions $\{r_i, n_i\}_{i=1}^{3}$ of the output and state components of a minimal Kalman representation of $\mathbf{y}$ in causal coordinated form. Algorithm 17 is based on the idea of applying Algorithm 16 with empirical inputs.

Algorithm 17 Estimating system matrices of minimal Kalman representation in causal coordinated form

Input $\{y(i)\}_{i=1}^{N}$ and $\{r_i, n_i\}_{i=1}^{3}$: Finite sample of $\mathbf{y}$ and output and state dimensions of a minimal Kalman representation of $\mathbf{y}$ in causal coordinated form

Output $\{A, K, C\}$: Estimate for system matrices of a minimal Kalman representation of $\mathbf{y}$ in causal coordinated form

for $i = 1 : 2$

Step 1 Estimate the system matrices of a minimal Kalman representation of $[\mathbf{y}_i^T, \mathbf{y}_3^T]$ by applying the MATLAB function `s4sid_CCA` from the SSGC toolbox (see (Barnett and Seth, 2015) for details on SSGC) with input $\{\Lambda_k^{y_{i,3}}\}_{k=0}^{2(n_i+n_3)}$, where $\{\Lambda_k^{y_{i,3}}\}_{k=0}^{2(n_i+n_3)}$ are the empirical covariances of $[\mathbf{y}_i^T, \mathbf{y}_3^T]^T$, see (7.4). Denote the estimated system matrices by $\{A_{i3}^*, K_{i3}^*, C_{i3}^*\}$.

Step 2 Apply Algorithm 14 with input $\{A_{i3}^*, K_{i3}^*, C_{i3}^*\}, \{\Lambda_k^{y_{i,3}}\}_{k=0}^{2n}$ and $\{r_i, r_3, n_i + n_3, n_3\}$, where in Steps 1 and 4, estimate the system matrices of minimal Kalman representations by using `s4sid_CCA` with the corresponding empirical covariances as its input. Denote its output by

$$\begin{bmatrix} A_{11}^{v_i} & A_{12}^{v_i} \\ 0 & A_{22}^{v_i} \end{bmatrix}, \quad \begin{bmatrix} C_{11}^{v_i} & C_{12}^{v_i} \\ 0 & C_{22}^{v_i} \end{bmatrix}, \quad \begin{bmatrix} K_{11}^{v_i} & K_{12}^{v_i} \\ 0 & K_{22}^{v_i} \end{bmatrix}.$$

end for

Step 3 Steps 4 and 5 of Algorithm 16.

We will rely on Algorithm 17 in the rest of the chapter to estimate system matrices of minimal Kalman representations in causal coordinated form.

7.1.3 Estimating system matrices of minimal Kalman representations with causal TADG-zero structure

The last class of representations that we present algorithms on are minimal Kalman representations with causal TADG-zero structure, where the TADG is given by $G = (V, E)$ with $V = \{1, 2, 3\}$ and $E = \{(3, 1), (3, 2), (2, 1)\}$, see Figure 7.5. First, we present a realization algorithm (Algorithm 18) which calculates system matrices of a minimal Kalman representation of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T]^T$ with causal G-zero structure if the representation exists, see conditions (i)–(ii) in Theorem 4.15. Note that if the representation does not exist then the output matrices of Algorithm 18 do not define a Kalman representation of $\mathbf{y}$. However, if the Kalman representation exists then Algorithm 18 is equivalent to Algorithm 11 in Chapter 4, see Remark 7.4. Algorithm 18 serves as a basis of an identification algorithm presented in Algorithm 19.

Note that Algorithm 11 calculates the system matrices of a minimal Kalman representation with causal TADG-zero structure for any general TADG, if the latter representation exists. We use Algorithm 18 and not Algorithm 11 because it exploits the possibility of distributed parameter estimation and returns matrices that are guaranteed to have zero blocks such that they define a minimal Kalman representation with causal G-zero structure. Algorithm 18 can be generalized to minimal Kalman representations with any TADG-zero structure, however, we do not deal with this general case.

Algorithm 18 Minimal Kalman representation with G-zero structure:

Input $\{A^*, K^*, C^*\}, \{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$ and $\{r_i, n_i\}_{i=1}^3$: System matrices of a minimal Kalman representation of $\mathbf{y}$, covariance sequence of $\mathbf{y}$ and dimensions of the components of $\mathbf{y}$ and the state of a minimal Kalman representation of $\mathbf{y}$ with causal G-zero structure

Output $\{A, K, C\}$: System matrices of a minimal Kalman representation of $\mathbf{y}$ with causal G-zero structure

Step 1 Calculate the system matrices $\{A_{23}^*, K_{23}^*, C_{23}^*\}$ of a minimal Kalman representation $(\hat{A}_{23}, \hat{K}_{23}, \hat{C}_{23}, I, [\mathbf{e}_2^T, \mathbf{e}_3^T], [\mathbf{y}_2^T, \mathbf{y}_3^T])$, e.g., by applying Algorithm 1 with input $\{\Lambda_k^{\mathbf{y}_{2,3}}\}_{k=0}^{2(n_2+n_3)}$, where $\Lambda_k^{\mathbf{y}_{2,3}} = E[[\mathbf{y}_2^T(t+k), \mathbf{y}_3^T(t+k)]^T[\mathbf{y}_2^T(t), \mathbf{y}_3^T(t)]]$ is obtained from $\Lambda_k^{\mathbf{y}}$.

Step 2 Apply Algorithm 14 with input $\{\hat{A}_{23}, \hat{K}_{23}, \hat{C}_{23}\}, \{\Lambda_k^{\mathbf{y}_{2,3}}\}_{k=0}^{2(n_2+n_3)}$ and $\{r_2, r_3, n_2 + n_3, n_3\}$ and denote its output by

$$\begin{bmatrix} A_{22}^{v_1} & A_{23}^{v_1} \\ 0 & A_{33}^{v_1} \end{bmatrix}, \begin{bmatrix} C_{22}^{v_1} & C_{23}^{v_1} \\ 0 & C_{33}^{v_1} \end{bmatrix}, \begin{bmatrix} K_{22}^{v_1} & K_{23}^{v_1} \\ 0 & K_{33}^{v_1} \end{bmatrix} \tag{7.5}$$

Step 3 Apply Algorithm 14 with input $\{A^*, K^*, C^*\}, \{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$ and $\{r_1, r_2 + r_3, n, n_2 + n_3\}$ and denote its output by

$$\begin{bmatrix} A_{11}^{v_2} & A_{12}^{v_2} & A_{13}^{v_2} \\ 0 & A_{22}^{v_2} & A_{23}^{v_2} \\ 0 & A_{32}^{v_2} & A_{33}^{v_2} \end{bmatrix}, \quad \begin{bmatrix} K_{11}^{v_2} & K_{12}^{v_2} & K_{13}^{v_2} \\ 0 & K_{22}^{v_2} & K_{23}^{v_2} \\ 0 & K_{32}^{v_2} & K_{33}^{v_2} \end{bmatrix}, \quad \begin{bmatrix} C_{11}^{v_2} & C_{12}^{v_2} & C_{13}^{v_2} \\ 0 & C_{22}^{v_2} & C_{23}^{v_2} \\ 0 & C_{32}^{v_2} & C_{33}^{v_2} \end{bmatrix} \tag{7.6}$$

Step 4 Denote the observability matrices of

$$\left(\begin{bmatrix} A_{22}^{v_1} & A_{23}^{v_1} \\ 0 & A_{33}^{v_2} \end{bmatrix}, \begin{bmatrix} C_{22}^{v_1} & C_{23}^{v_1} \\ 0 & C_{33}^{v_2} \end{bmatrix} \right), \quad \left(\begin{bmatrix} A_{22}^{v_2} & A_{23}^{v_2} \\ A_{32}^{v_2} & A_{33}^{v_2} \end{bmatrix}, \begin{bmatrix} C_{22}^{v_2} & C_{23}^{v_2} \\ C_{32}^{v_2} & C_{33}^{v_2} \end{bmatrix} \right)$$

up to $n_2 + n_3$ by $\mathcal{O}_{n_2+n_3}^{v_1}$ and $\mathcal{O}_{n_2+n_3}^{v_2}$, respectively and define the matrix $T^{v_1} = (\mathcal{O}_{n_2+n_3}^{v_2})^+ \mathcal{O}_{n_2+n_3}^{v_1}$. Then, define the matrices

$$A^{v_1} = \begin{bmatrix} A_{11}^{v_2} & [A_{12}^{v_2} \; A_{13}^{v_2}]T^{v_1} \\ 0 & A_{22}^{v_1} \; A_{23}^{v_1} \\ 0 & 0 \quad A_{33}^{v_1} \end{bmatrix} \; K^{v_1} = \begin{bmatrix} K_{11}^{v_2} & K_{12}^{v_2} & K_{13}^{v_2} \\ 0 & K_{22}^{v_1} & K_{23}^{v_1} \\ 0 & 0 & K_{33}^{v_1} \end{bmatrix} \; C^{v_1} = \begin{bmatrix} C_{11}^{v_2} & [C_{12}^{v_2} \; C_{13}^{v_2}]T^{v_1} \\ 0 & C_{22}^{v_1} \; C_{23}^{v_1} \\ 0 & 0 \quad C_{33}^{v_1} \end{bmatrix}.$$

Step 5 Let $\hat{A}_{ij}^{v_2}, \hat{K}_{ij}^{v_2}, \hat{C}_{ij}^{v_2}$, $i, j = 2, 3$ be such that

$$T^{v_1} \begin{bmatrix} A_{22}^{v_2} & A_{23}^{v_2} \\ A_{32}^{v_2} & A_{33}^{v_2} \end{bmatrix} (T^{v_1})^+ = \begin{bmatrix} \hat{A}_{22}^{v_2} & \hat{A}_{23}^{v_2} \\ \hat{A}_{32}^{v_2} & \hat{A}_{33}^{v_2} \end{bmatrix} \quad T^{v_1} \begin{bmatrix} K_{22}^{v_2} & K_{23}^{v_2} \\ K_{32}^{v_2} & K_{33}^{v_2} \end{bmatrix} = \begin{bmatrix} \hat{K}_{22}^{v_2} & \hat{K}_{23}^{v_2} \\ \hat{K}_{32}^{v_2} & \hat{K}_{33}^{v_2} \end{bmatrix}$$

$$\begin{bmatrix} C_{22}^{v_2} & C_{23}^{v_2} \\ C_{32}^{v_2} & C_{33}^{v_2} \end{bmatrix} (T^{v_1})^+ = \begin{bmatrix} \hat{C}_{22}^{v_2} & \hat{C}_{23}^{v_2} \\ \hat{C}_{32}^{v_2} & \hat{C}_{33}^{v_2} \end{bmatrix}$$

and define the matrices

$$A^{v_2} = \begin{bmatrix} A_{11}^{v_2} & [A_{12}^{v_2} \; A_{13}^{v_2}]T^{v_1} \\ 0 & \hat{A}_{22}^{v_2} \; \hat{A}_{23}^{v_2} \\ 0 & 0 \quad \hat{A}_{33}^{v_2} \end{bmatrix} \; K^{v_2} = \begin{bmatrix} K_{11}^{v_2} & K_{12}^{v_2} & K_{13}^{v_2} \\ 0 & \hat{K}_{22}^{v_2} & \hat{K}_{23}^{v_2} \\ 0 & 0 & \hat{K}_{33}^{v_2} \end{bmatrix} \; C^{v_2} = \begin{bmatrix} C_{11}^{v_2} & [C_{12}^{v_2} \; C_{13}^{v_2}]T^{v_1} \\ 0 & \hat{C}_{22}^{v_2} \; \hat{C}_{23}^{v_2} \\ 0 & 0 \quad \hat{C}_{33}^{v_2} \end{bmatrix}.$$

Step 6 Define the deterministic systems $\hat{S} = (A_{v_1}, K_{v_1}, C_{v_1}, I), S = (A_{v_2}, K_{v_2}, C_{v_2}, I)$ and $S^* = (A^*, K^*, C^*, I)$. If $\varepsilon_{sys}^{H_2}(\hat{S}, S^*) \leqslant \epsilon_{sys}^{H_2}(S, S^*)$ then define $A = A_{v_1}$, $K = K_{v_1}$ and $C = C_{v_1}$. Otherwise, define $A = A_{v_2}$, $K = K_{v_2}$ and $C = C_{v_2}$.

