DETECTING COUPLING DIRECTIONS WITH TRANSCRIPT MUTUAL INFORMATION: A COMPARATIVE STUDY

José M. Amigó
Centro de Investigación Operativa, Universidad Miguel Hernández, 03202 Elche, Spain

Beata Graff
Department of Hypertension and Diabetology, Medical University of Gdańsk, 80-952 Gdańsk, Poland
Orcid: 0000-0002-9181-9884

Grzegorz Graff
Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, 80-233 Gdańsk, Poland
Orcid: 0000-0001-5670-5729

Roberto Monetti
IngSoft GmbH. Irrerstrasse 17. 90403 Nuernberg, Germany

Katarzyna Tessmer
Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, 80-233 Gdańsk, Poland
Orcid: 0000-0003-4487-8037

Dedicated to Peter E. Kloeden on the occasion of his 70th birthday
(Communicated by Tomas Caraballo)

Abstract. Causal relationships are important to understand the dynamics of coupled processes and, moreover, to influence or control the effects by acting on the causes. Among the different approaches to determine cause-effect relationships and, in particular, coupling directions in interacting random or deterministic processes, we focus in this paper on information-theoretic measures. So, we study in the theoretical part the difference between directionality indicators based on transfer entropy as well as on its dimensional reduction via transcripts in algebraic time series representations. In the applications we consider specifically the lowest dimensional case, i.e., 3-dimensional transfer entropy, which is currently one of the most popular causality indicators.

2010 Mathematics Subject Classification. Primary: 37M10, 94A17; Secondary: 60G10.

Key words and phrases. Information directionality, transfer entropy, time series analysis, algebraic symbolic representations, transcripts, dimensional reduction.

J.M.A. was financially supported by the Spanish Ministry of Economy, Industry and Competitiveness, grant MTM2016-74921-P (AEI/FEDER, EU). B.G. was financially supported by the National Science Centre, Poland, grant MAESTRO UMO-2011/02/A/NZ5/00329. G.G. was financially supported by the National Science Centre, Poland, grant UMO-2014/15/B/ST1/01710.

* Corresponding author: jm.amigo@umh.es.
and the (2-dimensional) mutual information of transcripts. Needless to say, the lower dimensionality of the transcript-based indicator can make a difference in practice, where datasets are usually small. To compare numerically the performance of both directionality indicators, synthetic data (obtained with random processes) and real world data (in the form of biomedical recordings) are used. As happened in previous related work, we found again that the transcript mutual information performs as good as, and in some cases even better than, the lowest dimensional binned and symbolic transfer entropy, the symbols being ordinal patterns.

1. Introduction. Causality is a philosophical term that in natural science amounts usually to predictability improvement, an approach proposed by Norbert Wiener [24] and, therefore, called Wiener causality. Given two systems $A$ and $B$, $A$ is said to (Wiener-) cause $B$ if the knowledge of the past history of $A$ improves the prediction of the present value of $B$ (i.e., reduces the uncertainty about its present value) as compared to self-predictions. Wiener's proposal was materialized in the late 1960s by Granger ("Granger causality") in the realm of econometrics via autoregressive models [14]. In our case, the systems are deterministic or random processes whose evolution laws influence one another in only one direction (unidirectional coupling) or in both directional (bidirectional or mutual coupling). Therefore, we speak also of coupling directions or directionality when being specific about the systems meant.

The general subject of this paper is the detection of coupling directions in time series analysis. To this end we are going to follow an information-theoretic approach based on transfer entropy [21], which is a standard tool for quantifying coupling directionality in terms of “information flow”. By definition, transfer entropy is a conditional mutual information that measures the predictability improvement in a given coupling direction by means of entropies (see Definition 2.1). The other main ingredient of our approach is the dimensional reduction of transfer entropy in the form of (conditional or unconditional) mutual information of the so-called transcripts, meaning that under circumstances the former can be calculated by the latter with one variable less [18]. The concept of transcript requires that the elements of a time series or its symbolic representation belong to an algebraic group $G$; we speak then of algebraic time series. Examples include measurements ($G = \mathbb{R}$ or $\mathbb{Q}$), instantaneous phases ($G = ([0, 1) \mod 1)$), and symbolic representations by ordinal patterns (or rank vectors) of length $L$ ($G = S_L$, the symmetric group of degree $L \geq 2$). The dimensional reduction of algebraic transfer entropy (i.e., the transfer entropy of algebraic time series) was proved in [7].

Transcripts were introduced in [16] to characterize synchronization in time series analysis. Their application to the detection of coupling directions resulted from the aforementioned dimensional reduction of the (algebraic) transfer entropy since less variables amounts to a better estimation of the pertaining probabilities. For example, in the lowest dimensional case the transfer entropy is a conditional mutual information with three variables (one is conditioning), while its dimensional reduction in the sense meant here is a mutual information of two transcripts obtained from those three variables. The reduction in the number of variables holds under one condition that, in some cases (typically, with random or deterministic synthetic data) can be satisfied, at least approximately, by adjusting the delay time. The potential application of such transcript-based dimensional reductions to the detection of coupling directions prompted us to explore their performance, regardless of the one condition guaranteeing its perfect match with the corresponding transfer entropy. Because of its relevance for practical applications, this exploration has
been mainly carried out for the 2-dimensional transcript mutual information (the lowest dimensional case) with both numerical and real world data. The results were quite satisfactory so far [18, 4, 6, 7, 15].

The present paper is a follow-up of Amigó et al. [7] in that we continue studying the theoretical properties and practical applications of transcript mutual information. Regarding the theoretical contents, we derive new bounds for the difference between transfer entropy and its transcript-based dimensional reduction in terms of the entropies of the coupled processes being considered, both in the general case and in the lowest dimensional one. As for the practical part, in [7] we analyzed (i) two coupled chaotic systems, and (ii) heart inter-beat (RR) intervals as well as systolic blood pressure measurements of healthy individuals during both medical air breathing and 100% oxygen breathing, to detect the causality relationships in all cases via two directionality indicators: the 3-dimensional symbolic transfer entropy [23] (i.e., $G = S_L$) and the corresponding transcript mutual information. Here we extend this previous work in two directions. First we include also random processes in our analysis, specifically, coupled autoregressive processes. As for the biomedical data, we compare this time the performance of the transcript mutual information also with that of its most direct competitor: the 3-dimensional binned transfer entropy, by means of RR-data and blood pressure data of healthy individuals, hypertensive patients and hypertensive patients with transient ischemic attack.

This paper is organized as follows. Section 2 contains the mathematical tools needed in the subsequent sections, including the concepts of transcript and coupling complexity coefficient in algebraic time series, as well as the dimensional reduction of algebraic transfer entropy via transcripts (Theorem 2.2). The relationship between algebraic transfer entropy and its dimensional reduction is analyzed in Section 3.1 in the general case, and in Section 3.2 in the lowest dimensional case, in order to bound the difference between the corresponding directionality indicators (Theorems 3.2 and 3.3). The performance of the mutual information of transcripts as a coupling direction indicator is numerically tested in Section 4 with two study cases. The first study case (Section 4.1) consists of three delay-coupled autoregressive processes, the corresponding time series being represented by ordinal patterns of length 4. The second study case (Section 4.2) is devoted to the analysis of the biomedical data mentioned above. In Section 5 the results are briefly discussed and the main conclusions drawn.

2. Mathematical preliminaries. Henceforth $G$ denotes a finite or infinite group. For example, $G$ might be $\mathbb{R}$, the additive group of the real numbers. In practice, though, observations are discrete and finite. For this reason, we will use below discrete-state random variables. Moreover, all summations over the states of those variables (e.g., in the definitions of entropies) are actually finite, even if $|G|$ (the cardinality of $G$) is infinite, because only a finite number of states have a positive probability to occur.

2.1. Transcripts. Transcripts are perhaps the simplest way of exploiting the algebraic structure of group-valued random variables, including the usual case of random variables taking on real numbers.

Let $G$ be a group. Given $x, y \in G$, we call

$$t = yx^{-1}$$

the transcript from the element $x$ to the element $y$ [16]. For generality, we use a multiplicative notation, i.e., the inverse of $x \in G$ is written as $x^{-1}$, and the
composition or ‘product’ of two elements of $\mathcal{G}$ is denoted by concatenating them in the correct order (since $\mathcal{G}$ might be non-commutative). When the elements $x$ and $y$ whose transcript is $t$ are important for the discussion, we write $t_{x,y}$. For instance, $t_{y,x} = (t_{x,y})^{-1}$.

