

Article

Linearized Transfer Entropy for Continuous Second Order Systems

Jonathan M. Nichols ^{1,*}, Frank Bucholtz ¹ and Joe V. Michalowicz ^{2,†}

¹ U.S. Naval Research Laboratory, Optical Sciences Division, Washington DC 20375, USA;
E-Mail: frank.bucholtz@nrl.navy.mil

² U.S. Naval Research Laboratory, Optical Science Division, Washington DC 20375, USA

† Permanent Address: Global Strategies Group, Crofton, MD 21114, USA;
E-Mail: georgiamsa@yahoo.com

* Author to whom correspondence should be addressed; E-Mail: jonathan.nichols@nrl.navy.mil.

Received: 22 May 2013; in revised form: 5 July 2013 / Accepted: 18 July 2013 /

Published: 7 August 2013

Abstract: The transfer entropy has proven a useful measure of coupling among components of a dynamical system. This measure effectively captures the influence of one system component on the transition probabilities (dynamics) of another. The original motivation for the measure was to quantify such relationships among signals collected from a nonlinear system. However, we have found the transfer entropy to also be a useful concept in describing linear coupling among system components. In this work we derive the analytical transfer entropy for the response of coupled, second order linear systems driven with a Gaussian random process. The resulting expression is a function of the auto- and cross-correlation functions associated with the system response for different degrees-of-freedom. We show clearly that the interpretation of the transfer entropy as a measure of “information flow” is not always valid. In fact, in certain instances the “flow” can appear to switch directions simply by altering the degree of linear coupling. A safer way to view the transfer entropy is as a measure of the ability of a given system component to predict the dynamics of another.

Keywords: transfer entropy; joint entropy; coupling

1. Introduction

One of the biggest challenges in the modeling and analysis of dynamical systems is understanding coupling mechanisms among different system components. Whether one is studying coupling on a small scale (e.g., neurons in a biological system) or large scale (e.g. coupling among widely separated geographical locations due to climate), understanding the functional form, strength, and/or direction of the coupling between two or more system components is a non-trivial task. However, this understanding is necessary if we are to build accurate models of the coupled system and make predictions (our ultimate goal). Accurately assessing the functional form of the coupling is beyond the scope of this work. To do so would require positing various models for a particular coupled system and then testing the predictive power of those models against observed data. Rather, the focus here is on understanding the strength and direction of the coupling among two system components. This task can be accomplished by forming a general hypothesis about what it means for two system components to be coupled, and then testing that hypothesis against observation. It is in this framework that the transfer entropy is operated.

The transfer entropy (TE) is a scalar measure designed to capture both the magnitude and direction of coupling among two components of a dynamical system. This measure was posed initially for data described by discrete probability distributions [1] and was later extended to continuous random variables [2]. By construction, this measure quantifies a general definition of coupling that is appropriate for both linear and nonlinear systems. Moreover, TE is defined in such a way as to provide insight into the direction of the coupling (is component A driving component B or *vice-versa*?). Since its introduction, the TE has been applied to a diverse set of systems, including biological [1,3], chemical [4], economic [5], structural [6,7], and climate [8]. A number of papers in the Neurosciences also have focused on the TE as a useful way to draw inference about coupling [9–11]. In each case the TE provided information about the system that traditional linear measures of coupling (e.g., cross-correlation) could not.

The TE has also been linked to other concepts of coupling such as “Wiener-Granger Causality”. in fact, for the class of systems studied in this work the TE can be shown to be entirely equivalent to measures of Granger causality [12]. Linkages to other models and concepts of dynamical coupling such as conditional mutual information [13] and Dynamic Causal Modeling (DCM) [14], are also possible for certain special cases. The connectivity model assumed by DCM is fundamentally nonlinear (specifically bilinear), however as the degree of nonlinearity decreases the form of the DCM model approaches that of the model studied here.

Although the TE was designed as a way to gain insight into nonlinear system coupling, we have found the TE to be quite useful in the study of linear systems as well. In this special case, analytical expressions for the TE are possible and can be used to provide useful insight into the behavior of the TE. Furthermore, unlike in the general case, the linearized TE can be easily estimated from observed data. This work is therefore devoted to the understanding of TE as applied to coupled, driven linear systems. Specifically, we consider coupling among components of a general, second order linear structural system driven by a Gaussian random process. The particular model studied is used to describe numerous phenomena, including structural dynamics, electrical circuits, heat transfer, *etc.* [15]. As such, it presents an opportunity to better understand the properties of the TE for a broad class of dynamical

systems. Section 1 develops the general analytical expression for the TE in terms of the covariance matrices associated with different combinations of system response data. Section 2 specifies the general model under study and derives the TE for the model response data. Sections 3 and 4 present results and concluding remarks.

2. Mathematical Development

In what follows we assume that we have observed the signals $x_i(t_n)$, $i = 1 \cdots M$ as the output of a dynamical system and that we have sampled these signals at times t_n , $n = 1 \cdots N$. The system is assumed to be appropriately modeled as a mixture of deterministic and stochastic components, hence we choose to model each sampled value $x_i(t_n)$ as a random variable X_{in} . That is to say, for any particular observation time t_n we can define a function $P_{X_{in}}(x_i(t_n))$ that assigns a probability to the event that $X_{in} < x_i(t_n)$. We further assume that these are continuous random variables and that we may also define the probability density function (PDF) $p_{X_{in}}(x(t_n)) = dP_{X_{in}}/dx_n$.

The vector of random variables $\mathbf{X}_i \equiv (X_{i1}, X_{i2}, \dots, X_{iN})$ defines a random process and will be used to model the i^{th} signal $\mathbf{x}_i \equiv x_i(t_n)$, $n = 1 \cdots N$. Using this notation, we can also define the joint PDF $p_{\mathbf{X}_i}(\mathbf{x}_i)$ which specifies the probability of observing such a sequence. In this work we further assume that the random processes are strictly stationary, that is to say the joint PDF obeys $p_{\mathbf{X}_i}(x_i(t_1), x_i(t_2), \dots, x_i(t_N)) = p_{\mathbf{X}_i}(x_i(t_1 + \tau), x_i(t_2 + \tau), \dots, x_i(t_N + \tau))$ *i.e.* the joint PDF is invariant to a fixed temporal shift τ .

The joint probability density functions are models that predict the likelihood of observing a particular sequence of values. These same models can be extended to include dynamical effects by including conditional probability, $p_{X_{in}}(x_i(t_n)|x_i(t_{n-1}))$, which can be used to specify the probability of observing the value $x_i(t_n)$ given that we have already observed $x_i(t_{n-1})$. The idea that knowledge of past observations changes the likelihood of future events is certainly common in dynamical systems. A dynamical system whose output is a repeating sequence of 010101... is equally likely to be in state 0 or state 1 (probability 0.5) if the system is observed at a randomly chosen time. However, if we know the value at $t_1 = 0$ the value $t_2 = 1$ is known with probability 1. This concept lies at the heart of the \mathcal{P}^{th} order Markov model, which by definition obeys

$$\begin{aligned}
 p_{X_i}(x_i(t_{n+1})|x_i(t_n), x_i(t_{n-1}), x_i(t_{n-2}), \dots, x_i(t_{n-\mathcal{P}})) &= \\
 p_{X_i}(x_i(t_{n+1})|x_i(t_n), x_i(t_{n-1}), x_i(t_{n-2}), \dots, x_i(t_{n-\mathcal{P}}), x_i(t_{n-\mathcal{P}-1}), \dots) & \\
 \equiv p_{X_i}(x_i(t_n)^{(1)}|\mathbf{x}_i(t_n)^{(\mathcal{P})}). & \tag{1}
 \end{aligned}$$

That is to say, the probability of the random variable attaining the value $x_i(t_{n+1})$ is conditional on the previous \mathcal{P} values only. The shorthand notation used here specifies relative lags/advances as a superscript.