The idea of Algorithm 18 is as follows: If there exists a minimal Kalman representation of $\mathbf{y}$ with causal G-zero structure then Algorithm 18 first calculates system matrices of minimal Kalman representations of $[\mathbf{y}_2^T, \mathbf{y}_3^T]^T$ and of $[\mathbf{y}_1^T, [\mathbf{y}_2^T, \mathbf{y}_3^T]^T]^T$ in causal block triangular form by using Algorithm 14. Then, it calculates the system matrices of a Kalman representation of $\mathbf{y}$ with causal G-zero structure in

two ways: First, the system matrices of the Kalman representation of $[\mathbf{y}_2^T, \mathbf{y}_3^T]^T$ are combined with the Kalman representation of $[\mathbf{y}_1^T, [\mathbf{y}_2^T, \mathbf{y}_3^T]^T]^T$. Second, the Kalman representation of $[\mathbf{y}_1^T, [\mathbf{y}_2^T, \mathbf{y}_3^T]^T]^T$ is transformed to have causal G-zero structure. To decide which estimates to choose, we take a minimal Kalman representation $(A^*, K^*, C^*, I, \mathbf{e}, \mathbf{y})$, and compare the two sets of estimates to the matrices A^*, K^*, C^* using the relative H_2-error defined in (7.2). The latter step helps to reformulate Algorithm 18 as an identification algorithm and it defines the output matrices in a way that they correspond to the output matrices of Algorithm 7.

Remark 7.4. Assume that (i) and (ii) in Theorem 4.15 hold and let the input of Algorithm 18 be as follows: $\{A^*, K^*, C^*\}$ are system matrices of a minimal Kalman representation of $\mathbf{y}$, $\{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2N}$ is the covariance sequence of $\mathbf{y}$ and $\{r_i, n_i\}_{i=1}^3$ are the dimension of the output and state components of a minimal Kalman representation of $\mathbf{y}$ with causal G-zero structure. Then, the output matrices of Algorithm 18 define system matrices of a minimal Kalman representation of $\mathbf{y}$ in causal G-zero structure that correspond to the output matrices of Algorithm 11, see also Remark 7.1.

Next, we present Algorithm 19 on estimating system matrices of a minimal Kalman representation of $\mathbf{y}$ with G-zero structure from a finite sample $\{y(i)\}_{i=1}^N$ of $\mathbf{y}$ and the dimensions $\{r_i, n_i\}_{i=1}^3$. Algorithm 19 is based on the idea of applying Algorithm 18 with empirical inputs. We will rely on Algorithm 19 in the rest of the chapter to estimate system matrices of minimal Kalman representations with causal G-zero structure.

Algorithm 19 Estimating system matrices of minimal Kalman representation with causal G-zero structure

Input $\{y(i)\}_{i=1}^N$ and $\{r_i, n_i\}_{i=1}^3$: Finite sample of $\mathbf{y}$ and output and state dimensions of a minimal Kalman representation of $\mathbf{y}$ with causal G-zero structure
Output $\{A, K, C\}$: Estimate for system matrices of a minimal Kalman representation of $\mathbf{y}$ with causal G-zero structure

Step 1 Calculate the empirical covariances $\{\Lambda_k^y\}_{k=0}^{2n}$, by using the MATLAB function `cov`, see Remark 7.2.

Step 2 Estimate the system matrices of a minimal Kalman representation of $\mathbf{y}$ and $[\mathbf{y}_2^T, \mathbf{y}_3^T]$ by applying the MATLAB function `s4sid_CCA` with input $\{\Lambda_k^y\}_{k=0}^{2n}$ and $\{\Lambda_k^{y_{2,3}}\}_{k=0}^{2(n_2+n_3)}$, where $\Lambda_k^{y_{2,3}}$ are empirical covariances of $[\mathbf{y}_2^T, \mathbf{y}_3^T]^T$. Denote the output matrices by $\{A^*, K^*, C^*\}$ and $\{A_{23}^*, K_{23}^*, C_{23}^*\}$, respectively.

Step 3 Apply Algorithm 14 with input $\{A_{23}^*, K_{23}^*, C_{23}^*\}, \{\Lambda_k^{y_{2,3}}\}_{k=0}^{2(n_2+n_3)}$ and $\{r_2, r_3, n_2 + n_3, n_3\}$ such that for estimating system matrices of minimal Kalman representations in Step 1 and 4, use `s4sid_CCA` with the corresponding empirical

covariances as its input. Denote the output as in (7.5).
Step 4 Apply Algorithm 14 with input $\{A^*, K^*, C^*\}, \{\Lambda_k^y\}_{k=0}^{2n}$ and $\{r_1, r_2 + r_3, n, n_2 + n_3\}$ such that for estimating system matrices of minimal Kalman representations in Step 1 and 4, use `s4sid_CCA` with the corresponding empirical covariances as its input. Denote the output as in (7.6).
Step 5 Steps 4–5 and 6 of Algorithm 18.

7.2 Granger causality test

In this section, we introduce a statistical test for checking Granger causality among the components of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T \in \mathbb{R}^r$ based on an N-long sample $\{y(i)\}_{i=1}^N$ of $\mathbf{y}$, i.e., a sample of $\mathbf{y}(t), \ldots, \mathbf{y}(t+N-1)$ for a $t \in \mathbb{Z}$. Then the goal of this section is to present a method for deciding whether $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$ based on $\{y(i)\}_{i=1}^N$.

We assume that $\mathbf{y}_1 \in \mathbb{R}^{r_1}$ and $\mathbf{y}_2 \in \mathbb{R}^{r_2}$. Furthermore, assume that the dimension of a minimal Kalman representation of $\mathbf{y}$ is n, and that the dimension of a minimal Kalman representation of $\mathbf{y}_2$ is n_2. The *null hypothesis* H_0 of the test is that $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$, whereas the *alternative hypothesis* H_1 is that $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$. The *sample space* consists of all the possible N-long samples of $\mathbf{y}$, i.e., it is in $\mathbb{R}^{r\times N}$, and the actual sample that we apply the test for is an N-long sample $\{y(i)\}_{i=1}^N$ of $\mathbf{y}$.

The *parameter space* of the test is the parametrization of all n-dimensional minimal Kalman representations of Gaussian r-dimensional processes in the following way: Denoting the parameter space by $\Theta \subseteq \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times r} \times \mathbb{R}^{r\times n} \times \mathbb{R}^{r\times r}$, an element $\theta \in \Theta$ is a tuple $(A, K, C, Q^{\tilde{\mathbf{e}}})$ that parametrizes a minimal Kalman representation $(A, K, C, I, \tilde{\mathbf{e}}, \tilde{\mathbf{y}})$, such that $E[\tilde{\mathbf{e}}(t)\tilde{\mathbf{e}}^T(t)] = Q^{\tilde{\mathbf{e}}}$. Notice that these parameters of Kalman representations uniquely define the covariance sequence of the output process and every two parameters in Θ that correspond to isometric Kalman representations define the same covariance sequence of the output process. The null hypotheses partitions Θ into two disjoint subsets, the *null parameter space* Θ_0 and the *alternative parameter space* $\Theta_1 = \Theta \backslash \Theta_0$, where Θ_0 is as follows:

$\Theta_0 := \{n$-dimensional minimal Kalman representations that can be transformed by isomorphism into a minimal Kalman representation in causal block triangular form, where the state components have dimensions n_1 and n_2, such that n_2 is the dimension of a minimal Kalman representation of $\mathbf{y}_2$ and $n = n_1 + n_2\}$

In the parameter space we distinguish so-called *true parameters*: a parameter $\theta \in \Theta$ is called true if it defines parameters of a minimal Kalman representation of $\mathbf{y}$. It is important to note that due to the isomorphism between any two minimal Kalman

representations of $\mathbf{y}$, the set of true parameters are all in H_0 or they are all in H_1, it is not possible that one true parameter is in H_0 and another is in H_1. The null hypothesis holds if and only if the set of true parameter are all in Θ_0, i.e., when any minimal Kalman representation of $\mathbf{y}$ can be transformed into a causal block triangular form. Otherwise, the alternative hypothesis holds.

The core of hypothesis testing is to define numbers $\mathcal{F}_{\min}$ and $\mathcal{F}_{\max}$ and a measurable function $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2} : \mathcal{R}^{rN} \to [0,1]$ such that the probability that the random variable $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\})$ is out of the interval $(\mathcal{F}_{\min}, \mathcal{F}_{\max}) \subseteq \mathbb{R}$ under the assumption that the statistical hypothesis is true is less than a predefined number $\alpha \in (0,1)$, called *significance level*. If we knew the distribution $P_\theta(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) < x), x \in \mathbb{R}$ of $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\})$ for any fixed $\theta \in \Theta_0$, then this would mean to define numbers $\mathcal{F}_{\min}$ and $\mathcal{F}_{\max}$ such that

$$\sup_{\theta\in\Theta_0} P_\theta\left(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) \in \mathbb{R}\backslash(\mathcal{F}_{\min}, \mathcal{F}_{\max})\right) \leqslant \alpha. \tag{7.7}$$

The null hypothesis is accepted if $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) \in (\mathcal{F}_{\min}, \mathcal{F}_{\max})$, otherwise, it is rejected. The choice of $\mathcal{F}_{\min}$ and $\mathcal{F}_{\max}$ ensures that the probability of type I error, i.e., rejecting the null hypothesis while it is true, is less than α. For a good hypothesis testing, the type II error, i.e., accepting the null hypothesis while it is wrong, should also be small. This means that we have to choose the function $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ in such a way that

$$\sup_{\theta\in\Theta_1} P_\theta\left(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) \in \mathbb{R}\backslash(\mathcal{F}_{\min}, \mathcal{F}_{\max})\right)$$

is not too high.

In practice, the choice of $\mathcal{F}_{\min}$ and $\mathcal{F}_{\min}$ is further complicated by the fact that there is usually no analytic expression for

$$P_\theta\left(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) < x\right), \quad x \in \mathbb{R},$$

where $\theta \in \Theta_0$, so it has to be estimated from data.

In this thesis we estimate $P_{\theta*}\left(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) < x\right)$ for a particular $\theta^* \in \Theta_0$ instead of estimating the supremum for all $\theta \in \Theta_0$. This particular $\theta^* \in \Theta_0$ is defined by a function $\hat{\theta} : \mathbb{R}^{rN} \to \Theta_0$ such that $\hat{\theta}(\{y(i)\}_{i=1}^N) = \theta^*$ is intuitively close to the true parameter (i.e., to a parametrization of a minimal Kalman representation of $\mathbf{y}$) if the null hypothesis holds and N is large enough. Then,

$P_{\theta*}\left(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) < x\right)$ is estimated by

$$F_{\theta*}(x) = 1/N_{\mathcal{F}} \sum_{j=1}^{N_{\mathcal{F}}} \chi(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y^j(i)\}_{i=1}^N) < x),$$

where $\{y^j(i)\}_{i=1}^N$, $j = 1, \ldots, N_{\mathcal{F}}$ are N-long samples of the output process of the Kalman representation that θ* defines and $N_{\mathcal{F}}$ is an adjustable parameter. This estimation is expected to be close to $P_{\theta*}(\hat{\mathcal{F}}_{\mathbf{y}\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)\}_{i=1}^N) < x)$ for large enough N and $N_{\mathcal{F}}$.

Then, $\mathcal{F}_{\min}$ and $\mathcal{F}_{\max}$ are chosen so that $F_{\theta*}(\mathcal{F}_{\min}) + (1 - F_{\theta*}(\mathcal{F}_{\max})) < \alpha$, and the hypothesis is accepted, if $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y(i)\}_{i=1}^N) \in (\mathcal{F}_{\min}, \mathcal{F}_{\max})$. In the particular case of this paper, $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ takes values in $[0, 1]$, and $\mathcal{F}_{\min}$ is chosen to be minus infinity for practical reasons. It then follows that $\mathcal{F}_{\max}$ is chosen so that $F_{\theta*}(\mathcal{F}_{\max}) > 1 - \alpha$, i.e., so that at least $(1-\alpha)N$ numbers in the set $\{\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y^j(i)\}_{i=1}^N)\}_{j=1}^{N_{\mathcal{F}}}$ are smaller than $\mathcal{F}_{\max}$. Instead of choosing $\mathcal{F}_{\max}$ explicitly, we could also say that we accept the hypothesis, if there exists at least $(1-\alpha)N$ numbers in the set $\{\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y^j(i)\}_{i=1}^N)\}_{j=1}^{N_{\mathcal{F}}}$ which are greater than $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y(i)\}_{i=1}^N)$.