In the case that $\mathcal{G}$ has a finite cardinality $|\mathcal{G}|$, the map $\mathcal{G} \times \mathcal{G} \to \mathcal{G}$ defined by $(x, y) \mapsto t_{x,y}$ is $|\mathcal{G}|$-to-1 since the $|\mathcal{G}|$ distinct pairs $(x, tx)$ are sent to $t$ for all $x \in \mathcal{G}$. Reciprocally, any pair $(x,y) \in \mathcal{G} \times \mathcal{G}$ whose transcript is $t$ must have $y = tx$ by (1).

In the usual case, $\mathcal{G}$ is $\mathbb{R}$ endowed with addition, and $t_{x,y} = y - x$. A more interesting case occurs when a real-valued time series $x = (x_n)_{n \in \mathbb{N}}$ is represented by the elements of a finite algebraic group. In particular, an ordinal representation is a symbolic time series whose elements are ordinal patterns [8] or, for that matter, $\mathcal{G}$-valued random variables $X_i, \ldots, X_N$ such that

$$x_n + \rho_0 < x_n + \rho_1 < \ldots < x_n + \rho_{L-1}.$$ 

Other similar definitions can be also found in the literature. Thus, in this case, $\mathcal{G} = S_L$, the symmetric group of degree $L$. This group is non-commutative for $L \geq 3$. In the case of time series generated by map iteration (i.e., orbits of a deterministic dynamics), one can prove under mild conditions that not all ordinal patterns are allowed for $L$ sufficiently large [1, 2]. We are going to consider ordinal representations in Section 4.

2.2. Coupling complexity coefficients. The coupling complexity coefficient of $\mathcal{G}$-valued random variables $X_1, \ldots, X_N$ ($N \geq 2$), denoted by $C(X_1, \ldots, X_N)$, is directly linked to their transcripts. It is defined as [3, 17]

$$C(X_1, X_2, \ldots, X_N) = \min_{1 \leq n \leq N} H(X_n) - H(X_1, X_2, \ldots, X_N) + H(T_{x_1, x_2, x_3, \ldots, x_{N-1}, x_N}),$$

where $H(\ldots)$ is the (joint) entropy of the random variable(s) in the argument, and $T_{x_i, x_{i+1}}$ is a random variable that outputs the transcript $t_{x_i, x_{i+1}}$ whenever $X_i = x_i$ and $X_{i+1} = x_{i+1}$. In [3] it was shown that

$$0 \leq C(X_1, X_2, \ldots, X_N) \leq \min_{1 \leq n \leq N} H(X_n). \tag{2}$$

By convention,

$$C(X) = 0. \tag{3}$$

The coupling complexity coefficients were introduced in [3] to discriminate different synchronisation regimes in coupled dynamics.

The coefficients $C(X_1, \ldots, X_N)$ have a number of interesting properties [17, 4]. Along with (2) and (3), a property that we will use below is the invariance under permutation of their arguments.

2.3. Information-theoretic measures and the equivalence property. For further reference, recall that the mutual information between two (uni- or multivariate) random variables $X$ and $Y$ is given by [10]

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)},$$

where $p(x,y)$ is the joint probability density function of $X$ and $Y$, and $p(x)$ and $p(y)$ are the marginal probability density functions of $X$ and $Y$, respectively.
so that the conditional mutual information $I(X;Y|Z)$ satisfies

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z),$$

with the conditional entropy $H(X|Y)$ being defined as

$$H(X|Y) = H(X,Y) - H(Y).$$

Then one can show that

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

$$= -H(Z) + H(X, Z) + H(Y, Z) - H(X,Y, Z)$$

$$= \sum_{x,y,z} p(x,y,z) \log \frac{p(x|y,z)}{p(x)|z}. \tag{4}$$

When computing entropies and other derived quantities of $G$-valued processes, the equivalent property is a useful tool. By this we mean the fact that different sets of variables composed of group elements and their transcripts contain the same information just because the knowledge of the variables in one set univocally determines the variables in the other sets. Thus, given the triple $(x,y,t_{x,y})$, the knowledge of any pair of elements, i.e., $(x,y)$, $(x,t_{x,y})$, or $(y,t_{x,y})$, univocally determines the remaining element. The same happens with $(x,y,t_{y,x})$. This simple observation implies that

$$H(\ldots, X, Y, \ldots) = H(\ldots, X, T_{x,y}, \ldots) = H(\ldots, X, T_{y,x}, \ldots)$$

$$= H(\ldots, T_{x,y}, Y, \ldots) = H(\ldots, T_{y,x}, Y, \ldots) \tag{5}$$

because any of the random variable pairs explicitly shown in (5) can be determined from any other variable pair. A similar property holds with higher numbers of symbols and their transcripts.

### 2.4. Algebraic transfer entropy and its dimensional reduction via transcripts

Let $X = (X_n)_{n \in \mathbb{Z}}$ and $Y = (Y_n)_{n \in \mathbb{Z}}$ be two $G$-valued, stationary random processes. We remind next the standard definition of transfer entropy [21], adding thereby the term ‘algebraic’ to underline that the values of $X_n$ and $Y_n$ possess an algebraic structure. For $r, s \geq 1$, set $X_n^{(r)} := X_n, \ldots, X_{n-r+1}$, $Y_n^{(s)} := Y_n, \ldots, Y_{n-s+1}$, and write $x_n^{(r)}$, $y_n^{(s)}$ for the outcomes of the $r$-dimensional random variable $X_n^{(r)}$ and the $s$-dimensional random variable $Y_n^{(s)}$, respectively.

**Definition 2.1.** The algebraic transfer entropy from the process $Y$ to the process $X$ with coupling delay $\Lambda \geq 1$ is defined as

$$AT_{Y \rightarrow X}^{(s,r)}(\Lambda) = I(X_{n+\Lambda}; Y_n^{(s)} | X_n^{(r)})$$

$$= H(X_{n+\Lambda} | X_n^{(r)}) - H(X_{n+\Lambda} | X_n^{(r)}, Y_n^{(s)})$$

$$= \sum_{x_{n+\Lambda},x_n^{(r)},y_n^{(s)}} p(x_{n+\Lambda}, x_n^{(r)}, y_n^{(s)}) \log \frac{p(x_{n+\Lambda} | x_n^{(r)}, y_n^{(s)})}{p(x_{n+\Lambda} | x_n^{(r)})}. \tag{6}$$

In other words, $AT_{Y \rightarrow X}^{(s,r)}(\Lambda)$ is the uncertainty reduction in future values of $X$, given past values of $X$, due to the additional knowledge of past values of $Y$. Hence, transfer entropy is a direct implementation of Wiener causality which, as explained in the Introduction, is based on predictability improvement [24]. From (6) it follows that $AT_{Y \rightarrow X}^{(s,r)}(\Lambda) = 0$ if and only if $p(x_{n+\Lambda} | x_n^{(r)}, y_n^{(s)}) = p(x_{n+\Lambda} | x_n^{(r)})$ for all $x_{n+\Lambda}$,
\(x_n^{(r)}\), and \(y_n^{(s)}\). Otherwise, \(AT_{Y \rightarrow X}^{(s,r)}(\Lambda) > 0\) and we say that there is an information transfer or flow from the process \(Y\) to the process \(X\) with delay \(\Lambda\). Ideally one should set \(s = r = -\infty\) in (6) but, bearing in mind the applications, it is more convenient to consider rather finite histories. Note that \(AT_{Y \rightarrow X}^{(s,r)}(\Lambda)\) depends on \(s + r + 1\) variables (of which \(r\) are conditioning variables).

The next theorem equals the \((s + r + 1)\)-dimensional algebraic transfer entropy \(AT_{Y \rightarrow X}^{(s,r)}(\Lambda)\) to an \((s + r)\)-dimensional conditional (if \(r \geq 2\)) or unconditional (if \(r = 1\)) mutual information of transcripts, under a restriction involving coupling complexity coefficients. For this reason, we may aptly speak of a dimensional reduction of \(AT_{Y \rightarrow X}^{(s,r)}(\Lambda)\).