Armed with this notation we consider the work of Kaiser and Schreiber [2] and define the continuous transfer entropy between processes \mathbf{X}_i and \mathbf{X}_j as

$$TE_{j \rightarrow i}(t_n) = \int_{\mathbb{R}^{P+Q+1}} p_{\mathbf{X}_i} \left(x_i(t_n)^{(1)} | \mathbf{x}_i^{(P)}(t_n), \mathbf{x}_j^{(Q)}(t_n) \right) \times \log_2 \left(\frac{p_{\mathbf{X}_i}(x_i(t_n)^{(1)} | \mathbf{x}_i^{(P)}(t_n), \mathbf{x}_j^{(Q)}(t_n))}{p_{\mathbf{X}_i}(x_i(t_n)^{(1)} | \mathbf{x}_i^{(P)})} \right) dx_i(t_n)^{(1)} d\mathbf{x}_i(t_n)^{(P)} d\mathbf{x}_j(t_n)^{(Q)} \tag{2}$$

where $\int_{\mathbb{R}^N}$ is used to denote the N -dimensional integral over the support of the random variables. By definition, this measure quantifies the ability of the random process \mathbf{X}_j to predict the dynamics of the random process \mathbf{X}_i . To see why, we can examine the argument of the logarithm. In the event that the two random processes are *not* coupled, the dynamics will obey the Markov model in the denominator of Equation (2). However, should \mathbf{X}_j carry added information about the transition probabilities of \mathbf{X}_i , the numerator is a better model. The transfer entropy is effectively mapping the difference between these hypotheses to the scalar $TE_{j \rightarrow i}(t_n)$. In short, the transfer entropy measures deviations from the hypothesis that the dynamics of \mathbf{X}_i can be described entirely by its own past history and that no new information is gained by considering the dynamics of system \mathbf{X}_j .

Two simplifications are possible which will aid in the evaluation of Equation (2). First, recall that we assumed the processes were stationary such that the joint probability distributions are invariant to the particular temporal location t_n at which they are evaluated (only relative lags between observations matter). Hence, in what follows we may drop this index from the notation, *i.e.*, $TE_{j \rightarrow i}(t_n) \rightarrow TE_{j \rightarrow i}$. Secondly, we may use the law of conditional probability and expand Equation (2) as

$$\begin{aligned} TE_{j \rightarrow i} &= \int_{\mathbb{R}^{P+Q+1}} p_{\mathbf{X}_i^{(1)} \mathbf{X}_i \mathbf{X}_j} \left(x_i^{(1)}, \mathbf{x}_i^{(P)}, \mathbf{x}_j^{(Q)} \right) \log_2 \left(p_{\mathbf{X}_i^{(1)} \mathbf{X}_i \mathbf{X}_j} \left(x_i^{(1)}, \mathbf{x}_i^{(P)}, \mathbf{x}_j^{(Q)} \right) \right) \\ &\quad \times dx_i^{(1)} d\mathbf{x}_i^{(P)} d\mathbf{x}_j^{(Q)} \\ &- \int_{\mathbb{R}^{P+Q}} p_{\mathbf{X}_i \mathbf{X}_j} \left(\mathbf{x}_i^{(P)}, \mathbf{x}_j^{(Q)} \right) \log_2 \left(p_{\mathbf{X}_i \mathbf{X}_j} \left(\mathbf{x}_i^{(P)}, \mathbf{x}_j^{(Q)} \right) \right) d\mathbf{x}_i^{(P)} d\mathbf{x}_j^{(Q)} \\ &- \int_{\mathbb{R}^{P+1}} p_{\mathbf{X}_i^{(1)} \mathbf{X}_i} \left(x_i^{(1)}, \mathbf{x}_i^{(P)} \right) \log_2 \left(p_{\mathbf{X}_i^{(1)} \mathbf{X}_i} \left(x_i^{(1)}, \mathbf{x}_i^{(P)} \right) \right) dx_i^{(1)} d\mathbf{x}_i^{(P)} \\ &+ \int_{\mathbb{R}^P} p_{\mathbf{X}_i} \left(\mathbf{x}_i^{(P)} \right) \log_2 \left(p_{\mathbf{X}_i} \left(\mathbf{x}_i^{(P)} \right) \right) d\mathbf{x}_i^{(P)} \\ &= -h_{\mathbf{X}_i^{(1)} \mathbf{X}_i^{(P)} \mathbf{X}_j^{(Q)}} + h_{\mathbf{X}_i^{(P)} \mathbf{X}_j^{(Q)}} + h_{\mathbf{X}_i^{(1)} \mathbf{X}_i^{(P)}} - h_{\mathbf{X}_i^{(P)}} \end{aligned} \tag{3}$$

where the terms $h_{\mathbf{X}} = - \int_{\mathbb{R}^M} p_{\mathbf{X}}(\mathbf{x}) \log_2(p(\mathbf{x})) d\mathbf{x}$ are the joint differential entropies associated with the M -dimensional random variable \mathbf{X} . In the next section we evaluate Equation (3) among the outputs of a second-order linear system driven with a jointly Gaussian random process.

3. Transfer Entropy (TE) for Second Order Linear Systems

3.1. Time-Delayed TE

The only multivariate probability distribution that readily admits an analytical solution for the differential entropies is the jointly Gaussian distribution. Consider the general case of the two data vectors $\mathbf{x} \in \mathbb{R}^N$ and the $\mathbf{y} \in \mathbb{R}^M$. The jointly Gaussian model for these data vectors is

$$p_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{(N+M)/2} |C_{\mathbf{XY}}|^{1/2}} e^{-\frac{1}{2} \mathbf{x}^T C_{\mathbf{XY}}^{-1} \mathbf{y}} \tag{4}$$

where $C_{\mathbf{XY}}$ is the $N \times M$ covariance matrix and $|\cdot|$ takes the determinant. Substituting Equation (4) into the expression for the corresponding differential entropy yields

$$\begin{aligned} h_{\mathbf{XY}} &= - \int_{\mathbb{R}^{M \times N}} p_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) \log_2(p(\mathbf{x}, \mathbf{y})) \, dx dy \\ &= \frac{1}{2} \log_2(|C_{\mathbf{XY}}|). \end{aligned} \tag{5}$$

Therefore, assuming that both random processes \mathbf{X}_i and \mathbf{X}_j are jointly Gaussian distributed, we may substitute Equation (4) into Equation (3) for each of the differential entropies yielding

$$TE_{j \rightarrow i} = \frac{1}{2} \log_2 \left(\frac{|C_{\mathbf{X}_i^{(P)} \mathbf{X}_j^{(Q)}}| |C_{\mathbf{X}_i^{(1)} \mathbf{X}_i^{(P)}}|}{|C_{\mathbf{X}_i^{(1)} \mathbf{X}_i^{(P)} \mathbf{X}_j^{(Q)}}| |C_{\mathbf{X}_i}|} \right). \tag{6}$$

For P, Q large the needed determinants become difficult to compute. We therefore employ a simplification to the model that retains the spirit of the transfer entropy, but that makes an analytical solution more tractable. In our approach, we set $P = Q = 1$ *i.e.*, both random processes are assumed to follow a first order Markov model. However, we allow the time interval between the random processes to vary, just as is typically done for the mutual information and/or linear cross-correlation functions [6]. Specifically, we model $X_i(t)$ as the first order Markov model $p_{X_i}(x_i(t_n + \Delta_t) | x_i(t_n))$ and use the TE to consider the alternative $p_{X_i}(x_i(t_n + \Delta_t) | x_i(t_n), x_j(t_n + \tau))$. Note that in anticipation of dealing with measured data, sampled at constant time interval Δ_t , we have made the replacement $t_{n+1} = t_n + \Delta_t$. Although we are only using first order Markov models, by varying the time delay τ we can explore whether or not the random variable $X_j(t_n + \tau)$ carries information about the transition probability $p_{X_i}(x_i(t_n + \Delta_t) | x_i(t_n))$. Should consideration of $x_j(t_n + \tau)$ provide no additional knowledge about the dynamics of $x_i(t_n)$ the transfer entropy will be zero, rising to some positive value should $x_j(t_n + \tau)$ carry information not possessed in $x_j(t_n)$.