This section is organized as follows: In Subsection 7.2.1 we explain how we choose the statistics $\hat{\mathcal{F}}_{\mathbf{y}\to\mathbf{y}_2}$. Next, in Subsection 7.2.2 we explain how we estimate the probability distribution $P_\theta(\hat{\mathcal{F}}_{\mathbf{y}\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)\}_{i=1}^N) < x)$ from data. Finally, Subsection 7.2.3 describes in more details the decision algorithm on the hypothesis.

7.2.1 Theoretical and empirical Geweke–Granger causality

In this section, we define the statistics $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ used for hypothesis testing. We will call the function $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ the *empirical Geweke-Granger causality*. The reason for that is the following. The function $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$, to be defined later, can be viewed as an estimate of the so called Geweke-Granger causality. The latter is defined as follows.

Consider the process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T \in \mathbb{R}^r$, the innovation process $\mathbf{e}$ of $\mathbf{y}$ and the innovation process $\mathbf{e}_2$ of $\mathbf{y}_2 \in \mathbb{R}^{r_2}$. Denote the right-bottom $r_2 \times r_2$ sub-matrix of the variance matrix $E[\mathbf{e}(t)\mathbf{e}^T(t)]$ by $\Sigma_{22}^{\mathbf{e}}$. Then, the (time-domain) Geweke–Granger causality ((Geweke, 1984; Barnett and Seth, 2014)) is defined as follows:

Definition 7.5. Geweke–Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$ is given by

$$\mathcal{F}_{\mathbf{y}_1\to\mathbf{y}_2} = \ln \frac{|E[\mathbf{e}_2(t)\mathbf{e}_2^T(t)]|}{|\Sigma_{22}^{\mathbf{e}}|}, \tag{7.8}$$

where $|.|$ denotes the determinant of a matrix.

Geweke–Granger causality is a non-negative real number between zero and one, contrary to Granger causality, which is either zero or one. The connection between the two is that Geweke–Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$ equals zero if and only if $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ and is positive if and only if $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$. To check whether $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$ based on a finite sample $\{y(i)\}_{i=1}^N$ of $\mathbf{y}$, it is enough to check whether Geweke–Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$ is positive. Then the value $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y(i)\}_{i=1}^N)$, defined below, could be viewed as an approximation of $\mathcal{F}_{\mathbf{y}_1\to\mathbf{y}_2}$ calculated from samples. Accordingly, the proposed hypothesis testing boils down to deciding if the observed value $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y(i)\}_{i=1}^N)$ is small enough and whether observing this value is a statistically significant event.

The statistics $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ is defined as the output of Algorithm 20 below, applied to a finite sample $\{y(i)\}_{i=1}^N$ of $\mathbf{y}$.

Intuitively, Algorithm 20 estimates the system matrices of minimal Kalman representations $(A, K, C, I, \mathbf{e})$ of $\mathbf{y}$ and $(A_{22}, K_{22}, C_{22}, I, \mathbf{e}_2)$ of $\mathbf{y}_2$, including the variance matrices of $\mathbf{e}$ and $\mathbf{e}_2$. Then, it compares the variance matrices of $\mathbf{e}$ and $\mathbf{e}_2$ based on (7.8).

Algorithm 20 Empirical Geweke–Granger causality

Input $\{y(i)\}_{i=1}^N$, $\{r_1, r_2, n, n_2\}$: Finite sample of $\mathbf{y}$, dimensions of $\mathbf{y}_1, \mathbf{y}_2$ and dimensions of minimal Kalman representations of $\mathbf{y}$ and $\mathbf{y}_2$

Output $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y(i)\}_{i=1}^N)$: Empirical Geweke-Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$

Step 1 Calculate the empirical covariances $\{\Lambda_k^y\}_{k=0}^{2n}$ of $\mathbf{y}$ and $\{\Lambda_k^{y_2}\}_{k=0}^{2n_2}$ of $\mathbf{y}_2$, see (7.4).

Step 2 Estimate the system matrices of a minimal Kalman representation of $\mathbf{y}$ and of $\mathbf{y}_2$ by applying the MATLAB function `s4sid_CCA` from the SSGC toolbox (see (Barnett and Seth, 2015)) with input $\{\Lambda_k^y\}_{k=0}^{2n}$ and $\{\Lambda_k^{y_2}\}_{k=0}^{2n_2}$, respectively. Denote the system matrices by $\{\hat{A}, \hat{K}, \hat{C}, \hat{Q}\}$ and $\{\hat{A}_{22}, \hat{K}_{22}, \hat{C}_{22}, \hat{Q}_{22}\}$, respectively, where $\hat{Q}$ and $\hat{Q}_{22}$ are estimates of the noise variance matrices.

Step 3 Denote the $r_2 \times r_2$ right-bottom sub-matrix of the variance matrix $\hat{Q}$ by $\Sigma_{22}^{\mathbf{e}}$ and calculate $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2} = \ln\left(|\hat{Q}_{22}|/|\Sigma_{22}^{\mathbf{e}}|\right)$, see also (7.8).

We conjecture that the empirical Geweke–Granger causality $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ converges to Geweke–Granger causality $\mathcal{F}_{\mathbf{y}_1\to\mathbf{y}_2}$ as the number N of output samples go to infinity. This is because the more output samples are used, the more accurate estimate we obtain for a minimal Kalman representation of $\mathbf{y}$ in Step 2 of Algorithm 20. Notice that if in Step 2 of Algorithm 20 we use an exact minimal Kalman representation of $\mathbf{y}$ then Algorithm 20 calculates the real Geweke–Granger causality $\mathcal{F}_{\mathbf{y}_1\to\mathbf{y}_2}$.

7.2.2 Estimating the distribution of empirical Geweke–Granger causality

Next, we present an algorithm for estimating the probability distribution of the statistics $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$. More precisely, as it was explained above, we will define a function $\hat{\theta} : \mathbb{R}^{rN} \to \Theta_0$ such that $\hat{\theta}(\{y(i)\}_{i=1}^N) = \theta^*$ is intuitively close to the true parameter (i.e., to a parametrization of a minimal Kalman representation of $\mathbf{y}$) if the null hypothesis holds and N is large enough. Then, we estimate

$$P_{\theta*}\left(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{\mathbf{y}(t+i-1)_{i=1}^N\}) < x\right)$$

by a function

$$F_{\theta*}(x) = 1/N_{\mathcal{F}} \sum_{j=1}^{N_{\mathcal{F}}} \chi(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y^j(i)\}_{i=1}^N) < x),$$

where $\{y^j(i)\}_{i=1}^N$, $j = 1,\dots,N_{\mathcal{F}}$ are N-long samples of an output process of the Kalman representation that θ^* defines and $N_{\mathcal{F}}$ is an adjustable parameter. If the true parameter belongs to Θ_0, i.e., the null hypothesis holds, and N, $N_{\mathcal{F}}$ are large enough, then $F_{\theta*}(x)$ is close to $P_{\theta*}(\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}(\{y(i)\}_{i=1}^N < x)$.

The main idea behind calculating $F_{\theta*}(x)$ is the following. In general, if the null hypothesis was true, and a Kalman representation of $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ in causal block triangular form and the distribution of its innovation process $\mathbf{e}$ were known, then we would calculate the empirical distribution of $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ as follows: From random samples of $\mathbf{e}$ we generate N-long samples of $\mathbf{y}$, using its known Kalman representation. Then, we apply Algorithm 20 to N-long samples of $\mathbf{y}$ to obtain $N_{\mathcal{F}}$ number of new empirical Geweke–Granger causalities $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$. These numbers define the empirical distribution

$$F(x) = 1/N_{\mathcal{F}} \sum_{i=1}^{N_{\mathcal{F}}} \chi(\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2} < x)$$

of the empirical Geweke–Granger causality $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$. Since in practice, neither a Kalman representation of $\mathbf{y}$ nor the distribution of $\mathbf{e}$ is known, we rely on an estimated Kalman representation $\hat{\theta}(\{y(i)\}_{i=1}^N) = \theta^*$ of $\mathbf{y}$, which we choose to be in causal block triangular form, and an estimated distribution of $\mathbf{e}$. For the distribution of $\mathbf{e}$, we restrict ourselves to Gaussian innovation processes, i.e., $\mathbf{e}$ is a Gaussian white noise. Note that if $\mathbf{e}$ were not Gaussian and its distribution were unknown then the distribution of $\mathbf{e}$ could be estimated from $\{y(i)\}_{i=1}^N$ by calculating sam-

ples of $\mathbf{e}$ based on the equation $\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t)|\mathcal{H}^{\mathbf{y}}_{t-}]$, where the projection $E_l[\mathbf{y}(t)|\mathcal{H}^{\mathbf{y}}_{t-}]$ can be approximated with a projection on a Hilbert space generated by the finite past $\mathbf{y}(t-1), \ldots, \mathbf{y}(t-K)$ of $\mathbf{y}$ for some $K > 0$.

The intuition explained above is summarized in Algorithm 21. The inputs of Algorithm 21 are $\{y(i)\}_{i=1}^{N}, N_{\mathcal{F}}, N_{\text{training}}$ and $\{r_1, r_2\}$, where $N_{\mathcal{F}}$ and N_{training} are adjustable parameters. The result of Algorithm 21 is the distribution $F_{\theta *}(x) = 1/N_{\mathcal{F}} \sum_{i=1}^{N_{\mathcal{F}}} \chi(\hat{\mathcal{F}}^i_{\mathbf{y}_1 \to \mathbf{y}_2} < x)$ that is obtained from the output $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ of Algorithm 21.

Algorithm 21 Calculating samples for the empirical distribution of zero Geweke–Granger causality

Input $\{y(i)\}_{i=1}^{N}, N_{\mathcal{F}}, N_{\text{training}}$ and $\{r_1, r_2\}$: Finite sample of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T \in \mathbb{R}^r$, required number of estimates of Geweke–Granger causality in the output, length of training samples, dimensions of $\mathbf{y}_1$, and $\mathbf{y}_2$

Output $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$: Estimated Geweke-Granger causalities from $\mathbf{y}_1$ to $\mathbf{y}_2$

Step 1 Estimate minimal Kalman representations of $\mathbf{y}$ and $\mathbf{y}_2$ by applying the MATLAB function `s4sid_CCA` to the empirical covariances of $\mathbf{y}$ and $\mathbf{y}_2$, respectively. Let $\Sigma^{\hat{\mathbf{e}}}$ be the estimated variance of the innovation process of $\mathbf{y}$. Denote the dimensions of the estimated minimal Kalman representations of $\mathbf{y}$ and $\mathbf{y}_2$ by n and n_2, respectively.

Step 2 Apply Algorithm 15 with input $\{y(i)\}_{i=1}^{N}$ and $\{r_1, r_2, n, n_2\}$ and denote its output by $\hat{A}, \hat{K}, \hat{C}$.

for $i = 1 : N_{\mathcal{F}}$

Step 3 Generate $N + N_{\text{training}}$ independent samples from a Gaussian distribution with zero mean and $\Sigma^{\hat{\mathbf{e}}}$ variance. Denote this sample by $\{e(i)\}_{i=1}^{N+N_{\text{training}}}$.

Step 4 By using a zero initial state vector $x(1) = 0$ and the first N_{training} generated samples of $\mathbf{e}$ as input, calculate the vectors $x(2), \ldots, x(N_{\text{training}})$ from the equations $x(i+1) = \hat{A}x(i) + \hat{K}e(i)$, $i = 1, \ldots, N_{\text{training}}$.

Step 5 Calculate $\{y(t)\}_{t=N_{\text{training}}+1}^{N_{\text{training}}+N}$ based on the equations $x(t+1) = \hat{A}x(t) + \hat{K}e(t)$ and $\hat{y}(t) = \hat{C}x(t) + e(t)$, $t = N_{\text{training}}, \ldots, N + N_{\text{training}}$ and let $y(i) = \hat{y}(N_{\text{training}} + i)$ for $i = 1, \ldots, N$.

