

On the spectral formulation of Granger causality

D. Chicharro

the date of receipt and acceptance should be inserted later

Abstract Spectral measures of causality are used to explore the role of different rhythms in the causal connectivity between brain regions. We study several spectral measures related to Granger causality, comprising the bivariate and conditional Geweke measures, the directed transfer function and the partial directed coherence. We derive the formulation of dependence and causality in the spectral domain from the more general formulation in the information theory framework. We argue that the transfer entropy, the most general measure derived from the concept of Granger causality, lacks a spectral representation in terms of only the processes associated with the recorded signals.

For all the spectral measures we show how they are related to mutual information rates when explicitly considering the parametric autoregressive representation of the processes. In this way we express the conditional Geweke spectral measure in terms of a multiple coherence involving innovation variables inherent to the autoregressive representation. We also link partial directed coherence with Sims criterion of causality.

Given our results we discuss the causal interpretation of the spectral measures related to Granger causality and stress the necessity to explicitly consider their specific formulation based on modeling the signals as linear Gaussian stationary autoregressive processes.

Keywords Granger causality, Partial directed coherence, Directed transfer function, Transfer entropy, Sims causality, multivariate processes

1 Introduction

The concept of Granger causality (Granger, 1963, 1980) provides a data-driven approach to study causal interactions from time series. In neuroscience, measures based on Granger causality have been used to study the *effective connectivity* between brain areas (Friston, 1994), that is, the causal influence that one brain area exerts on another. (See Pereda et al., 2005; Gourevitch et al., 2006; Bressler and Seth, 2011, for a review of Granger causality measures applied to neural data).

In its original formulation Granger causality infers a causal interaction from a process Y to a process X relying on the reduction of the prediction error of X when including the past of Y (Wiener, 1956; Granger, 1963). Given this criterion it is easy to implement a Granger causality measure if linear Gaussian stationary processes are assumed. This prediction error criterion can be generalized to a criterion of conditional independence on probability distributions (Granger, 1980) that is generally applicable to stationary and non-stationary stochastic processes. This general criterion is naturally implemented in the framework of information theory (e. g. Schreiber, 2000; Solo, 2008; Amblard and Michel, 2011), resulting in the information theoretic measure of Granger causality called *transfer entropy*. Despite its generality, in practice the criterion of conditional independence is difficult to apply because it involves the estimation of high-dimensional probability distributions. Therefore it has been a common practice in applications to neural data to use the Granger causality measure specific for linear Gaussian stationary processes (e. g. Bernasconi and König, 1999; Brovelli et al., 2004) to avoid this estimation problem.

For the study of neural causal interactions it is especially appealing to examine how the causal interactions

D. Chicharro
Dept. of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain E-mail: chicharro31@yahoo.es

are associated with different brain rhythms which are functionally relevant (Buzsáki, 2006). For this purpose different spectral measures related to Granger causality have been introduced, comprising the Geweke spectral measures of Granger causality (GSC) (Geweke, 1982, 1984), the Directed Transfer Function (DTF) (Kaminski and Blinowska, 1991), and the Partial Directed Coherence (PDC) (Baccalá and Sameshima, 2001). These spectral measures have been used to analyze in which frequency bands the causal interactions between the brain regions predominantly occur (e. g. Bernasconi and König, 1999; Bernasconi et al., 2000; Kaminski et al., 2001; Brovelli et al., 2004; Bressler et al., 2008; Ladrone et al., 2009; Besserve et al., 2010).

Furthermore, Geweke (1982) showed that the measures of Granger causality form part of a decomposition of a measure of the total dependence between the processes X and Y into causal and instantaneous dependence measures. This decomposition was first formulated specifically for linear Gaussian stationary processes in the time and spectral domain using the framework of autoregressive models, for bivariate (Geweke, 1982) and multivariate (Geweke, 1984) systems. More generally, for stationary and non-stationary stochastic processes, the corresponding decomposition has also been formulated in the framework of information theory (Rissanen and Wax, 1987; Solo, 2008).

The connection between the linear Gaussian and the information theoretic decompositions in the time domain has been shown (Barnett et al., 2009; Amblard and Michel, 2011) and relies on the specific form of the entropy for Gaussian variables, as we will see below. However, the derivation of the decomposition in the spectral domain from the more general decomposition in the information theory framework has not been considered yet. Takahashi et al. (2010) showed that the DTF and the PDC are also associated with some mutual information rates that involve an explicit consideration of the autoregressive model representation of the recorded processes. We here will complete the connection between the measures in the information theory formulation and the spectral domain formulation. Based on this connection we will discuss the interpretation of the spectral measures of Granger causality questioning to which degree they are informative about the role of each frequency band in the causal interactions between brain regions.

This paper is organized as follows: in Sect. 2 we review the three frameworks of dependence and causality measures: the information theory formulation, and the one for linear Gaussian stationary processes in the time and frequency domain. In Sect. 3 we examine for the bivariate case the connection of the measures in the

spectral domain to information theoretic measures. We provide a nonparametric derivation of a spectral representation of the information theoretic measure of total dependence showing that it corresponds to the spectral measure of total dependence proposed by Geweke (1982). However, we indicate that the same type of derivation does not lead to a spectral representation of the transfer entropy. Our arguments suggest that Granger causality cannot be expressed in the spectral domain only in terms of the observed processes, without explicitly considering a parametrical autoregressive representation. In Sect. 4 we extend these results to the multivariate case.

In Sect. 5 we examine for the bivariate case how the GSC, DTF, and PDC are connected to information theoretic measures when explicitly considering the autoregressive representation. In this way we prove that, for bivariate processes, PDC is associated with generalized Sims causality (Sims, 1972), an alternative criterion to infer causality. We then extend the analysis to the multivariate case (Sect. 6). We prove that the conditional GSC can be expressed in terms of a multiple coherence in the same way that for the bivariate case GSC is related to a coherence (Takahashi et al., 2010).

Finally in Sect. 7 we discuss the implications that the lack of a spectral representation of Granger causality in terms only of the observed processes has for the interpretation of the spectral measures that are related to Granger causality. We review previous applications of these measures to study causal interactions in the brain focusing on their methodological procedures.

2 Bivariate framework of dependence and causality measures

We start reviewing the decomposition of the total dependence between two processes into causal and instantaneous dependence terms. We will review this decomposition and the measures of dependence and causality involved for linear Gaussian stationary processes in the time domain (Sect. 2.1), in the general information theoretic formulation (Sect. 2.2), and for linear Gaussian stationary processes in the spectral domain (Sect. 2.3). Notice that these decompositions have been introduced for each type of processes without considering the specific nature of the dynamics they represent. For brain signals, the linear Granger causality measure in the time domain and its information theoretic counterpart, the transfer entropy, have been widely applied (see Bressler and Seth, 2011, for a review), but the joint consideration of all the terms appearing in the decomposition of the total dependence has mainly been applied in the spectral domain (e. g. Ding et al., 2006).

Accordingly, focusing on the frequency domain decomposition, we will discuss the meaning of the measures and how they are expected to be informative about the causal interactions between brain regions.

2.1 Decomposition for linear Gaussian stationary processes in the time domain

Consider a system formed by X and Y , which are stationary stochastic processes. We follow Geweke (1982) considering the projection of X_{i+1} on his own past:

$$X_{i+1} = \sum_{s=0}^{\infty} a_{xs}^{(x)} X_{i-s} + \epsilon_{xi+1}^{(x)}, \quad \text{var}(\epsilon_x^{(x)}) = \Sigma_x^{(x)}, \quad (1)$$

and its projection on the past of both X and Y :

$$X_{i+1} = \sum_{s=0}^{\infty} a_{xys}^{(xy)} X_{i-s} + a_{xys}^{(xy)} Y_{i-s} + \epsilon_{xi+1}^{(xy)} \quad (2)$$

$$Y_{i+1} = \sum_{s=0}^{\infty} a_{yxs}^{(xy)} X_{i-s} + a_{yys}^{(xy)} Y_{i-s} + \epsilon_{yi+1}^{(xy)}$$

$$\Sigma^{(xy)} = \begin{pmatrix} \Sigma_{xx}^{(xy)} & \Sigma_{xy}^{(xy)} \\ \Sigma_{yx}^{(xy)} & \Sigma_{yy}^{(xy)} \end{pmatrix} \quad (3)$$

where $\Sigma_{xx}^{(xy)} = \text{var}(\epsilon_x^{(xy)})$, $\Sigma_{yy}^{(xy)} = \text{var}(\epsilon_y^{(xy)})$, $\Sigma_{xy}^{(xy)} = \text{cov}(\epsilon_x^{(xy)}, \epsilon_y^{(xy)})$, and $\Sigma_{yx}^{(xy)} = \Sigma_{xy}^{(xy)T}$. Notice that while the subindexes are used to refer to the corresponding variable or to components of a matrix, the superindexes refer to the particular projection.

The Geweke measure of Granger causality is defined as:

$$F_{Y \rightarrow X} = \ln\left(\frac{\Sigma_x^{(x)}}{\Sigma_{xx}^{(xy)}}\right). \quad (4)$$

The measure reflects the principle first formulated by Wiener (1956) and formalized by Granger (1963) that causality from Y to X implies an improvement in predictability when using information from the past of Y together with from the past of X , and thus reducing the variance of the innovation variables ($\Sigma_{xx}^{(xy)}$) with respect to the variance obtained when only information from the own past is used ($\Sigma_x^{(x)}$). The Geweke measure of Granger causality from X to Y , $F_{X \rightarrow Y}$, is defined analogously to (4) using the projections of Y_{i+1} .

Apart from these measures of causality, two other measures were proposed by Geweke (1982) to characterize the interactions between the processes. A measure of instantaneous causality is defined as

$$F_{X \cdot Y} = \ln\left(\frac{\Sigma_{xx}^{(xy)} \Sigma_{yy}^{(xy)}}{|\Sigma^{(xy)}|}\right), \quad (5)$$

which is zero for $\Sigma^{(xy)}$ diagonal, that is, if the innovations are uncorrelated. Finally, the measure of total dependence between X and Y is (Geweke, 1982):

$$F_{X,Y} = \ln\left(\frac{\Sigma_x^{(x)} \Sigma_y^{(y)}}{|\Sigma^{(xy)}|}\right), \quad (6)$$

where $\Sigma_y^{(y)}$ is obtained from the analogous projection of Y_{i+1} like in (1). This measure compares the prediction error under the assumption that the processes are independent with the prediction error in the bivariate projection, thus accounting for any source of interdependence. From the definition of the measures it is straightforward to check that:

$$F_{X,Y} = F_{X \rightarrow Y} + F_{Y \rightarrow X} + F_{X \cdot Y}. \quad (7)$$

This decomposition shows that the total dependence arise from the causal interactions in both directions plus a term of instantaneous interactions. It is important to notice that none of the terms in (7) was defined by Geweke (1982) *ad hoc* to enforce this decomposition. Oppositely, each measure has a clear interpretation on its own and the existence of this decomposition only results *a posteriori* from the meaningful definition of the measures.