**Theorem 2.2.** (Dimensional reduction of the algebraic transfer entropy) If

\[
C(X_{n+1},X_n,\ldots,X_{n-r+1},Y_{n},\ldots,Y_{n-s+1}) - C(X_{n+1},X_n,\ldots,X_{n-r+1}) = C(X_n,\ldots,X_{n-r+1},Y_{n},\ldots,Y_{n-s+1}) - C(X_n,\ldots,X_{n-r+1})
\]

then

\[
AT_{Y \rightarrow X}^{(s,1)}(\Lambda) = I(T_{x_{n+\Lambda},x_n};T_{x_{n+r}y_n},T_{y_n,y_{n-1}},\ldots,T_{y_{n-s+2},y_{n-s+1}}).
\]

Otherwise, if \(r \geq 2\), then

\[
AT_{Y \rightarrow X}^{(s,r)}(\Lambda) = I(T_{x_{n+\Lambda},x_n};T_{x_{n-r+1}y_n},T_{y_n,y_{n-1}},\ldots,T_{y_{n-s+2},y_{n-s+1}}|T_{x_{n-r+2},x_{n-r+1}}).
\]

See [7] for a proof. The dimensional reductions (8) and (9) are good news for practitioners for two reasons: less probabilities to estimate and faster computation.

As a conditional mutual information, \(AT_{Y \rightarrow X}^{(s,r)}(\Lambda)\) is not symmetric under the exchange of the processes \(X\) and \(Y\). Therefore, the directionality indicator

\[
\Delta AT_{Y \rightarrow X}^{(s,r)}(\Lambda) = AT_{Y \rightarrow X}^{(s,r)}(\Lambda) - AT_{X \rightarrow Y}^{(r,s)}(\Lambda) = -\Delta AT_{X \rightarrow Y}^{(r,s)}(\Lambda)
\]

measures the net transfer of information between the processes \(X\) and \(Y\), the dominant coupling direction being determined by its sign. For example, if \(\Delta AT_{Y \rightarrow X}^{(s,r)}(\Lambda) > 0\) then \(Y\) is the dominant driving process with a coupling delay \(\Lambda\).

By Theorem 2.2, if the restrictions (7) and

\[
C(Y_{n+\Lambda},Y_n,\ldots,Y_{n-r+1},X_n,\ldots,X_{n-s+1}) - C(Y_{n+\Lambda},Y_n,\ldots,Y_{n-r+1}) = C(Y_n,\ldots,Y_{n-r+1},X_n,\ldots,X_{n-s+1}) - C(Y_n,\ldots,Y_{n-r+1})
\]

hold, then

\[
AT_{Y \rightarrow X}^{(s,r)}(\Lambda) = TI_{Y \rightarrow X}^{(s,r)}(\Lambda), \quad AT_{X \rightarrow Y}^{(r,s)}(\Lambda) = TI_{X \rightarrow Y}^{(r,s)}(\Lambda)
\]

where \(TI_{Y \rightarrow X}^{(s,r)}(\Lambda)\) and \(TI_{X \rightarrow Y}^{(r,s)}(\Lambda)\) are the corresponding dimensional reductions given in Theorem 2.2. Here \(TI\) stands for transcript mutual information. According to Theorem 2.2, \(TI_{Y \rightarrow X}^{(s,r)}(\Lambda)\) is an unconditional mutual information if \(r = 1\); otherwise, it is a conditional mutual information. Hence, under the above assumptions,

\[
\Delta AT_{Y \rightarrow X}^{(s,r)}(\Lambda) = \Delta TI_{Y \rightarrow X}^{(s,r)}(\Lambda)
\]

where (see (10) and (12))

\[
\Delta TI_{Y \rightarrow X}^{(s,r)}(\Lambda) = TI_{Y \rightarrow X}^{(s,r)}(\Lambda) - TI_{X \rightarrow Y}^{(r,s)}(\Lambda) = -\Delta TI_{X \rightarrow Y}^{(r,s)}(\Lambda).
\]
3. Transcript mutual information as a directionality indicator. If restrictions (7) and (11) hold, then (13) shows that the net information flow between two coupled processes, $\Delta AT_{Y \rightarrow X}(\Lambda)$, can be actually obtained with a quantity of lower dimensionality, $\Delta TI_{Y \rightarrow X}(\Lambda)$, that therefore requires less data for its estimation. However, the conditions (7) and (11) are seldom satisfied, in particular when dealing with real observations [7]. This leads to the question of how good $\Delta TI_{Y \rightarrow X}(\Lambda)$ performs as a coupling directionality indicator per se. In this section we derive some theoretical results related to that question. The performance of $\Delta TI_{Y \rightarrow X}(\Lambda)$ in practice will be tested in Section 4.

3.1. General case. As a first observation, note that $\Delta TI_{Y \rightarrow X}(\Lambda) = 0$ if $X$ and $Y$ are uncoupled because then the transcript $T_{x_n \Lambda, x_n}$ is independent of the other random variables in the argument of $TI_{Y \rightarrow X}(\Lambda)$ (see the right hand sides of Eqs. (8) and (9)) and, likewise, $T_{y_n \Lambda, y_n}$ is independent of the other random variables in the argument of $TI_{X \rightarrow Y}(\Lambda)$. This is a consistency condition for $\Delta TI_{Y \rightarrow X}(\Lambda)$ to be a directionality indicator.

Next we are going to bound the difference between $TI_{Y \rightarrow X}(\Lambda)$ and $AT_{Y \rightarrow X}(\Lambda)$. The following theorem gives an answer in terms of $H(X_n)$ and $H(Y_n)$.

**Theorem 3.1.** It holds:

$$-H(X_n) - \min\{H(X_n), H(Y_n)\} \leq AT_{Y \rightarrow X}(\Lambda) - TI_{Y \rightarrow X}(\Lambda) \leq \min\{H(X_n), H(Y_n)\}. \quad (15)$$

Otherwise, if $r \geq 2$, then

$$\left| AT_{Y \rightarrow X}(\Lambda) - TI_{Y \rightarrow X}(\Lambda) \right| \leq H(X_n) + \min\{H(X_n), H(Y_n)\}. \quad (16)$$

A proof is given in the Appendix. Note that the bounds in (15) are independent of $s$, and the bounds in (16) are independent of $s$ and $r$, i.e., the difference between an algebraic transfer entropy and its transcript-based dimensional reduction is bounded by quantities that do not depend on the dimension (except for the distinction between the cases $r = 1$ and $r > 1$).

Use now Theorem 3.1 and definitions (10) and (14) to bound readily the difference between the coupling directionality indicators $\Delta AT_{Y \rightarrow X}(\Lambda)$ and $\Delta TI_{Y \rightarrow X}(\Lambda)$ as follows.

**Theorem 3.2.** It holds:

$$-H(X_n) - 2\min\{H(X_n), H(Y_n)\} \leq \Delta AT_{Y \rightarrow X}(\Lambda) - \Delta TI_{Y \rightarrow X}(\Lambda) \leq H(Y_n) + 2\min\{H(X_n), H(Y_n)\}. \quad (17)$$

Otherwise, if $r \geq 2$,

$$\left| \Delta AT_{Y \rightarrow X}(\Lambda) - \Delta TI_{Y \rightarrow X}(\Lambda) \right| \leq H(X_n) + H(Y_n) + 2\min\{H(X_n), H(Y_n)\}. \quad (18)$$

From Theorems 3.1 and 3.2 we obtain the following immediate results.

**Corollary 1.** (i) If $H(X_n) \simeq 0$ then

$$AT_{Y \rightarrow X}(\Lambda) \simeq TI_{Y \rightarrow X}(\Lambda).$$
(ii) If \( H(X_n) \simeq H(Y_n) \simeq 0 \), then
\[
\Delta AT_{Y \to X}^{(s,r)}(\Lambda) \simeq \Delta T_{Y \to X}^{(s,r)}(\Lambda).
\]

Small entropy means small complexity. Such is the case for a deterministic process with small positive Lyapunov exponents.

3.2. Lowest dimensional case. Since the probabilities in (6) are usually estimated by the relative frequencies of the corresponding symbols or via nearest neighbors [21], the lowest dimensional case \( r = s = 1 \) is the preferred choice in the event that time series are short.

Let us emphasize at this point that the term ‘short’ time series or ‘small’ data set depends both on the number of data available, \( N \), and the number of probabilities to be estimated, \( P \). Needless to say, in a numerical simulation \( N \) can be taken as large as necessary; this is why numerical simulation is the ideal tool when it comes to test statistical properties. On the contrary, real world data can be scarce for a number of reasons, in particular stationarity and availability. Clearly, \( N \) has to be much greater than \( P \) to avoid undersampling. As a rule of thumb, one can use \( N \lesssim 10P \). Thus, given \( N \), a data set can be sufficient or insufficient for a good estimation of a directionality indicator depending on its dimension. We come back to this issue in Sections 4.1 and 4.2.