Step 6 Apply Algorithm 20 with input $\{y(i)\}_{i=1}^{N}$ and denote the output by $\hat{\mathcal{F}}^i_{\mathbf{y}_1 \to \mathbf{y}_2}$.

end for

Algorithm 21 uses a so-called bootstrapping method where a distribution is estimated by random sampling. The idea of using bootstrapping method for estimating the distribution of Geweke–Granger causality in LTI–SS representations is not new,

see (Barnett and Seth, 2015). However, to the best of the authors' knowledge, this is the first attempt to use Kalman representation in causal block triangular form to estimate empirical values of zero Geweke–Granger causality (Geweke–Granger causality that equals zero) and the corresponding empirical distribution. Note that the method proposed in this section relies on Theorem 2.5.

Algorithm 21 is computationally demanding. A faster method for estimating the empirical distribution of the zero Geweke–Granger causality could be the so-called permutation method (Barnett and Seth, 2014). The latter method permutes the data $\{y(i)\}_{i=1}^{N}$ in time for re-sampling $\mathbf{y}$, instead of generating new data from a representation whose output has the property that the corresponding Geweke-Granger causality is zero. However, after permuting a sample of a stationary process, the new data might lead to an estimated minimal Kalman representation that has completely different properties than a minimal Kalman representation of $\mathbf{y}$, e.g., the dimension of the state. We believe that the above mentioned phenomenon can effect the distribution of the empirical values of Geweke–Granger causality and thus the permutation method might lead to false results. Our method works with an estimated minimal Kalman representation that is calculated from a finite sample of $\mathbf{y}$, and thus the parameters that might influence the distribution of the Geweke–Granger causality are close to the corresponding parameters of minimal Kalman representations of $\mathbf{y}$.

7.2.3 Hypothesis testing for Granger causality

Below, we summarize the procedure for testing the hypothesis of Granger non-causality. The required calculations to prepare the proposed statistical test are:

(i) Apply Algorithm 20 with input $\{y(i)\}_{i=1}^{N}$ and $\{r_1, r_2, n, n_2\}$, where $\{r_1, r_2\}$ are the dimensions of $\mathbf{y}_1, \mathbf{y}_2$ and $\{n, n_2\}$ are the dimensions of minimal Kalman representations of $\mathbf{y}$ and $\mathbf{y}_2$, respectively. Denote the output by $\hat{\mathcal{F}}_{\mathbf{y}_1 \to \mathbf{y}_2}$ which was called empirical Geweke–Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$.

(ii) Let $N_{\mathcal{F}}, N_{\text{training}}$ be positive integers and apply Algorithm 21 with input $\{y(i)\}_{i=1}^{N}$, $N_{\mathcal{F}}, N_{\text{training}}$ and $\{r_1, r_2\}$. Denote the output by $\{\hat{\mathcal{F}}^{i}_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$.

With the help of $\{\hat{\mathcal{F}}^{i}_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$, we test whether $\mathbf{y}_1$ Granger causes $\mathbf{y}_2$ in the following way: Denote the significance level by $\alpha \in (0, 1)$. Then, the null hypothesis is accepted if $\hat{\mathcal{F}}_{\mathbf{y}_1 \to \mathbf{y}_2}$ (see (i) above) is smaller than $\lfloor N/(1-\alpha) \rfloor$ numbers from the set $\{\hat{\mathcal{F}}^{i}_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ (see (ii) above). Otherwise, the null hypothesis is rejected.

Intuitively, we can say that if $\hat{\mathcal{F}}_{\mathbf{y}_1 \to \mathbf{y}_2}$ is smaller than $\lfloor N/(1-\alpha) \rfloor$ numbers from the set $\{\hat{\mathcal{F}}^{i}_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ then the condition that $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ (or equivalently that $\mathcal{F}_{\mathbf{y}_1 \to \mathbf{y}_2} = 0$) is accepted because the empirical Geweke–Granger causal-

ity $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ from $\mathbf{y}_1$ to $\mathbf{y}_2$ is not large enough to be accepted as a statistical proof against the null hypothesis

Test 1 Statistical test of Geweke–Granger causality from $\mathbf{y}_1$ to $\mathbf{y}_2$

Input $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$, $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$, α: Empirical Geweke-Granger causality, samples for the empirical distribution of empirical Geweke–Granger causality of zero Geweke–Granger causality, significance level

Output D: Decision on the null hypothesis

If $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ is smaller than $\lfloor N/(1-\alpha)\rfloor$ numbers from the set $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ then $D =$ Accept.

Else $D =$ Reject.

There are two types of error of Test 1 that we will consider in the coming sections: the first one is called type I error and it refers to the case when H_0 is rejected even though the true parameters are in Θ_0. The second one is called type II error and it refers to the case when H_0 is accepted even though the true parameters are in Θ_1.

For large N, the estimation of the empirical Geweke–Granger causality $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ from $\mathbf{y}_1$ to $\mathbf{y}_2$ is close to the Geweke–Granger causality $\mathcal{F}_{\mathbf{y}_1\to\mathbf{y}_2}$ from $\mathbf{y}_1$ to $\mathbf{y}_2$. Furthermore, since we calculate $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ from N-long samples of an output process of a minimal Kalman representation in causal block triangular form, that is, they are empirical values of zero Geweke–Granger causality, the larger N is, the closer the numbers $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ are to zero. Based on this, we expect the following: if the true parameters are from Θ_0 then both $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ and $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ are empirical values of the same zero Geweke–Granger causality (although $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ are calculated in an indirect way). Therefore, $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ gives samples for the empirical distribution of $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$. In this case, we expect that the probability of the occurrence of type I error of the test approximates α for large enough N and $N_{\mathcal{F}}$. Note that for the simulations presented later on in this chapter, the significance level α is always 0.01. If the true parameters are from Θ_1 then $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ and $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ are empirical values of different Geweke–Granger causalities, $\hat{\mathcal{F}}_{\mathbf{y}_1\to\mathbf{y}_2}$ estimates a non-zero and $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1\to\mathbf{y}_2}\}_{i=1}^{N_{\mathcal{F}}}$ a zero Geweke–Granger causality. In this case, we expect that the empirical value of the zero Geweke–Granger causality is smaller for larger enough N than the empirical values of the non-zero (positive) Geweke–Granger causality. In other words, we expect that the distribution of the two overlap less (or does not overlap) as we increase N and thus that the type II error is small.

7.3 Evaluation of the estimated Kalman representations

In this section, we discuss how we evaluate the accuracy of the estimates of the Kalman representations with different network graphs that are calculated from Algorithms 15, 17, and 19. More precisely, we introduce measures that help us in the evaluation and which compare the above-mentioned estimates to estimates that are obtained by using classical methods (subspace and prediction error methods).

Consider a process $\mathbf{y} \in \mathbb{R}^r$, an n-dimensional minimal Kalman representation $S = (A, K, C, I, \mathbf{e})$ of $\mathbf{y}$ and a covariance sequence $\{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$ of $\mathbf{y}$. Furthermore, let $\{y(i)\}_{i=1}^N$ be an N-long sample of $\mathbf{y}$ and $\{\Lambda^y\}_{k=0}^{2n}$ be the empirical covariances of $\mathbf{y}$ calculated from $\{y(i)\}_{i=1}^N$, see (7.4). Apply the MATLAB function `s4sid_CCA` from the SSGC toolbox with input $\{\Lambda^y\}_{k=0}^{2n}$ and denote its output matrices by $\hat{A}, \hat{K}, \hat{C}$ and $Q^{\hat{\mathbf{e}}}$. Define an estimate $\hat{S} = (\hat{A}, \hat{K}, \hat{C}, I, \hat{\mathbf{e}})$ of S, where $E[\hat{\mathbf{e}}(t)(\hat{\mathbf{e}}(t))^T] = Q^{\hat{\mathbf{e}}}$. For a second estimate of S, apply one of the algorithms from Algorithms 15, or 17 or 19 in order to obtain matrices $\bar{A}, \bar{K}, \bar{C}$ of an estimated Kalman representation in causal block triangular form, in causal coordinated form or with causal G-zero structure, where $G = (V, E)$ is a TADG graph with $V = \{1, 2, 3\}$ and $E = \{(3, 1), (3, 2), (2, 1)\}$. Define the second estimate for S by $\bar{S} = (\bar{A}, \bar{K}, \bar{C}, I, \hat{\mathbf{e}})$. Then, the output covariances of $\hat{S}$ and $\bar{S}$ for lags $k = 0, \ldots, 2n$ are given as follows

$$\begin{aligned} \hat{\Lambda}_k^y &= \hat{C}\hat{A}^{k-1}\left(\hat{K}Q^{\hat{\mathbf{e}}} + \hat{A}\hat{P}\hat{C}\right) \\ \bar{\Lambda}_k^y &= \bar{C}\bar{A}^{k-1}\left(\bar{K}Q^{\hat{\mathbf{e}}} + \bar{A}\bar{P}\bar{C}\right), \end{aligned} \tag{7.9}$$

where $\hat{P}$ and $\bar{P}$ are the solutions of the Lyapunov equations

$$\hat{P} = \hat{A}\hat{P}\hat{A}^T + Q^{\hat{\mathbf{e}}}\hat{K}^T \qquad \bar{P} = \bar{A}\bar{P}\bar{A}^T + Q^{\hat{\mathbf{e}}}\bar{K}^T, \tag{7.10}$$

respectively. By using the covariances $\{\Lambda_k^{\mathbf{y}}\}_{k=0}^{2n}$, $\{\hat{\Lambda}_k^y\}_{k=0}^{2n}$ and $\{\bar{\Lambda}_k^y\}_{k=0}^{2n}$, we define the following measures

$$\hat{\varepsilon}_{\text{sys}}^{\Lambda^{\mathbf{y}}} = \frac{1}{H_{2n+1}} \sum_{k=0}^{2n} \frac{1}{k+1} \frac{||\hat{\Lambda}_k^y - \Lambda_k^{\mathbf{y}}||_F}{||\Lambda_k^{\mathbf{y}}||_F} \tag{7.11}$$

$$\bar{\varepsilon}_{\text{sys}}^{\Lambda^{\mathbf{y}}} = \frac{1}{H_{2n+1}} \sum_{k=0}^{2n} \frac{1}{k+1} \frac{||\bar{\Lambda}_k^y - \Lambda_k^{\mathbf{y}}||_F}{||\Lambda_k^{\mathbf{y}}||_F}, \tag{7.12}$$

where $||.||_F$ denotes the Frobenius norm of a matrix and $H_{2n+1} = \sum_{k=0}^{2n} \frac{1}{k+1}$ is the $(2n+1)$th Harmonic number. The numbers $\hat{\varepsilon}_{\text{sys}}^{\Lambda^{\mathbf{y}}}$ and $\bar{\varepsilon}_{\text{sys}}^{\Lambda^{\mathbf{y}}}$ serve to measure how well

the representations $\hat{S}$ and $\bar{S}$ reproduce the true covariances of $\mathbf{y}$, or in other words, how well they estimate S.

Besides $\hat{S}$ and $\bar{S}$ we also estimate the system matrices of minimal Kalman representations that are obtained from prediction error method. Correspondingly, for evaluating the accuracy of these estimate Kalman representations, we define similar measures to $\hat{\varepsilon}^{\Lambda^{\mathbf{y}}}_{\text{sys}}$ and $\bar{\varepsilon}^{\Lambda^{\mathbf{y}}}_{\text{sys}}$, see (7.13) below. The prediction error method is elaborated as follows: we apply the MATLAB function `pem` (see (Ljung, 1999, Chapter 7)) with input being $\{y(i)\}_{i=1}^{N}$ and as an initial representation, we set $\hat{S}$, i.e., the estimate Kalman representation that is obtained from the MATLAB function `s4sid_CCA`. Denote the output of the function `pem` by $\dot{A}, \dot{K}, \dot{C}$ and $Q^{\dot{\mathbf{e}}}$ and define the estimate of S by $\dot{S} = (\dot{A}, \dot{K}, \dot{C}, I, \dot{\mathbf{e}})$, where $\dot{\mathbf{e}}$ is a Gaussian white noise with variance matrix $Q^{\dot{\mathbf{e}}}$. Furthermore, denote the corresponding output covariances by $\{\dot{\Lambda}^{y}_{k}\}_{k=0}^{2n}$ that are calculated in the same manner as the output covariances corresponding to S and $\hat{S}$ in (7.9) and (7.10). The following measure is considered

$$\dot{\varepsilon}^{\Lambda^{\mathbf{y}}}_{\text{sys}} = \frac{1}{H_{2n+1}} \sum_{k=0}^{2n} \frac{1}{k+1} \frac{||\dot{\Lambda}^{y}_{k} - \Lambda^{\mathbf{y}}_{k}||_F}{||\Lambda^{\mathbf{y}}_{k}||_F}. \tag{7.13}$$

In addition to $\hat{\varepsilon}^{\Lambda^{\mathbf{y}}}_{\text{sys}}$, $\bar{\varepsilon}^{\Lambda^{\mathbf{y}}}_{\text{sys}}$ and $\dot{\varepsilon}^{\Lambda^{\mathbf{y}}}_{\text{sys}}$, we use the measures below that define relative H_2-norms between the deterministic state-space systems $S^{\text{det}} = (A, K, C, I)$, $\hat{S}^{\text{det}} = (\hat{A}, \hat{K}, \hat{C}, I)$, $\bar{S}^{\text{det}} = (\bar{A}, \bar{K}, \bar{C}, I)$ and $\dot{S}^{\text{det}} = (\dot{A}, \dot{K}, \dot{C}, I)$, see also (7.1).

$$\hat{\varepsilon}^{H_2}_{\text{sys}} = \frac{||\hat{S}^{\text{det}} - S^{\text{det}}||_{H_2}}{||S^{\text{det}}||_{H_2}} \quad \bar{\varepsilon}^{H_2}_{\text{sys}} = \frac{||\bar{S}^{\text{det}} - S^{\text{det}}||_{H_2}}{||S^{\text{det}}||_{H_2}} \quad \dot{\varepsilon}^{H_2}_{\text{sys}} = \frac{||\dot{S}^{\text{det}} - S^{\text{det}}||_{H_2}}{||S^{\text{det}}||_{H_2}}. \tag{7.14}$$

The measures defined in (7.11), (7.12), (7.13) and (7.14) will help us in evaluating the accuracy of the different estimations of minimal Kalman representations that are calculated from Algorithms 15, 17 or 19.

7.4 Simulation

In this section, we present simulation results on estimating Kalman representations with different network graphs. Before presenting the results, we first summarize the method of simulation and the calculations on the simulated data.

7.4.1 Calculation step-by-step

Reference representation: A representation that is used for simulating data is called *reference representation*. We consider reference representations that are randomly gen-

erated minimal Kalman representations in causal block triangular form, in causal coordinated form and with causal G-zero structure, where $G = (V, E)$, such that $V = \{1, 2, 3\}$, $E = \{(3, 1), (3, 2), (2, 1)\}$. To obtain a reference representation, we first decide on the dimensions $\{r_i, n_i\}_{i=1}^{l}$ of the components of the output and state processes (where $l = 2$ if the reference representation is set to be in block triangular form and $l = 3$ otherwise) and choose a random variance matrix $\Lambda_0^{\mathbf{e}}$ for the innovation process by the MATLAB function `cov_rand` from the SSGC toolbox. Note that the innovation correlation factor that is required as an input for the function `cov_rand` is set to 10. By these parameters, we generate a random minimal Kalman representation with the help of the MATLAB function `iss_rand` from the SSGC toolbox. The state transition matrix spectral norm that is required as an input for `iss_rand` is chosen to be 0.9. Denote the randomly generated minimal Kalman representation by S.

The reference representation is calculated from S as follows: Depending on the network graph that we aim to have for the reference representation, we apply Algorithm 14, 16, or 18 with the input being the output covariances of S (see (7.9)) and $\{r_i, n_i\}_{i=1}^{l}$. Denote the output matrices of the algorithm that was applied by A, K and C. If the matrices A, K and C define a stable, minimum phase minimal Kalman representation then the reference representation is $(A, K, C, I, \mathbf{e}, \mathbf{y})$. Otherwise we regenerate S and recalculate A, K and C.

Data simulation: Let $\mathcal{S} = (A, K, C, I, \mathbf{e}, \mathbf{y})$ be the reference representation, where $\mathbf{e}$ is a Gaussian white noise process with variance matrix $\Lambda_0^{\mathbf{e}}$ and let $\{e(t)\}_{t=1}^{10000}$ be a finite sample of $\mathbf{e}$. Then, for $t = 0, \ldots, 10000$ and with $x(0) = 0$ we calculate

$$\begin{aligned} x(t+1) &= \hat{A}x(t) + \hat{K}e(t) \\ y(t) &= \hat{C}x(t) + e(t), \end{aligned} \tag{7.15}$$

and we set $x(10000)$ as the initial state $x^* = x(10000)$. The data $\{y(s)\}_{s=1}^{N}$ is then generated based on (7.15) for $t = 0, \ldots, N$ with $x(0) = x^*$ and with an independently generated N-long sample $\{e(t)\}_{t=1}^{N}$ of $\mathbf{e}$.

Stationarity test: The simulated data is tested for stationarity by the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test with the help of the function `mvgc_kpss` of the MVGC toolbox. If the null hypothesis that the data is a finite sample of a stationary process is rejected on significance level 0.01 then the data is regenerated independently from the previous simulation(s).

Estimation of the reference representation: Depending on the network graph of $\mathcal{S}$, we apply the corresponding algorithm, i.e., Algorithm 15, 17 or 19, to obtain the system matrices of an estimate for $\mathcal{S}$ that has the same network graph as $\mathcal{S}$. We also apply two classical methods by using the MATLAB function `s4sid_CCA` from

the SSGC toolbox and `pem` to obtain estimates for $\mathcal{S}$. The three estimates are then compared based on the measures introduced in Section 7.3.

Granger causality test: The existence of the reference Kalman representation requires certain (conditional) Granger non-causality conditions to hold, see Theorems 2.5–3.5 and 4.15. By Remark 3.7 and Lemma 4.13, we can always test these conditions by Granger non-causality conditions, substituting the potential conditional Granger non-causality conditions. For their practical importance, we test these Granger causality conditions based on the method explained in Section 7.2, in particular in Section 7.2.3 and we show results on the performance of the tests.

7.4.2 Simulation results

Below, we present results on applying Algorithms 15, 17, and 19 on data, simulated from several reference representations that are respectively minimal Kalman representations in causal block triangular form, in causal coordinated form and with causal G-zero structure.

Minimal Kalman representation in causal block triangular form

We consider minimal Kalman representations in causal block triangular form as reference representations for data simulation, see Section 7.4.1. Figure 7.1 illustrates the network graph of these reference representation.

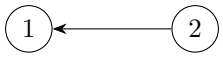

Figure 7.1: Network graph of the reference representation in block triangular form.

Parameter settings: Denote the dimensions of the components of the output process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T \in \mathbb{R}^r$ and the state process $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T \in \mathbb{R}^n$ of the reference representation by r_1, r_2, n_1, n_2, where $\mathbf{y}_1 \in \mathbb{R}^{r_1}$, $\mathbf{y}_2 \in \mathbb{R}^{r_2}$, $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{n_2}$. We consider nine families of representations, $S_1, S_2, \ldots, S_9$, where the dimensions r_1, r_2, n_1, and n_2 are as in Table 7.1 below.

For each family of representations S_i, $i = 1, \ldots, 9$, i.e., settings of (r_1, r_2, n_1, n_2), we consider 55 independently generated reference representations. Then, from each of these 9 times 55 reference representations, we independently simulate data with length 1000, 5000 and 10000.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
r_1	1	1	1	1	1	2	2	3	3
r_2	1	1	1	1	2	2	3	3	4
n_1	1	1	2	3	3	3	4	4	5
n_2	1	2	3	4	4	4	5	5	5

Table 7.1: Dimension settings for reference representation in block triangular form.

Error of the Granger causality test: The reference representations are chosen such that $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ and $\mathbf{y}_2$ Granger causes $\mathbf{y}_1$. Accordingly, we apply two statistical tests (see Section 7.2 for more details on these tests):

$$\begin{aligned} &\text{Test 1} \quad H_0 : \mathcal{F}_{\mathbf{y}_1 \to \mathbf{y}_2} = 0 \\ &\text{Test 2} \quad H_0 : \mathcal{F}_{\mathbf{y}_2 \to \mathbf{y}_1} = 0. \end{aligned}$$

We illustrate Tests 1 in Figure 7.2: Apply Algorithm 21 with input being the simulated data, $N_{\mathcal{F}} = 100$, $N_{training} = 10000$ and denote its output by $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{100}$. Apply Algorithm 21 with the same input, however, in Step 1 set the estimated system matrices $\hat{A}, \hat{K}, \hat{C}$ and $\Sigma^{\hat{\mathbf{e}}}$ to be the (real) system matrices of the reference representation and the (real) noise variance matrix. Denote the output by $\{\mathcal{F}^i_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{100}$. Then, in Figure 7.2 we compare the histograms of $\{\hat{\mathcal{F}}^i_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{100}$ and $\{\mathcal{F}^i_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=1}^{100}$ for one-one simulations from each dimension settings (r_1, r_2, n_1, n_2), see Table 7.1 and data lengths $N = 1000, 5000, 10000$.

We illustrate Tests 2 in Figure 7.3: We repeat the same calculation as for Figure 7.2, however, taking the reference representation and the simulated data for $[\mathbf{y}_2^T, \mathbf{y}_1^T]^T$ instead of $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$.

For both Figures 7.2 and 7.3, the rows are for the different dimension settings (r_1, r_2, n_1, n_2), see Table 7.1 and the columns are for the data lengths $N = 1000, 5000, 10000$.

To evaluate the performance of Tests 1 and 2, we consider the type I error of Test 1, denoted by $\varepsilon^{\mathbf{y}_1 \nrightarrow \mathbf{y}_2}_{\text{GC-test}}$, when the null hypothesis that $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ is rejected and the type II error of Test 2, denoted by $\varepsilon^{\mathbf{y}_2 \nrightarrow \mathbf{y}_1}_{\text{GC-test}}$, when the null hypothesis that $\mathbf{y}_2$ does not Granger cause $\mathbf{y}_1$ is accepted.

$$\begin{aligned} &\text{Type I error of Test 1:} \quad \varepsilon^{\mathbf{y}_1 \nrightarrow \mathbf{y}_2}_{\text{GC-test}} \\ &\text{Type II error of Test 2:} \quad \varepsilon^{\mathbf{y}_2 \nrightarrow \mathbf{y}_1}_{\text{GC-test}}. \end{aligned}$$

The percentage of the occurrence of these errors among the tests run for the 55 ref-

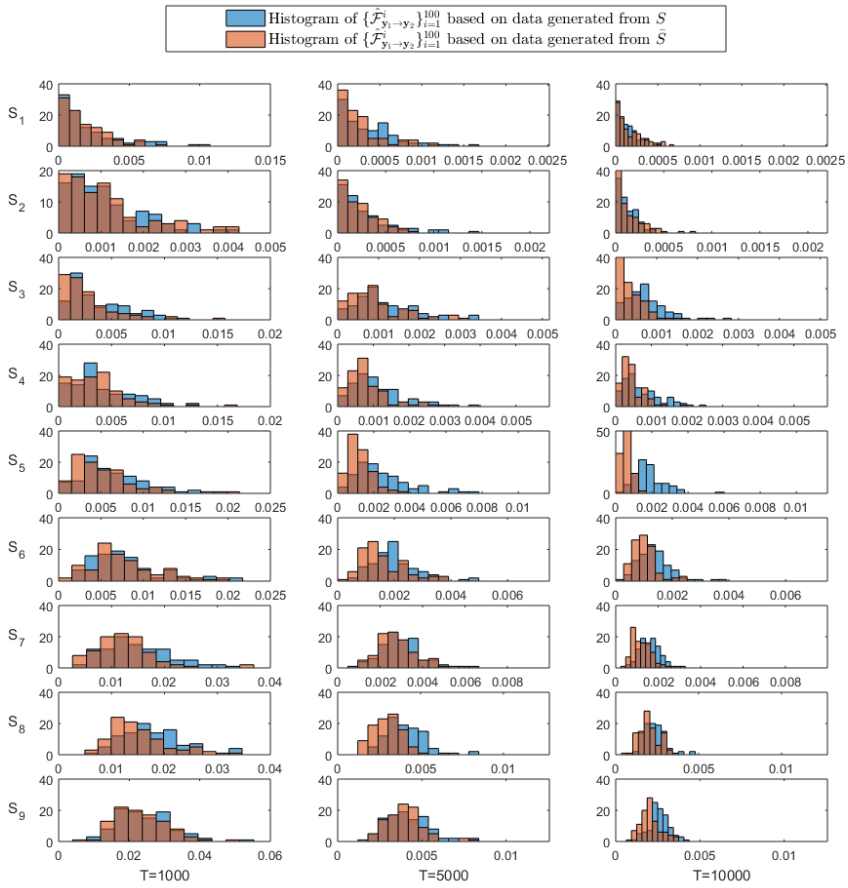

Figure 7.2: Illustration of Test 1: Empirical zero Geweke–Granger causalities from $\mathbf{y}_1$ to $\mathbf{y}_2$ were calculated from data generated from the reference representation S (blue) and from data generated from the estimated representation $\bar{S}$ in block triangular form (red).

erence representations for each parameter settings (r_1, r_2, n_1, n_2) and data length $N = 1000, 5000, 10000$ can be found in Table 7.2.