2.2 Decomposition for stochastic processes in the information theory framework

Measures analogous to the ones above but for general stochastic processes have also been introduced (Rissanen and Wax, 1987; Solo, 2008). The principle of Granger based on predictability is generally formulated in terms of conditional independence of specific probability distributions: the general criterion of Granger noncausality is

$$p(X_{i+1}|X^i) = p(X_{i+1}|X^i, Y^i), \quad (8)$$

where $X^i = \{X_i, X_{i-1}, \dots, X_{i-n}\}$ for $n \rightarrow \infty$, and analogously for Y^i . An information theoretic measure of causality based on this criterion has been repeatedly proposed in different fields (e. g. Marko, 1973; Gourioux et al., 1987; Schreiber, 2000). We follow Schreiber (2000) and we call this information theoretic measure of causality from Y to X the *transfer entropy* rate defined as:

$$\begin{aligned} T_{Y \rightarrow X} &= \sum_{X_{i+1}, X^i, Y^i} p(X_{i+1}, X^i, Y^i) \log \frac{p(X_{i+1}|X^i, Y^i)}{p(X_{i+1}|X^i)} \\ &= H(X_{i+1}|X^i) - H(X_{i+1}|X^i, Y^i) \\ &= I(X_{i+1}; Y^i|X^i). \end{aligned}$$

(9)

Here $H(\cdot)$ denotes the entropy and $I(\cdot)$ the mutual information (Cover and Thomas, 2006). The transfer entropy rate corresponds to the Kullback-Leibler distance (Cover and Thomas, 2006) between the partial probability distributions $p(X_{i+1}|X^i, Y^i)$ and $p(X_{i+1}|X^i)$. Accordingly, $I(X_{i+1}; Y^i|X^i) = 0$ if and only if the equality in (8) holds.

The correspondence between $T_{Y \rightarrow X}$ and $F_{Y \rightarrow X}$ can be seen (Barnett et al., 2009) taking into account that the entropy of a N -variate Gaussian distribution is completely determined by its covariance matrix Σ :

$$H(X_{\text{Gaussian}}^N) = \frac{1}{2} \ln((2\pi e)^N |\Sigma|). \quad (10)$$

Accordingly, the two measures are such that:

$$F_{Y \rightarrow X} = 2 T_{Y \rightarrow X}. \quad (11)$$

Based also on the specific form of the entropy for Gaussian variables (10), it is also easy to infer the information theoretic measures corresponding to the other terms in the decomposition of (7). In particular, we have that for linear Gaussian stationary processes

$$F_{X,Y} = 2 (I(X_{i+1}X^i; Y_{i+1}Y^i) - I(X^i; Y^i)), \quad (12)$$

where

$$I(X_{i+1}X^i; Y_{i+1}Y^i) - I(X^i; Y^i) = H(Y_{i+1}|Y^i) + H(X_{i+1}|X^i) - H(X_{i+1}; Y_{i+1}|X^iY^i). \quad (13)$$

Similarly,

$$F_{X \cdot Y} = 2I(X_{i+1}; Y_{i+1}|X^iY^i), \quad (14)$$

where

$$I(X_{i+1}; Y_{i+1}|X^iY^i) = -H(X_{i+1}; Y_{i+1}|X^iY^i) + H(Y_{i+1}|X^iY^i) + H(X_{i+1}|X^iY^i). \quad (15)$$

Altogether this results in the general decomposition

$$I(X_{i+1}X^i; Y_{i+1}Y^i) - I(X^i; Y^i) = T_{X \rightarrow Y} + T_{Y \rightarrow X} + I(X_{i+1}; Y_{i+1}|X^iY^i). \quad (16)$$

This decomposition subsumes the one in (7), which is specific for linear Gaussian stationary processes. Furthermore, although we assumed stationarity to connect with (7), the measures in (16) can also be used for non-stationary processes considering i as a particular point in time, so that the probability distributions such as the ones in (8) are defined in a time-resolved way.

The correspondence between the information theoretic formulation and the one for linear Gaussian stationary processes goes beyond the level of the four measures appearing in the decompositions in (7) and (16). There is also a one-to-one relation between the components of the measures. In particular, each variance appearing in the linear Gaussian stationary measures has a direct relation to an entropy term appearing in the information theoretic measures. For example, given (1) and (10)

$$H(X_{i+1}|X^i) = \frac{1}{2} \ln 2\pi e + \frac{1}{2} \ln \Sigma_x^{(x)}, \quad (17)$$

and given (2) and (10)

$$H(X_{i+1}|X^iY^i) = \frac{1}{2} \ln 2\pi e + \frac{1}{2} \ln \Sigma_{xx}^{(xy)}. \quad (18)$$

Analogous equalities exist for the other individual entropies in (13) and (15).

2.3 Decomposition for linear Gaussian stationary processes in the frequency domain

Geweke (1982) also proposed a spectral decomposition of the time domain Granger causality measure (4) and of the other terms in (7). Geweke derived the spectral measures requiring the fulfillment of some properties: First, the measures had to be nonnegative. Second for the spectral measure of causality from Y to X , $f_{Y \rightarrow X}(\omega)$, it was required that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f_{Y \rightarrow X}(\omega) d\omega = F_{Y \rightarrow X}, \quad (19)$$

and analogously for the other spectral measures. These two conditions together imply that

$$F_{Y \rightarrow X} = 0 \Leftrightarrow f_{Y \rightarrow X}(\omega) = 0 \quad \forall \omega. \quad (20)$$

As a third condition, the spectral measures should have an intuitive interpretation so that the spectral decomposition is useful for empirical applications.

We here will describe the Geweke spectral measure of Granger causality (GSC) for the simplified case in which the covariance matrix in (3) is diagonal so that there is no instantaneous correlation between X and Y . This enormously simplifies its derivation and suffices for our purpose of characterizing the spectral measure of Granger causality examining its derivation from an information theoretic measure. The GSC is obtained from the spectral representation of the bivariate autoregressive process as follows. Fourier transforming (2) leads to:

$$\begin{pmatrix} A_{xx}^{(xy)}(\omega) & A_{xy}^{(xy)}(\omega) \\ A_{yx}^{(xy)}(\omega) & A_{yy}^{(xy)}(\omega) \end{pmatrix} \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} \epsilon_x^{(xy)}(\omega) \\ \epsilon_y^{(xy)}(\omega) \end{pmatrix}, \quad (21)$$

where we have $A_{xx}^{(xy)}(\omega) = 1 - \sum_{s=1}^{\infty} a_{xxs}^{(xy)} e^{-i\omega s}$, as well as $A_{xy}^{(xy)}(\omega) = -\sum_{s=1}^{\infty} a_{xys}^{(xy)} e^{-i\omega s}$, and analogously for $A_{yy}^{(xy)}(\omega)$, $A_{yx}^{(xy)}(\omega)$. The coefficients matrix $\mathbf{A}^{(xy)}(\omega)$ can be inverted into the transfer function $\mathbf{H}^{(xy)}(\omega) = (\mathbf{A}^{(xy)})^{-1}(\omega)$, so that

$$\begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} H_{xx}^{(xy)}(\omega) & H_{xy}^{(xy)}(\omega) \\ H_{yx}^{(xy)}(\omega) & H_{yy}^{(xy)}(\omega) \end{pmatrix} \begin{pmatrix} \epsilon_x^{(xy)}(\omega) \\ \epsilon_y^{(xy)}(\omega) \end{pmatrix}. \quad (22)$$

Accordingly, the spectral matrix can be expressed as:

$$\mathbf{S}^{(xy)}(\omega) = \mathbf{H}^{(xy)}(\omega) \mathbf{\Sigma}^{(xy)} (\mathbf{H}^{(xy)})^*(\omega) \quad (23)$$

where $*$ denotes complex conjugate and matrix transpose. Given the lack of instantaneous correlations

$$S_{xx}(\omega) = \Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2 + \Sigma_{yy}^{(xy)} |H_{xy}^{(xy)}(\omega)|^2. \quad (24)$$

The GSC from Y to X at frequency ω is defined as:

$$f_{Y \rightarrow X}(\omega) = \ln \frac{S_{xx}(\omega)}{\Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2}. \quad (25)$$

This definition fulfills the requirement of being nonnegative since, given (24), $S_{xx}(\omega)$ is always higher than $\Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2$. It also fulfills the requirement of being intuitive since $f_{Y \rightarrow X}(\omega)$ quantifies the portion of the power spectrum which is associated with the intrinsic innovation process of X . We will show in Sect. 3 that it also fulfills the requirement of (19), which establishes the link between the time domain and frequency domain measures.

Geweke did not explicitly complete a decomposition analogous to (7) and (16) in the spectral domain. However, he took into account that, for bivariate linear Gaussian stationary processes, a spectral decomposition of the mutual information rate $\mathcal{I}(X^i; Y^i)$ already had been introduced by Gelfand and Yaglom (1959) as:

$$\begin{aligned} \mathcal{I}(X^i; Y^i) &= \frac{-1}{4\pi} \int_{-\pi}^{\pi} \ln \frac{|\mathbf{S}^{(xy)}(\omega)|}{S_{xx}(\omega) S_{yy}(\omega)} d\omega \\ &= \frac{-1}{4\pi} \int_{-\pi}^{\pi} \ln (1 - |C(X, Y)|^2) d\omega, \end{aligned} \quad (26)$$

where $|C(X, Y)|^2$ is the squared coherence of X and Y (Priestley, 1981). Considering this, it is possible (Ding et al., 2006) to introduce a spectral measure of total dependence between X and Y at a frequency ω :

$$\begin{aligned} f_{X,Y}(\omega) &= \mathcal{I}(X(\omega); Y(\omega)) = \ln \frac{S_{xx}(\omega) S_{yy}(\omega)}{|\mathbf{S}^{(xy)}(\omega)|} \\ &= -\ln (1 - |C(X, Y)|^2), \end{aligned} \quad (27)$$

where $\mathcal{I}(X(\omega); Y(\omega))$ is the mutual information rate between the spectral variables $X(\omega)$ and $Y(\omega)$. The requirement analogous to (19) is thus fulfilled by definition. This definition also fulfills the nonnegativity condition and has a clear interpretation, since the coherence is a measure which is widely applied to quantify the interdependence of the processes in the spectral domain.

Defining $f_{X \rightarrow Y}(\omega)$ analogously to (25), the definition of the instantaneous causality spectral measure $f_{X,Y}(\omega)$ is chosen *ad hoc* to satisfy

$$f_{X,Y}(\omega) = f_{X \rightarrow Y}(\omega) + f_{Y \rightarrow X}(\omega) + f_{X \cdot Y}(\omega). \quad (28)$$

This results, for the case of no instantaneous causality, in

$$f_{X \cdot Y}(\omega) = \ln \frac{\Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2 \Sigma_{yy}^{(xy)} |H_{yy}^{(xy)}(\omega)|^2}{|\mathbf{S}^{(xy)}(\omega)|}. \quad (29)$$

Being introduced *ad hoc* to complete the decomposition, the spectral measure $f_{X \cdot Y}(\omega)$ does not fulfill the requirements of Geweke. It may be negative for some frequencies and has no clear physical meaning (Ding et al., 2006). The lack of nonnegativity can be seen considering that the integration over frequencies of the decomposition in (28) has to be consistent with the decomposition in (7). For the lack of instantaneous causality in the time domain $F_{X,Y} = 0$, but, if a bidirectional coupling exists, $f_{X \cdot Y}(\omega)$ is generally nonzero even if there is no instantaneous causality. Since the integral of $f_{X \cdot Y}(\omega)$ has to cancel when $F_{X,Y} = 0$, not being zero for all frequencies, this implies the violation of nonnegativity. This difference in the properties of $f_{X \cdot Y}(\omega)$ and $F_{X,Y}$ also indicates that also the other terms in the time and the spectral domain decompositions cannot play equivalent roles.