In what follows we refer to this particular form of the TE as the time-delayed transfer entropy, or, TDTE. In this simplified situation the needed covariance matrices are

$$\begin{aligned}
 C_{X_i X_j}(\tau) &= \begin{bmatrix} E[(x_i(t_n) - \bar{x}_i)^2] & E[(x_i(t_n) - \bar{x}_i)(x_j(t_n + \tau) - \bar{x}_j)] \\ E[(x_j(t_n + \tau) - \bar{x}_j)(x_i(t_n) - \bar{x}_i)] & E[(x_j(t_n + \tau) - \bar{x}_j)^2] \end{bmatrix} \\
 C_{X_i^{(1)} X_i X_j}(\tau) &= \begin{bmatrix} E[(x_i(t_n + \Delta_t) - \bar{x}_i)^2] & E[(x_i(t_n + \Delta_t) - \bar{x}_i)(x_i(t_n) - \bar{x}_i)] \\ E[(x_i(t_n) - \bar{x}_i)(x_i(t_n + \Delta_t) - \bar{x}_i)] & E[(x_i(t_n) - \bar{x}_i)^2] \\ E[(x_j(t_n + \tau) - \bar{x}_j)(x_i(t_n + \Delta_t) - \bar{x}_i)] & E[(x_j(t_n + \tau) - \bar{x}_j)(x_i(t_n) - \bar{x}_i)] \\ E[(x_i(t_n + \Delta_t) - \bar{x}_i)(x_j(t_n + \tau) - \bar{x}_j)] & \\ E[(x_i(t_n) - \bar{x}_i)(x_j(t_n + \tau) - \bar{x}_j)] & \\ E[(x_j(t_n + \tau) - \bar{x}_j)^2] & \end{bmatrix} \\
 C_{X_i^{(1)} X_i} &= \begin{bmatrix} E[(x_i(t_n + \Delta_t) - \bar{x}_i)^2] & E[(x_i(t_n + \Delta_t) - \bar{x}_i)(x_i(t_n) - \bar{x}_i)] \\ E[(x_i(t_n) - \bar{x}_i)(x_i(t_n + \Delta_t) - \bar{x}_i)] & E[(x_i(t_n) - \bar{x}_i)^2] \end{bmatrix} \tag{7}
 \end{aligned}$$

and $C_{X_i X_i} = E[(x_i(t_n) - \bar{x}_i)^2] \equiv \sigma_i^2$ is simply the variance of the random process X_i and \bar{x}_i its mean. The assumption of stationarity also allows to write $E[(x_i(t_n + \Delta_t) - \bar{x}_i)^2] = \sigma_i^2$ and $E[(x_j(t_n + \tau) - \bar{x}_j)^2] = \sigma_j^2$. Making these substitutions into Equation (6) yields the expression

$$TE_{j \rightarrow i}(\tau) = \frac{1}{2} \log_2 \left[\frac{(1 - \rho_{ii}^2(\Delta_t)) (1 - \rho_{ij}^2(\tau))}{1 - \rho_{ij}^2(\tau) - \rho_{ij}^2(\tau - \Delta_t) - \rho_{ii}^2(\Delta_t) + 2\rho_{ii}(\Delta_t)\rho_{ij}(\tau)\rho_{ij}(\tau - \Delta_t)} \right] \tag{8}$$

where we have defined particular expectations in the covariance matrices using the shorthand $\rho_{ij}(\tau) \equiv E[(x_i(t_n) - \bar{x}_i)(x_j(t_n + \tau) - \bar{x}_j)]/\sigma_i\sigma_j$. This particular quantity is referred to in the literature as the cross-correlation function [16]. Note that the covariance matrices are positive-definite matrices and that the determinant of a positive definite matrix is positive [17]. Thus the quantity inside the logarithm will always be positive and the logarithm will exist.

Now, the hypothesis that the TE was designed to test is whether or not past values of the process X_j carry information about the transition probabilities of the second process X_i . Thus, if we are to keep with the original intent of the measure we would only consider $\tau < 0$. However, this restriction is only necessary if one implicitly assumes a non-zero TE means X_j is *influencing* the transition $p_{X_i}(x_i(t_n + \Delta_t)|x_i(t_n))$ as opposed to simply carrying additional information *about* the transition. Again, this latter statement is a more accurate depiction of what the TE is really quantifying and we have found it useful to consider both negative and positive delays τ in trying to understand coupling among system components.

It is also interesting to note the bounds of this function. Certainly for constant signals (*i.e.* $x_i(t_n), x_j(t_n)$ are single-valued for all time) we have $\rho_{X_i X_i}(\Delta_t) = \rho_{X_i X_j}(\tau) = 0 \forall \tau$ and the transfer entropy is zero for any choice of time-scales τ defining the Markov processes. Knowledge of X_j does not aid in forecasting X_i simply because the transition probability in going from $x_i(t_n)$ to $x_i(t_n + \Delta_t)$ is always unity. Likewise, if there is no coupling between system components we have $\rho_{X_i X_j}(\tau) = 0$ and the TDTE becomes $TE_{j \rightarrow i}(\tau) = \frac{1}{2} \log_2 \left[\frac{1 - \rho_{X_i X_i}^2(\Delta_t)}{1 - \rho_{X_i X_i}^2(\Delta_t)} \right] = 0$. At the other extreme, for *perfectly* coupled systems *i.e.* $X_i = X_j$, consider $\tau \rightarrow 0$. In this case, we have $\rho_{X_i X_j}^2(\tau) \rightarrow 1$, and $\rho_{X_i X_j}(\tau - \Delta_t) \rightarrow \rho_{X_i X_i}(-\Delta_t) = \rho_{X_i X_i}(\Delta_t)$ (in this last expression we have noted the symmetry of the function $\rho_{X_i X_i}(\tau)$ with respect to the time-delay). The transfer entropy then becomes

$$TE_{j \rightarrow i}(0) = \frac{1}{2} \log_2 \left[\frac{0}{0} \right] \rightarrow 0 \tag{9}$$

and the random process X_j at $\tau = 0$ is seen to carry no *additional* information about the dynamics of X_i simply due to the fact that in this special case we have $p_{X_i}(x_i(t_n + \Delta_t)|x_i(t_n)) = p_{X_i}(x_i(t_n + \Delta_t)|x_i(t_n), x_i(t_n))$. These extremes highlight the care that must be taken in interpreting the transfer entropy. Because the TDTE is zero for both the perfectly coupled and uncoupled case we must not interpret the measure to quantify the coupling strength between two random processes. Rather, the TDTE measures the additional information provided by one random process about the dynamics of another.

We should point out that the average mutual information function can resolve the ambiguity in the TDTE as a measure of coupling strength. For two Gaussian random processes the time-delayed mutual information is known to be $I_{X_i X_j}(\tau) = -\frac{1}{2} \log_2 [1 - \rho_{ij}^2(\tau)]$. Hence, for perfect coupling $I_{X_i X_j}(0) \rightarrow \infty$ whereas for uncoupled systems $I_{X_i X_j}(0) \rightarrow 0$. Estimating both time-delayed mutual information and transfer entropies can therefore permit stronger inference about dynamical coupling.

3.2. Analytical Cross-Correlation Function

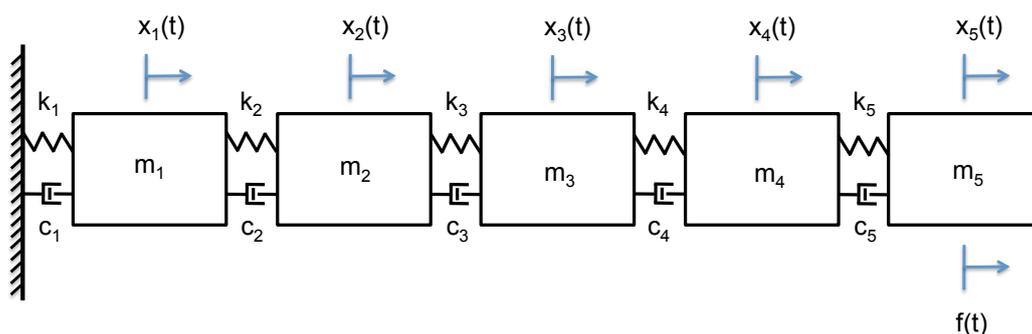
To fully define the TDTE, the auto- and cross-correlation functions $\rho_{ii}(T)$, $\rho_{ij}(T)$ are required. They are derived here for a general class of linear system found frequently in the modeling and analysis of physical processes. Consider the system

$$M\ddot{\mathbf{x}}(t) + C\dot{\mathbf{x}}(t) + K\mathbf{x}(t) = \mathbf{f}(t) \tag{10}$$

where $\mathbf{x}(t) \equiv (x_1(t), x_2(t), \dots, x_M(t))^T$ is the system’s response to the forcing function(s) $\mathbf{f}(t) \equiv (f_1(t), f_2(t), \dots, f_M(t))^T$ and M , C , K are $M \times M$ constant coefficient matrices that capture the system’s physical properties. Thus, we are considering a second-order, constant coefficient, M -degree-of-freedom (DOF) linear system. It is assumed that we may measure the response of this system at any of the DOFs and/or the forcing functions.

One physical embodiment of this system is shown schematically in Figure 1. Five masses are coupled together via restoring elements k_i (springs) and dissipative elements, c_i (dash-pots). The first mass is fixed to a boundary while the driving force is applied at the end mass. If the response data $\mathbf{x}(t)$ are each modeled as a stationary random process we may use the analytical TDTE to answer questions about shared information between any two masses. We can explore this relationship as a function of coupling strength and also which particular mass response data we choose to analyze.

