Relationship between causality of stochastic processes and zero blocks of their joint innovation transfer matrices

Mónika Józsa * Mihály Petreczky ** M. Kanat Camlibel ***

* Unité de Recherche en Informatique et Automatique, Ecole des Mines de Douai and Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen (e-mail: m.jozsa@rug.nl).
** Centre de Recherche en Informatique, Signal et Automatique de Lille (CRISiAL) CNRS, Centrale Lille (e-mail: mihaly.petreczky@ec-lice.fr)
*** Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen (e-mail: m.k.camlibel@rug.nl)

Abstract: We consider output processes which are realizable by stochastic linear time-invariant (LTI) systems. Such processes can always be realized by LTI systems in forward innovation form, and we study the transfer matrices of such LTI realizations. We show that such a transfer matrix is consistent with an acyclic directed graph if and only if the edges of this graph represent Granger-causality relations among the components of the output process. By consistency we mean that if there is no edge between two vertices of the graph, then the corresponding block of the transfer matrix is zero. Under this assumption, conditional Granger non-causality between the components of the process is equivalent with a zero block in the transfer matrix.

© 2017, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

1. INTRODUCTION

Many complex systems arise by interconnecting several smaller subsystems which communicate with each other. The resulting network structure and its consequences for the global behavior of the system are of interest both for control and analysis of such systems. In addition, reverse engineering of this network structure is a major challenge in several applications. Understanding the relationship between the network structure and the global observed behavior is essential for addressing all these problems. Unfortunately, this relationship is far from obvious: even if the number of subsystems is known, and each subsystem generates observations, it is not clear if the interaction between any two subsystems induces an intrinsic relationship between their observed behavior. If the observed behavior is modelled as a collection of stochastic processes, then various notions from probability theory can be used to formalize the interactions among them. For example, the notion of (conditional) Granger causality Granger (1963) can be used. Informally, a process $y_1$ does not conditionally Granger cause a process $y_2$ with respect to $y_3$, if using the past values of $y_1$, $y_2$ and $y_3$ do not allow to predict the future values of $y_2$ with a higher accuracy than using only the past values of $y_2$ and $y_3$. The concept of Granger causality has been used in systems biology, neuroscience and economics Roebroeck et al. (2011); Valdes-Sosa et al. (2011). Although there are several ways to represent a stochastic process (auto-regressive, moving average, state-space models), the relationship between Granger causalities and the network structure of these representations is not evident.

In this paper, we consider discrete-time multivariate stochastic processes with a proper rational spectrum, i.e., stochastic processes which can be interpreted as outputs of linear-time-invariant stochastic state-space representations, shortly LTI state-space representation, driven by a white noise process. Consider such a process $y$. It is well known that there exists an LTI state-space representation whose output is $y$ and whose noise process is the innovation process of $y$. Furthermore, if this state-space representation is minimal then it is unique up to isomorphism, and hence its transfer matrix is uniquely determined by $y$. We will call this transfer matrix the innovation transfer matrix of $y$. It is well known that the LTI state-space representation with innovation noise (and hence the innovation transfer matrix) can be computed from the covariances of $y$, or estimated from a sample path of $y$ using subspace identification methods (Lindquist and Picci (2015)).

Contribution. We show that the innovation transfer matrix of a process $y$ is consistent with a transitive acyclic graph, if and only if the components of $y$ are related by conditional Granger non-causality in a way determined by that graph. By consistency with a graph we mean that the edges of the graph correspond to potentially non-zero blocks of the innovation transfer matrix. That is, we relate the graph structure of the innovation transfer matrix with a graph formed by conditional Granger non-causality relations of the components of $y$. Note that each block of the innovation transfer matrix can be viewed as a transfer matrix of a subsystem. Hence, the graph with which the innovation transfer matrix is consistent can be interpreted as a description of interconnections among various subsystems. That is, the results of the paper relate intrinsic properties of a process with the interconnection structure of a finite representation (innovation transfer matrix) of this process. In addition to providing insights into fundamental theoretical problems, the result of the paper could serve as a starting point for testing (conditional) Granger causality.