Enforcing the definition of $f_{X \cdot Y}(\omega)$ to fulfill the decomposition in the spectral domain impairs the usefulness of such decomposition. The value of the decompositions in the time domain and information theoretic framework relies on the possibility to physically interpret them as quantifying how any interdependence between two processes results from either causal interactions or from instantaneous dependencies. Having a measure of causality embedded in a such decomposition may be helpful when studying a system to, for example, better analyze the contribution of the causal effects to the level of synchronization between the processes.

For the study of neural data, given that it has been hypothesized that neural coherence plays an important role in neural communication (Fries, 2005), the connection with $f_{X,Y}(\omega)$, which is directly related to the coherence, helps to interpret the spectral measure of causality $f_{X \rightarrow Y}(\omega)$. In fact, it is common when causal

analysis between brain regions is carried out in the spectral domain to examine how the location of the peaks in the spectrum of the causal measures is related to the location of the peaks in the power spectrum and the coherence spectrum (Brovelli et al., 2004; Chen et al., 2006; Bressler et al., 2007).

Given the requirements imposed by Geweke, the measure $f_{X \rightarrow Y}(\omega)$ has a clear interpretation as the portion of the power spectrum associated with the innovation process $\epsilon_x^{(xy)}$. However, this interpretation relies on the use of the autoregressive representation, because the innovation process is only well-defined in the context of this representation and does not correspond to any process recorded from a given brain region. Since the spectral measures of causality should serve us to study how the causal effects between two regions from which we record are channeled in a given brain rhythm, we would like to know how, if possible, they can be expressed in terms of the X and Y variables, like the time domain and information theoretic Granger causality measures. This is because the processes X and Y directly reflect the activity of the brain regions and do not depend on the specific autoregressive representation.

For that purpose, given the generality of the information theoretic framework, we will examine how the different spectral measures can be derived from information theoretic measures. We do that to connect the spectral formulation (28) with the information theoretic formulation (16), but more specifically our main interest is to better characterize the Granger causality spectral measure.

3 The connection between the information theoretic framework and the spectral framework of dependence and causality measures

In the information theoretic formulation all the measures are expressed involving only variables corresponding to the processes X and Y , and thus are independent of any particular representation, like the autoregressive representation. In the time domain, the linear stationary Gaussian measures result to be the specific form that take the information theoretic measures for linear Gaussian stationary processes. As we said, the correspondence not only holds for each measure but also for their different components, so that each variance of the innovations is related to an entropy involving only variables of the processes X and Y , as shown for example in (17) and (18).

The connection between the time domain and frequency domain Granger causality measures is based on the requirements expressed in (19) and (20) for GSC.

We want to examine if it is possible to derive the GSC and the other spectral measure in (28) from the information theoretic measures in the same way that it has been done for the time domain measures. In particular, we want to see if there is a one-to-one correspondence between some terms derived from the four terms in the information theoretic representation of the decomposition of the total dependence (16) and the four terms in spectral representation of the decomposition in (28). Furthermore, we want to examine if this connection also holds for the components of the spectral measures, that is, we want to determine if, for example, $\Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2$, which appears in the definition of $f_{Y \rightarrow X}(\omega)$ (25), can be related to an information theoretic quantity similarly to the relation shown in (18) for $\Sigma_{xx}^{(xy)}$, which appears in the definition of $F_{Y \rightarrow X}$.

To do so we will start by considering an important equality between two alternative definitions of the entropy rate of a stochastic process under the existence of stationarity. The first definition is:

$$\mathcal{H}(X^i) = \lim_{n \rightarrow \infty} \frac{H(X_1, X_2, \dots, X_n)}{n}, \quad (30)$$

while the second is:

$$H(X_{i+1}|X^i) = \lim_{n \rightarrow \infty} H(X_{n+1}|X_n, X_{n-1}, \dots, X_1). \quad (31)$$

The existence of the limit of (31) is assured by the existence of the autoregressive representation in (1), and corresponds to the expression in (17). It can then be shown that for a stationary stochastic process these limits are equal (Theorem 4.2.1, Cover and Thomas (2006)):

$$\mathcal{H}(X^i) = H(X_{i+1}|X^i). \quad (32)$$

Below in this section we will consider in more detail why this equality holds. For the moment it is important to note that the reason why this equality is relevant for our arguments is that while $H(X_{i+1}|X^i)$ explicitly considers a distinction between the past X^i , and the value one step in the future X_{i+1} , the other definition $\mathcal{H}(X^i)$ does not distinguish between future and past variables. This lack of distinction is convenient to express the information theoretic measure in terms of frequency variables. In (17) we showed the relation between $H(X_{i+1}|X^i)$ and the variance of $\epsilon_x^{(x)}$, the innovation term in the projection in (1). Similarly, for linear stationary Gaussian processes, we have that

$$\mathcal{H}(X^i) = \frac{1}{2} \ln 2\pi e + \lim_{n \rightarrow \infty} \frac{1}{2n} \ln |\Sigma(X^n)|. \quad (33)$$

where $\Sigma(X^n)$ is the covariance matrix of X_1, X_2, \dots, X_n . In general, this covariance matrix is not diagonal, however, for a linear Gaussian stationary process, the frequency modes are independent and thus the matrix is

diagonal, consistently with the additivity of the entropy for independent variables (Cover and Thomas, 2006). Therefore, changing to the basis of the frequency modes we get

$$\begin{aligned}\mathcal{H}(X^i) &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathcal{H}(X(\omega_k)) \\ &= \frac{1}{2} \ln 2\pi e + \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{k=1}^n \ln \Sigma(\omega_k),\end{aligned}\quad (34)$$

and considering a continuous spectrum in the limit $n \rightarrow \infty$

$$\begin{aligned}\mathcal{H}(X^i) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{H}(X(\omega)) d\omega = \frac{1}{2} \ln 2\pi e \\ &\quad + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln S_{xx}(\omega) d\omega.\end{aligned}\quad (35)$$

Accordingly, the component of $f_{Y \rightarrow X}(\omega)$ that contains $S_{xx}(\omega)$ in (25) is related to an entropy involving only X variables in the same way that in (17) $\Sigma_x^{(x)}$ is related to $H(X_{i+1}|X^i)$:

$$\mathcal{H}(X(\omega)) = \frac{1}{2} \ln 2\pi e + \frac{1}{2} \ln S_{xx}(\omega). \quad (36)$$

Combining (17), (32), and (35), we obtain that

$$\ln \Sigma_x^{(x)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S_{xx}(\omega) d\omega, \quad (37)$$

which is the well-known equality of Kolmogorov (1939). We also have that

$$H(X_{i+1}|X^i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{H}(X(\omega)) d\omega, \quad (38)$$

so that $H(X_{i+1}|X^i)$ has a spectral representation in terms of $S_{xx}(\omega)$. This means that one of the two entropies appearing in the definition of the transfer entropy (9) has a spectral representation corresponding to one of the components of $f_{Y \rightarrow X}(\omega)$. This entropy $H(X_{i+1}|X^i)$ also appears in the definition of the total dependence in the information theoretic framework (13).

We continue examining which other entropies used to define the information theoretic measures have a spectral representation. Following the same line of arguments and substituting X by the bivariate stochastic process XY the same theorem (Theorem 4.2.1, Cover and Thomas (2006)) assures that for stationary processes

$$\mathcal{H}(X^i Y^i) = H(X_{i+1} Y_{i+1} | X^i Y^i). \quad (39)$$

For linear Gaussian stationary processes these entropy rates can be expressed as:

$$H(X_{i+1} Y_{i+1} | X^i Y^i) = \ln 2\pi e + \frac{1}{2} \ln |\Sigma^{(xy)}|, \quad (40)$$

in terms of the covariance matrix of the innovations in the joint autoregressive representation in (2), and as

$$\begin{aligned}\mathcal{H}(X^i Y^i) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H(X(\omega) Y(\omega)) d\omega \\ &= \ln 2\pi e + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln |\mathbf{S}^{(xy)}(\omega)| d\omega,\end{aligned}\quad (41)$$

where we have taken again into account that in the basis of the frequency modes all the modes are independent for X and Y except for the same frequency of X and Y . Accordingly, we have that

$$H(X(\omega) Y(\omega)) = \ln 2\pi e + \frac{1}{2} \ln |\mathbf{S}^{(xy)}(\omega)|, \quad (42)$$

and

$$H(X_{i+1} Y_{i+1} | X^i Y^i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(X(\omega) Y(\omega)) d\omega, \quad (43)$$

which means that another of the entropies appearing in the information theoretic measure of total dependence (13) as well as in the information theoretic measure of instantaneous causality (15) has a spectral representation corresponding to a component of $f_{X,Y}(\omega)$ (27) and $f_{X \cdot Y}(\omega)$ (29).

The existence of the relations between $H(X_{i+1}|X^i)$ and $S_{xx}(\omega)$ (38) and between $H(X_{i+1} Y_{i+1} | X^i Y^i)$ and $|\mathbf{S}^{(xy)}(\omega)|$ (43) suffices to derive the decomposition of the total dependence into terms of causality and instantaneous dependence in the spectral domain (28) from the information theoretic decomposition (16). This is because the equality in (16) is obtained after the cancellation of the terms $H(Y_{i+1}|X^i Y^i)$ and $H(X_{i+1}|X^i Y^i)$, which appear in $T_{X \rightarrow Y} + T_{Y \rightarrow X}$ and $I(X_{i+1}; Y_{i+1} | X^i Y^i)$ with opposite sign. Therefore, at the level of the decompositions, we do not have to further examine how $H(Y_{i+1}|X^i Y^i)$ and $H(X_{i+1}|X^i Y^i)$ are linked to any quantity in the frequency domain.

We will now consider if the individual terms in (16) have a spectral representation. Combining (35) and (41) we see that the components of the mutual information rate proposed by Gelfand and Yaglom (1959) (26) correspond to what we expect from its definition in terms of entropies (Cover and Thomas, 2006):

$$\mathcal{I}(X^i; Y^i) = \mathcal{H}(X^i) + \mathcal{H}(Y^i) - \mathcal{H}(X^i Y^i). \quad (44)$$

Furthermore, using the equalities between the entropy rates in (32) and (39), and comparing (13) with (44), we see that the equality between the entropy rates also leads to the equality between the alternative definitions of mutual information rate, corresponding one of these definitions to the definition of the information theoretic measure of total dependence:

$$I(X_{i+1} X^i; Y_{i+1} Y^i) - I(X^i; Y^i) = \mathcal{I}(X^i; Y^i). \quad (45)$$

Therefore the information theoretic measure of total dependence has a spectral representation and, as happen for the connection with the time domain measures, the correspondence with the spectral measure also holds for the different entropies that compound the information theoretic measure of total dependence. These entropies, under the assumption of having linear Gaussian stationary processes and Fourier transforming X^i and Y^i , naturally take the form shown in (36) and (42).