Figure 1. Physical system modeled by Equation (10). Here, an $M = 5$ DOF structure is represented by masses coupled together via both restoring and dissipative elements. Forcing is applied at the end mass.



However, before proceeding we require a general expression for the cross-correlation between any two DOFs, $i, j \in [1, M]$. In other words, we require the expectation $E[x_i(n)x_j(n+T)]$ for any combination of i, j . Such an expression can be obtained by first transforming coordinates. Let $\mathbf{x}(t) = \mathbf{u}\eta(t)$ where the matrix \mathbf{u} contain the non-trivial solutions to the eigen-value problem $|\mathbf{M}^{-1}\mathbf{K} - \omega_i^2\mathbf{I}|\mathbf{u}_i = 0$ as its columns [18]. Here the eigen-values are the natural frequencies of the system, denoted $\omega_i, i = 1 \cdots M$. Making the above coordinate transformation, substituting into Equation (10) and then pre-multiplying both sides by \mathbf{u}^T allows the equations of motion to be uncoupled and written separately as

$$\ddot{\eta}_i(t) + 2\zeta_i\omega_i\dot{\eta}_i(t) + \omega_i^2\eta_i(t) = \mathbf{u}_i^T \mathbf{f}(t) \equiv q_i(t). \tag{11}$$

where the eigenvectors have been normalized such that $\mathbf{u}^T\mathbf{M}\mathbf{u} = \mathbf{I}$ (the identity matrix). In the above formulation we have also made the assumption that $\mathbf{C} = \alpha\mathbf{K}$ *i.e.*, the dissipative coupling $\mathbf{C}\dot{\mathbf{x}}(t)$ is of the same form as the restoring term, albeit scaled by the constant $\alpha \ll 1$ (*i.e.*, a lightly damped system). To obtain the form shown in Equation (11) we introduce the dimensionless damping coefficient $\zeta_i = \frac{\alpha}{2}\omega_i$.

The general solution to these un-coupled, linear equations is well-known [18] and can be written as the convolution

$$\eta_i(t) = \int_0^\infty h_i(\theta)q_i(t - \theta)d\theta \tag{12}$$

where $h(\theta)$ is the impulse response function

$$h_i(\theta) = \frac{1}{\omega_{di}}e^{-\zeta_i\omega_i\theta} \sin(\omega_{di}\theta) \tag{13}$$

and $\omega_{di} \equiv \omega_i\sqrt{1 - \zeta_i^2}$. In general terms, we therefore have

$$\begin{aligned} x_i(t) &= \sum_{l=1}^M u_{il}\eta_l(t) \\ &= \int_0^\infty \sum_{l=1}^M u_{il}h_l(\theta)q_l(t - \theta)d\theta \end{aligned} \tag{14}$$

If we further consider the excitation $\mathbf{f}(t)$ to be a zero-mean random process, so too will be $q_l(t)$. Using this model, we may construct the covariance

$$\begin{aligned} E[x_i(t)x_j(t + \tau)] &= \\ E \left[\int_0^\infty \int_0^\infty \sum_{l=1}^M \sum_{m=1}^M u_{il}u_{jm}h_l(\theta_1)h_m(\theta_2)q_l(t - \theta_1)q_m(t + \tau - \theta_2)d\theta_1d\theta_2 \right] \\ &= \int_0^\infty \int_0^\infty \sum_{l=1}^M \sum_{m=1}^M u_{il}u_{jm}h_l(\theta_1)h_m(\theta_2)E[q_l(t - \theta_1)q_m(t + \tau - \theta_2)]d\theta_1d\theta_2 \end{aligned} \tag{15}$$

which is a function of the eigen-vectors \mathbf{u}_i , the impulse response function $h(\cdot)$ and the covariance of the modal forcing matrix. Knowledge of this covariance matrix can be obtained from knowledge of the forcing covariance matrix $R_{F_l F_m}(\tau) \equiv E[f_l(t)f_m(t + \tau)]$. Recalling that

$$q_l(t) = \sum_{p=1}^M u_{lp}f_p(t) \tag{16}$$

we write

$$E[q_l(t - \theta_1)q_m(t + \tau - \theta_2)] = \sum_{p=1}^M \sum_{q=1}^M u_{lp}u_{mq}E[f_q(t - \theta_1)f_p(t + \tau - \theta_2)] \tag{17}$$

It is assumed that the random vibration inputs are uncorrelated, *i.e.* $E[f_q(t)f_p(t)] = 0 \forall q \neq p$, with variance $\sigma_{F_p}^2 = E[f_p(t)f_p(t)]$. Thus, the above can therefore be simplified as

$$E[q_l(t - \theta_1)q_m(t + \tau - \theta_2)] = \sum_{p=1}^M u_{lp}u_{mp}E[f_p(t - \theta_1)f_p(t + \tau - \theta_2)] \tag{18}$$

The most common linear models assume the input is applied at a single DOF, *i.e.* $f_p(t)$ is non-zero only for $p = P$. For a load applied at DOF P , the auto-covariance becomes

$$\begin{aligned} E[x_i(t)x_j(t + \tau)] &= \int_0^\infty \int_0^\infty \sum_{l=1}^M \sum_{m=1}^M u_{il}u_{jm}u_{lP}u_{mP}h_l(\theta_1)h_m(\theta_2)E[f_P(t - \theta_1)f_P(t + \tau - \theta_2)]d\theta_1d\theta_2 \\ &= \sum_{l=1}^M \sum_{m=1}^M u_{il}u_{jm}u_{lP}u_{mP} \int_0^\infty h_l(\theta_1) \int_0^\infty h_m(\theta_2)E[f_P(t - \theta_1)f_P(t + \tau - \theta_2)]d\theta_2d\theta_1. \end{aligned} \tag{19}$$

The inner integral can be further evaluated as

$$\int_0^\infty h_m(\theta_2)E[f_P(t - \theta_1)f_P(t + \tau - \theta_2)]d\theta_2 = \int_0^\infty h_m(\theta_2) \int_{-\infty}^\infty S_{FF}(\omega)e^{i\omega(\tau - \theta_2 + \theta_1)}d\omega d\theta_2. \tag{20}$$

Note that we have re-written the forcing auto-covariance as the inverse Fourier transform of the associated power spectral density function, denoted $S_{FF}(\omega)$, via the well-known Wiener-Khinchine relation [16]. We have already assumed the forcing is comprised of independent, identically distributed values, in which case the forcing power spectral density $S_{FF}(\omega) = const \forall \omega$. Denoting this constant $S_{FF}(0)$, we note that the Fourier Transform of a constant is simply $\int_{-\infty}^\infty S_{FF}(0) \times e^{i\omega t} dt = S_{FF}(0) \times \delta(t)$, hence our integral becomes

$$\begin{aligned} &\int_0^\infty h_m(\theta_2)E[f_P(t - \theta_1)f_P(t + \tau - \theta_2)]d\theta_2 \\ &= \int_0^\infty h_m(\theta_2)S_{FF}(0)\delta(\tau - \theta_2 + \theta_1)d\theta_2 = h(\tau + \theta_1)S_{FF}(0). \end{aligned} \tag{21}$$

Returning to Equation (19) we have

$$E[x_i(t)x_j(t + \tau)] = \int_0^t \sum_{l=1}^M \sum_{m=1}^M u_{lP}u_{mP}u_{il}u_{jm}h_l(\theta_1)h_m(\theta_1 + \tau)S_{FF}(0)d\theta_1. \tag{22}$$

At this point we can simplify the expression by carrying out the integral.

Substituting the expression for the impulse response in Equation (13), the needed expectation in Equation (22) becomes [19,20]

$$R_{X_iX_j}(\tau) = \frac{S_{FF}(0)}{4} \sum_{l=1}^M \sum_{m=1}^M u_{lP}u_{mP}u_{il}u_{jm} [A_{lm}e^{-\zeta_m\omega_m\tau} \cos(\omega_{dm}\tau) + B_{lm}e^{-\zeta_m\omega_m\tau} \sin(\omega_{dm}\tau)] \tag{23}$$

where

$$\begin{aligned}
 A_{lm} &= \frac{8(\omega_l \zeta_l + \omega_m \zeta_m)}{\omega_l^4 + \omega_m^4 + 4\omega_l^3 \omega_m \zeta_l \zeta_m + 4\omega_m^3 \omega_l \zeta_l \zeta_m + 2\omega_m^2 \omega_l^2 (-1 + 2\zeta_l^2 + 2\zeta_m^2)} \\
 B_{lm} &= \frac{4(\omega_l^2 + 2\omega_l \omega_m \zeta_l \zeta_m + \omega_m^2 (-1 + 2\zeta_m^2))}{\omega_{dm}(\omega_l^4 + \omega_m^4 + 4\omega_l^3 \omega_m \zeta_l \zeta_m + 4\omega_m^3 \omega_l \zeta_l \zeta_m + 2\omega_m^2 \omega_l^2 (-1 + 2\zeta_l^2 + 2\zeta_m^2))}
 \end{aligned}
 \tag{24}$$

We can further normalize this function to give

$$\rho_{ij}(\tau) = R_{ij}(\tau) / \sqrt{R_{ii}(0)R_{jj}(0)}
 \tag{25}$$

for the normalized auto- and cross-correlation functions.