In the lowest dimensional case \( r = s = 1 \), Theorem 2.2 states that if
\[
C(X_{n+\Lambda}, X_n, Y_n) = C(X_{n+\Lambda}, X_n) + C(X_n, Y_n)
\]
then (see (12))
\[
AT_{Y \to X}^{(1,1)}(\Lambda) = T_{Y \to X}^{(1,1)}(\Lambda)
\]
so that \( AT_{Y \to X}^{(1,1)}(\Lambda) = I(X_{n+\Lambda}; Y_n \mid X_n) \), a conditional mutual information, equals the unconditional mutual information of transcripts \( T_{Y \to X}^{(1,1)}(\Lambda) = I(T_{x_{n+\Lambda}, x_n}; T_{x_n, y_n}) \). Note that \( r = s = 1 \) is the only case in which the use of transcripts instead of transfer entropy makes a difference in practice too concerning the dimensionality (i.e., the number of probabilities to be estimated), otherwise \( AT_{Y \to X}^{(s-1,r)}(\Lambda) \) or \( AT_{Y \to X}^{(s,r-1)}(\Lambda) \) would do as well.

Let us mention in passing that sometimes (19) can be satisfied, at least approximately, by fine tuning the time delay or by choosing it sufficiently large [18, 4]. The dependence of \( AT_{Y \to X}^{(1,1)}(\Lambda) - T_{Y \to X}^{(1,1)}(\Lambda) \) on the time delay (hence, on how well the restriction (19) is satisfied) was studied in [18] for bidirectionally delayed-coupled logistic maps (see Figure 1 of [18]). Also, by Corollary 1, \( AT_{Y \to X}^{(1,1)}(\Lambda) \simeq T_{Y \to X}^{(1,1)}(\Lambda) \) if \( H(X_n) \simeq 0 \).

On the above grounds, we focus henceforth on the case \( r = s = 1 \). Set
\[
AT_{Y \to X}(\Lambda) \equiv AT_{Y \to X}^{(1,1)}(\Lambda)
\]
\[
= \sum_{x_{n+\Lambda}, x_n, y_n} p(x_{n+\Lambda}, x_n, y_n) \log \frac{p(x_{n+\Lambda}, x_n, y_n)}{p(x_{n+\Lambda}, x_n)p(x_n, y_n)}
\]
for the 3-dimensional algebraic transfer entropy in the direction \( Y \to X \), and likewise,
\[ TI_{Y \rightarrow X}(\Lambda) = TI_{Y \rightarrow X}^{(1,1)}(\Lambda) \]
\[ = \sum_{t_{x_n+A}, t_{x_n}, t_{x_n}, y_n} p(t_{x_n+A}, t_{x_n}, t_{x_n+y_n}) \log \frac{p(t_{x_n+A}, t_{x_n}, t_{x_n+y_n})}{p(t_{x_n+A}, t_{x_n})p(t_{x_n}, y_n)} \]  
\[ = \sum_{x_{n+1}^{A}, y_{n+1}, y_{n}, y_{n-1}} p(x_{n+1}^{A}, y_{n+1}, y_{n-1}) \log \frac{p(x_{n+1}^{A}, y_{n+1}, y_{n-1})}{p(x_{n+1}^{A})p(y_{n+1})p(y_{n-1})}. \]

for its dimensional reduction, \( I(T_{x_n+A}; T_{x_n}, y_n) \). We call \( TI_{Y \rightarrow X}(\Lambda) \) the (2-dimensional) transcript mutual information in the direction \( Y \rightarrow X \). Note that \( TI_{Y \rightarrow X}(\Lambda) \) is not symmetric under the exchange \( X \leftrightarrow Y \) (although, of course, it is symmetric under the exchange \( T_{x_n+A}, x_n \leftrightarrow T_{x_n}, y_n \)).

Of course, if the restrictions (19) and
\[ C(Y_{n+1}, Y_n, X_n) = C(Y_{n+1}, Y_n) + C(Y_n, X_n) \]  
are satisfied, then \( \Delta AT_{Y \rightarrow X}(\Lambda) = \Delta TI_{Y \rightarrow X}(\Lambda) \), where \( \Delta AT_{Y \rightarrow X} = \Delta AT_{Y \rightarrow X}^{(1,1)} \) and \( \Delta TI_{Y \rightarrow X} = \Delta TI_{Y \rightarrow X}^{(1,1)} \). The next theorem shows that the bounds (17) can be made sharper in the case \( r = s = 1 \).

**Theorem 3.3.** It holds
\[ |\Delta AT_{Y \rightarrow X}(\Lambda) - \Delta TI_{Y \rightarrow X}(\Lambda)| \leq H(X_n) + H(Y_n). \]  

A proof of Theorem 3.3 can be found in the Appendix. This proof illustrates also the application of the equivalence property (Section 2.3).

Regarding the sign of \( \Delta AT_{Y \rightarrow X}(\Lambda) \), that determines the dominant driving process (should \( \Delta AT_{Y \rightarrow X}(\Lambda) \) perform flawless), Eq. (24) leads only to the following weak conclusions:
\[ \Delta AT_{Y \rightarrow X}(\Lambda) > 0 \Rightarrow \Delta TI_{Y \rightarrow X}(\Lambda) > -(H(X_n) + H(Y_n)) \]
\[ \Delta AT_{Y \rightarrow X}(\Lambda) < 0 \Rightarrow \Delta TI_{Y \rightarrow X}(\Lambda) < H(X_n) + H(Y_n) \]

Therefore, \( \Delta AT_{Y \rightarrow X}(\Lambda) \) and \( \Delta TI_{Y \rightarrow X}(\Lambda) \) may have opposite signs even if \( H(X_n) \approx H(Y_n) \approx 0 \) (in which case \( \Delta AT_{Y \rightarrow X}(\Lambda) \approx \Delta TI_{Y \rightarrow X}(\Lambda) \) by Corollary 1(ii)). In actually, though, \( \Delta AT_{Y \rightarrow X}(\Lambda) \) can perform poorly due to a number of factors, including hidden drivers, small datasets, noise, symbolization, etc. [22, 9]. Intriguingly, numerical and experimental results show that \( \Delta TI_{Y \rightarrow X}(\Lambda) \) performs as good as (if not better than) \( \Delta AT_{Y \rightarrow X}(\Lambda) \) or other directionality indicators, especially when the dimensionality is important due to the dataset size [18, 4, 6, 7, 15].

**4. Comparative study.** We are going to extend in this section the case study in [7] as follows. First we use random processes (instead of chaotic oscillators) to compare the performances of the symbolic transfer entropy (i.e., the transfer entropy of symbolic time series representations by ordinal patterns [23]) and the 2-dimensional transcript mutual information. For this we will choose three unidirectionally, delay-coupled autoregressive processes. Second, we use biomedical data (heart inter-beat interval and systolic blood pressure) to compare the performances of the binned transfer entropy, the symbolic transfer entropy, and the transcript mutual information. By binned and symbolic transfer entropy we mean in both cases their lowest dimensional versions (i.e., \( s = r = 1 \) in Eq. (6)). See Section 2.1 for the concepts of ordinal pattern and transcript. The ordinal methodology in general biomedical applications can be found in [5]. For specific applications to cardiology, the interested reader is referred to [19, 12, 13].
4.1. Study case 1: Autoregressive processes. Since our present objective is to compare the performance of $\Delta AT_{Y \rightarrow X}(\Lambda)$ and $\Delta TI_{Y \rightarrow X}(\Lambda)$ as directionality indicators, we are going to consider three unidirectionally coupled systems. For them we choose the three delay-coupled, autoregressive processes

$$\begin{align*}
x_{n+1} &= 0.6x_n + 0.8y_{n-3} + \eta^x_n \\
y_{n+1} &= 0.4y_n + 0.9z_{n-1} + \eta^y_n \\
z_{n+1} &= 0.9z_n + \eta^z_n
\end{align*}$$

for $n \geq 3$, where $\eta^x, \eta^y, \eta^z$ are normal random variables satisfying $\langle \eta^r(t_i)\eta^s(t_j) \rangle = \delta_{ij}\delta_{rs}$ with $r, s \in \{x, y, z\}$. Therefore, $Z = (z_n)_{n \geq 3}$ drives $Y = (y_n)_{n \geq 3}$, and, in turn, $Y$ drives $X = (x_n)_{n \geq 3}$, or, in more graphical terms, their causal connectivity is $Z \rightarrow Y \rightarrow X$.

For the numerical simulation, $10^5$ points of each time series were generated and a symbolic representation of the time series with ordinal patterns of length $L = 4$ was used. To be more specific, let $\hat{x}_n = (\xi_0, \xi_1, \xi_2, \xi_3) \in S_4$ be the ordinal 4-pattern of the time delay vector $v(x_n) = (x_n, x_{n+T}, x_{n+2T}, x_{n+3T})$ (i.e., $x_n + \xi_i T < x_n + \xi_j T < x_n + \xi_k T$, with $\xi_i \in \{0, 1, 2, 3\}$), where $T \geq 1$ is the time delay and $3 \leq n \leq 10^5 - 3T$. Proceed similarly with the time series $(y_n)$ and $(z_n)$ to obtain the corresponding ordinal representations $(\hat{y}_n)$ and $(\hat{z}_n)$. The data analyzed were the $S_4$-valued time series $(\hat{x}_n)$, $(\hat{y}_n)$, and $(\hat{z}_n)$.

In Figures 1-3 we plot the values of the directionality indicators $\Delta AT_{D \rightarrow R}(\Lambda)$ (solid line) and $\Delta TI_{D \rightarrow R}(\Lambda)$ (dash-dotted line) vs $\Lambda$ for 8 choices of $T$. Here $\hat{D}$ stands for the driving system and $\hat{R}$ for the response system in each case considered, namely, $\hat{Z} \rightarrow \hat{Y}$ (Figure 1), $\hat{Y} \rightarrow \hat{X}$ (Figure 2), and $\hat{Z} \rightarrow \hat{X}$ (Figure 3), where $\hat{X}$ is the $S_4$-valued random process which output $\hat{x}_n$, etc. Therefore, $\Delta AT_{D \rightarrow R}(\Lambda) > 0$ and $\Delta TI_{D \rightarrow R}(\Lambda) > 0$ are the correct results. To avoid spurious correlations due to overlaps of the ordinal 4-patterns we chose $1 \leq \Lambda \leq T - 1$. The points $(\Delta AT_{D \rightarrow R}(\Lambda), \Lambda)$ and $(\Delta TI_{D \rightarrow R}(\Lambda), \Lambda)$ have been linearly interpolated for a better visualization. A summary of the numerical results follows.

Figure 1 shows the results for the indicators $\Delta AT_{Z \rightarrow Y}(\Lambda)$ and $\Delta TI_{Z \rightarrow Y}(\Lambda)$. Note that $\Delta AT_{Z \rightarrow Y}(\Lambda) \leq 0$ for $T = 4$ and $\Lambda = 3$; for $T = 5$ and $\Lambda = 4$; for $T = 6$ and $\Lambda = 4.5$; for $T = 7$ and $\Lambda = 5, 6$; for $T = 8$ and $\Lambda = 5, 6, 7, 8$; and for $T = 10$ and $\Lambda = 6, 7, 8, 9$. In all these cases, $\Delta AT_{Z \rightarrow Y}(\Lambda)$ indicates the wrong connectivity $\hat{Y} \rightarrow \hat{Z}$. On the contrary, $\Delta TI_{Z \rightarrow Y}(\Lambda) > 0$ in all cases.

Figure 2 shows the results for indicators $\Delta AT_{Y \rightarrow X}(\Lambda)$ and $\Delta TI_{Y \rightarrow X}(\Lambda)$. This time $\Delta AT_{Y \rightarrow X}(\Lambda) \leq 0$ for $T = 4$ and $\Lambda = 1$; for $T = 5$ and $\Lambda = 1, 2$; for $T = 6$ and $\Lambda = 1, 2$; for $T = 7$ and $\Lambda = 1, 2$; for $T = 9$ and $\Lambda = 6, 7, 8$; and for $T = 10$ and $\Lambda = 6, 7, 8$. In all these cases, $\Delta AT_{Y \rightarrow X}(\Lambda)$ fails to indicate the right connectivity $\hat{Y} \rightarrow \hat{X}$. Again, $\Delta TI_{Z \rightarrow Y}(\Lambda) > 0$ without exception.

Finally, Figure 3 shows the results for $\Delta AT_{Z \rightarrow X}(\Lambda)$ and $\Delta TI_{Z \rightarrow X}(\Lambda)$. We see that mistakenly $\Delta AT_{Z \rightarrow X}(\Lambda) \leq 0$ for $T = 6$ and $\Lambda = 1, 2$; for $T = 7$ and $\Lambda = 1, 2, 3$; for $T = 8$ and $\Lambda = 1, 2, 3$; for $T = 9$ and $\Lambda = 1, 2, 3$; and for $T = 10$ and $\Lambda = 1, 2, 3$. Exceptionally, $\Delta TI_{Z \rightarrow X}(\Lambda) = 0$ for $T = 3$ and $\Lambda = 1$.

In sum, even though $X$, $Y$, and $Z$ are unidirectionally coupled, $\Delta AT_{D \rightarrow R}(\Lambda)$ wrongly detects bidirectional couplings between all the processes for plenty of delay times and coupling delays, while $\Delta TI_{D \rightarrow R}(\Lambda)$ performs flawless. It is worth noting that undersampling is not an issue here, then $N/P \approx 10^5/(2 \times 4^3) \approx 781.2$ for $\Delta AT_{D \rightarrow R}(\Lambda)$. We conclude that the transcript mutual information outperforms
clearly the symbolic transfer entropy (with $L = 4$) when determining the coupling directions of the processes (25).

To wrap up the above numerical analysis, we tested also the consistency and sensitivity of $\Delta T_{\hat{Z} \rightarrow \hat{X}}(\Lambda)$. To this end we let $k_{zy}$, the coupling constant between the processes $\hat{Z}$ and $\hat{Y}$ (i.e., the coefficient of $z_{n-1}$ on the second line of (25)), to take the values 0.0, 0.1, 0.5 and, for comparison, 0.9 (the same value as in Figures 1-3), while keeping fixed the values of the other coupling constants. If consistency holds, we expect $\Delta T_{\hat{Z} \rightarrow \hat{Y}}(\Lambda) = \Delta T_{\hat{Z} \rightarrow \hat{X}}(\Lambda) = 0$ when $k_{zy} = 0.0$, i.e., when $\hat{Z}$ and $\hat{Y}$ are uncoupled. If sensitivity holds, we expect $\Delta T_{\hat{Z} \rightarrow \hat{Y}}(\Lambda)$ and $\Delta T_{\hat{Z} \rightarrow \hat{X}}(\Lambda)$ to take
on slightly positive values when $k_{zy} = 0.1$, i.e., when $\hat{Z}$ and $\hat{Y}$ are weakly coupled, these values increasing as $k_{zy}$ increases. Figure 4 confirms all those expectations for $T = 8$ and $1 \leq \Lambda \leq 7$. We see again that $\Delta T I_{\hat{D} \rightarrow \hat{R}}(\Lambda)$ outperforms $\Delta A T_{\hat{D} \rightarrow \hat{R}}(\Lambda)$, meaning that $\Delta T I_{\hat{D} \rightarrow \hat{R}}(\Lambda) > 0$ whenever there is a coupling $\hat{D} \rightarrow \hat{R}$, as well as $\Delta T I_{\hat{D} \rightarrow \hat{R}}(\Lambda) > \Delta A T_{\hat{D} \rightarrow \hat{R}}(\Lambda)$ whenever $\Delta A T_{\hat{D} \rightarrow \hat{R}}(\Lambda) > 0$. Similar results are also obtained for other values of $T$, $1 \leq \Lambda \leq T - 1$ (not shown).

4.2. Study case 2: Biomedical data. Cardiovascular disease is a major death cause around the world. Alone in Europe it causes 3.9 million deaths every year, thereof over 1.8 million in the European Union. The risk of suffering a cardiovascular disease is directly related to both systolic and diastolic blood pressure levels. Hypertension, defined as systolic blood pressure $\geq 140$ mmHg and/or diastolic blood pressure $\geq 90$ mmHg, is the leading cause of death and disability-adjusted life years worldwide [11]. Complications of hypertension include coronary artery disease, heart failure, stroke, chronic kidney disease, and peripheral artery disease. Transient ischemic attack (TIA) is a transient episode of neurologic dysfunction caused by ischemia. Symptoms are the same as in a stroke (difficulty of speaking, vision loss, weakness on one side of the body, etc.) but resolve within 24 hours.