We continue considering the spectral representation of $T_{X \rightarrow Y}$, $T_{Y \rightarrow X}$, and $I(X_{i+1}; Y_{i+1} | X^i Y^i)$. This representation would require that also the terms canceled in (16), that is, $H(Y_{i+1} | X^i Y^i)$ and $H(X_{i+1} | X^i Y^i)$ could be represented in the spectral domain. We will show that it is not possible to do so using the same procedure used for $H(X_{i+1} | X^i)$ and $H(X_{i+1} Y_{i+1} | X^i Y^i)$. For that purpose we will first review how the equality between the entropy rates in (32) is obtained (Theorem 4.2.1, Cover and Thomas (2006)). Remember that this equality allowed us above to establish the link between entropies like $H(X_{i+1} | X^i)$, that appears in the definition of the information theoretic measures of dependence and causality, and a spectral representation derived from the Fourier transformation of X^i in $\mathcal{H}(X^i)$. Using the chain rule for entropies (Cover and Thomas, 2006) we have that the term inside the limit used to define $\mathcal{H}(X^i)$ can be rewritten as:

$$\frac{H(X_1, X_2, \dots, X_n)}{n} = \frac{1}{n} \sum_{i=0}^{n-1} H(X_{i+1} | X_i, X_{i-1}, \dots, X_1), \quad (46)$$

so that the entropy rate is the time average of conditional entropies that have the same form as the one used for the alternative definition of the entropy rate (31). Theorem 4.2.3 in Cover and Thomas (2006) (called the *Cesáro mean*) states that if a series of terms a_n has a limit $a_n \rightarrow a$, then another series $b_n = \frac{1}{n} \sum_{i=0}^{n-1} a_{i+1}$ also has the limit $b_n \rightarrow a$. Therefore, the possibility to use the chain rule to express $H(X_1, X_2, \dots, X_n)$ as conditional entropies with the form appearing in the alternative definition of the entropy rate is necessary to apply the *Cesáro mean*, which establishes the equality of the limits.

We will now apply the same logic used in the proof of the equality of the entropy rates in (32) to the entropy rates $H(Y_{i+1} | X^i Y^i)$ and $H(X_{i+1} | X^i Y^i)$ and consider how they arise as a limit for $n \rightarrow \infty$. $H(X_{i+1} | X^i Y^i)$ corresponds to the entropy rate:

$$H(X_{i+1} | X^i Y^i) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, Y_{n-1}, X_{n-2}, Y_{n-2}, \dots, X_1, Y_1), \quad (47)$$

and analogously for $H(Y_{i+1} | X^i Y^i)$.

We now have to consider the alternative definition of the entropy rate as an average of the joint entropy, analogously to (30):

$$\mathcal{H}(X_{i+1} X^i Y^i) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, Y_1, X_2, Y_2, \dots, X_n, Y_n, X_{n+1}). \quad (48)$$

However, we see that the definition of such entropy rate is problematic for our aims. When changing from $H(X_{i+1} | X^i)$ to $\mathcal{H}(X^i)$ the distinction between past and future disappears, and so does when changing from $H(X_{i+1} Y_{i+1} | X^i Y^i)$ to $\mathcal{H}(X^i Y^i)$. But in $H(X_{i+1} | X^i Y^i)$ there is an asymmetry between X and Y based on the fact that only the future of X appears. Since this asymmetry intrinsically depends on the distinction between future and past, we cannot get rid of this distinction in the definition of the alternative entropy rate $\mathcal{H}(X_{i+1} X^i Y^i)$ in (48). In fact, the variable X_{i+1} is what distinguishes $\mathcal{H}(X_{i+1} X^i Y^i)$ from $\mathcal{H}(X^i Y^i)$, which is equal to $H(X_{i+1} Y_{i+1} | X^i Y^i)$. The variable X_{i+1} is needed to maintain the asymmetry but prevents from expressing the variables in the frequency domain. We cannot Fourier transform $X_{i+1} X^i Y^i$ in the same way as X^i or $X^i Y^i$ because the spectral space is not compatible with maintaining such temporal distinction. Therefore, establishing the equivalence with an alternative entropy rate such as $\mathcal{H}(X_{i+1} X^i Y^i)$ would not allow us to connect to the spectral representation in the same way we did before. Furthermore, even the equivalence of these alternative entropy rates is problematic. The chain rule should be applied in a way that conditional entropies of the form $H(X_n | X_{n-1}, Y_{n-1}, \dots, X_1, Y_1)$ appear, analogously to (46). But then it would not be possible to avoid the appearance of other terms with the form $H(Y_n | X_n, X_{n-1}, Y_{n-1}, \dots, X_1, Y_1)$. Therefore, in this case the *Cesáro mean* could not be used to equate the two alternative definitions of entropy rate.

These arguments suggest that $H(Y_{i+1} | X^i Y^i)$ and $H(X_{i+1} | X^i Y^i)$ do not have a spectral representation analogous to the one that have $H(X_{i+1} | X^i)$ (38) and $H(X_{i+1} Y_{i+1} | X^i Y^i)$ (43). Accordingly, also the transfer entropy would be incompatible with a spectral representation derived without explicitly considering the parametric autoregressive representation (see Sect. 5.1 for such derivation). Notice that we are not pointing to any inconsistency of the spectral measure of Granger causality proposed by Geweke with the requirements imposed in Geweke (1982). These requirements, in particular (19), are weaker than the existence of a spectral representation for the transfer entropy. This is because (19) only links the time domain measure with the integral across all frequencies. It does not require that, for each frequency, an information theoretic measure in-

volving variables from the observed processes X and Y takes the form of the spectral measure when expressed under the assumption of having linear Gaussian stationary processes. The difference is important because, when using the spectral measure of Granger causality, we specifically aim to interpret it at each frequency in terms of causal interactions between the observed processes, and not just to know that integrated across all frequencies they are consistent with the time domain Granger causality measure.

Geweke (1982) showed that the fulfillment of (19) relies on Theorem 4.2 Rozanov (1967). Considering the conventional definition of the lag operator such that $Lx_i = x_{i-1}$, this theorem states that for a square matrix lag operator $D(L)$ with a leading term 1 and all roots outside the unit circle:

$$\int_{-\pi}^{\pi} \ln |D(\omega)|^2 d\omega = 0. \quad (49)$$

This theorem is applied to show that the integration of the component in the denominator of $f_{Y \rightarrow X}(\omega)$ (25) results in:

$$\int_{-\pi}^{\pi} \ln (\Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2) d\omega = 2\pi \ln \Sigma_{xx}^{(xy)}. \quad (50)$$

Therefore, it is clear that the use of the integration across frequencies is necessary to fulfill the requirement of (19), without deriving from $H(X_{i+1}|X^i Y^i)$ any spectral representation like the ones of (36) and (42).

To derive the linear Gaussian stationary measures in the time domain from the information theoretic measures it suffices to consider the form of the entropy for Gaussian variables (10). To derive the linear Gaussian stationary measures in the frequency domain we used the equalities between the different definitions of entropy rates (32) and (39) and the Fourier transformation of X^i and Y^i . This procedure in the frequency domain allowed us to show the correspondence between the spectral and information theory formulations at the level of the decomposition of the total dependence into causal and instantaneous terms. At the level of the components of the measures, this procedure only allowed us to derive the spectral form of some of entropies in (36) and (42), which suffice for a spectral representation of the total dependence (13), but not of the transfer entropy (9) and the instantaneous causality (15).

The fact that the same procedure that allows us to establish the connection at the level of the decompositions and to express $H(X_{i+1}|X^i)$, $H(Y_{i+1}|Y^i)$, and $H(X_{i+1}Y_{i+1}|X^i Y^i)$ in their spectral representation is not valid to derive a spectral representation for the entropy rates $H(Y_{i+1}|X^i Y^i)$ and $H(X_{i+1}|X^i Y^i)$ does not constitute a proof of inexistence of their spectral representation. Therefore we have not provided a proof of the

inexistence of a spectral representation of the transfer entropy containing only variables of the observed processes X and Y . However, the way the procedure fails in comparison with how it is valid for the other entropies is illustrative of the main impediment to obtain such a spectral representation. The asymmetry between the role of X variables and Y variables is based on the distinction between the future and the past. A valid spectral representation of the transfer entropy would need to reflect this asymmetry in the spectral space. If this is possible remains as an open question.

4 The connection between the formulations for the multivariate framework of dependence and causality measures

Above we showed for the bivariate case the problems to obtain a spectral representation for the transfer entropy. Here we deal with the spectral representation in the case of multivariate processes. We will not review the multivariate formulation for linear Gaussian stationary processes in the time domain, in the frequency domain, and the general formulation in the information theory framework as we did for the bivariate case (Sect. 2). Here we will only indicate what differs from the bivariate case in relation to the connection between the different formulations.

For a multivariate process W , to assess the existence of causality from Y to X it is necessary to consider also the statistical dependence between X and Y that is mediated by $Z = W \setminus XY$. Accordingly, in the information theoretic formulation, in all the partializations appearing in (9), (13), and (15) one has to include Z^i . With this inclusion the decomposition in (16) is converted to:

$$I(X_{i+1}X^i; Y_{i+1}Y^i|Z^i) - I(X^i; Y^i|Z^i) = T_{X \rightarrow Y|Z} + T_{Y \rightarrow X|Z} + I(X_{i+1}; Y_{i+1}|X^i Y^i Z^i), \quad (51)$$

which given the terms that cancel between $T_{X \rightarrow Y|Z}$, $T_{Y \rightarrow X|Z}$ and $I(X_{i+1}; Y_{i+1}|X^i Y^i Z^i)$ results in

$$I(X_{i+1}X^i; Y_{i+1}Y^i|Z^i) - I(X^i; Y^i|Z^i) = H(X_{i+1}|X^i Z^i) + H(Y_{i+1}|Y^i Z^i) - H(X_{i+1}Y_{i+1}|X^i Y^i Z^i). \quad (52)$$

The terms appearing in the r.h.s. of (52) have a structure analogous to $H(Y_{i+1}|X^i Y^i)$ and $H(X_{i+1}|X^i Y^i)$. For example, in $H(X_{i+1}|X^i Z^i)$ there is an asymmetry between the variables of X and Z based on the distinction between future and past. Accordingly, the logic used in Sect. 3 to derive the spectral representation for $H(X_{i+1}|X^i)$, which plays the same role of $H(X_{i+1}|X^i Z^i)$ in the bivariate case (see (13)), is not

valid in the multivariate case. The lack of an equality analogous to (32) for $H(X_{i+1}|X^i Z^i)$ also implies that there is not an equality for alternative partial mutual information rates, analogously to (45). The conditional total dependence $I(X_{i+1}X^i; Y_{i+1}Y^i|Z^i) - I(X^i; Y^i|Z^i)$ cannot be associated with the partial mutual information rate $\mathcal{I}(X^i; Y^i|Z^i)$, like the total dependence was associated with the mutual information rate in (45). Accordingly, in the conditional case, also for the information theoretic measures of total dependence a spectral representation is not obtained. We cannot obtain this spectral representation for the conditional transfer entropies $T_{X \rightarrow Y|Z}$ and $T_{Y \rightarrow X|Z}$ either.

5 The interpretation of the spectral measures related to Granger causality in bivariate processes

The arguments in Sect. 3 suggested that the transfer entropy, which is the generalized measure of Granger causality in the framework of information theory, does not have a spectral representation for linear Gaussian stationary processes that involves only variables from the observed processes. We argued that (19) is a weaker requirement that only constraints the integration of the spectral measure across all frequencies. We now want to examine, at the level of each frequency, how to interpret the Geweke spectral measure of Granger causality (GSC) (Geweke, 1982) and other spectral measures related to Granger causality, namely the Directed Transfer Function (DTF) and the Partial Directed Coherence (PDC). In this section we discuss the bivariate case and we address the multivariate case in Sect. 6.

5.1 Definition of the bivariate Directed Transfer Function and the Partial Directed Coherence

We start reviewing the definition of these two spectral measure for the bivariate case. We here will focus on the extended version of DTF and PDC introduced by Takahashi et al. (2010), namely the information directed transfer function (iDTF) and the information partial directed coherence (iPDC), which are related to mutual information rates.