It will also prove instructive to study the TDTE between the drive and response. This requires $R_{X_i F_P}(\tau) \equiv E[x_i(t)f_P(t + \tau)]$. Following the same procedure as above results in the expression

$$R_{X_i F_P}(\tau) = \begin{cases} \mathbf{S}_{FF}(0) \sum_{m=1}^M u_{im} u_{mP} h_m(-\tau) & : \tau \leq 0 \\ 0 & : \tau > 0 \end{cases}
 \tag{26}$$

Normalizing by the variance of the random process X_i and assuming $\sigma_{F_P}^2 = 1$ yields the needed correlation function $\rho_{if}(\tau)$. This expression may be substituted into the expression for the transfer entropy to yield the TDTE between drive and response. At this point we have completely defined the analytical TDTE for a broad class of second order linear systems. The behavior of this function is described next. Before concluding this section we note that it also may be possible to derive expressions for the TDTE for different types of forcing functions. Impulse excitation and also non-Gaussian inputs where the marginal PDF can be described as a polynomial transformation of a Gaussian random variable (see e.g., [21]) are two such possibilities.

4. Behavior of the TDTE

Before proceeding with an example, we first require a means of estimating $TE_{j \rightarrow i}(\tau)$ from observed data. Assume we have recorded the signals $x_i(n\Delta_t)$, $x_j(n\Delta_t)$, $n = 1 \dots N$ with a fixed sampling interval Δ_t . In order to estimate the TDTE we require a means of estimating the normalized correlation functions $\rho_{ij}(\tau)$ which can be substituted into Equation (8). While different estimators of correlation functions exist (see e.g., [16]), we use a frequency domain estimator. This estimator relies on the assumption that the observed data are the output of an ergodic (therefore stationary) random process. If we further assume that the correlation functions are absolute integrable, e.g., $\int |R_{ij}(\tau)| d\tau < \infty$, the Wiener-Khinchin Theorem tells us that the cross-spectral density and cross-covariance functions are related via Fourier transform as [16].

$$\int_{-\infty}^{\infty} E[x_i(t)x_j(t + \tau)]e^{-i2\pi f\tau} d\tau = S_{X_j X_i}(f) \equiv \lim_{T \rightarrow \infty} E \left[\frac{X_i^*(f)X_j(f)}{2T} \right].
 \tag{27}$$

where $X_i(f)$ denotes the Fourier transform of the signal $x_i(t)$. One approach is to therefore estimate the spectral density $\hat{S}_{X_j X_i}(f)$ and then inverse Fourier transform to give $\hat{R}_{X_i X_j}(\tau)$. We further rely on the ergodic theorem of Birkhoff ([22]) which (when applied to probability) allows one to write expectations defined over multiple realizations to be well-approximated temporally averaging over a finite number

of samples. More specifically, we divide the temporal sequences $x_i(n)$, $x_j(n)$, $n = 1 \cdots N$ into S segments of length N_s (possibly) overlapping by L points. Taking the discrete Fourier transform of each segment, e.g., $X_{is}(k) = \sum_{n=0}^{N_s-1} x_i(n + sN_s - L)e^{-i2\pi kn/N_s}$, $s = 0 \cdots S - 1$ and averaging gives the estimator

$$\hat{S}_{X_j X_i}(k) = \frac{\Delta t}{N_s S} \sum_{s=0}^{S-1} \hat{X}_{is}^*(k) \hat{X}_{js}(k) \tag{28}$$

at discrete frequency k . This quantity is then inverse discrete Fourier transformed to give

$$\hat{R}_{X_i X_j}(n) = \sum_{k=0}^{N_s-1} \hat{S}_{X_j X_i}(k) e^{i2\pi kn/S}. \tag{29}$$

Finally, we may normalize the estimate to give the cross-correlation coefficient

$$\hat{\rho}_{X_i X_j}(n) = \hat{R}_{X_i X_j}(n) / \sqrt{\hat{R}_{X_i X_i}(0) \hat{R}_{X_j X_j}(0)}. \tag{30}$$

This estimator is asymptotically consistent and unbiased and can therefore be substituted into Equation (8) to produce very accurate estimates of the TE (see examples to follow). In the general (nonlinear) case, kernel density estimators are typically used but are known to be poor in many cases, particularly when data are scarce (see e.g., [6,23]). We also point out that for this study stationarity (and ergodicity) only up to second order (covariance) is required. In general the TDTE is a function of all joint moments hence higher-order ergodicity must be assumed.

As an example, consider a five-DOF system governed by Equation (10), where:

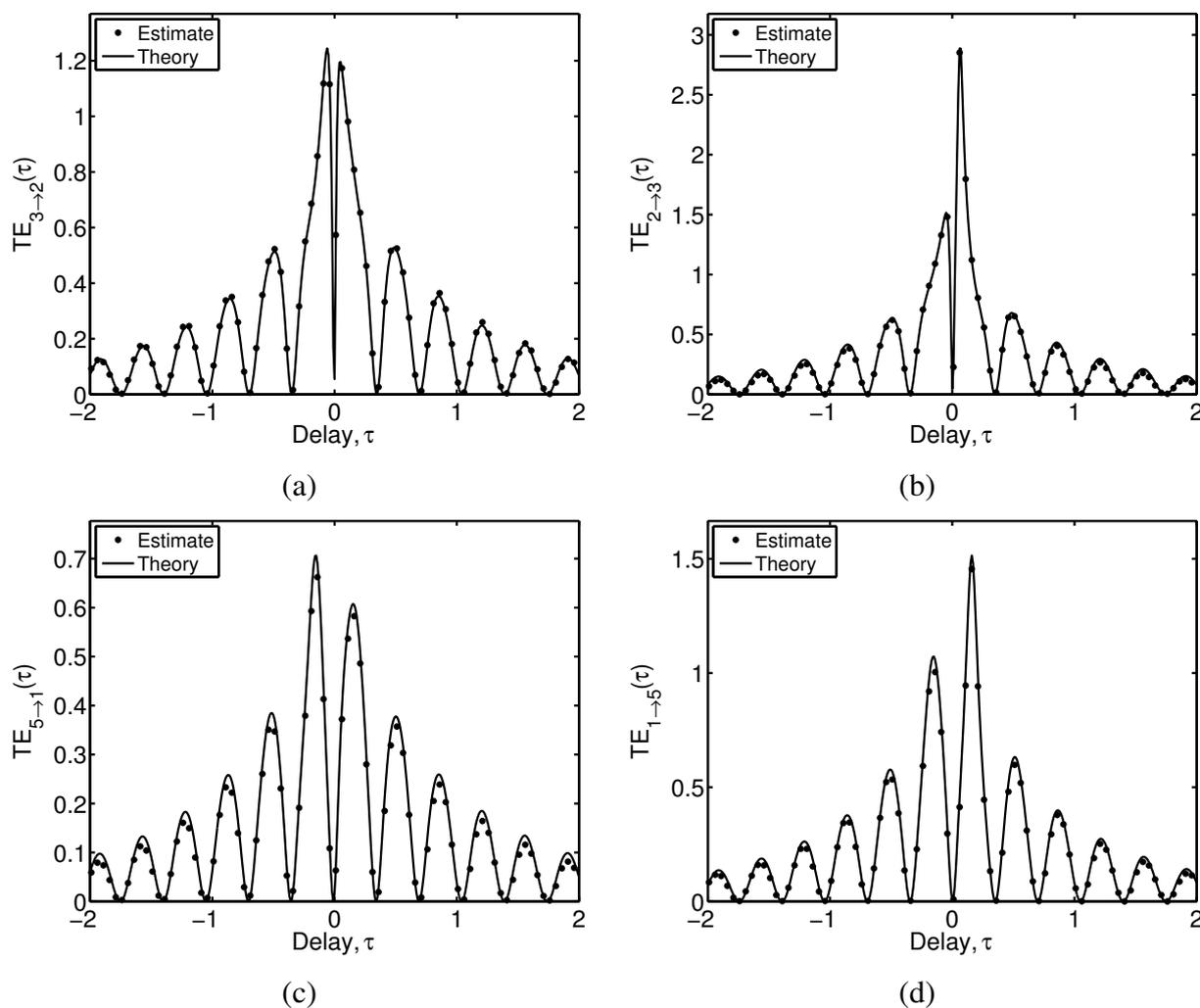
$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} m_1 & 0 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 & 0 \\ 0 & 0 & m_3 & 0 & 0 \\ 0 & 0 & 0 & m_4 & 0 \\ 0 & 0 & 0 & 0 & m_5 \end{bmatrix} \\ \mathbf{C} &= \begin{bmatrix} c_1 + c_2 & -c_2 & 0 & 0 & 0 \\ -c_2 & c_2 + c_3 & -c_3 & 0 & 0 \\ 0 & -c_3 & c_3 + c_4 & -c_4 & 0 \\ 0 & 0 & -c_4 & c_4 + c_5 & -c_5 \\ 0 & 0 & 0 & -c_5 & c_5 \end{bmatrix} \\ \mathbf{K} &= \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & 0 & 0 \\ -k_2 & k_2 + k_3 & -k_3 & 0 & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 & 0 \\ 0 & 0 & -k_4 & k_4 + k_5 & -k_5 \\ 0 & 0 & 0 & -k_5 & k_5 \end{bmatrix} \end{aligned} \tag{31}$$

are constant coefficient matrices commonly used to describe structural systems. In this case, these particular matrices describe the motion of a cantilevered structure where we assume a joint normally

distributed random process applied at the end mass, *i.e.* $\mathbf{f}(t) = (0, 0, 0, 0, \mathcal{N}(0, 1))$. In this first example we examine the TDTE between response data collected from two different points on the structure. We fix $m_i = 0.01 \text{ kg}$, $c_i = 0.1 \text{ N} \cdot \text{s}/\text{m}$, and $k_i = 10 \text{ N}/\text{m}$ for each of the $i = 1 \dots 5$ degrees of freedom (thus we are using $\alpha = 0.01$ in the modal damping model $\mathbf{C} = \alpha \mathbf{K}$). The system response data $x_i(n\Delta_t)$, $n = 1 \dots 2^{15}$ to the stochastic forcing is then generated via numerical integration. For simulation purposes we used a time-step of $\Delta_t = 0.01 \text{ s}$ which is sufficient to capture all five of the system natural frequencies (the lowest of which is $\omega_1 = 9.00 \text{ rad/s}$). Based on these parameters, we generated the analytical expressions $TE_{3 \rightarrow 2}(\tau)$ and $TE_{2 \rightarrow 3}(\tau)$ and also $TE_{5 \rightarrow 1}(\tau)$ and $TE_{1 \rightarrow 5}(\tau)$ for illustrative purposes. These are shown in Figure 2 along with the estimates formed using the Fourier transform-based procedure. In forming the estimates we used $L = 0$, $S = 2^3$, $N_s = 2^{12}$, resulting in low bias and variance, and providing curves that are in very close agreement with theory.

With Figure 2 in mind, first consider negative delays only where $\tau < 0$. Clearly, the further the random variable $X_j(t_n + \tau)$ is from $X_i(t_n)$, the less information it carries about the probability of X_i transitioning to a new state Δ_t seconds into the future. This is to be expected from a stochastically driven system and accounts for the decay of the transfer entropy to zero for large $|\tau|$. However, we also see periodic returns to the point $TE_{j \rightarrow i}(\tau) = 0$ for even small temporal separation. Clearly this is a reflection of the periodicity observed in second order linear systems. In fact, for this system the dominant period of oscillation is $2\pi/\omega_1 = 0.698$ seconds. It can be seen that the argument of the logarithm in Equation (8) periodically reaches a minimum value of unity at precisely half this period, thus we observe zeros of the TDTE at times $(i-1) \times \pi/\omega_1$, $i = 1 \dots$. In this case the TDTE is going to zero *not* because the random variables $X_j(t_n + \tau)$, $X_i(t_n)$ are unrelated, but because knowledge of one allows us to exactly predict the position of the other (no additional information is present). We believe this is likely to be a feature of most systems possessing an underlying periodicity and is one reason why using the TE as a measure of coupling must be done with care.

Figure 2. Time delay transfer entropy between masses two and three (top row) and one and five (bottom row) of a 5 DOF system driven at mass, $P = 5$.

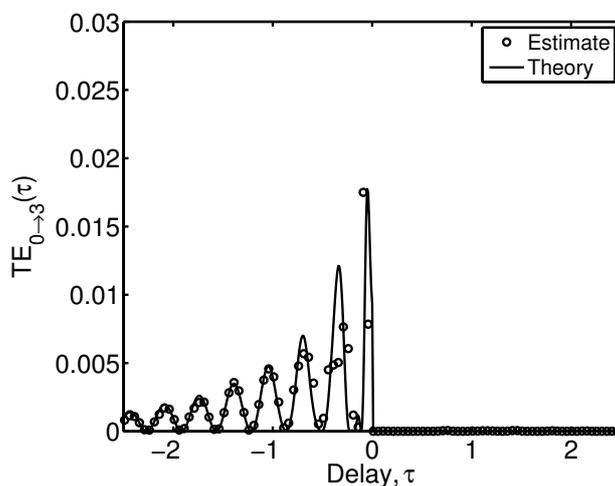


One possible way to eliminate this feature is to condition the measure on more of the signal's past history. In fact, several papers (see e.g., [9,13]) mention the importance of conditioning on the full state vector $X_j(t_n - \tau_1), X_j(t_n - \tau_2), \dots, X_j(t_n - \tau_d)$ where d is the dimensionality (in a loose sense, the number of dynamical degrees of freedom) of the random process X_j . Building in more past history would almost certainly remove the oscillations as some of the past observations would always be providing additional predictive power. However, building in more history significantly complicates the ability to derive closed-form expressions. Moreover, for this simple linear system the basic envelope of the TDTE curves would not likely be effected by altering the model in this way.

We also point out that values of the TDTE are non-zero for positive delays as well. Again, so long as we interpret the TE as a measure of predictive power this makes sense. That is to say, future values X_j can aid in predicting the current dynamics of X_i . Interestingly, the asymmetry in the TE peaks near $\tau = 0$ may provide the largest clue as to the location of the forcing signal. Consistently we have found that the TE is larger for negative delays when mass closest the driven end plays the role of X_j ; conversely it is larger for positive delays when the mass furthest from the driven end plays this role. So long as the coupling is bi-directional, results such as those shown in Figure 2 can be expected in general.

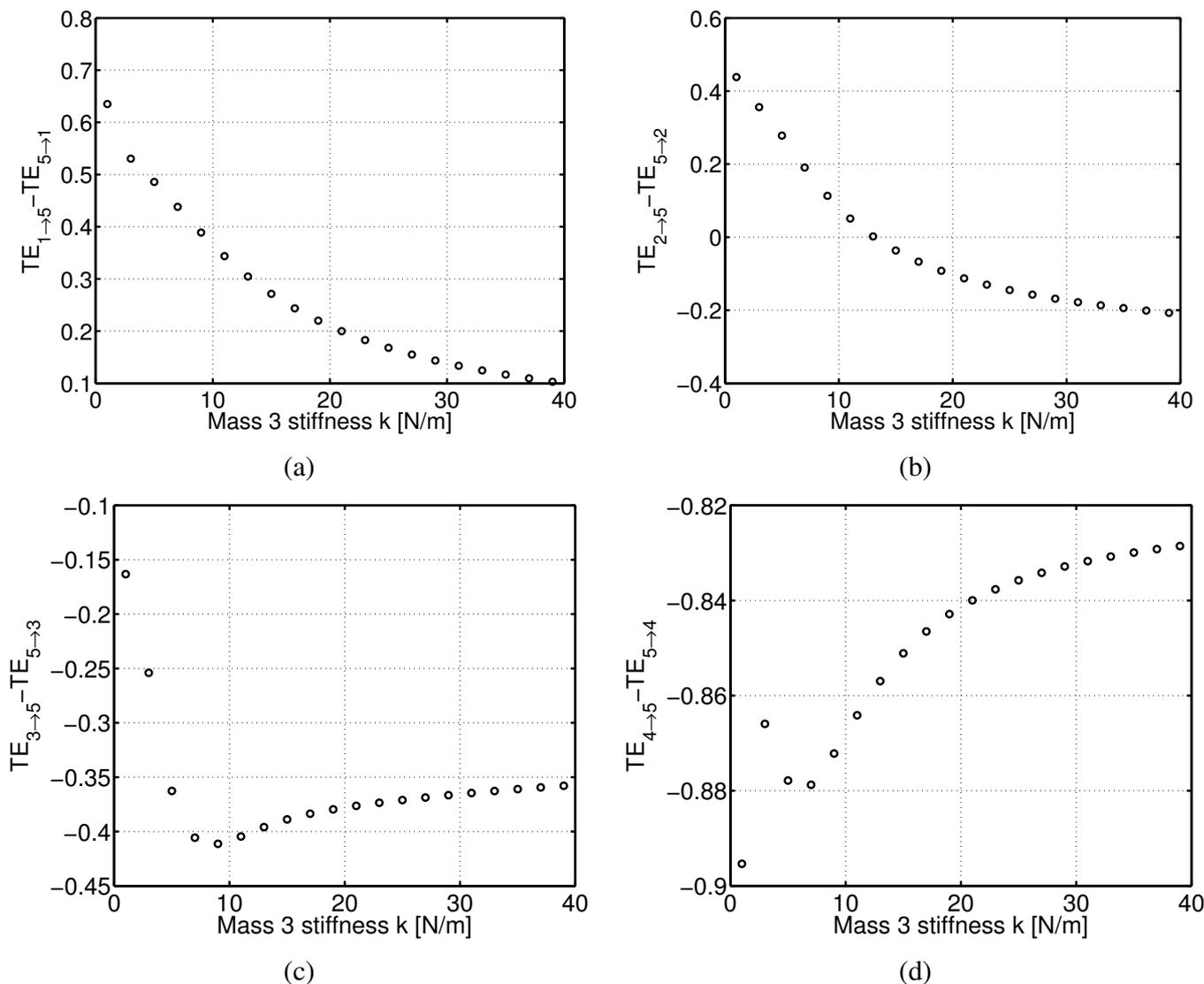
However, the situation is quite different if we consider the case of uni-directional coupling. For example, we may consider $TE_{f \rightarrow i}(\tau)$, *i.e.* the TDTE between the forcing signal and response variable i . This is a particularly interesting case as, unlike in previous examples, there is no feedback from DOF i to the driving signal. Figure 3 shows the TDTE between drive and response and clearly highlights the directional nature of the coupling. Past values of the forcing function clearly help in predicting the dynamics of the response. Conversely, future values of the forcing say nothing about transition probabilities for the mass response simply because the mass has not “seen” that information yet. Thus, for uni-directional coupling, the TDTE can easily diagnose whether X_j is driving X_i or vice-versa. It can also be noticed from these plots that the drive signal is not that much help in predicting the response as the TDTE is much smaller in magnitude that when computed between masses. We interpret this to mean that the response data are dominated by the physics of the structure (e.g., the structural modes), which is information not carried in the drive signal. Hence, the drive signal offers little in the way of additional predictive power. While the drive signal puts energy into the system, it is not very good at predicting the response. It should also be pointed out that the kernel density estimation techniques are not able to capture these small values of the TDTE. The error in such estimates is larger than these subtle fluctuations. Only the “linearized” estimator is able to capture the fluctuations in the TDTE for small ($O(10^{-2})$) values.

Figure 3. Time delay transfer entropy between the forcing (denoted as DOF “0”) and mass three for the 5 DOF system driven at mass, $P = 5$. The plot is consistent with the interpretation of information moving from the forcing to mass three.



It has been suggested that the main utility of the TE is to, given a sequence of observations, assess the direction of information flow in a coupled system. More specifically, one computes the difference $TE_{i \rightarrow j} - TE_{j \rightarrow i}$ with a positive difference suggesting information flow from i to j (negative differences indicating the opposite) [2,4]. In the system modeled by Equation (10) one would heuristically understand the information as flowing from the drive signal to the response. This is certainly reinforced by Figure 3. However, by extension it might seem probable that information would similarly flow from the mass closest the drive signal to the mass closest the boundary (e.g., DOF 5 to DOF 1).

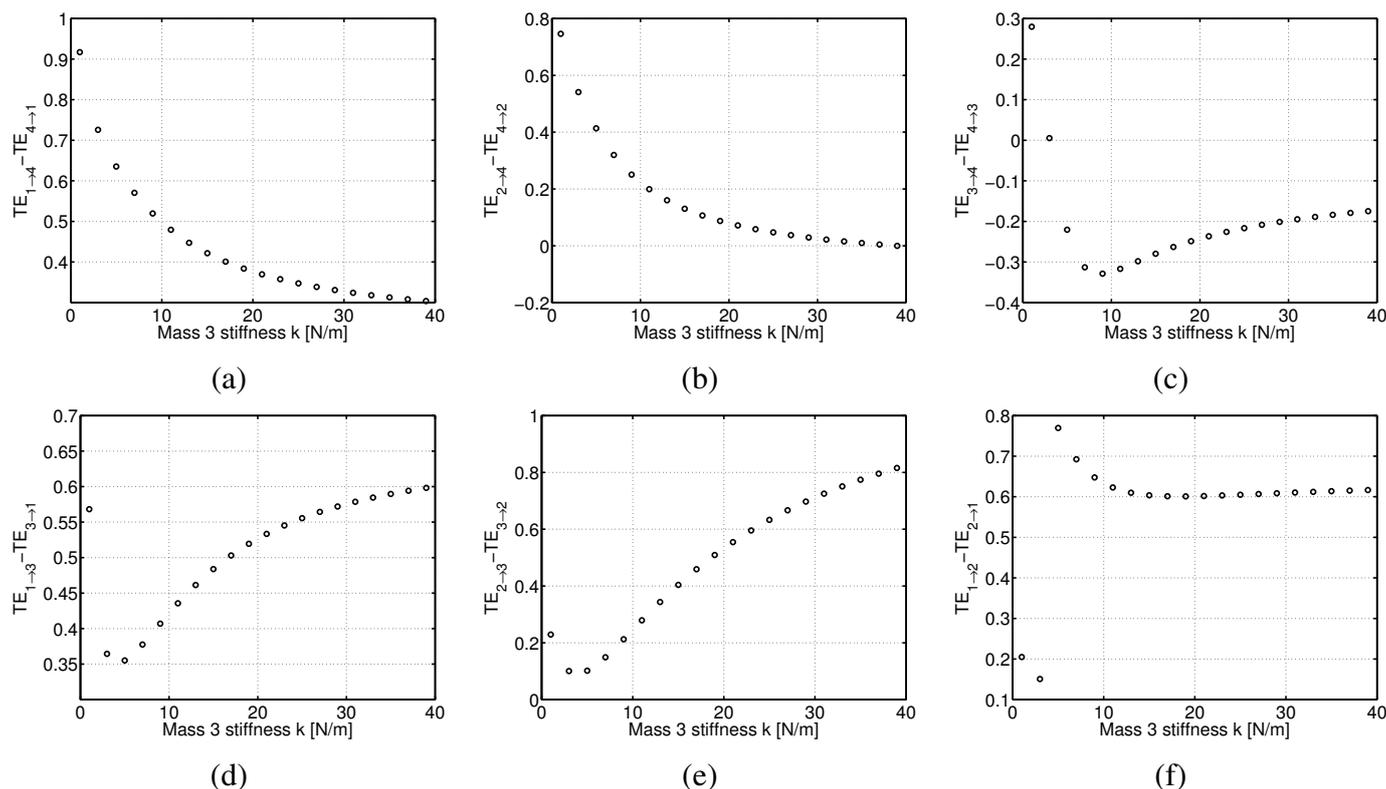
Figure 4. Difference in time delay transfer entropy between the driven mass five and each other DOF as a function of k_3 . A positive difference indicates $TE_{i \rightarrow j} > TE_{j \rightarrow i}$ and is commonly used to indicate that information is moving from mass i to mass j . Based on this interpretation, negative values indicate information moving from the driven end to the base; positive values indicate the opposite. Even for this linear system, choosing different masses in the analysis can produce very different results. In fact, $TE_{2 \rightarrow 5} - TE_{5 \rightarrow 2}$ implies a different direction of information transfer, depending on the strength of the coupling, k_3



We test this hypothesis as a function of the coupling strength between masses. Fixing each stiffness and damping coefficient to the previously used values, we vary k_3 from 1 N/m to 40 N/m and examine the quantity $TE_{i \rightarrow j} - TE_{j \rightarrow i}$ evaluated at τ^* , taken as the delay at which the TDTE reaches its maximum. Varying k_3 slightly alters the dominant period of the response. By accounting for this shift we eliminate the possibility of capturing the TE at one of its nulls (see Figure 2). For example, in Figure 2 we see that $\tau^* = -0.15$ in the plot of $TE_{3 \rightarrow 2}(\tau)$. Figure 4 shows the difference in TDTE as a function of the coupling strength. The result is non-intuitive if one assumes information would move from driven end toward the non-driven end of the system. For certain DOFs this interpretation holds, for others, it does

not. Herein lies the difficulty in interpreting the TE when bi-directional coupling exists. This was also pointed out by Schreiber [1] who noted “Reducing the analysis to the identification of a “drive” and a “response” may not be useful and could even be misleading”. The above results certainly reinforce this statement.

Figure 5. Difference in time-delayed transfer entropy (TDTE) among different combinations of masses. By the traditional interpretation of TE, negative values indicate information moving from the driven end to the base; positive values indicate the opposite.



Rather than being viewed as a measure of information flow, we find it more useful to interpret the difference measure as simply one of predictive power. That is to say, does knowledge of system j help predict system i more so than i helps predict j . This is a slightly different question. Our analysis suggests that if X_i and X_j are both near the driven end but with DOF i the closer of the two, then knowledge of X_j is of more use in predicting X_i than vice-versa. This interpretation also happens to be consistent with the notion of information moving from the driven end toward the base. However as i and j become de-coupled (physically separated) it appears the reverse is true. The random process X_i is better at predicting X_j than X_j is in predicting X_i . Thus, for certain pairs of masses information seems to be traveling from the base toward the drive. One possible explanation is that because the mass X_i is further removed from the drive signal it is strongly influenced by the vibration of each of the other masses. By contrast, a mass near the driven end is strongly influenced only by the drive signal. Because the dynamics X_i are influenced heavily by the structure (as opposed to the drive), X_i does a good job in helping to predict the dynamics everywhere. The main point of this analysis is that the difference in TE is not at all an unambiguous measure of the direction of information flow.

To further explore this question, we have repeated this numerical experiment for all possible combinations of masses. These results are displayed in Figure 5 where the same basic phenomenology is observed. If both masses being analyzed are near the driven end, the mass closest the drive is a better predictor of the one that is further away. However again, as i and j become decoupled the reverse is true. Our interpretation is that the further the process is removed from the drive signal, the more it is dominated by the other mass dynamics and the boundary conditions. Because such a process is strongly influenced by the other DOFs, it can successfully predict the motion for these other DOFs.

It is also interesting to note how the strength, and even directionality (sign) of the difference in TDTE changes with variations in a single stiffness element. Depending on the value of k_3 we see changes in which of the two masses is a better predictor. In some cases we even see zero TDTE difference, implying that the dynamics of the constituent signals are equally useful in predicting one another. Again, this does not support our intuitive notion of what it means for information to travel through a structural system. Only in the case of uni-directional coupling can we unambiguously use the TE to indicate directionality of information transport.

One of the strengths of our analysis is that these conclusions are not influenced by estimation error. In studying heart and breath rate interactions, for example, the ambiguity in information flow was assigned to difficulties in the estimation process [2]. We have shown here that even when estimation error is not a factor the ambiguity remains. We would imagine a similar result would hold for more complex systems, however such systems are beyond our ability to develop analytical expressions. The difference in TDTE is, however, a useful indicator of which system component carries the most predictive power about the rest of the system dynamics.

In short, the TDTE can be a very useful descriptor of system dynamics and coupling among system components. However any real understanding is only likely to be obtained in the context of a particular system model, or class of models (e.g., linear). Absent physical insight into the process that generates the observations, understanding results of a TDTE analysis can be challenging at best.

However, it is perhaps worth mentioning that the expressions derived here might permit inference about the general form of the underlying “linearized” system model. Different linear system models yield different expressions for $\rho_{ij}(\tau)$, hence different expressions for the TDTE. One could then conceivably use estimates of the TDTE as a means to select among this class of models given observed data. Whether or not the TDTE is of use in the context of model selection remains to be seen.

5. Conclusions

In this work we have derived an analytical expression for the time-delayed transfer entropy (TDTE) among components of a broad class of second order linear systems driven by a jointly Gaussian input. This solution has proven particularly useful in understanding the behavior of the TDTE as a measure of dynamical coupling. In particular, when the coupling is uni-directional, we have found the TDTE to be an unambiguous indicator of the direction of information flow in a system. However, for bi-directional coupling the situation is significantly more complicated, even for linear systems. We have found that a heuristic understanding of information flow is not always accurate. For example, one might expect information to travel from the driven end of a system toward the non-driven end. In fact, we have

shown precisely the opposite to be true. Simply varying a linear stiffness element can cause the apparent direction of flow to change. It would seem a safer interpretation is that a positive difference in the transfer entropy between two system components tells the practitioner which component has the greater predictive power.

Acknowledgements

The authors would like to thank the Naval Research Laboratory for providing funding for this work.

Conflict of Interest

The authors declare no conflict of interest

References

1. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.
2. Kaiser, A.; Schreiber, T. Information transfer in continuous processes. *Phys. D: Nonlinear Phenom.* **2002**, *166*, 43–62.
3. Moniz, L.J.; Cooch, E.G.; Ellner, S.P.; Nichols, J.D.; Nichols, J.M. Application of information theory methods to food web reconstruction. *Ecol. Model.* **2007**, *208*, 145–158.
4. Bauer, M.; Cox, J.W.; Caveness, M.H.; Downs, J.J.; Thornhill, N.F. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Trans. Control Syst. Technol.* **2007**, *15*, 12–21.
5. Marschinski, R.; Kantz, H. Analysing the information flow between financial time series: An improved estimator for transfer entropy. *Eur. Phys. J. B* **2002**, *30*, 275–281.
6. Nichols, J.M. Examining structural dynamics using information flow. *Probab. Eng. Mech.* **2006**, *21*, 420–433.
7. Nichols, J.M.; Seaver, M.; Trickey, S.T.; Salvino, L.W.; Pecora, D.L. Detecting impact damage in experimental composite structures: An information-theoretic approach. *Smart Mater. Struct.* **2006**, *15*, 424–434.
8. Moniz, L.J.; Nichols, J.D.; Nichols, J.M. Mapping the information landscape: Discerning peaks and valleys for ecological monitoring. *J. Biol. Phys.* **2007**, *33*, 171–181.
9. Vicente, R.; Wibral, M.; Lindner, M.; Pipa, G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **2001**, *30*, 45–67.
10. Wibral, M.; Rahm, B.; Rieder, M.; Lindner, M.; Vicente, R.; Kaiser, J. Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. *Progr. Biophys. Mol. Biol.* **2011**, *105*, 80–97.
11. Vakorin, V.A.; Krakovska, O.A.; McIntosh, A.R. Confounding effects of indirect connections on causality estimation. *J. Neurosci. Method.* **2009**, *184*, 152–160.
12. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701.
13. Palus, M.; Vejmelka, M. Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections. *Phys. Rev. E* **2007**, *75*, 056211.

14. Friston, K.J.; Harrison, L.; Penny, W. Dynamic causal modelling. *NeuroImage* **2003**, *19*, 1273–1302.
15. Greenberg, M.D. *Advanced Engineering Mathematics*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, USA, 1988.
16. Bendat, J.S.; Piersol, A.G. *Random Data Analysis and Measurement Procedures*, 3rd ed.; Wiley & Sons: New York, NY, USA, 2000.
17. Kay, S.M. *Fundamentals of Statistical Signal Processing: Volume I, Estimation Theory*; Prentice Hall: New Jersey, NJ, USA, 1993.
18. Meirovitch, L. *Introduction to Dynamics and Control*, 1st ed.; Wiley & Sons.: New York, NY, USA, 1985.
19. Crandall, S.H.; Mark, W.D. *Random Vibration in Mechanical Systems*; Academic Press: New York, NY, USA, 1963.
20. Benaroya, H. *Mechanical Vibration: Analysis, Uncertainties, and Control*; Prentice Hall: New Jersey, NJ, USA, 1998.
21. Nichols, J.M.; Olson, C.C.; Michalowicz, J.V.; Bucholtz, F. The bispectrum and bicoherence for quadratically nonlinear systems subject to non-gaussian inputs. *IEEE Trans. Signal Process.* **2009**, *57*, 3879–3890.
22. Birkhoff, G.D. Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA* **1931**, *17*, 656–660.
23. Hahs, D.W.; Pethel, S.D. Distinguishing Anticipation from Causality: Anticipatory Bias in the Estimation of Information Flow. *Phys. Rev. Lett.* **2011**, *107*, 128701.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).