Related work. Reverse engineering of the network structure of deterministic linear systems has been investigated in i.e.,
Yuan et al. (2015, 2011); Nordling and Jacobsen (2011) and the references therein. In contrast to the cited papers, we consider stochastic systems and we relate their network structure to Granger causality of their outputs. In Dufour and Renault (1998); Caines (1976); Gevers and Anderson (1982) the relationship between Granger causality of two processes and their Wold decomposition was investigated. Granger causality for state-space representation was studied by using transfer matrix approach in Barnett and Seth (2015). Contrary to those papers, we consider a more general graph of Granger causality relations, which involves more than two processes, and we relate it to the zero blocks of the innovation transfer matrix of the joint process. The notion of conditional Granger causality is a type of spurious causality in Hsiao (1982), where Hsiao related causality relation with representations but did not discuss multiple causality conditions. More complex causality structure was also studied for state space representation, see Caines et al. (2003); Caines and Wynn (2007); Caines et al. (2009). In comparison with Caines and Wynn (2007); Caines et al. (2009) we consider conditional Granger non-causality, while in the cited papers (non-conditional) Granger causality and conditional orthogonality has been studied where the conditional orthogonality condition has no trivial translation into (conditional) Granger non-causality. Regarding innovation transfer matrices, the results in Caines and Wynn (2007) and Caines et al. (2009) form special cases of the results in this paper since we have arbitrary covariance of the innovation process. For representing different kinds of causality relations Eichler (2005) worked on the approach of causality graphs, however Eichler (2005) did not link the causality properties of the process with its LTI representations.

Outline. Before presenting our results, in §2 we introduce the terminology and the basic tools, such as the stochastic processes of interest, Hilbert spaces generated by stochastic processes and transfer matrices. Then, in §3.1 we characterized Granger non-causality between two components of a process with the help of transfer matrices. As a generalization, in §3.2 we present our main result for conditional Granger non-causality between several components of a process. In §3.3, we provide an example for our main result.

2. PRELIMINARIES

2.1 Notation and terminology

We will use standard terminology from theory of stochastic processes, see for example Lindquist and Picci (2015). In particular, we consider discrete-time stochastic processes whose values are vectors with real entries, i.e. by a stochastic process \( z \) taking values in \( \mathbb{R}^k \) we mean a sequence \( \{ z(t) \}_{t \in \mathbb{Z}} \) of random variables taking values in \( \mathbb{R}^k \), where \( z(t) \) is referred as the value of \( z \) at time \( t \). Here, as usual, \( \mathbb{Z} \) denotes the set of integers. If \( \nu \) is a random variable with values in \( \mathbb{R}^k \), then by the coordinates of \( \nu \) we will mean the random variables \( \nu^i, i = 1, \ldots, k \) taking values in \( \mathbb{R} \), such that \( \nu = [\nu^1, \ldots, \nu^k]^T \). We denote by \( E[\nu] \) the mathematical expectation of a random variable \( \nu \). For standard notions of stochastic processes, such as wide-sense stationarity, spectral density, etc. we refer to Lindquist and Picci (2015). The class of processes studied in this paper is defined below:

\[ \text{Definition 1. [ZMSIR]} \] A stochastic process \( z \) is called zero-mean square-integrable with rational spectrum (abbreviated by ZMSIR), if it is a zero mean, square-integrable, wide-sense stationary, purely non-deterministic and full rank process whose spectral density is rational and strictly positive definite on the unit circle.