We report next results obtained from patients studied at the Department of Hypertension and Diabetology Medical University of Gdańsk, Poland. This work was supported by the National Science Centre, grant MAESTRO UMO-2011/02/A/NZ5/00329. The study protocol was approved by the Ethics Committee of the Medical University of Gdańsk (NKEBN/422/2011). All participants were informed about the study merits and signed a written consent. The groups of patients considered in our data analysis were the following:

- Group I consisted of 22 healthy subjects (45.1 $\pm$ 13.1 years, 12 men).
- Group II consisted of 41 patients (53.7 $\pm$ 12.3 years, 25 men) with hypertension but no other cardiovascular disease.
Figure 4. Plots of $\Delta T_{\hat{D} \rightarrow \hat{R}}(\Lambda)$ (continuous lines) and $\Delta T_{\hat{I} \rightarrow \hat{R}}(\Lambda)$ (dash-dotted lines) for $T = 8$, $1 \leq \Lambda \leq 7$, and $k_{zy} = 0.0$ (first column), $k_{zy} = 0.1$ (second column), $k_{zy} = 0.5$ (third column), and $k_{zy} = 0.9$ (fourth column). Top row corresponds to the coupling direction $\hat{Z} \rightarrow \hat{Y}$, middle row to $\hat{Z} \rightarrow \hat{X}$, and bottom row to $\hat{Y} \rightarrow \hat{X}$.

- **Group II** consisted of 15 patients (49.0 ± 15.1 years, 10 men) with both hypertension and a history of transient ischemic attack.

All patients underwent short-term electrocardiographic recording (ECG) using PowerLab system with Lab Chart software (ADInstruments, Australia). The sampling rate was 1000 Hz. Non-invasive beat-to-beat blood pressure (BP) was recorded by a FINOMETER device (Finapres Medical Systems). Recordings of 1000 inter-beat (RR) intervals during rest in the supine position were used for further analyses. The number of artifacts was less than 5% of all RR intervals.

Consider first the RR data. The raw data consisted of one time series per patient and Group of the form

$$\left(RR_n\right)^{1000}_{n=1} = (1000(\tau_{n+1} - \tau_n))^{1000}_{n=1}$$

where $\tau_n$ are the times in seconds when the heart beats and, hence, $RR_n$ are the consecutive inter-beat time intervals in milliseconds. The maximum value of the arterial pressure waveform inside the $n$th RR interval, $RR_n$, was taken as the $n$th systolic blood pressure $BP_n$. As a result, we obtained the time series

$$(y_n)^{1000}_{n=1} = (RR_n)^{1000}_{n=1}$$
and

\[(x_n)_{n=1}^{1000} = (BP_n)_{n=1}^{1000}\]

for each patient in each Group. Therefore, the data sample consisted of 22 such time series for Group I (healthy individuals), 41 time series for Group II (hypertensive patients), and 15 time series for Group IIIB (hypertensive patients with TIA).

The objective of our analysis is to ascertain the coupling direction between the RR time series \((y_n)_{n=1}^{1000}\) and the BP time series \((x_n)_{n=1}^{1000}\) in all three patient Groups. Actually, owing to the way in which the directionality indicators were the following:

\[(y_n)_{n=1}^{1000} \text{ and } (x_n+1)_{n=1}^{999} \text{ when studying the direction } RR \rightarrow BP, \text{ while we used } (x_n)_{n=1}^{1000} \text{ and } (y_n)_{n=1}^{1000} \text{ when studying the direction } BP \rightarrow RR \text{ below [20].} \]

Intuitively, in the former direction \(RR_n\) cannot influence \(BP_n\) but \(BP_n\) is measured before the time interval \(RR_n\) has elapsed. With this proviso, the directionality indicators were the following:

(i) \(\Delta T E_{RR \rightarrow BP}(\Lambda) = T_{RR \rightarrow BP}^{(1,1)}(\Lambda) - T_{BP \rightarrow RR}^{(1,1)}(\Lambda), \)

where \(T_{RR \rightarrow BP}^{(1,1)}(\Lambda) = \sum_{x_n+1, x_n+1, y_n} p(x_{n+1}, x_{n+1}, y_n) \log \left( \frac{p(x_{n+1})p(x_{n+1}, y_n)}{p(x_{n+1}, y_n)} \right) \)

and \(T_{BP \rightarrow RR}^{(1,1)}(\Lambda) = \sum_{y_n+1, x_n, y_n} p(y_{n+1}, y_n, x_n) \log \left( \frac{p(y_{n+1}, y_n, x_n)}{p(y_{n+1}, x_n)} \right) \)

are 3-dimensional, binned transfer entropies with coupling delay \(\Lambda\), all the observations being quantized with 6 bins;

(ii) \(\Delta STE_{RR \rightarrow BP}(\Lambda) = AT_{RR \rightarrow BP}^{(1,1)}(\Lambda) - AT_{BP \rightarrow RR}^{(1,1)}(\Lambda), \)

where \(AT_{RR \rightarrow BP}^{(1,1)}(\Lambda)\) is the 3-dimensional algebraic transfer entropy (21) computed with ordinal patterns of length 3, i.e., with \((x_n)_{n=1}^{1000}\) and \((y_n)_{n=1}^{1000}\) being represented by ordinal patterns of length 3 (elements of the permutation group \(S_3\)); and

(iii) \(\Delta TMI_{RR \rightarrow BP}(\Lambda) = TI_{RR \rightarrow BP}(\Lambda) - TI_{BP \rightarrow RR}(\Lambda), \)

where \(TI_{RR \rightarrow BP}(\Lambda)\) is the transcript mutual information (22), also computed with ordinal patterns of length 3.

Thus, we represent next all the time series by the ordinal patterns of sliding time delay vectors of length \(L = 3\) and time delay \(T \geq 1\). That is, if \(v(x_n) = (x_n, x_{n+1}, x_{n+2T})\) and \(v(y_n) = (y_n, y_{n+1}, y_{n+2T})\) are such vectors, \(1 \leq n \leq 1000 - 2T\), and \(\hat{x}_n = BP_n \in S_3, \hat{y}_n = RR_n \in S_3\) are the ordinal patterns of \(v(x_n)\) and \(v(y_n)\), respectively, then the symbolic time series to be analyzed are

\[(\hat{x}_n)_{n=1}^{1000-2T} = (BP_n)_{n=1}^{1000-2T}, \quad (\hat{y}_n)_{n=1}^{1000-2T} = (RR_n)_{n=1}^{1000-2T}.\]

To avoid spurious correlations due to overlaps of the ordinal 3-patterns, we take \(1 \leq \Lambda \leq T - 1\).

As a matter of fact, the estimations of \(\Delta T E_{RR \rightarrow BP}(\Lambda)\) and \(\Delta STE_{RR \rightarrow BP}(\Lambda)\) are going to be undersampled. Indeed, the ratios of the dataset size to the number of probabilities to be estimated are \(N/P \approx 1000/(2 \times 6)^3 = 2.3\) for \(\Delta T E_{RR \rightarrow BP}(\Lambda)\) and \(\Delta STE_{RR \rightarrow BP}(\Lambda)\). However, \(N/P \approx 1000/(2 \times 6^2) = 13.8\) for \(\Delta TMI_{RR \rightarrow BP}(\Lambda)\).

Figure 5 shows the results of the directionality indicators \(\Delta T E_{RR \rightarrow BP}(\Lambda)\) (top row), \(\Delta STE_{RR \rightarrow BP}(\Lambda)\) (middle row), and \(\Delta TMI_{RR \rightarrow BP}(\Lambda)\) (bottom row) for the patient Groups I (left column), II (middle column), and IIIB (right column) with \(T = 6\) and \(1 \leq \Lambda \leq 5\).

Consider first Groups I and II (two leftmost columns of Figure 5). Then we see that \(\Delta TMI_{RR \rightarrow BP}(1) > 0\) for both Groups (which is in concordance with previous
Figure 5. The directionality indicators $\Delta T_{RR\rightarrow BP}(\Lambda)$ (top row), $\Delta STE_{RR\rightarrow BP}(\Lambda)$ (middle row), and $\Delta TMI_{RR\rightarrow BP}(\Lambda)$ (bottom row) for the patient Groups I (left column), II (middle column), and IIB (right column) with $T=6$ and $1 \leq \Lambda \leq 5$. For convenience, here $TE$ stands for binned transfer entropy, $STE$ for symbolic transfer entropy, and $TMI$ for transcript mutual information. See the text for more details, the description of the patient Groups, and the data ($RR_n$) and ($BP_n$).

findings), while $\Delta T_{ER\rightarrow BP}(1) \geq 0$ for Groups I and II, and $\Delta STE_{RR\rightarrow BP}(1) \geq 0$ only for Group II. For other values of $\Lambda$, $\Delta T_{ER\rightarrow BP}(\Lambda)$ performs marginally better (although with error bars including negatives values), with $\Delta TMI_{RR\rightarrow BP}(\Lambda)$ performing better than $\Delta STE_{RR\rightarrow BP}(\Lambda)$.

As for Group IIB, none of the three indicators signalizes a clear direction and their performance is comparable. This means that the sampled data are too noisy for the indicators to converge.

As a final remark, no clinical implications from the results of our analysis are claimed.

5. Discussion and conclusion. The scope of this paper was two-fold. First, to derive new theoretical results on the relation between the coupling indicators $\Delta AT_{Y\rightarrow X}(\Lambda)$ and $\Delta TI_{Y\rightarrow X}(\Lambda)$ in general, and between $\Delta AT_{Y\rightarrow X} = \Delta AT_{Y\rightarrow X}^{(1)}(\Lambda)$ and $\Delta TI_{Y\rightarrow X} = \Delta TI_{Y\rightarrow X}^{(1,1)}(\Lambda)$ in particular. (For the time being, we drop the dependence on $\Lambda$.) By Theorem 2.2, $\Delta AT_{Y\rightarrow X}^{(s,r)} = \Delta TI_{Y\rightarrow X}^{(s,r)}$ under the restrictions (7) and (11) but, otherwise, these indicators may be different (although their difference...
is bounded by Theorems 3.2 and 3.3). Note that 3 is the lowest dimension of a transfer entropy-based directionality indicator, corresponding to $\Delta AT_Y \rightarrow X$ (for the choice $r = s = 1$), while $\Delta TI_Y \rightarrow X$ is 2-dimensional (Theorem 2.2). Our choice of these particular indicators for the comparative study in Section 4 has precisely to do with their relevance in practical applications, where such a dimensional reduction can make a difference if the data set is too small for a good estimation of $\Delta AT_Y \rightarrow X$.

The second scope of this paper was to explore the performance of $\Delta TI_Y \rightarrow X$ as directionality indicator regardless of whether restrictions (19) and (23) are fulfilled. To this end we compared the performance of $\Delta TI_Y \rightarrow X$ against $\Delta AT_Y \rightarrow X$ via numerical simulations with random processes (Section 4.1) and bivariate biomedical data (Section 4.2), to wit: time series of heart inter-beat intervals (RR) and systolic blood pressures (BP). Specifically, in Section 4.1 we computed $\Delta TI_{\hat{D} \rightarrow \hat{R}}$ with ordinal patterns of length 4, where $\hat{D}$ stands for the driving symbolic time series and $\hat{R}$ for the response symbolic time series in a unidirectionally coupled chain of three autoregressive processes; we also checked the consistency and sensitivity of both indicators. In Section 4.2, we computed $\Delta TI_{RR \rightarrow BP}$ with ordinal patterns of length 3, while for $\Delta AT_{RR \rightarrow BP}$ we used both ordinal patterns of length 3 (denoted as $\Delta TE_{RR \rightarrow BP}$) and binned values (denoted as $\Delta TE_{RR \rightarrow BP}$), the number of bins being 6. Both study cases extend in several ways the survey initiated in [7]. In the first case (synthetic data), $\Delta TI_{\hat{D} \rightarrow \hat{R}}$ outperformed clearly $\Delta AT_{\hat{D} \rightarrow \hat{R}}$. In the second case (real world data), $\Delta TMI_{RR \rightarrow BP}$ outperformed $\Delta TE_{RR \rightarrow BP}$ and $\Delta TE_{RR \rightarrow BP}$ in patient groups I and II, and performed similarly to the other two indicators in the remaining patient group IIb. These results are in line with the results found in previous, related works [18, 4, 6, 7, 15], i.e., the transcript-based coupling directionality indicator $\Delta TI_Y \rightarrow X$ performs as good as (and usually better than) its most direct competitor $\Delta AT_Y \rightarrow X$.

This having been said, the main advantage of $\Delta AT_Y \rightarrow X(\Lambda)$ from a theoretical point of view is that it implements Wiener causality via information-theoretic concepts in a direct manner, so its interpretation as a causality indicator is immediate. On the other hand, it only takes into account self-dependencies in $X$ and cross-dependencies from $Y$ that are effective $\Lambda$ time steps apart.

As for $\Delta TI_Y \rightarrow X(\Lambda)$, a clear practical advantage of this indicator is its lower dimensionality –one dimension less than $\Delta AT_Y \rightarrow X(\Lambda)$. Therefore it may happen when studying small datasets (i.e., $N/P \lesssim 10$, see Section 3.2) that the probabilities in $\Delta AT_Y \rightarrow X(\Lambda)$ are undersampled while the probabilities in $\Delta TI_Y \rightarrow X(\Lambda)$ are sufficiently well estimated, as it was the case in Section 4.2. A theoretical disadvantage of $\Delta TI_Y \rightarrow X(\Lambda)$, though, is that, unlike $\Delta AT_Y \rightarrow X(\Lambda)$, it lacks a simple translation in terms of predictability improvement (except when restrictions (19) and (23) hold).

Acknowledgments. We thank very much our referees for their constructive criticism.

Appendix
Proof of Theorem 3.1. Consider first the case $r = 1$. According to Eqs. (13) and (11) of Ref. [7],
\begin{align*}
AT_{Y \rightarrow X}(s,1)(\Lambda) - TI_{Y \rightarrow X}(s,1)(\Lambda) &= C(X_{n+\Lambda}, X_n, Y_n, ..., Y_{n-s+1}) - C(X_n, Y_n, ..., Y_{n-s+1}) \\
&- C(X_{n+\Lambda} | X_n)
\end{align*}

(26)
\[
\leq C(X_{n+A}, X_n, Y_{n-s+1}),
\]

where the last inequality follows from the non-negativity of the coupling complexity coefficients (lower bound in (2)). In turn, from the upper bound in (2) we derive
\[
AT_y^{(s,r)}(\Lambda) - TI_{Y \rightarrow X}^{(s,r)}(\Lambda) \leq \min\{H(X_n), H(Y_n)\}. \tag{27}
\]

Similarly, change signs in (26) and use again the bounds in (2) to derive
\[
TI_{Y \rightarrow X}^{(s,r)}(\Lambda) - AT_y^{(s,r)}(\Lambda) = C(X_{n+A}, Y_{n-s+1}) - C(X_{n+A}, Y_{n-s+1}) - C(X_{n+A}, Y_{n-s+1}) \tag{28}
\]

Inequalities (27) and (28) prove (15).

If \( r \geq 2 \), use Eqs. (12) and (11) of Ref. [7] to derive
\[
AT_y^{(s,r)}(\Lambda) - TI_{Y \rightarrow X}^{(s,r)}(\Lambda) = C(X_{n+A}, Y_{n-s+1}, Y_{n-s+1}) - C(X_{n+A}, Y_{n-s+1}, Y_{n-s+1}) - C(X_{n+A}, Y_{n-s+1}, Y_{n-s+1}) \tag{29}
\]

Similarly, by changing signs in (29),
\[
TI_{Y \rightarrow X}^{(s,r)}(\Lambda) - AT_y^{(s,r)}(\Lambda) = C(X_{n+A}, Y_{n-s+1}) - C(X_{n+A}, Y_{n-s+1}) \tag{30}
\]

Inequalities (29) and (30) prove (16). \( \square \)

**Proof of Theorem 3.3.** From Eq. (6) with \( r = s = 1 \), and Eq. (4) it follows
\[
\Delta AT_{Y \rightarrow X}(\Lambda) = I(X_{n+A}; Y_n | X_n) - I(Y_{n+A}; X_n | Y_n) = -H(X_n) + H(X_{n+A}, X_n) + H(Y_n, X_n) - H(X_{n+A}, Y_n, X_n) + H(Y_n) - H(Y_{n+A}, Y_n) - H(X_{n+A}, Y_n) + H(X_{n+A}, X_n, Y_n) = -H(X_n) + H(Y_n) + H(X_{n+A}, X_n) - H(Y_{n+A}, Y_n) - H(X_{n+A}, X_n, Y_n) + H(Y_{n+A}, Y_n, X_n).
\]