Error of the reference representation estimation: Consider a reference representation S that is a minimal Kalman representation in causal block triangular form with some parameters (r_1, r_2, n_1, n_2), see Table 7.1. Denote the estimated minimal Kalman representations, whose system matrices are the outputs of the MATLAB functions `s4sid_CCA` from the SSGC toolbox (see (Barnett and Seth, 2015) and `pem`

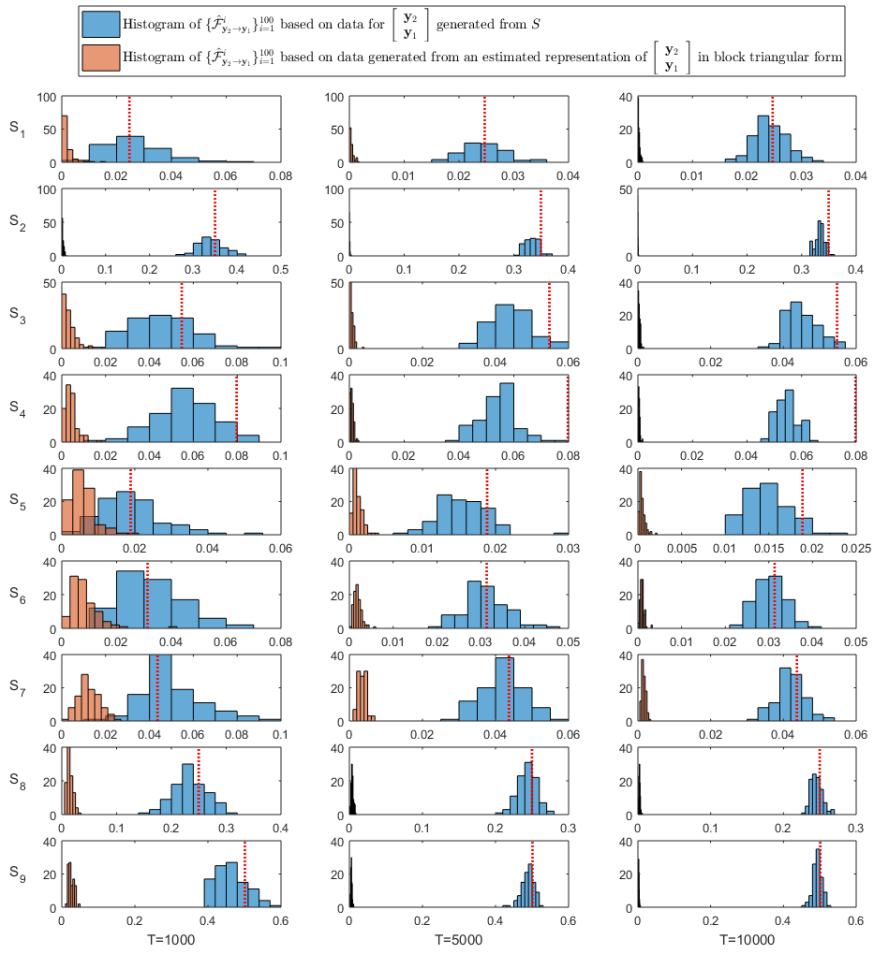

Figure 7.3: Illustration of Test 2: Empirical Geweke–Granger causalities from $\mathbf{y}_2$ to $\mathbf{y}_1$ were calculated from data generated from the reference representation S (blue). Empirical zero Geweke–Granger causalities from $\mathbf{y}_2$ to $\mathbf{y}_1$ were calculated from data generated from an estimated minimal Kalman representation of $[\mathbf{y}_2^T, \mathbf{y}_1^T]^T$ in causal block triangular form (red). The red vertical dashed lines denote the true Geweke–Granger causality from $\mathbf{y}_2$ to $\mathbf{y}_1$.

(see (Ljung, 1999, Chapter 7)), by $\hat{S}$ and $\dot{S}$, respectively. Furthermore, denote the estimated minimal Kalman representation in causal block triangular form, whose system matrices are the outputs of Algorithm 15, by $\bar{S}$. Then, we consider the sys-

tem errors (see Section 7.3):

$$\bar{\varepsilon}^{\Lambda^y}_{\text{sys}}, \quad \hat{\varepsilon}^{\Lambda^y}_{\text{sys}}, \quad \dot{\varepsilon}^{\Lambda^y}_{\text{sys}}, \quad \bar{\varepsilon}^{H_2}_{\text{sys}}, \quad \hat{\varepsilon}^{H_2}_{\text{sys}}, \quad \dot{\varepsilon}^{H_2}_{\text{sys}}.$$

Before evaluating how well $\bar{S}, \hat{S}$ and $\dot{S}$ estimate S, to avoid the dominance of the bad estimates, at each dimension settings and data length we dropped 5 reference representations from the 55 randomly generated reference representations, where the sum of the H_2-errors $\bar{\varepsilon}^{H_2}_{\text{sys}} + \hat{\varepsilon}^{H_2}_{\text{sys}} + \dot{\varepsilon}^{H_2}_{\text{sys}}$ were the largest. Furthermore, only those systems were considered, where the condition that $\mathbf{y}_1$ does not Granger cause $\mathbf{y}_2$ was verified by Test 1, i.e., where there was no type I error of Test 1. The averaged system errors for the reference representations can be found in Table 7.2.

Description of Table 7.2: The columns are as follows: percentage of type I error of Tests 1, percentage of type II error of Test 2, average covariance based system errors and average H_2-system errors of the estimated representations $\bar{S}$, $\hat{S}$ and $\dot{S}$. The rows are as follows: rows $1-9$ show the results for data length $N = 1000$ and for the nine parameter setting of (r_1, r_2, n_1, n_2), see Table 7.1; in a similar way, rows $10-18$ and $19-27$ show the results for data length $N = 5000$ and 10000, respectively.

Conclusions on Table 7.2: In the first two columns we can see the percentage of the occurrence of type I error $\varepsilon^{\mathbf{y}_1 \nrightarrow \mathbf{y}_2}_{\text{GC-test}}$ of Test 1 and type II error $\varepsilon^{\mathbf{y}_2 \nrightarrow \mathbf{y}_1}_{\text{GC-test}}$ of Test 2: The results do not show consistent change in terms of the parameter settings of (r_1, r_2, n_1, n_2). However, it can be seen that with increasing the data length, type I error of Test 1 occurred more and type II error of Test 2 occurred less times. It can be explained with the phenomenon that the empirical value of the zero Geweke-Granger causality is closer to zero for longer data, see Figures 7.2 and 7.3. In Figure 7.2, we can also see that for Test 1 the empirical distribution defined by $\{\mathcal{F}^i_{\mathbf{y}_1 \to \mathbf{y}_2}\}_{i=0}^{100}$ is slightly underestimated, which indicates more type I errors. In case of Test 2, the better estimations we have for the distribution of the zero and the non-zero Geweke-Granger causality, the less overlap occur for the two estimated distribution, see Figure 7.3. To increase the performance of Tests 1 and 2, we can increase the number of samples $N_{\mathcal{F}}$ for estimating the empirical distribution, see Section 7.2. For our calculations, it was set to 100. Furthermore, type I error of Test 1 can be improved by decreasing and type II error of Test 2 can be improved by increasing the significance level which was set to $\alpha = 0.01$ for our calculation.

Regarding the system errors, we can see that, Algorithm 15 gives estimates ($\bar{S}$) for the reference representations that are as good as and in some cases better (especially for short data $N = 1000$) than the two other estimates ($\hat{S}$ and $\dot{S}$) from classical methods on estimating minimal Kalman representations. From this, we can conclude that Algorithm 15 possibly reduces the estimation error of estimated minimal Kalman representations in causal block triangular form compared to the considered

	$\varepsilon^{\mathbf{y}_1 \nrightarrow \mathbf{y}_2}_{\text{GC-test}}$	$\varepsilon^{\mathbf{y}_2 \nrightarrow \mathbf{y}_1}_{\text{GC-test}}$	$\bar{\varepsilon}^{\Lambda^y}_{\text{sys}}$	$\hat{\varepsilon}^{\Lambda^y}_{\text{sys}}$	$\dot{\varepsilon}^{\Lambda^y}_{\text{sys}}$	$\bar{\varepsilon}^{H_2}_{\text{sys}}$	$\hat{\varepsilon}^{H_2}_{\text{sys}}$	$\dot{\varepsilon}^{H_2}_{\text{sys}}$	
S_1	2%	29%	0.55	0.89	1.65	0.06	0.06	0.07	$N = 1000$
S_2	0%	33%	1.10	1.27	1.41	0.09	0.09	0.10	
S_3	7%	25%	0.54	0.65	0.76	0.10	0.10	0.13	
S_4	4%	20%	0.93	0.98	2.03	0.12	0.12	0.16	
S_5	5%	33%	0.86	0.98	1.35	0.10	0.12	0.16	
S_6	7%	22%	0.54	0.71	1.17	0.11	0.12	0.16	
S_7	0%	24%	0.91	1.35	2.36	0.12	0.14	0.18	
S_8	4%	24%	0.54	0.66	1.23	0.13	0.14	0.18	
S_9	4%	22%	1.60	2.77	8.37	0.14	0.15	0.20	
S_1	5%	11%	0.25	0.41	0.42	0.03	0.03	0.03	$N = 5000$
S_2	16%	9%	0.20	0.28	0.27	0.05	0.05	0.04	
S_3	15%	7%	0.19	0.21	0.22	0.07	0.06	0.06	
S_4	16%	9%	0.18	0.16	0.17	0.07	0.07	0.07	
S_5	7%	7%	0.22	0.24	0.28	0.06	0.06	0.07	
S_6	4%	13%	0.60	0.78	1.03	0.06	0.06	0.07	
S_7	4%	16%	0.51	0.64	1.16	0.06	0.06	0.08	
S_8	7%	9%	0.29	0.34	0.53	0.07	0.07	0.07	
S_9	7%	7%	0.57	0.72	1.49	0.07	0.07	0.08	
S_1	7%	9%	0.30	0.58	0.90	0.03	0.03	0.02	$N = 10000$
S_2	11%	15%	0.20	0.24	0.27	0.04	0.04	0.03	
S_3	11%	15%	0.23	0.27	0.30	0.05	0.05	0.04	
S_4	20%	9%	1.48	0.30	0.33	0.07	0.06	0.05	
S_5	15%	4%	0.25	0.29	0.34	0.04	0.05	0.05	
S_6	5%	2%	0.15	0.15	0.17	0.04	0.04	0.04	
S_7	15%	4%	0.66	0.74	1.06	0.05	0.05	0.05	
S_8	11%	5%	0.31	0.37	0.76	0.05	0.05	0.05	
S_9	7%	4%	0.36	0.48	0.75	0.05	0.05	0.05	

Table 7.2: Results on estimating the system matrices of minimal Kalman representation in causal block triangular form.

classical methods. Note that the estimation of the representations did not involve the estimation of the state and output dimensions r, n and r_1, r_2, n_1, n_2. If these parameters have to be estimated then, since r_1, r_2, n_1, n_2 are only necessary for the calculation of $\bar{S}$, it potentially generates additional errors in the calculation of $\bar{S}$.

Minimal Kalman representation in causal coordinated form

The reference representations for data simulation considered in this section are min-

imal Kalman representations in causal coordinated form where the output process has three components. See Figure 7.4 for the network graph of such representations. The reference representations are randomly generated as explained in Section 7.4.1.

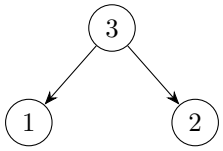

Figure 7.4: Network graph of the reference representation in coordinated form.