Given the joint autoregressive model of X and Y (2), and assuming for simplicity that $\Sigma^{(xy)}$ (3) is diagonal, the iDTF from Y to X is defined as (Takahashi et al., 2010):

$$i\gamma_{xy}^{(xy)}(\omega) = \frac{H_{xy}^{(xy)}(\omega) \sqrt{\Sigma_{yy}^{(xy)}}}{\sqrt{S_{xx}(\omega)}} = C(X, \epsilon_y^{(xy)}), \quad (53)$$

where $H_{xy}^{(xy)}(\omega)$ is a component of the transfer function in (22) and $C(X, \epsilon_y^{(xy)})$ is the coherence between X and $\epsilon_y^{(xy)}$. The iDTF is equal to the DTF (Kaminski and Blinowska, 1991) for $\Sigma^{(xy)}$ the identity matrix and equal to the directed coherence (Baccalá et al., 1999) for $\Sigma^{(xy)}$ diagonal (as considered here). In the bivariate case, considering the definition of the bivariate GSC $f_{Y \rightarrow X}(\omega)$ (25), there is a direct connection between the bivariate GSC and the bivariate iDTF:

$$\begin{aligned} f_{Y \rightarrow X}(\omega) &= -\ln(1 - |C(X, \epsilon_y^{(xy)})|^2) \\ &= -\ln(1 - |i\gamma_{xy}^{(xy)}(\omega)|^2), \end{aligned} \quad (54)$$

where $|C(X, \epsilon_y^{(xy)})|^2$ is the squared coherence of X with the innovations $\epsilon_y^{(xy)}$ of (2). Given the expression of $f_{Y \rightarrow X}(\omega)$ in (54) and the general relation of the mutual information rate with the squared coherence (26), Takahashi et al. (2010) showed that:

$$\begin{aligned} F_{Y \rightarrow X} &= 2 \mathcal{I}(X^i; \epsilon_y^{(xy)i}) \\ &= \frac{-1}{2\pi} \int_{-\pi}^{\pi} \ln(1 - |i\gamma_{xy}^{(xy)}(\omega)|^2) d\omega. \end{aligned} \quad (55)$$

The information partial directed coherence (iPDC) from Y to X (Takahashi et al., 2010) is defined as:

$$\begin{aligned} i\pi_{xy}^{(xy)}(\omega) &= \frac{A_{xy}^{(xy)}(\omega)}{\sqrt{\Sigma_{xx}^{(xy)}} \sqrt{(a_y^{(xy)})^*(\omega) (\Sigma^{(xy)})^{-1} a_y^{(xy)}(\omega)}} \\ &= C(\epsilon_x^{(xy)}, \eta_y^{(xy)}) = \frac{A_{xy}^{(xy)}(\omega) \sqrt{S_{yy|X}}}{\sqrt{\Sigma_{xx}^{(xy)}}}, \end{aligned} \quad (56)$$

where $\mathbf{A}^{(xy)}(\omega)$ is the spectral representation of the autoregressive coefficients matrix of (2), $a_y^{(xy)}(\omega)$ is the y column of $\mathbf{A}^{(xy)}(\omega)$, and $S_{yy|X}$ is the partial spectrum (Priestley, 1981) of the Y process when partialized on process X . Furthermore, $\eta_y^{(xy)}$ refers to the partialized process resulting from the Y process when partialized on X . Like in the case of the iDTF, a mutual information rate is associated with iPDC (Takahashi et al., 2010):

$$\mathcal{I}(\epsilon_x^{(xy)i}; \eta_y^{(xy)i}) = \frac{-1}{4\pi} \int_{-\pi}^{\pi} \ln(1 - |i\pi_{xy}^{(xy)}(\omega)|^2) d\omega. \quad (57)$$

5.2 The interpretation of Geweke measure and the directed transfer function for bivariate processes

The results of Takahashi et al. (2010), in particular (54), indicate that the Geweke spectral measure of Granger causality can be associated, at each frequency, with a

mutual information rate. This is because it is possible to express $\mathcal{I}(X^i; \epsilon_y^{(xy)i})$ (55) as an integral across frequencies, in the same way that we did for $\mathcal{H}(X^i)$ and $\mathcal{H}(X^i Y^i)$ in (35) and (41):

$$\mathcal{I}(X^i; \epsilon_y^{(xy)i}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{I}(X(\omega); \epsilon_y^{(xy)}(\omega)) d\omega, \quad (58)$$

so that

$$\begin{aligned} f_{Y \rightarrow X}(\omega) &= 2 \mathcal{I}(X(\omega); \epsilon_y^{(xy)}(\omega)) \\ &= 2(\mathcal{H}(X(\omega)) - \mathcal{H}(X(\omega) | \epsilon_y^{(xy)}(\omega))), \end{aligned} \quad (59)$$

where from (36) we know that $\mathcal{H}(X(\omega))$ is related to $S_{xx}(\omega)$. The conditional entropy $\mathcal{H}(X(\omega) | \epsilon_y^{(xy)}(\omega))$ can be seen to be related to $\Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2$, the term of $f_{Y \rightarrow X}(\omega)$ in (25) for which in Sect. 3 we were looking for its correspondence with an information theoretic measure. In particular we have that

$$S_{xx|\epsilon_y^{(xy)}} = S_{xx} - \frac{S_{x\epsilon_y^{(xy)}} S_{\epsilon_y^{(xy)}x}}{S_{\epsilon_y^{(xy)}\epsilon_y^{(xy)}}} = \Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2. \quad (60)$$

Recognizing $\Sigma_{xx}^{(xy)} |H_{xx}^{(xy)}(\omega)|^2$ as a partial spectrum is enough to relate it to a conditional entropy, since in general for the linear Gaussian stationary processes the conditional entropy can be expressed in terms of the partial spectrum (Brillinger, 1981):

$$\mathcal{H}(X(\omega) | \epsilon_y^{(xy)}(\omega)) = \frac{1}{2} \ln 2\pi e + \frac{1}{2} \ln S_{xx|\epsilon_y^{(xy)}}. \quad (61)$$

Therefore, in contrast to the spectral measure of total dependence $f_{X,Y}(\omega)$, which corresponds to the mutual information rate $\mathcal{I}(X(\omega); Y(\omega))$, (59) shows that, at each frequency, we are only able to express the GSC from Y to X in terms of the dependence of X with the innovation process $\epsilon_y^{(xy)}$, but not in terms of only variables from the observed processes X and Y . Nonetheless, with the integration across all the frequencies the relation in terms of X and Y is recovered. This means that to interpret the spectral measure at each frequency one has to consider explicitly the representation of the processes X and Y with an autoregressive model, since the process $\epsilon_y^{(xy)}$ is only meaningful under this representation.

5.3 The interpretation of the partial directed coherence

The PDC has been shown to be a measure to test for Granger causality (Schelter et al., 2006; Takahashi et al., 2010) because

$$F_{Y \rightarrow X} = 0 \Leftrightarrow i\pi_{xy}(\omega) = 0 \quad \forall \omega. \quad (62)$$

However, there is no direct relation between the iPDC and GSC, like it exists with iDTF (54). Here we will show that the mutual information rate associated with iPDF (57) is in fact testing another principle of conditional independence alternative to the general criterion of Granger causality (8). This principle is the generalized Sims condition of strict exogeneity of X (Sims, 1972) (also referred as Sims non-causality from Y to X). For the simplified case of noninstantaneous dependencies it corresponds to the following condition:

$$p(Y_{i+1} | Y^i X_1, \dots, X_\infty) = p(Y_{i+1} | X^i Y^i), \quad (63)$$

that is, if no causal interaction exists from Y to X , Y_{i+1} is conditionally independent of X_{i+1}, \dots, X_∞ given $X^i Y^i$. It has been proven (Chamberlain, 1982) that Granger non-causality and Sims non-causality are equivalent for bivariate systems (see also Kuersteiner, 2008, for a review of the relation between the two criteria of causality), that is, the equality in (63) is fulfilled if and only if the one in (8) is fulfilled.

To establish the relation between iPDC and Sims non-causality we consider that

$$\begin{aligned} (a_y^{(xy)})^*(\omega) (\Sigma^{(xy)})^{-1} a_y^{(xy)}(\omega) &= \frac{|A_{xy}^{(xy)}(\omega)|^2}{\Sigma_{xx}^{(xy)}} \\ &+ \frac{|A_{yy}^{(xy)}(\omega)|^2}{\Sigma_{yy}^{(xy)}} \end{aligned} \quad (64)$$

so that we can rewrite (57) as:

$$\mathcal{I}(\epsilon_x^{(xy)i}; \eta_y^{(xy)i}) = \frac{-1}{4\pi} \int_{-\pi}^{\pi} \ln \frac{|A_{yy}^{(xy)}(\omega)|^2 S_{yy|X}}{\Sigma_{yy}^{(xy)}} d\omega. \quad (65)$$

We now examine how the integral of the different components in the r.h.s. of (65) are related to information theoretic measures in terms of X and Y variables. For the partial spectrum $S_{yy|X}$ it can be shown (Brillinger, 1981), given (35) and (41) that

$$\begin{aligned} \mathcal{H}(Y^i | X^i) &= \mathcal{H}(X^i Y^i) - \mathcal{H}(X^i) = \frac{1}{2} \ln 2\pi e \\ &+ \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \frac{|\mathbf{S}^{(xy)}(\omega)|}{S_{xx}(\omega)} d\omega \end{aligned} \quad (66)$$

where

$$\frac{|\mathbf{S}^{(xy)}(\omega)|}{S_{xx}(\omega)} = S_{yy} - \frac{S_{xy} S_{yx}}{S_{xx}} = S_{yy|X}. \quad (67)$$

Furthermore, we indicated in Sect. 2.2 that $\Sigma_{yy}^{(xy)}$ is related to $H(Y_{i+1} | X^i Y^i)$ (18). Finally, the square matrix lag operator $A_{yy}^{(xy)}(L)$ has a leading term 1, and if all the roots are outside the unit circle, it follows from

(49) that its integral across frequencies cancels. Altogether we obtain that:

$$\mathcal{I}(\epsilon_x^{(xy)i}; \eta_y^{(xy)i}) = H(Y_{i+1}|X^i Y^i) - \mathcal{H}(Y^i|X^i). \quad (68)$$

This means that only after the integration across frequencies, the iPDC is associated with an information theoretic measure that depends only on X and Y variables, as it occurs for GSC. To better appreciate the parallelism with GSC and iDTF lets combine (9), (11), and (55) to get:

$$\mathcal{I}(X^i; \epsilon_y^{(xy)i}) = H(X_{i+1}|X^i) - \mathcal{H}(X_{i+1}|X^i Y^i). \quad (69)$$

In (69) it is evident that the difference of entropy rates corresponds to the comparison of the probability distribution equated in the general criterion of Granger causality (8). We will now show that in an analogous way the difference of entropy rates in (68) compares the probability distributions equated in the Sims non-causality criterion (63). Consider $\mathcal{H}(Y^i|X^i)$ as the limit

$$\begin{aligned} \mathcal{H}(Y^i|X^i) &= \lim_{n \rightarrow \infty} \frac{1}{n} H(Y_1, \dots, Y_n | X_1, \dots, X_n) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{\infty} H(Y_i | Y_{i-1}, \dots, Y_1, X_1, \dots, X_n). \end{aligned} \quad (70)$$

If the equality (63) of Sims non-causality from Y to X holds

$$H(Y_i | Y_{i-1}, \dots, Y_1, X_1, \dots, X_n) = H(Y_i | Y_{i-1}, \dots, Y_1, X_{i-1}, \dots, X_1). \quad (71)$$

Considering the definition of $H(Y_{i+1}|X^i Y^i)$ analogous to (47), for $n \rightarrow \infty$ we get $\mathcal{H}(Y^i|X^i) = H(Y_{i+1}|X^i Y^i)$.

Apart from the theoretical interest of the link of iPDC with the criterion of Sims non-causality, the argument above is important in practice for the interpretation of the spectral measures because it indicates, given (68) and (69), that independently of which criterion of causality is adopted, the appearance of terms like $H(Y_{i+1}|X^i Y^i)$ and $H(X_{i+1}|X^i Y^i)$ prevents from deriving a spectral representation in terms of X and Y variables in the same way that it is done for the mutual information rate $\mathcal{I}(X^i; Y^i)$. Like for the relation of $f_{Y \rightarrow X}(\omega)$ and $i\gamma_{xy}^{(xy)}(\omega)$ (54) one can consider a measure

$$g_{Y \rightarrow X}(\omega) = -\ln(1 - |i\pi_{xy}(\omega)|^2), \quad (72)$$

so that

$$\begin{aligned} g_{Y \rightarrow X}(\omega) &= 2 \mathcal{I}(\epsilon_x^{(xy)}(\omega); \eta_y^{(xy)}(\omega)) \\ &= 2(\mathcal{H}(\eta_y^{(xy)}(\omega)) - \mathcal{H}(\eta_y^{(xy)}(\omega)|\epsilon_x^{(xy)}(\omega))). \end{aligned} \quad (73)$$

Given that $\eta_y^{(xy)}$ is the process obtained after partializing Y on X , $\mathcal{H}(\eta_y^{(xy)}(\omega))$ is related to $S_{yy|X}$ (Brillinger, 1981). It can then be seen that the rest of the terms apart from $S_{yy|X}$ inside the logarithm in the r.h.s of (65) correspond to the partial spectrum $S_{\eta_y^{(xy)}|\epsilon_x^{(xy)}}$, and thus are related to $\mathcal{H}(\eta_y^{(xy)}(\omega)|\epsilon_x^{(xy)}(\omega))$. Therefore, like for the bivariate GSC in Sect. 5.2, there is a one-to-one connection between components of the PDC measure and entropies involving innovation variables. However, at each frequency it is not possible to express the measures of causal interactions only using variables from the observed processes. Therefore the results presented in Sects. 5.2 and 5.3 complement the ones in Sect. 3 showing that the information theoretic measures related to the spectral measures of Granger causality involve innovation variables and not only variables from the observed processes X and Y .

6 The interpretation of the spectral measures related to Granger causality in multivariate processes

6.1 Definition of the spectral measures in the multivariate case

We now extend the analysis to the case of multivariate systems. We start reviewing the definition of the spectral measures. For that purpose, we consider the fully multivariate autoregressive representation of the system $W = \{X, Y, Z\}$:

$$\begin{aligned} X_{i+1} &= \sum_{s=0}^{\infty} a_{xxs}^{(xyz)} X_{i-s} + a_{xys}^{(xyz)} Y_{i-s} + a_{zxs}^{(xyz)} Z_{i-s} \\ &\quad + \epsilon_{xi+1}^{(xyz)} \\ Y_{i+1} &= \sum_{s=0}^{\infty} a_{yxs}^{(xyz)} X_{i-s} + a_{yys}^{(xyz)} Y_{i-s} + a_{yzs}^{(xyz)} Z_{i-s} \\ &\quad + \epsilon_{yi+1}^{(xyz)} \\ Z_{i+1} &= \sum_{s=0}^{\infty} a_{zxs}^{(xyz)} X_{i-s} + a_{zys}^{(xyz)} Y_{i-s} + a_{zzs}^{(xyz)} Z_{i-s} \\ &\quad + \epsilon_{zi+1}^{(xyz)} \end{aligned} \quad (74)$$

$$\Sigma^{(xyz)} = \begin{pmatrix} \Sigma_{xx}^{(xyz)} & \Sigma_{xy}^{(xyz)} & \Sigma_{xz}^{(xyz)} \\ \Sigma_{yx}^{(xyz)} & \Sigma_{yy}^{(xyz)} & \Sigma_{yz}^{(xyz)} \\ \Sigma_{zx}^{(xyz)} & \Sigma_{zy}^{(xyz)} & \Sigma_{zz}^{(xyz)} \end{pmatrix}. \quad (75)$$

Like for the bivariate case we will from now on assume that $\Sigma^{(xyz)}$ is diagonal. This joint autoregressive representation suffices to define the iDTF and the iPDC, which are straightforward extensions of the bivariate

case. In that case the iDTF from Y to X corresponds to (Takahashi et al., 2010):

$$i\gamma_{xy}^{(xyz)}(\omega) = \frac{H_{xy}^{(xyz)}(\omega)\sqrt{\Sigma_{yy}^{(xyz)}}}{\sqrt{S_{xx}(\omega)}} = C(X, \epsilon_y^{(xyz)}), \quad (76)$$

where $C(X, \epsilon_y^{(xyz)})$ is the coherence between X and $\epsilon_y^{(xyz)}$, and $H_{xy}^{(xyz)}$ is a component of the transfer function $\mathbf{H}^{(xyz)}$ associated with the multivariate autoregressive process in (74) analogous to $\mathbf{H}^{(xy)}$ in the bivariate case (22).

The information partial directed coherence (iPDC) from Y to X (Takahashi et al., 2010) is:

$$\begin{aligned} i\pi_{xy}^{(xyz)}(\omega) &= C(\epsilon_x^{(xyz)}, \eta_y^{(xyz)}) = \frac{A_{xy}^{(xyz)}(\omega)\sqrt{S_{yy|W\setminus y}}}{\sqrt{\Sigma_{xx}^{(xyz)}}} \\ &= \frac{A_{xy}^{(xyz)}(\omega)}{\sqrt{\Sigma_{xx}^{(xyz)}}\sqrt{(a_y^{(xyz)})^*(\omega)(\Sigma^{(xyz)})^{-1}a_y^{(xyz)}(\omega)}}, \end{aligned} \quad (77)$$

where $\mathbf{A}^{(xyz)}(\omega)$ is the spectral representation of the autoregressive coefficients matrix of (74), $a_y^{(xyz)}(\omega)$ is the y column of $\mathbf{A}^{(xyz)}(\omega)$, and $S_{yy|W\setminus y}$ is the partial spectrum (Priestley, 1981) of the Y process when partialized on all the other processes in the multivariate process W . Furthermore, $\eta_y^{(xyz)}$ refers to the partialized process resulting from the Y process when partialized on all the others.

In contrast to iDTF and iPDC, the definition of the conditional GSC (Geweke, 1984) is not a straightforward extension of the bivariate case. Apart from the joint autoregressive representation of W in (74) to calculate the conditional GSC from Y to X it is also needed the projection of X_{i+1} only on the past of X and Z :

$$X_{i+1} = \sum_{s=0}^{\infty} a_{xxs}^{(xz)} X_{i-s} + a_{xzs}^{(xz)} Z_{i-s} + \epsilon_{xi+1}^{(xz)} \quad (78)$$

$$\begin{aligned} Z_{i+1} &= \sum_{s=0}^{\infty} a_{zxs}^{(xz)} X_{i-s} + a_{zzs}^{(xz)} Z_{i-s} + \epsilon_{zi+1}^{(xz)} \\ \Sigma^{(xz)} &= \begin{pmatrix} \Sigma_{xx}^{(xz)} & \Sigma_{xz}^{(xz)} \\ \Sigma_{zx}^{(xz)} & \Sigma_{zz}^{(xz)} \end{pmatrix}. \end{aligned} \quad (79)$$

The conditional GSC (Geweke, 1984) is defined in the time domain analogously to $F_{Y \rightarrow X}$ (4):

$$F_{Y \rightarrow X|Z} = \ln\left(\frac{\Sigma_{xx}^{(xz)}}{\Sigma_{xx}^{(xyz)}}\right). \quad (80)$$

It is straightforward to see that given the form of the entropy for Gaussian variables (10) and the definition of

the conditional transfer entropy $T_{Y \rightarrow X|Z}$ as discussed in Sect. 4

$$F_{Y \rightarrow X|Z} = 2 T_{Y \rightarrow X|Z}. \quad (81)$$

To derive the spectral representation of $F_{Y \rightarrow X|Z}$ for simplicity we assume that there is no instantaneous causality and $\Sigma^{(xyz)}$ and $\Sigma^{(xz)}$ are diagonal (see for example Ding et al., 2006, for a detailed derivation when instantaneous correlations exist). We rewrite (78) after Fourier transforming as:

$$\begin{pmatrix} \epsilon_x^{(xz)}(\omega) \\ \epsilon_z^{(xz)}(\omega) \end{pmatrix} = \begin{pmatrix} A_{xx}^{(xz)}(\omega) & A_{xz}^{(xz)}(\omega) \\ A_{zx}^{(xz)}(\omega) & A_{zz}^{(xz)}(\omega) \end{pmatrix} \begin{pmatrix} X(\omega) \\ Z(\omega) \end{pmatrix}. \quad (82)$$

Furthermore we rewrite (74) using the transfer function $\mathbf{H}^{(xyz)}$:

$$\begin{pmatrix} X(\omega) \\ Y(\omega) \\ Z(\omega) \end{pmatrix} = \mathbf{H}^{(xyz)} \begin{pmatrix} \epsilon_x^{(xyz)}(\omega) \\ \epsilon_y^{(xyz)}(\omega) \\ \epsilon_z^{(xyz)}(\omega) \end{pmatrix}. \quad (83)$$

Geweke (1984) showed that

$$F_{Y \rightarrow X|Z} = F_{Y_{\epsilon_z^{(xz)} \rightarrow \epsilon_x^{(xz)}}}. \quad (84)$$

Accordingly, (82) and (83) are combined to express Y , $\epsilon_z^{(xz)}$ and $\epsilon_x^{(xz)}$ in terms of the innovations of the fully multivariate process:

$$\begin{pmatrix} \epsilon_x^{(xz)}(\omega) \\ Y(\omega) \\ \epsilon_z^{(xz)}(\omega) \end{pmatrix} = \mathbf{G}\mathbf{H}^{(xyz)} \begin{pmatrix} \epsilon_x^{(xyz)}(\omega) \\ \epsilon_y^{(xyz)}(\omega) \\ \epsilon_z^{(xyz)}(\omega) \end{pmatrix}, \quad (85)$$

where

$$\mathbf{G} = \begin{pmatrix} A_{xx}^{(xz)}(\omega) & 0 & A_{xz}^{(xz)}(\omega) \\ 0 & 1 & 0 \\ A_{zx}^{(xz)}(\omega) & 0 & A_{zz}^{(xz)}(\omega) \end{pmatrix}. \quad (86)$$

Considering $\mathbf{Q} = \mathbf{G}\mathbf{H}^{(xyz)}$, the spectrum matrix of Y , $\epsilon_z^{(xz)}$ and $\epsilon_x^{(xz)}$ is:

$$\hat{\mathbf{S}}(\omega) = \mathbf{Q}(\omega)\Sigma^{(xyz)}\mathbf{Q}^*(\omega), \quad (87)$$

and in particular

$$\begin{aligned} S_{\epsilon_x^{(xz)}\epsilon_x^{(xz)}}(\omega) &= |Q_{xx}(\omega)|^2 \Sigma_{xx}^{(xyz)} + |Q_{xy}(\omega)|^2 \Sigma_{yy}^{(xyz)} \\ &\quad + |Q_{xz}(\omega)|^2 \Sigma_{zz}^{(xyz)}. \end{aligned} \quad (88)$$

The conditional GSC from Y to X given Z is defined (Geweke, 1984) as the portion of the power spectrum associated with $\epsilon_x^{(xyz)}$, in analogy to (25):

$$f_{Y \rightarrow X|Z}(\omega) = f_{Y_{\epsilon_z^{(xz)} \rightarrow \epsilon_x^{(xz)}}}(\omega) = \ln \frac{S_{\epsilon_x^{(xz)}\epsilon_x^{(xz)}}(\omega)}{|Q_{xx}(\omega)|^2 \Sigma_{xx}^{(xyz)}}.$$

(89)

This measure also fulfills the requirements that Geweke (1982) imposed to the spectral measures. Apart from the interpretation stated above it is nonnegative and again the link with the time domain measure analogous to (19) is fulfilled given the cancelation of the integral across frequencies of $|Q_{xx}(\omega)|^2$, following (49).

6.2 The interpretation of the conditional Geweke spectral measure of Granger causality

In the same way that Takahashi et al. (2010) indicated the link of the bivariate GSC to a mutual information rate, we here want to examine if it is possible to establish this link for the conditional GSC. In fact it is easy to see that given its definition (89) the conditional GSC corresponds to a multiple coherence (Priestley, 1981):

$$\begin{aligned} f_{Y \rightarrow X|Z}(\omega) &= -\ln(1 - |C(\epsilon_x^{(xz)}, \epsilon_y^{(xyz)} \epsilon_z^{(xyz)})|^2) \\ &= 2 \mathcal{I}(\epsilon_x^{(xz)}(\omega); \epsilon_y^{(xyz)}(\omega) \epsilon_z^{(xyz)}(\omega)), \end{aligned} \quad (90)$$

where $|C(\epsilon_x^{(xz)}, \epsilon_y^{(xyz)} \epsilon_z^{(xyz)})|^2$ is the squared multiple coherence. This equality results from the direct application of the definition of the squared multiple coherence. For the variables here involved, defining $\mathbf{V} \equiv (\epsilon_y^{(xyz)}, \epsilon_z^{(xyz)})$, the squared multiple coherence is defined as:

$$|C(\epsilon_x^{(xz)}, \mathbf{V})|^2 = \frac{\mathbf{S}_{\epsilon_x^{(xz)} \mathbf{V}} \mathbf{S}_{\mathbf{V} \mathbf{V}}^{-1} \mathbf{S}_{\epsilon_x^{(xz)} \mathbf{V}}^*}{\mathbf{S}_{\epsilon_x^{(xz)} \epsilon_x^{(xz)}}(\omega)}, \quad (91)$$

where $\mathbf{S}_{\mathbf{V} \mathbf{V}}$ is the spectral matrix:

$$\mathbf{S}_{\mathbf{V} \mathbf{V}} = \begin{pmatrix} \Sigma_{yy}^{(xyz)} & 0 \\ 0 & \Sigma_{zz}^{(xyz)} \end{pmatrix}, \quad (92)$$

and

$$\mathbf{S}_{\epsilon_x^{(xz)} \mathbf{V}} = \begin{pmatrix} Q_{xy} \Sigma_{yy}^{(xyz)}, & Q_{xz} \Sigma_{zz}^{(xyz)} \end{pmatrix}. \quad (93)$$

These cross-spectra are obtained considering (85) and the general definition of cross-spectrum, which for two spectral variables $X_1(\omega)$, $X_2(\omega)$ is (Priestley, 1981):

$$S_{12} d\omega = \mathbb{E}[X_1(\omega) X_2^*(\omega)]. \quad (94)$$

Given the definition of $f_{Y \rightarrow X|Z}(\omega)$ in terms of the squared multiple coherence it is clear that, analogously to $F_{Y \rightarrow X}$ (55):

$$F_{Y \rightarrow X|Z} = 2 \mathcal{I}(\epsilon_x^{(xz)i}; \epsilon_y^{(xyz)i} \epsilon_z^{(xyz)i}). \quad (95)$$

The identification of the conditional GSC as directly related to a particular application of a standard well-known spectral measure, namely the multiple coherence

is useful from a theoretical point of view because it allows us to relate the conditional GSC to a mutual information rate. Furthermore, in practice, this connection is helpful because the properties of the multiple coherence as well as its estimation have been already widely studied (Priestley, 1981). How this knowledge could be used to improve the estimation of the bias and confidence intervals for conditional GSC is left for a future contribution. Also from the perspective of the practical interpretation of the measure, we see that, like it occurs for $f_{Y \rightarrow X}(\omega)$ in the bivariate case (54), at each frequency we can express the spectral measure only in terms of the innovations. It is after integrating across frequencies that the representation in terms of innovations and in terms of the processes X , Y , and Z are equivalent. This stresses again the role in the spectral causal measures of the innovation variables intrinsically related to the autoregressive representation, which cannot be interpreted in terms of a causal interaction between the actual observed processes.

6.3 The interpretation of the directed transfer function

The most relevant difference for the iDTF in the multivariate case concerns the lack of a relation to GSC like in the bivariate case (54). This relation relied on the specific form of S_{xx} (24) for bivariate processes. For the multivariate case the DTF is associated with the total causal effect of one process on another (Eichler, 2006) and not to the direct causal effect like conditional GSC. Apart from this difference, analogously to the bivariate case we have that

$$\begin{aligned} \mathcal{I}(X(\omega); \epsilon_y^{(xyz)}(\omega)) &= \mathcal{H}(X(\omega)) - \mathcal{H}(X(\omega) | \epsilon_y^{(xyz)}(\omega)) \\ &= -\frac{1}{2} \ln(1 - |i\gamma_{xy}^{(xyz)}(\omega)|^2), \end{aligned} \quad (96)$$

where given (76)

$$\begin{aligned} -\ln(1 - |i\gamma_{xy}^{(xyz)}(\omega)|^2) &= \ln S_{xx}(\omega) \\ -\ln(|H_{xx}^{(xyz)}(\omega)|^2 \Sigma_{xx}^{(xyz)} + |H_{xz}^{(xyz)}(\omega)|^2 \Sigma_{zz}^{(xyz)}). \end{aligned} \quad (97)$$

The connection at each frequency between the two terms in the r.h.s of (97) with the entropies in (96) is the same as in the bivariate case. However, when integrating across frequencies, while the term with $S_{xx}(\omega)$ is related to $H(X_{i+1} | X^i)$ (38), the integration of second term on the r.h.s. cannot be expressed in terms of X , Y , Z in general. This is because this term does not correspond to the spectral matrix of any subgroup of W and furthermore theorem 4.2 of Rozanov (1967) (49) does not apply in this case. Accordingly, in the multivariate case, not even the integration across frequencies

of the iDTF can be expressed in terms of the observed variables X, Y, Z . The equality (Takahashi et al., 2010)

$$\mathcal{I}(X^i; \epsilon_y^{(xyz)i}) = \frac{-1}{4\pi} \int_{-\pi}^{\pi} \ln(1 - |i\gamma_{xy}^{(xyz)}(\omega)|^2) d\omega \quad (98)$$

analogous to the one of the bivariate case (55), provides only an expression which involves the innovation process $\epsilon_y^{(xyz)}$.

6.4 The interpretation of the partial directed coherence

The most relevant difference for iPDC in the multivariate case concerns the lack of a relation with the principle of Sims non-causality. It has been shown that the criterion of Sims non-causality is not equivalent to the criterion of Granger non-causality in the multivariate case (Florens, 2003). Since the equivalence between a nonzero partial directed coherence and the existence of Granger causality (62) is not restricted to the bivariate case (Schelter et al., 2006), it would be contradictory that the link between iPDC and Sims non-causality holds for the multivariate case.

In particular, analogously to the bivariate case

$$\begin{aligned} g_{Y \rightarrow X}(\omega) &= 2 \mathcal{I}(\epsilon_x^{(xyz)}(\omega); \eta_y^{(xyz)}(\omega)) \\ &= 2(\mathcal{H}(\eta_y^{(xyz)}(\omega)) - \mathcal{H}(\eta_y^{(xyz)}(\omega) | \epsilon_x^{(xyz)}(\omega))) \\ &= -\ln(1 - |i\pi_{xy}^{(xyz)}(\omega)|^2), \end{aligned} \quad (99)$$

and given (77)

$$\begin{aligned} g_{Y \rightarrow X}(\omega) &= -\ln\left(\frac{|A_{yy}^{(xyz)}(\omega)|^2}{\Sigma_{yy}^{(xyz)}} + \frac{|A_{zy}^{(xyz)}(\omega)|^2}{\Sigma_{zz}^{(xyz)}}\right) \\ &\quad - \ln S_{yy|XZ}. \end{aligned} \quad (100)$$

The connection between the two terms in the r.h.s of (100) with the entropies in (99) is the same as in the bivariate case. Regarding the integration across frequencies, the integration of the term $\ln S_{yy|XZ}$ is related to $\mathcal{H}(Y^i | X^i Z^i)$, in analogy to (66). Oppositely, the integration of the other term on the r.h.s. cannot be in general expressed in terms of the variables X, Y , and Z , since theorem 4.2 of Rozanov (1967) is not applicable for the same reason like for the term in the multivariate iDTF (97). This implies that we cannot follow the same procedure that we used in Sect. 5.3 to link iPDC with Sims non-causality. Accordingly, like it occurs for the iDTF, in the multivariate case not even after the integration across frequencies the iPDC can be expressed

in terms of the variables of the observed processes X, Y, Z . The equality (Takahashi et al., 2010)

$$\mathcal{I}(\epsilon_x^{(xyz)i}; \eta_y^{(xyz)i}) = \frac{-1}{4\pi} \int_{-\pi}^{\pi} \ln(1 - |i\pi_{xy}^{(xyz)}(\omega)|^2) d\omega \quad (101)$$

analogous to the one of the bivariate case (57), provides only an expression which involves the innovation processes $\epsilon_x^{(xyz)}$ and $\eta_y^{(xyz)}$.

7 Discussion

We studied the relation that exists between the formulation of Granger causality in the spectral domain and the formulation in the more general framework of information theory. For the bivariate case we derived the decomposition in the frequency domain (Geweke, 1982) of the total dependence into causal and instantaneous dependence terms from the decomposition in the information theory framework (Sect. 3). In a nonparametric way we proved the connection between the measures of total dependence in the two frameworks showing that they can be expressed in terms of variables from the observed processes X and Y . These measures are related to the coherence between X and Y and thus quantify an important physical property that can be functionally relevant (Fries, 2005).

In contrast to the total dependence, we showed that the same type of derivation does not lead to a spectral representation of the transfer entropy in terms of the observed processes (Sect. 3). Although our arguments do not constitute a complete proof of nonexistence we indicated the impediments that such a spectral representation of Granger causality needs to deal with. We extended these results to the multivariate case in Sect. 4.

To further support the arguments in Sect. 3 and to complete the picture of the connection between the spectral and the information theory framework, we examined for the bivariate case (Sect. 5) and for the multivariate case (Sect. 6) how the spectral measures related to Granger causality are connected to information theoretic measures when explicitly considering the parametric autoregressive representation. In this way we connected bivariate PDC to the generalized criterion of Sims non-causality and we expressed the conditional GSC in terms of a multiple coherence.

More generally we showed that the spectral measures of Granger causality examined are associated with mutual information rates that involve innovation variables inherent to the autoregressive representation, and only after integrating across frequencies they can, some

of them, be expressed only in terms of the observed processes. Although not considered here, our arguments for the lack of a spectral representation of Granger causality in terms of the observed processes also hold for other measures that have been proposed to study causal interactions in the spectral domain. They comprise a variation to improve the robustness of the estimation of the Geweke measure for the multivariate case (Chen et al., 2006), measures for the study of causal interactions between groups of processes (Guo et al., 2008a; Ladrone et al., 2009), and to deal with the effect of latent variables (Guo et al., 2008b).

The lack of a spectral representation of Granger causality in terms only of variables from the observed processes indicates that the spectral measures related to Granger causality should be interpreted considering explicitly the autoregressive model. The innovation variables are intrinsic to this model and do not have a physical meaning. This prevents from interpreting the value at each frequency as quantifying the degree to which the causal interactions occur in a given frequency band. This does not mean that the spectral measures are not informative about the dynamics resulting from the causal interactions, but one should be cautious when using them to associate the causal interactions with specific functionally relevant rhythms. Usually the role played by the innovation processes in the definition of the measures remains implicit and the peaks in the Granger causality spectrum are directly interpreted in terms of coherent oscillatory activity that serves to carry causal influences between the observed processes associated with different brain regions (e. g. Bernasconi and König, 1999; Winterhalder et al., 2005; Bressler et al., 2007).

Considered in terms of the autoregressive representation the measures have a clear interpretation: for example, the GSC and DTF quantify the relative contribution of a particular innovation process to the power spectrum. Oppositely, for the bivariate case the interpretation in terms of Granger causality only holds as a condition for the existence of a direct causal interaction, since a zero value of the Granger causality time domain measure implies the cancelation of the spectral measures at all frequencies. For the multivariate case this constraint holds for the conditional GSC and the PDC, while the DTF is related to the existence of a total causal effect (Eichler, 2006).

Accepting that the spectral measures should be interpreted considering the autoregressive model in an explicit way highlights the necessity to validate this model. This implies that checking the fulfillment of the linear Gaussian stationary condition is a requirement for the correct application of the measures. The vali-

dation of the model with the corresponding tests for stationarity and Gaussianity was discussed in detail in one of the first applications Bernasconi and König (1999), but has been discussed with less detail in other posterior applications to neural data. This validation is still relevant even if nonparametric calculations are used (Dhamala et al., 2008; Nedungadi et al., 2009) for the measures. Only if the autoregressive model is a good model the spectral measures of Granger causality, which involve innovation processes, will be meaningful to characterize the dynamics of the system.

Furthermore, considering the explicit autoregressive representation helps us to appreciate the sensitivity and specificity of the measures to different changes in the properties of the processes. For example, the GSC from Y to X may change not due to a change in the coupling between the two processes, but simply because the autocorrelation of X changes. Being aware of which properties affect the measures is important, for example, to compare the value obtained for a particular pair of regions in two different experimental conditions. This means that for a better understanding of the processes and the interactions, it is better to examine not only directly the spectral measures of Granger causality but also the individual components of the transfer functions and the power spectra, to identify which properties change. For example, the bivariate Geweke spectral measure of causality from Y to X corresponds to the relative contribution to the power spectrum of X of the innovation process associated with X in the bivariate autoregressive process. Therefore, displaying together with the Granger causality measure the power spectrum and the transfer function from the innovation process associated with X to the actual process X can help to understand how the predominant rhythms arise and contribute to the spectral measure.

Altogether our results provide a complete framework that integrates the information theory formulation of dependence and causality, and the time and spectral domain formulation for linear Gaussian stationary processes. The relationship between the spectral measures and mutual information rates involving innovation processes suggests that the parametric autoregressive representation should be considered to interpret the measures. Here we have not addressed how useful these measure can be in practice to study causal interactions in the brain. Our results are independent of the nature of the processes and not specific to study neural data. However, given that these techniques have been applied quite often to study the causal interactions between brain regions, and the especial relevance of the spectral analysis for the neural data, we believe that our results are relevant for future analysis of brain con-

nectivity. In particular, they provide a theoretical background to appreciate the necessity of validating the autoregressive representation, to consider the specificity of the measures to different parameters in the model which are not only related to the coupling between the processes, and to be cautious for interpreting the spectral measures as indicative of causal interactions related to specific brain rhythms.

Finally, we should note that in this study it was generally assumed the lack of instantaneous causality or, equivalently for the linear Gaussian stationary processes, that the correlation matrices were diagonal. We restricted ourselves to this case to avoid the so called normalization of the correlation matrix (Ding et al., 2006), which makes more complicated and less intuitive the derivations without adding new qualitative insights.

8 Acknowledgments

DC was supported by the grant 2010FI-B2 00079 of the "Comissionat per a Universitats i Recerca del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya i del Fons Social Europeu" and is supported by the BrainScaleS EU FET-Proactive FP7 funded research project (Project number 269921). DC thanks Anders Ledberg and Ralph G. Andrzejak for fruitful discussions, and the reviewers for comments that improved the quality of this contribution.

References

- Amblard PO, Michel O (2011) On directed information theory and granger causality graphs. *J Comput Neurosci* 30(1):7–16.
- Baccalá L, Sameshima K, Ballester G, Do Valle A, Timo-Iaria C (1999) Studying the interaction between brain structures via directed coherence and granger causality. *Appl Sig Process* 5:40–48.
- Baccalá L, Sameshima K (2001) Partial directed coherence: a new concept in neural structure determination. *Biol Cybern* 84(1):463–474.
- Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for gaussian variables. *Phys Rev Lett* 103(23):238701.
- Bernasconi C, König P (1999) On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. *Biol Cybern* 81(3):199–210.
- Bernasconi C, von Stein A, Chiang C, König P (2000) Bi-directional interactions between visual areas in the awake behaving cat. *Neuroreport* 11(4):689–692.
- Besserve M, Schoelkopf B, Logothetis NK, Panzeri S (2010) Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. *J Comput Neurosci* 29(3):547–566.
- Bressler SL, Richter CG, Chen Y, Ding M (2007) Cortical functional network organization from autoregressive modeling of local field potential oscillations. *Stat Med* 26(21):3875–3885.
- Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M (2008) Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J Neurosci* 28(40):10056–10061.
- Bressler SL, Seth AK (2011) Wiener granger causality: A well established methodology. *Neuroimage* 58(2):323–329.
- Brillinger D (1981) Time series. Data analysis and theory. Holden-Day, San Francisco, USA.
- Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality. *P Natl Acad Sci USA* 101:9849–9854.
- Buzsáki G (2006) Rhythms of the brain. Oxford University Press, New York, USA.
- Chamberlain G (1982) The general equivalence of granger and Sims causality. *Econometrica* 50(3):569–581.
- Chen Y, Bressler S, Ding M (2006) Frequency decomposition of conditional granger causality and application to multivariate neural field potential data. *J Neurosci Meth* 150(2):228–237.
- Cover TM, Thomas JA (2006) Elements of Information Theory. John Wiley and Sons, 2nd edition.
- Dhamala M, Rangarajan G, Ding M (2008) Estimating granger causality from fourier and wavelet transforms of time series data. *Phys Rev Lett* 100(1):018701.
- Ding M, Chen Y, Bressler SL (2006) Granger causality: Basic theory and application to neuroscience. In *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. pages 437–460. Wiley-VCH Verlag.
- Eichler M (2006) On the evaluation of information flow in multivariate systems by the directed transfer function. *Biol Cybern* 94(6):469–482.
- Florens J (2003) Some technical issues in defining causality. *J Econometrics* 112:127–128.
- Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9(10):474–480.
- Friston KJ (1994) Functional and effective connectivity in neuroimaging: A synthesis. *Hum Brain Mapp*

- 2(5):56–78.
- Gelfand I, Yaglom A (1959) Calculation of the amount of information about a random function contained in another such function. *Am Math Soc Transl Ser* 2(12):199–246.
- Geweke JF (1982) Measurement of linear dependence and feedback between multiple time series. *J Am Stat Assoc* 77(378):304–313.
- Geweke JF (1984) Measures of conditional linear dependence and feedback between time series. *J Am Stat Assoc* 79(388):907–915.
- Gourevitch B, Le Bouquin-Jeannes R, Faucon G (2006) Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biol Cybern* 95(4):349–369.
- Gourieroux C, Monfort A, Renault E (1987) Kullback causality measures. *Annales d'Economie et de Statistique* (6-7):470–410.
- Granger CWJ (1963) Economic processes involving feedback. *Information and Control* 6:28–48.
- Granger CWJ (1980) Testing for causality : A personal viewpoint. *J Econ Dynamics and Control* 2(1):329–352.
- Guo S, Seth AK, Kendrick KM, Zhou C, Feng J (2008a) Partial granger causality - eliminating exogenous inputs and latent variables. *J Neurosci Meth* 172(1):79–93.
- Guo S, Wu J, Ding M, Feng J (2008b) Uncovering interactions in the frequency domain. *PLoS Comput Biol* 4(5).
- Kaminski M, Blinowska K (1991) A new method of the description of the information flow in the brain structures. *Biol Cybern* 65(3):203–210.
- Kaminski M, Ding M, Truccolo W, Bressler S (2001) Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol Cybern* 85(2):145–157.
- Kolmogorov A (1939) Sur l'interpolation et extrapolation des suites stationnaires. *Comp Rend Acad Sci Paris* 208:2043–2045.
- Kuersteiner G (2008) Granger-Sims causality. *The New Palgrave Dictionary of Economics*, 2nd edition.
- Ladroue C, Guo S, Kendrick K, Feng J (2009) Beyond element-wise interactions: Identifying complex interactions in biological processes. *PLoS one* 4(9):e6899.
- Marko H (1973) Bidirectional communication theory - generalization of information-theory. *IEEE T Commun* 12:1345–1351.
- Nedungadi AG, Rangarajan G, Jain N, Ding M (2009) Analyzing multiple spike trains with nonparametric granger causality. *J Comput Neurosci* 27(1):55–64.
- Pereda E, Quiñan Quiroga R, Bhattacharya J (2005) Nonlinear multivariate analysis of neurophysiological signals. *Prog Neurobiol* 77:1–37.
- Priestley M (1981) Spectral analysis and time series. Academic Press Inc., San Diego.
- Rissanen J, Wax M (1987) Measures of mutual information and causal dependence between 2 time-series. *IEEE Transactions on information theory* 33(4):598–601.
- Rozanov YA (1967) Stationary random processes. Holden-Day, San Francisco, USA.
- Schelter B, Winterhalder M, Eichler M, Peifer M, Hellwig B, Guschlbauer B, Lücking C, Dahlhaus R, Timmer J (2006) Testing for directed influences among neural signals using partial directed coherence. *J Neurosci Meth* 152(1-2):210–219.
- Schreiber T (2000) Measuring information transfer. *Phys Rev Lett* 85:461–464.
- Sims C (1972) Money, income, and causality. *American Economic Rev* 62(4):540–552.
- Solo V (2008) On causality and mutual information. In *Proceedings of the 47th IEEE Conference on Decision and Control* pages 4639–4944.
- Takahashi DY, Baccala LA, Sameshima K (2010) Information theoretic interpretation of frequency domain connectivity measures. *Biol Cybern* 103(6):463–469.
- Wiener N (1956) The theory of prediction. In *Modern Mathematics for Engineers*. McGraw-Hill, New York.
- Winterhalder M, Schelter B, Hesse W, Schwab K, Leistriz L, Klan D, Bauer R, Timmer J, Witte H (2005) Comparison directed of linear signal processing techniques to infer interactions in multivariate neural systems. *Sig Process* 85(11):2137–2160.