Use now the equivalence property, Eq. (5), to derive
\[
\Delta AT_{Y \rightarrow X}(\Lambda) = -H(X_n) + H(Y_n) + H(T_{x_{n+A}, x_n}, X_n) - H(T_{y_{n+A}, y_n}, X_n) - H(T_{x_{n+A}, x_n}, T_{x_n, y_n}, Y_n) + H(T_{y_{n+A}, y_n}, T_{y_n, x_n}, X_n).
\]
Since $H(X, Y) \leq H(X) + H(Y)$ and $H(X, Y) \geq H(X)$ (i.e., $-H(X, Y) \leq -H(X)$), we can upper bound $\Delta A T_{Y \rightarrow X}$ as follows:

$$
\Delta A T_{Y \rightarrow X}(\Lambda) \leq -H(X_n) + H(Y_n) + H(T_{x_{n+\Lambda}, x_n}) + H(X_n) - H(T_{y_{n+\Lambda}, y_n}) - H(T_{x_{n+\Lambda}, x_n} T_{x_n, y_n}) + H(T_{y_{n+\Lambda}, y_n} T_{y_n, x_n}) + H(X_n)
$$

$$
= H(X_n) + H(T_{x_{n+\Lambda}, x_n}) - H(T_{y_{n+\Lambda}, y_n}) - H(T_{x_{n+\Lambda}, x_n} T_{x_n, y_n}) + H(T_{y_{n+\Lambda}, y_n} T_{y_n, x_n}).
$$

On the other hand, see (22),

$$
\Delta T I_{Y \rightarrow X}(\Lambda) = T I_{Y \rightarrow X}(\Lambda) - T I_{X \rightarrow Y}(\Lambda)
$$

$$
= I(T_{x_{n+\Lambda}, x_n}, T_{x_n, y_n}) - I(T_{y_{n+\Lambda}, y_n}, T_{y_n, x_n})
$$

$$
= H(T_{x_{n+\Lambda}, x_n}) + H(T_{x_n, y_n}) - H(T_{x_{n+\Lambda}, x_n} T_{x_n, y_n}) - H(T_{y_{n+\Lambda}, y_n}) - H(T_{y_{n+\Lambda}, y_n} T_{y_n, x_n})
$$

$$
= H(T_{x_{n+\Lambda}, x_n}) + H(T_{x_n, y_n}) - H(T_{y_{n+\Lambda}, y_n}) - H(T_{x_{n+\Lambda}, x_n} T_{x_n, y_n}) + H(T_{y_{n+\Lambda}, y_n} T_{y_n, x_n}).
$$

Comparison of (31) and (32) yields then

$$
\Delta A T_{Y \rightarrow X}(\Lambda) \leq H(X_n) + H(Y_n) + \Delta T I_{Y \rightarrow X}(\Lambda).
$$

(33)

Interchange $X$ and $Y$ to obtain

$$
\Delta A T_{X \rightarrow Y}(\Lambda) \leq H(X_n) + H(Y_n) + \Delta T I_{X \rightarrow Y}(\Lambda)
$$

hence

$$
\Delta A T_{Y \rightarrow X}(\Lambda) = -\Delta A T_{X \rightarrow Y}(\Lambda) \geq -H(X_n) - H(Y_n) - \Delta T I_{X \rightarrow Y}(\Lambda)
$$

(34)

$$
= -H(X_n) - H(Y_n) + \Delta T I_{Y \rightarrow X}(\Lambda).
$$

Eq. (24) follows then from (33) and (34). □

REFERENCES

[1] J. M. Amigó and M. B. Kennel, Forbidden ordinal patterns in higher dimensional dynamics, *Physica D*, 237 (2008), 2893–2899.

[2] J. M. Amigó, S. Zambrano and M. A. F. Sanjuán. Detecting determinism with ordinal patterns: A comparative study, *Int. J. Bifurcation and Chaos*, 20 (2010), 2915–2924.

[3] J. M. Amigó, R. Monetti, T. Aschenbrenner and W. Bunk, Transcripts: An algebraic approach to coupled time series, *Chaos*, 22 (2012), 013105, 13pp.

[4] J. M. Amigó, T. Aschenbrenner, W. Bunk and R. Monetti, Dimensional reduction of conditional algebraic multi-information via transcripts, *Inform. Sci.*, 278 (2014), 298–310.

[5] J. M. Amigó, K. Keller and V. Unakafova, Ordinal symbolic analysis and its application to biomedical recordings, *Phil. Trans. R. Soc. A*, 373 (2015), 20140091, 18pp.

[6] J. M. Amigó, R. Monetti, N. Tort-Colet and M. V. Sanchezes-Vives, Infragranular layers lead information flow during slow oscillations according to information directionality indicators, *J. Comput. Neurosci.*, 39 (2015), 53–62.

[7] J. M. Amigó, R. Monetti, B. Graff and G. Graff, Computing algebraic transfer entropy and coupling directions via transcripts, *Chaos*, 26 (2016), 113115, 12pp.

[8] C. Bandt and B. Pompe, Permutation entropy: A natural complexity measure for time series, *Phys. Rev. Lett.*, 88 (2002), 174102.

[9] C. Cafaro, W. M. Lord, J. Sun and E. M. Bollt, Causation entropy from symbolic representations of dynamical systems, *Chaos*, 25 (2015), 043106, 10pp.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, (second ed.), John Wiley & Sons, Hoboken, 2006.

[11] European-Heart-Network, *Cardiovascular Disease Statistics*, 2017.
[12] G. Graff, B. Graff, A. Kaczkowska, D. Makowiec, J. M. Amigó, J. Piskorski, K. Narkiewicz and P. Guzik, Ordinal pattern statistics for the assessment of heart rate variability, *Eur. Phys. J. Special Topics*, **222** (2013), 525–534.

[13] B. Graff, G. Graff, D. Makowiec, A. Kaczkowska, D. Wejer, S. Budrejko, D. Kozłowski and K. Narkiewicz, Entropy measures in the assessment of heart rate variability in patients with cardiodepressive vasovagal syncope, *Entropy*, **17** (2015), 1007–1022.

[14] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37** (1969), 424–438.

[15] Y. Hirata, J. M. Amigó, Y. Matsuzaka, R. Yokota, H. Mushiake and K. Aihara, Detecting causality by combined use of multiple methods: Climate and brain examples, *PLoS One*, **11** (2016), e0158572.

[16] R. Monetti, W. Bunk, T. Aschenbrenner and F. Jamitzky, Characterizing synchronization in time series using information measures extracted from symbolic representations, *Phys. Rev. E*, **79** (2009), 046207.

[17] R. Monetti, J. M. Amigó, T. Aschenbrenner and W. Bunk, Permutation complexity of interacting dynamical systems, *Eur. Phys. J. Special Topics*, **222** (2013), 421–436.

[18] R. Monetti, W. Bunk, T. Aschenbrenner, S. Springer and J. M. Amigó, Information directionality in coupled time series using transcripts, *Phys. Rev. E*, **88** (2013), 022911.

[19] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths and N. Wessel, Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics, *Comput. Biol. Med.*, **42** (2012), 319–327.

[20] A. Porta, A. Catai, A. Takahashi, V. Magagnin, T. Bassani, E. Tobaldini, P. van de Borne and N. Montano, Causal relationships between heart period and systolic arterial pressure during graded head-up tilt, *Am. J. Physiol. Regul. Integr. Comp. Physiol*, **300** (2011), R378–R386.

[21] T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.*, **85** (2000), 461–464.

[22] D. Smirnov, Spurious causalities with transfer entropy, *Phys. Rev. E*, **87** (2013), 042917.

[23] M. Staniek and K. Lehnertz, Symbolic transfer entropy, *Phys. Rev. Lett.*, **100** (2008), 158101.

[24] N. Wiener, *Modern Mathematics for Engineers*, McGraw-Hill, New York, 1956.

Received March 2018; revised July 2018.

E-mail address: jm.amigo@umh.es
E-mail address: bgraff@gumed.edu.pl
E-mail address: grzegorz.graff@pg.edu.pl
E-mail address: r_monetti@gmail.com
E-mail address: katarzyna.tessmer@pg.edu.pl