Parameter settings: The dimensions of the components of the output process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T]^T \in \mathbb{R}^r$ of the reference representations are such that $\mathbf{y}_i \in \mathbb{R}^{r_i}$ and of the state process $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{x}_3^T]^T \in \mathbb{R}^n$ are such that $\mathbf{x}_i \in \mathbb{R}^{n_i}$ for some $r_i > 0, n_i > 0$, $i = 1, 2, 3$. We consider nine families of representations, $S_1, S_2, \ldots, S_9$ with different parameters for the dimensions $(r_1, r_2, r_3, n_1, n_2, n_3)$, see Table 7.3.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
r_1	1	1	1	1	1	2	2	2	2
r_2	1	1	1	1	2	2	2	2	3
r_3	1	1	1	2	2	2	2	2	3
n_1	1	1	2	2	2	2	3	3	3
n_2	1	2	3	3	3	3	3	3	3
n_3	1	3	4	4	4	4	4	4	4

Table 7.3: Dimension settings for reference representation in coordinated form.

For each set of settings for the parameters $(r_1, r_2, r_3, n_1, n_2, n_3)$, we consider 55 independently generated reference representations. Then, from each of these 9 times 55 reference representation we independently simulate data with length 1000, 5000 and 10000.

Error of the Granger causality test: By Theorem 3.5 and Remark 3.7, the existence of the reference representations of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T]^T$ in coordinate form implies the following conditions: $\mathbf{y}_i$ does not Granger cause $\mathbf{y}_3$ for $i = 1, 2$ and $\mathbf{y}_i$ does not Granger cause $[\mathbf{y}_j^T, \mathbf{y}_3^T]^T$ for $i \neq j$, $i, j \in \{1, 2\}$. Furthermore, by Lemma 3.12, we have that $\mathbf{y}_i$ does not Granger cause $\mathbf{y}_3$ for $i = 1, 2$ is equivalent to saying that $[\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ does not Granger cause $\mathbf{y}_3$. Accordingly, we apply three statistical tests

with null hypotheses that $[\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ does not Granger cause $\mathbf{y}_3$, that $\mathbf{y}_1$ does not Granger cause $[\mathbf{y}_2^T, \mathbf{y}_3^T]^T$ and that $\mathbf{y}_2$ does not Granger cause $[\mathbf{y}_1^T, \mathbf{y}_3^T]^T$, see also Section 7.2 for more details on the tests:

$$\begin{aligned}
&\text{Test 1} \quad H_0 : \mathcal{F}_{[\mathbf{y}_1^T, \mathbf{y}_2^T]^T \to \mathbf{y}_3} = 0 \\
&\text{Test 2} \quad H_0 : \mathcal{F}_{\mathbf{y}_1 \to [\mathbf{y}_2^T, \mathbf{y}_3^T]^T} = 0 \\
&\text{Test 3} \quad H_0 : \mathcal{F}_{\mathbf{y}_2 \to [\mathbf{y}_1^T, \mathbf{y}_3^T]^T} = 0.
\end{aligned}$$

For each simulation, we consider the type I errors of Test 1, Test 2 and Test 3 i.e., when the null hypothesis is rejected. Denote the type I error of Test 1, Test 2 and Test 3 by $\varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2} \nrightarrow \mathbf{y}_3}$, $\varepsilon_{\text{GC-test}}^{\mathbf{y}_1 \nrightarrow \mathbf{y}_{2,3}}$ and $\varepsilon_{\text{GC-test}}^{\mathbf{y}_2 \nrightarrow \mathbf{y}_{1,3}}$, respectively.

$$\begin{aligned}
&\text{Type I error of Test 1:} \quad \varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2} \nrightarrow \mathbf{y}_3} \\
&\text{Type I error of Test 2:} \quad \varepsilon_{\text{GC-test}}^{\mathbf{y}_1 \nrightarrow \mathbf{y}_{2,3}} \\
&\text{Type I error of Test 3:} \quad \varepsilon_{\text{GC-test}}^{\mathbf{y}_2 \nrightarrow \mathbf{y}_{1,3}}
\end{aligned}$$

The percentage of the occurrence of these errors among the tests run for the 55 reference representations for each parameter settings $(r_1, r_2, r_3, n_1, n_2, n_3)$ and data length $N = 1000, 5000, 10000$ can be found in Table 7.4.

Error of the reference representation estimation: Consider a reference representation, i.e., a minimal Kalman representation in causal coordinated form with a set of parameters $(r_1, r_2, r_3, n_1, n_2, n_3)$, see Table 7.3 and denote it by S. In addition, denote the estimated minimal Kalman representation, whose system matrices are the outputs of the MATLAB function `s4sid_CCA` from the SSGC toolbox, by $\hat{S}$, the estimated minimal Kalman representation in causal coordinated form, whose system matrices are the outputs of Algorithm 17 by $\bar{S}$, and the estimated minimal Kalman representation, whose system matrices are the outputs of the MATLAB function `pem`, by $\dot{S}$ (see Section 7.3). Then, we consider the following system errors (see Section 7.3):

$$\hat{\varepsilon}_{\text{sys}}^{\Lambda^y}, \quad \bar{\varepsilon}_{\text{sys}}^{\Lambda^y}, \quad \dot{\varepsilon}_{\text{sys}}^{\Lambda^y}, \quad \hat{\varepsilon}_{\text{sys}}^{H_2}, \quad \bar{\varepsilon}_{\text{sys}}^{H_2}, \quad \dot{\varepsilon}_{\text{sys}}^{H_2}.$$

As in the previous section, we dropped the 5 reference representations where the estimations were the worst regarding the H_2-errors, i.e., where $\bar{\varepsilon}_{\text{sys}}^{H_2} + \hat{\varepsilon}_{\text{sys}}^{H_2} + \dot{\varepsilon}_{\text{sys}}^{H_2}$ were the highest. Furthermore, only those reference representations were considered where the null hypothesis of Tests 1, 2 and 3 were verified. The averaged system errors over the 55 reference representations for each parameter settings $(r_1, r_2, r_3, n_1, n_2, n_3)$ and data length N can be found in Table 7.4.

Description of Table 7.4: The columns are as follows: percentage of type I er-

	$\varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2}\nrightarrow\mathbf{y}_3}$	$\varepsilon_{\text{GC-test}}^{\mathbf{y}_1\nrightarrow\mathbf{y}_{2,3}}$	$\varepsilon_{\text{GC-test}}^{\mathbf{y}_2\nrightarrow\mathbf{y}_{1,3}}$	$\bar{\varepsilon}_{\text{sys}}^{\Lambda^y}$	$\hat{\varepsilon}_{\text{sys}}^{\Lambda^y}$	$\dot{\varepsilon}_{\text{sys}}^{\Lambda^y}$	$\bar{\varepsilon}_{\text{sys}}^{H_2}$	$\hat{\varepsilon}_{\text{sys}}^{H_2}$	$\dot{\varepsilon}_{\text{sys}}^{H_2}$	
S_1	0%	2%	5%	1.32	5.24	15.52	0.06	0.07	0.11	$N = 1000$
S_2	0%	2%	0%	0.60	1.03	1.72	0.08	0.10	0.15	
S_3	5%	2%	5%	0.42	0.64	1.22	0.09	0.11	0.16	
S_4	4%	4%	4%	0.86	1.44	2.71	0.10	0.12	0.21	
S_5	2%	0%	0%	1.03	2.01	4.70	0.10	0.13	0.21	
S_6	4%	2%	4%	1.13	1.56	3.94	0.11	0.14	0.20	
S_7	2%	2%	4%	1.17	2.12	4.18	0.11	0.14	0.22	
S_8	4%	2%	2%	1.41	2.67	6.25	0.11	0.14	0.22	
S_9	4%	4%	2%	1.61	3.23	10.22	0.11	0.14	0.22	
S_1	2%	7%	2%	1.05	5.89	15.76	0.04	0.04	0.05	$N = 5000$
S_2	0%	0%	0%	0.60	1.16	1.77	0.04	0.05	0.06	
S_3	0%	2%	0%	0.23	0.30	0.49	0.05	0.05	0.07	
S_4	2%	4%	4%	0.34	0.44	0.68	0.05	0.06	0.08	
S_5	7%	2%	2%	0.80	1.28	2.53	0.06	0.06	0.09	
S_6	11%	7%	7%	0.91	2.27	7.01	0.05	0.06	0.09	
S_7	5%	11%	4%	0.81	1.55	3.23	0.06	0.07	0.09	
S_8	5%	5%	4%	0.70	1.01	3.05	0.06	0.07	0.09	
S_9	11%	7%	5%	0.52	0.85	2.02	0.06	0.07	0.09	
S_1	7%	4%	9%	0.59	0.56	2.19	0.03	0.03	0.03	$N = 10000$
S_2	5%	4%	2%	0.24	0.30	0.52	0.03	0.03	0.04	
S_3	5%	4%	7%	0.44	0.70	1.55	0.03	0.04	0.05	
S_4	5%	2%	4%	0.19	0.24	0.35	0.04	0.04	0.06	
S_5	7%	9%	7%	0.33	0.51	0.88	0.04	0.04	0.06	
S_6	5%	4%	5%	0.31	0.39	0.64	0.04	0.05	0.06	
S_7	5%	9%	2%	0.43	0.71	1.18	0.04	0.05	0.06	
S_8	7%	5%	4%	0.40	0.64	1.31	0.04	0.05	0.06	
S_9	15%	4%	7%	0.70	1.60	4.54	0.04	0.05	0.06	

Table 7.4: Results on estimating minimal Kalman representation in causal coordinated form.

rors ($\varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2}\nrightarrow\mathbf{y}_3}, \varepsilon_{\text{GC-test}}^{\mathbf{y}_1\nrightarrow\mathbf{y}_{2,3}}, \varepsilon_{\text{GC-test}}^{\mathbf{y}_2\nrightarrow\mathbf{y}_{1,3}}$) of Tests 1–2 and 3, respectively, average covariance based system errors ($\bar{\varepsilon}_{\text{sys}}^{\Lambda^y}, \hat{\varepsilon}_{\text{sys}}^{\Lambda^y}, \dot{\varepsilon}_{\text{sys}}^{\Lambda^y}$) and average H_2-system errors $\bar{\varepsilon}_{\text{sys}}^{H_2}, \hat{\varepsilon}_{\text{sys}}^{H_2}, \dot{\varepsilon}_{\text{sys}}^{H_2}$ of the estimations $\bar{S}$, $\bar{S}$ and $\dot{S}$ of the reference representations. The rows are as follows: rows $1-9$ show the results for data length $N = 1000$ and for the nine parameter setting of (r_1, r_2, n_1, n_2), see Table 7.3; in a similar way, rows $10-18$ and rows $19-27$ show the results for data length $N = 5000$ and 10000, respectively.

Conclusions on Table 7.4: As in Table 7.2, we can see that there is no consistent change in the test results for the different parameter settings of $\{r_i, n_i\}_{i=1}^3$. At the

conclusions on Table 7.2 we mentioned that for longer data, the empirical distribution of empirical value of the zero Geweke-Granger causality is slightly underestimated which causes more type I errors. This can be seen on Table 7.3 as well. Note that the significance level was set to be $\alpha = 0.01$ and the number of samples for estimating the distribution of the empirical value of the zero Geweke–Granger causality was chosen to be $N_{\mathcal{F}} = 100$.

Regarding the system errors, we can see that, in general, $\bar{S}$ estimates the reference representation better than the other two estimations $\hat{S}$ and $\dot{S}$. This is due to that when $\bar{S}$ estimates the reference representation, there is possibility for estimating the system matrices in a distributed way, see Algorithm 17. This suggest that, Algorithm 17 possibly reduces the estimation error when a minimal Kalman representation in causal coordinated form is estimated compared to the considered classical methods. Note that the estimation of the representations did not involve the estimation of the state and output dimensions r, n and $r_1, r_2, r_3, n_1, n_2, n_3$. If these parameters are estimated then, since $r_1, r_2, r_3, n_1, n_2, n_3$ are only necessary for the calculation of $\bar{S}$, it generates additional potential errors in the calculation of $\bar{S}$.

Minimal Kalman representation with causal G-zero structure

Let $G = (V, E)$ be a TADG, where the set of nodes is $V = \{1, 2, 3\}$ and the set of edges is $E = \{(3, 1), (3, 2), (2, 1)\}$. Then, the reference representations considered in this section are minimal Kalman representations with causal G-zero structure. See Figure 7.5 for the network graph of representations with G-zero structure. The reference representations are randomly generated as explained in Section 7.4.1.

Parameter settings: The components of the output process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T]^T \in \mathbb{R}^r$ of the reference representations have dimensions $\mathbf{y}_i \in \mathbb{R}_i^r$ and the components of the state process $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{x}_3^T]^T \in \mathbb{R}^n$ have dimensions $\mathbf{x}_i \in \mathbb{R}^{n_i}$ for some $r_i > 0, n_i > 0$, $i = 1, 2, 3$. We consider nine families of representations, $S_1, S_2, \ldots, S_9$ with different parameters for the dimensions r_1, r_2, r_3, n_1, n_2, and n_3, see Table 7.3. Notice that the parameters $(r_1, r_2, r_3, n_1, n_2, n_3)$ are chosen from the same sets of parameter as it was chosen for reference representations in coordinated form in the previous section. Also, as for the reference representations in block triangular and in coordinated form in the previous sections, for each parameter settings of $(r_1, r_2, r_3, n_1, n_2, n_3)$, we consider 55 independently generated reference representations and from each reference representation we independently simulate data with length 1000, 5000 and 10000.

Error of the Granger causality test: By Theorem 4.15 and Lemma 4.13, the existence of the reference representations of a process $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T]^T$ with G-zero structure implies the following conditions: $\mathbf{y}_i$ does not Granger cause $\mathbf{y}_3$ for $i = 1, 2$, $\mathbf{y}_1$ does

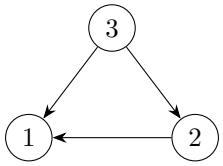

Figure 7.5: Network graph of the reference representation with G-zero structure.

not Granger cause $[\mathbf{y}_2^T, \mathbf{y}_3^T]^T$ and $[\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ does not Granger cause $\mathbf{y}_3$. Note that by Lemma3.12, $\mathbf{y}_i$ does not Granger cause $\mathbf{y}_3$ for $i = 1, 2$ if and only if $[\mathbf{y}_1^T, \mathbf{y}_2^T]^T$ does not Granger cause $\mathbf{y}_3$. Accordingly, we apply two statistical tests with the following null hypotheses, see Section 7.2 for more details on the tests:

$$\begin{aligned} &\text{Test 1} \quad H_0 : \mathcal{F}_{[\mathbf{y}_1^T, \mathbf{y}_2^T]^T \to \mathbf{y}_3} = 0 \\ &\text{Test 2} \quad H_0 : \mathcal{F}_{\mathbf{y}_1 \to [\mathbf{y}_2^T, \mathbf{y}_3^T]^T} = 0. \end{aligned}$$

For each simulation, we consider the type I errors of Test 1 and Test 2. Denote the type I errors of Test 1 and Test 2 by $\varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2} \nrightarrow \mathbf{y}_3}$ and $\varepsilon_{\text{GC-test}}^{\mathbf{y}_1 \nrightarrow \mathbf{y}_{2,3}}$, respectively.

$$\begin{aligned} &\text{Type I error of Test 1:} \quad \varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2} \nrightarrow \mathbf{y}_3} \\ &\text{Type I error of Test 2:} \quad \varepsilon_{\text{GC-test}}^{\mathbf{y}_1 \nrightarrow \mathbf{y}_{2,3}} \end{aligned}$$

The percentage of the occurrence of these errors among the tests run for the 55 reference representations for each parameter settings $(r_1, r_2, r_3, n_1, n_2, n_3)$ and data length $N \in \{1000, 5000, 10000\}$ can be found in Table 7.5.

Error of the reference representation estimation: Fix a set of parameters $(r_1, r_2, r_3, n_1, n_2, n_3)$ and denote a corresponding reference representation by S (that is a minimal Kalman representation with causal G-zero structure). In addition, denote the estimated Kalman representation, whose system matrices are the outputs of the MATLAB function `s4sid_CCA`, by $\hat{S}$, the estimated minimal Kalman representation with causal G-zero structure, whose system matrices are the outputs of Algorithm 15 by $\bar{S}$ and the estimated minimal Kalman representation, whose system matrices are the outputs of the MATLAB function `pem`, by $\dot{S}$, see also Section 7.3. Then, we consider the following system errors, see Section 7.3:

$$\hat{\varepsilon}_{\text{sys}}^{\Lambda^y}, \quad \bar{\varepsilon}_{\text{sys}}^{\Lambda^y}, \quad \dot{\varepsilon}_{\text{sys}}^{\Lambda^y}, \quad \hat{\varepsilon}_{\text{sys}}^{H_2}, \quad \bar{\varepsilon}_{\text{sys}}^{H_2}, \quad \dot{\varepsilon}_{\text{sys}}^{H_2}.$$

As in the previous sections, we dropped the 5 worst reference representations where the estimations were the worst regarding the H_2-errors i.e., where $\bar{\varepsilon}_{\text{sys}}^{H_2} + \hat{\varepsilon}_{\text{sys}}^{H_2} + \dot{\varepsilon}_{\text{sys}}^{H_2}$ were the highest. Furthermore, only those reference representations were consid-

	$\varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2}\nrightarrow\mathbf{y}_3}$	$\varepsilon_{\text{GC-test}}^{\mathbf{y}_1\nrightarrow\mathbf{y}_{2,3}}$	$\bar{\varepsilon}_{\text{sys}}^{\Lambda^y}$	$\hat{\varepsilon}_{\text{sys}}^{\Lambda^y}$	$\dot{\varepsilon}_{\text{sys}}^{\Lambda^y}$	$\bar{\varepsilon}_{\text{sys}}^{H_2}$	$\hat{\varepsilon}_{\text{sys}}^{H_2}$	$\dot{\varepsilon}_{\text{sys}}^{H_2}$	
S_1	0%	4%	1.73	11.79	29.41	0.07	0.08	0.11	$N = 1000$
S_2	4%	7%	0.41	0.45	0.63	0.11	0.10	0.14	
S_3	0%	0%	0.52	0.99	2.08	0.10	0.10	0.16	
S_4	4%	2%	1.34	2.31	3.89	0.12	0.12	0.20	
S_5	2%	2%	0.96	1.64	3.72	0.11	0.13	0.20	
S_6	4%	5%	0.97	2.18	6.31	0.12	0.13	0.21	
S_7	2%	0%	0.69	1.01	2.30	0.12	0.13	0.22	
S_8	5%	5%	1.32	3.18	10.15	0.13	0.14	0.22	
S_9	4%	2%	1.91	5.66	18.58	0.13	0.15	0.22	
S_1	5%	5%	0.96	1.46	4.37	0.04	0.04	0.04	$N = 5000$
S_2	4%	4%	0.40	0.51	0.79	0.05	0.05	0.06	
S_3	5%	5%	0.31	0.40	0.89	0.05	0.05	0.07	
S_4	5%	4%	0.56	0.74	1.09	0.07	0.06	0.09	
S_5	4%	5%	0.34	0.42	0.68	0.06	0.06	0.08	
S_6	4%	5%	0.52	0.70	1.58	0.07	0.06	0.08	
S_7	2%	7%	0.45	0.59	1.21	0.07	0.06	0.09	
S_8	9%	4%	1.01	1.59	3.95	0.06	0.07	0.09	
S_9	15%	13%	0.59	0.92	2.11	0.07	0.07	0.09	
S_1	7%	9%	0.33	0.38	0.57	0.03	0.03	0.03	$N = 10000$
S_2	9%	15%	0.37	0.47	0.83	0.04	0.04	0.04	
S_3	15%	5%	0.34	0.48	0.91	0.04	0.04	0.05	
S_4	13%	7%	0.40	0.55	0.96	0.05	0.05	0.06	
S_5	9%	9%	0.62	1.06	1.83	0.05	0.04	0.06	
S_6	11%	9%	0.33	0.33	0.55	0.06	0.05	0.06	
S_7	15%	16%	0.30	0.36	0.57	0.05	0.05	0.06	
S_8	13%	7%	0.67	1.95	7.71	0.05	0.05	0.06	
S_9	11%	7%	0.34	0.40	0.75	0.05	0.05	0.06	

Table 7.5: Results on estimating minimal Kalman representation with causal G-zero structure.

ered where the null hypothesis of Tests 1 and Test 2 were verified. The averaged system errors over the remaining reference representations for each parameter settings $(r_1, r_2, r_3, n_1, n_2, n_3)$ and data length N can be found in Table 7.5.

Description of Table 7.5: The columns are as follows: percentage of type I errors $(\varepsilon_{\text{GC-test}}^{\mathbf{y}_{1,2}\nrightarrow\mathbf{y}_3}, \varepsilon_{\text{GC-test}}^{\mathbf{y}_1\nrightarrow\mathbf{y}_{2,3}})$ of Tests 1 and Tests 2, average covariance based system errors $(\bar{\varepsilon}_{\text{sys}}^{\Lambda^y}, \hat{\varepsilon}_{\text{sys}}^{\Lambda^y}, \dot{\varepsilon}_{\text{sys}}^{\Lambda^y})$ and average H_2-system errors $(\dot{\varepsilon}_{\text{sys}}^{\Lambda^y}, \bar{\varepsilon}_{\text{sys}}^{H_2}, \hat{\varepsilon}_{\text{sys}}^{H_2}, \dot{\varepsilon}_{\text{sys}}^{H_2})$ of the estimated representations $\bar{S}$, $\hat{S}$ and $\bar{S}$. The rows are as follows: rows $1-9$ show the results for data length $N = 1000$ and for the nine parameter setting of (r_1, r_2, n_1, n_2), see

Table 7.3; in a similar way, rows $10-18$ and $19-27$ show the results for data length $N = 5000$ and 10000, respectively.

Conclusions on Table 7.5: The results of Table 7.5 show that there is no consistent change in the test results for the different parameter settings of $\{r_i, n_i\}_{i=1}^3$, see Table 7.3. As we saw in Tables 7.2 and 7.4, for longer data, the empirical distribution of the empirical value of the zero Geweke-Granger causality is slightly underestimated which causes more type I errors in the tests. Note that as before, the significance level was set to be $\alpha = 0.01$ and the number of samples for estimating the distribution of the empirical value of the zero Geweke–Granger causality was set to 100.

The average relative covariance and H_2 system errors of $\bar{S}$, $\hat{S}$ and $\dot{S}$ show similar results as in previous cases, see Tables 7.2 and 7.4: in general, $\bar{S}$ estimates the reference representation better than the other two estimates $\hat{S}$ and $\dot{S}$. This is due to that when $\bar{S}$ estimates the reference representation, there is possibility for distributed estimation of the system matrices, for details on the estimation see Section 7.1.2. This suggests that Algorithm 19 possibly reduces the estimation error when a minimal Kalman representation in causal coordinated form is estimated compared to the considered classical methods. Note that, as before, the estimation of the representations did not involve the estimation of the state and output dimensions r, n and $r_1, r_2, r_3, n_1, n_2, n_3$. If these parameters have to be estimated then, since $r_1, r_2, r_3, n_1, n_2, n_3$ are only necessary for the calculation of $\bar{S}$, it generates additional potential errors in the calculation of $\bar{S}$.

7.5 Conclusions

In this chapter, we have proposed methods for estimating minimal Kalman representations with different network graphs and to test whether the necessary Granger non-causality conditions for its existence hold. Using the results of Chapters 2–4, we introduced algorithms that estimate the system matrices and reconstruct the network graph of the considered Kalman representations in block triangular form, in coordinated form and with TADG-zero structure based on output data. These algorithms were compared to two other classical methods, the state-space subspace method and prediction error method (that are implemented in MATLAB as `s4sid_CCA` and `pem` functions) by applying them to simulated data. The results show that the proposed algorithms that calculate a Kalman representation with the same network graph as the original one, provided as good, and in many cases better estimation than the estimations from classical methods.

Besides the above mentioned algorithms, we also proposed a new way of testing

Granger non-causality. This test is essential for statistically verifying the existence of the Kalman representations that the proposed algorithms estimate. The test results show that applying the test on simulated data for several random examples, in the vast majority of the cases, it correctly verified the appropriate Granger non-causality conditions.

The methods introduced in this chapter on estimating minimal Kalman representations with different network graphs and on testing Granger causality support the results and algorithms in Chapters 2–4 by proposing a way of using them in practice. However, they should be used with caution because the results are experimental. A complete analysis on them requires further research.