In the sequel, we will use various Hilbert-spaces generated by stochastic processes. The zero-mean square-integrable random variables taking values in \( \mathbb{R} \) form a Hilbert space with the covariance as the inner product, Lindquist and Picci (2015). We denote this Hilbert-space by \( \mathcal{H} \). By the Hilbert-space generated by a set \( S \) of elements of \( \mathcal{H} \) we will mean the smallest (with respect to set inclusion) closed subspace of \( \mathcal{H} \) which contains \( S \). Consider a ZMSIR process \( x \) taking values in \( \mathbb{R}^k \). Then for each \( \ell \in \mathbb{R}^k, t \in \mathbb{Z}, \{ t^\ell \ z(t) \} \) is an element of \( \mathcal{H} \) and we denote by \( \mathcal{H}^x, \mathcal{H}^y_1, \ldots, \mathcal{H}^y_n, \mathcal{H}^{x \mid t} \) the Hilbert-spaces generated by the sets \( \{ \{ t^\ell \ z(s) \} \mid s \in \mathbb{Z}, \ell \in \mathbb{Z}, s \leq t \} \), \( \{ \{ t^\ell \ z(s) \} \mid s \in \mathbb{Z}, s \geq t \} \), and \( \{ \{ t^\ell \ z(t) \} \} \). Informally, \( \mathcal{H}^x \) is the Hilbert-space generated by the coordinates of all the values (past and future) of \( z \), \( \mathcal{H}^y_1 \) is the Hilbert-space generated by the coordinates of the past values \( \{ z(s) \}_{s \leq t} \) of \( z \) up to time \( t \), \( \mathcal{H}^{x \mid t} \) is the Hilbert-space generated by the coordinates of the future values \( \{ z(s) \}_{s \geq t} \) of \( z \), \( \mathcal{H}^{x \mid t} \) is the Hilbert-space generated by the coordinates of \( z(t) \). If \( z_1, \ldots, z_n \) are vector valued processes then \( \begin{bmatrix} z_1^T & \cdots & z_n^T \end{bmatrix} \) will denote the process defined by \( z(t) = \begin{bmatrix} z_1^T(t) & \cdots & z_n^T(t) \end{bmatrix}^T, t \in \mathbb{Z} \). In this case, we sometimes will denote \( \mathcal{H}^{x_1, \ldots, x_n} \), \( \mathcal{H}^{x_1} \), \( \mathcal{H}^y \) by \( \mathcal{H}^{x_1, \ldots, x_n} \), \( \mathcal{H}^{x_1} \), \( \mathcal{H}^y \) respectively.

If \( \eta \in \mathcal{H} \) and \( \mathcal{H} \) is a closed subspace of \( \mathcal{H} \), then we denote by \( E[\eta | \mathcal{H}] \) the orthogonal projection of \( \eta \) onto \( \mathcal{H} \). The orthogonal projection onto \( \mathcal{H} \) of a random variable \( \nu \) taking values in \( \mathbb{R}^k \) is defined as follows. Assume that \( \nu = [\nu^1, \ldots, \nu^k]^T \), i.e. \( \nu^i \in \mathcal{H} \), \( i = 1, \ldots, k \) are the coordinates of \( \nu \). Then the orthogonal projection of \( \nu \) onto \( \mathcal{H} \), denoted by \( E_\mathcal{H}[\nu | \mathcal{H}] \), is defined as \( E_\mathcal{H}[\nu | \mathcal{H}] := [\nu^1, \ldots, \nu^k]^T, \) where \( \nu^i = E_\mathcal{H}[\nu^i | \mathcal{H}] \), \( i = 1, \ldots, k \). That is, \( E_\mathcal{H}[\nu | \mathcal{H}] \) is the random variable with values in \( \mathbb{R}^k \) obtained by projecting the coordinates of \( \nu \) onto \( \mathcal{H} \). By a slight abuse of terminology and notation, we will say that a random variable \( \nu \) taking values in \( \mathbb{R}^k \) belongs to a closed subspace \( \mathcal{H} \) of \( \mathcal{H} \), denoted by \( \nu \in \mathcal{H} \), if every coordinate of \( \nu \) is an element of \( \mathcal{H} \). This is equivalent to saying that \( E_\mathcal{H}[\nu] \) is an element of \( \mathcal{H} \) for all \( \ell \in \mathbb{R}^k \). A random variable \( \nu \) is said to be orthogonal to a closed subspace \( \mathcal{H} \) of \( \mathcal{H} \), denoted by \( \nu \perp \mathcal{H} \), if \( E[\nu | \mathcal{H}] = 0 \) for all \( \eta \in \mathcal{H} \).

2.2 Innovation transfer matrix

Consider a transfer matrix \( G(z) \) of a finite-dimensional discrete-time stable deterministic LTI system (Anderson and Moore, 1979, Appendix C & D). Consider its Laurent series expansion, i.e. let \( G_k \in \mathbb{R}^{n \times m}, k \geq 0 \) be such that \( G(z) = \sum_{k=0}^{\infty} G_k z^{-k} \) for all \( z \in \mathbb{C} \) such that \( |z| \geq 1 \). If \( y \) is a ZMSIR process, then the expression \( \sum_{k=0}^{\infty} G_k y(t-k) \) converges in the topology of \( \mathcal{H}_1^y \) (Anderson and Moore, 1979, Theorem 4.1). In the sequel, we will write

\[ G(z)y(t) = \sum_{k=0}^{\infty} G_k y(t-k). \]

That is, \( G(z) \) can be interpreted as a causal linear filter which transforms \( y \) to the process \( \{ G(z)y(t) \}_{t=-\infty}^{\infty} \).
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات