

Comparison of nonlinear Granger causality extensions for low-dimensional systems

Katsuhiko Ishiguro,^{1,*} Nobuyuki Otsu,^{1,2,†} Max Lungarella,^{1,‡} and Yasuo Kuniyoshi^{1,§}

¹Graduate School of Information Science and Technology, University of Tokyo, 113-8656 Tokyo, Japan

²National Institute of Advanced Industrial Science and Technology, 305-8568 Tsukuba, Japan

(Received 5 April 2007; revised manuscript received 4 November 2007; published 25 March 2008)

The identification of hidden interdependences among the parts of a complex system is a fundamental issue. Typically, the objective is, given only a sequence of scalar measurements, to infer as much as possible about the internal dynamics of the system and about the interactions between its subsystems. In general, such interactions are not only nonlinear but also asymmetric. Constraints on the estimation of hidden relationships are further posed by noise and by the length of signals sampled from real world systems. The focus of this paper is causal dependences between bivariate time series. We especially focus on the nonlinear extension of Granger causality with polynomial terms of the conventional embedding vector. In this paper, we study the performance of this measure in comparison with three alternative methods proposed recently in low-dimensional and low-order-nonlinearity systems. Those methods are tested with three different artificial chaotic maps with several noise contamination setups. As a result, we find that the polynomial embedding technique successfully detects asymmetric (causal) dependences between bivariate time series in many low-dimensional cases.

DOI: [10.1103/PhysRevE.77.036217](https://doi.org/10.1103/PhysRevE.77.036217)

PACS number(s): 05.45.Tp, 02.50.Sk

I. INTRODUCTION

The detection and characterization of dependences among interacting components of a complex system can give novel insights into its functioning and lead to a better understanding of its dynamics. In most natural as opposed to laboratory settings, however, we are not granted direct access to the system's components, but possess instead only a set of strategically selected and typically simultaneously recorded variables. The problem of inferring and quantifying relationships between the parts of the system (e.g., how they contribute to the generation of information and at what rate they exchange information) is thus mapped onto the one of studying the relationships between multivariate time series corresponding to the recorded variables.

A popular way of evaluating the statistical dependency between time series is cross correlation. This method is not only computationally efficient but also has an appealing and natural interpretation. Cross correlation, however, is a second-order statistics and handles merely linear dependences. Mutual information in the information theory provides an attractive way of circumventing the restrictions of cross correlation because it is sensitive to higher-order relationships. Yet mutual information is a symmetric measure, and fails to detect directional (asymmetric) flow of information unless one of the time series is delayed. Measures of interdependence or "causality structure" aimed at overcoming the limitations of mutual information have recently been

introduced [1–5]. Although all these measures are asymmetric by construction and can be shown to work in specific instances, their general application is not a straightforward subject (for a review, see [6]). Real world signals (e.g., neurophysiological signals) are typically noisy, nonstationary, and of finite sample size; moreover, their dependences are often characterized by nonlinearities.

Here, we focus on another type of extension of Granger causality. We augment the time-delayed embedding vector by incorporating multivariate polynomial terms, and use the resulting nonlinear embedding vector to extend linear Granger causality [7]. Polynomial extension of Granger causality was reported in [8], but intensive comparison with other nonlinear extensions of Granger causality was not made. In this paper, we provide a quantitative study of several methods' performance on some numerical data sets. As a first step, we test low-dimensional and low-order-nonlinearity systems in this paper.

In what follows, we first describe three previously introduced methods. We then introduce the notion of the polynomial embedding vector and show how it can be used to extend Granger causality. We proceed by exposing the results of a set of numerical experiments with chaotic maps. Finally, we discuss possible extensions and alternatives to the proposed method.

II. EXISTING METHODS

Given a complex system, it is most likely the case that its internal dynamics is *a priori* unknown. There is also no knowledge of whether the system is deterministic or stochastic. Here, we assume that all we can do is to measure two or more variables observed separately from different parts of the complex system.

A. Embedding vector

For simplicity, let us consider the bivariate case, and let us assume that $X = \{x_1, x_2, \dots, x_N\}$ and $Y = \{y_1, y_2, \dots, y_N\}$ are

*Corresponding author. Present address: NTT Communication Science Laboratories, 619-0237 Kyoto, Japan. ishiguro@cslab.kecl.ntt.co.jp

†otsu.n@aist.go.jp

‡Present address: Artificial Intelligence Laboratory, University of Zurich, Switzerland. lunga@ifi.uzh.ch

§kuniyosh@isi.imi.i.u-tokyo.ac.jp

two simultaneously measured time series each consisting of N scalar quantities. The time-delayed embedding vector reconstructing the state (or phase) space X of time series X is expressed as

$$\mathbf{x}_t^{m,\tau} = (x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(m-1)\tau})^T, \quad (1)$$

where m is the embedding dimension and τ is the time delay (or lag) between successive elements of the state vector [9,10]. The choice of both parameters depends on the dynamics underlying the data. Here, first guesses for the embedding dimension and the time lag were obtained by the false nearest neighbor technique [11] and as the first minimum of the mutual information function [12], respectively. The time-delayed embedding vector reconstructing the state space Y of time series Y can be defined analogously. If not otherwise stated in this paper, $\tau=1$, and we use the following expression:

$$\mathbf{x}_t = \mathbf{x}_t^m = (x_t, x_{t-1}, x_{t-2}, \dots, x_{t-m+1})^T. \quad (2)$$

Similarly, we can define \mathbf{y}_t .

Note that many of the existing causality measures (including the ones described in this paper) are based on the notion of the embedding vector.

B. Granger causality and its extensions

The idea of Granger causality applied to bivariate time series [7] relies on the notion that the prediction of one time series through linear regression can be improved by incorporating information about the past values of the other time series. To assess the dependence of X on Y (subsequently also denoted as $Y \Rightarrow X$), consider the following set of autoregressive predictions:

$$x_{t+1} = \boldsymbol{\alpha}^T \mathbf{x}_t + \pi + \epsilon_t^{(x)}, \quad (3)$$

$$x_{t+1} = \boldsymbol{a}^T \mathbf{x}_t + \boldsymbol{b}^T \mathbf{y}_t + p + \epsilon_t^{(x|y)}, \quad (4)$$

where $\epsilon_t^{(x)}$ and $\epsilon_t^{(x|y)}$ represent the prediction errors, $\boldsymbol{\alpha}$, \boldsymbol{a} , and \boldsymbol{b} denote regression coefficient vectors, and π and p are constants. The coefficient vectors and constants are determined so as to minimize the variance of $\epsilon_t^{(x)}$ and $\epsilon_t^{(x|y)}$. Once the coefficient vectors have been calculated, the causal influence of Y on X can be expressed as

$$G_{Y \Rightarrow X} = 1 - \frac{\text{var}(\epsilon_t^{(x|y)})}{\text{var}(\epsilon_t^{(x)})}. \quad (5)$$

Intuitively, the larger $G_{Y \Rightarrow X}$ the stronger is the estimated causal influence of Y on X . In other words, if the prediction of x_t improves by incorporating the past values of y_t , then y_t is said to ‘‘Granger cause’’ x_t . In a similar way, it is possible to define $\epsilon_t^{(y)}$, $\epsilon_t^{(y|x)}$, and $G_{X \Rightarrow Y}$.

Granger causality is formulated for linear models, and its application to nonlinear systems may not be appropriate in the general case. Here, we describe two recently introduced nonlinear extensions of Granger causality. The first one is nonlinear Granger causality (NLGC) [1]. NLGC employs a nonlinear kernel autoregression scheme instead of a linear

autoregression to estimate the causal influence of the two time series on each other. The autoregression models of Eqs. (3) and (4) are replaced by the following expressions:

$$x_{t+1} = \boldsymbol{a}^T \boldsymbol{\Phi}(\mathbf{x}_t) + \pi + \eta_t^{(x)}, \quad (6)$$

$$x_{t+1} = \boldsymbol{a}^T \boldsymbol{\Phi}(\mathbf{x}_t) + \boldsymbol{b}^T \boldsymbol{\Phi}(\mathbf{y}_t) + p + \eta_t^{(x|y)}, \quad (7)$$

where $\boldsymbol{\Phi}$ is a P -dimensional vector whose elements are nonlinear radial based functions (RBFs) centered at $\mathbf{x}_{\rho 1}, \dots, \mathbf{x}_{\rho P}$,

$$\boldsymbol{\Phi}(\mathbf{x}_t) \triangleq \begin{pmatrix} -\frac{|\mathbf{x}_t - \mathbf{x}_{\rho 1}|}{\sigma^2} \\ -\frac{|\mathbf{x}_t - \mathbf{x}_{\rho 2}|}{\sigma^2} \\ \vdots \\ -\frac{|\mathbf{x}_t - \mathbf{x}_{\rho P}|}{\sigma^2} \end{pmatrix} \quad (8)$$

with σ a parameter defined *a priori*, and the center vectors chosen using fuzzy c -means clustering (other clustering algorithms can also be applied). The nonlinear Granger causal influence of Y on X or X on Y can be expressed as the difference between the variances of the two error terms

$$c_{Y \Rightarrow X} = \text{var}(\eta_t^{(x)}) - \text{var}(\eta_t^{(x|y)}), \quad (9)$$

$$c_{X \Rightarrow Y} = \text{var}(\eta_t^{(y)}) - \text{var}(\eta_t^{(y|x)}). \quad (10)$$

If $c_{Y \Rightarrow X} > 0$ then the prediction of x_t improves by including y_t , and it can be said that Y causally affects X . Analogously, if $c_{X \Rightarrow Y} > 0$ then X has a causal influence on Y .

Another extension of Granger causality, called the extended Granger causality index (EGCI), was proposed in [4]. The rationale of this technique is to divide the phase space into a set of small neighborhoods and approximate the globally nonlinear dynamics by local linear regression models. Let us express the joint dynamics of state space X and state space Y as

$$\mathbf{z}_t = \begin{pmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{pmatrix}. \quad (11)$$

The resulting state vector $\mathbf{z}_t \in R^{2m}$ is a point in the $2m$ -dimensional reconstructed phase space Z . Given $L < N$ points in Z , we define L clusters of local neighbors $\Theta_k = \{\mathbf{z}_t : |\mathbf{z}_t - \mathbf{z}_{t(k)}| < d\}$, where d is the size of the neighborhood and $t(k)$ ($k=1, \dots, L$) is the index of an arbitrarily selected centroid vector $\mathbf{z}_{t(k)}$. The parameter d is the size of the neighborhood and is chosen of the order of the resolution of the measurements. If the d neighborhood contains no points, d is increased.

Once the entire reconstructed phase space Z has been segmented, the linear Granger ‘‘causalities’’ $G_{Y \Rightarrow X}(k)$ and $G_{X \Rightarrow Y}(k)$ are calculated for each local neighborhood Θ_k . By averaging over all neighborhoods sampling Z (i.e., L neighboring clusters), we obtain two extended Granger causality indices

$$\delta_{Y \Rightarrow X} = \frac{1}{L} \sum_{k=1}^L G_{Y \Rightarrow X}(k), \quad (12)$$

$$\delta_{X \Rightarrow Y} = \frac{1}{L} \sum_{k=1}^L G_{X \Rightarrow Y}(k). \quad (13)$$

One relevant problem of both nonlinear extensions of Granger causality is their dependency on at least two parameters which need to be estimated and tuned to the time series at hand (e.g., size and number of the local neighborhoods, number of radial basis functions, and their parameter σ). Another drawback is their respective computational costs. In the case of NLGC, the computational cost of the least-mean-square regression is polynomial in the number of RBFs. For the EGCI, it is necessary to exhaustively search the state space to determine the local neighborhood of every centroid vector.

C. Transfer entropy

Another approach to obtaining knowledge about asymmetric (causal) dependences between coupled systems is by measuring to what extent its individual components contribute to information production and at what rate they exchange information among each other. Transfer entropy is an information theoretic measure which not only shares some of the properties of mutual information but also takes the dynamics of information transport into account [3]. It is a specific version of the mutual information for conditional probabilities, but unlike mutual information it is designed to detect the directed exchange of information between two systems, separate for both directions, and conditional to common history and input signals.

Transfer entropy measures the deviation from the following conditional independency assumption about the Markov process:

$$p(x_{t+1} | \mathbf{x}_t^k) = p(x_{t+1} | \mathbf{x}_t^k, \mathbf{y}_t^l) \quad (14)$$

where p denotes the transition (conditional) probability. If the deviation is small, then we can safely assume that the state of space Y has no (or little) relevance on the transition probabilities of the state vectors of space X . If the deviation is large, however, then the conditional independency assumption can be quantified by the transfer entropy, which is formulated as the Kullback-Leibler entropy between $p(x_{t+1} | \mathbf{x}_t^k)$ and $p(x_{t+1} | \mathbf{x}_t^k, \mathbf{y}_t^l)$,

$$T_{Y \Rightarrow X} = \sum_t p(x_{t+1}, \mathbf{x}_t^k, \mathbf{y}_t^l) \log \frac{p(x_{t+1} | \mathbf{x}_t^k, \mathbf{y}_t^l)}{p(x_{t+1} | \mathbf{x}_t^k)}. \quad (15)$$

A similar expression exists for $T_{X \Rightarrow Y}$. Transfer entropy represents the information about a future observation of variable x_t obtained from the simultaneous observation of past values of both x_t and y_t , after discarding the information about the future of x_t obtained from the past of x_t alone. If not stated otherwise, in the subsequent part of this paper $k=l=1$. The conditional probabilities of Eq. (15) are calculated from the

joint probability $p(x_{t+1}, \mathbf{x}_t^k, \mathbf{y}_t^l)$ which, in turn, is estimated using the histograms of the embedding vectors. The only parameter is the number of bins, r .

III. POLYNOMIAL EMBEDDING VECTOR AND NONLINEAR GRANGER CAUSALITY

Here, we present an alternative way of extending linear Granger causality, which not only performs better than NLGC and the EGCI but solves the problem of parameter tuning. We first extend the time-delayed embedding vector by incorporating multivariate polynomial terms. We then generalize Granger's idea by using the resulting "polynomial" embedding vector.

A. Polynomial embedding vector

Let us consider all possible combinations of order R of different monomial terms generated from the elements of \mathbf{x}_t ,

$$\tilde{\mathbf{x}}_t(i, j, R) = \prod_{k=i}^{k=j} x_{t-k}^{R_k}, \quad (16)$$

where $0 \leq i \leq j < m$, $R_k \geq 0$, and $1 \leq R_i + \dots + R_j \leq R$. By arranging all possible $\tilde{\mathbf{x}}_t$ in an array, we can define a nonlinear extension of the time-delayed embedding vector, subsequently called the polynomial embedding vector (PEV):

$$\tilde{\mathbf{x}}_t \triangleq (x_t, \dots, x_{t-m+1}, x_t^2, x_t x_{t-1}, \dots, x_{t-m+1}^R)^T. \quad (17)$$

B. PEV-extended Granger causality

The PEV can be used to extend Granger causality. One naive formulation is to replace the conventional delayed embedding vectors \mathbf{x}_t and \mathbf{y}_t with the PEVs $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{y}}_t$, that is,

$$x_{t+1} = \boldsymbol{\alpha}^T \tilde{\mathbf{x}}_t + \pi + \xi_t^{(x)}, \quad (18)$$

$$y_{t+1} = \boldsymbol{\beta}^T \tilde{\mathbf{y}}_t + \theta + \xi_t^{(y)}, \quad (19)$$

with $\boldsymbol{\alpha}$, π , $\boldsymbol{\beta}$, and θ determined by autoregression so as to minimize the variance of $\xi_t^{(x)}$ and $\xi_t^{(y)}$. The "mixed" autoregression equations are also modified according to

$$x_{t+1} = \boldsymbol{a}^T \tilde{\mathbf{x}}_t + \boldsymbol{b}^T \tilde{\mathbf{y}}_t + p + \xi_t^{(x|y)}, \quad (20)$$

$$y_{t+1} = \boldsymbol{c}^T \tilde{\mathbf{x}}_t + \boldsymbol{d}^T \tilde{\mathbf{y}}_t + q + \xi_t^{(y|x)}. \quad (21)$$

In this paper, however, we consider the polynomials of the mixed-state embedding vector \mathbf{z}_t as in [8]. The polynomially extended mixed-state embedding vector $\tilde{\mathbf{z}}_t$ incorporates the higher-order cross-correlation terms of x_t and y_t as well as the elements of $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{y}}_t$ [8].

In this case, the mixed information regression would be

$$x_{t+1} = \boldsymbol{a}^T \tilde{\mathbf{z}}_t + p + \xi_t^{(x|y)}, \quad (22)$$

$$y_{t+1} = \boldsymbol{c}^T \tilde{\mathbf{z}}_t + q + \xi_t^{(y|x)}, \quad (23)$$

and the resulting causality indices are

$$C_{Y \Rightarrow X} = \text{var}(\xi_t^{(x)}) - \text{var}(\xi_t^{(x|y)}), \quad (24)$$

$$C_{X \Rightarrow Y} = \text{var}(\xi_t^{(y)}) - \text{var}(\xi_t^{(y|x)}). \quad (25)$$

Hereafter, we call this method PEV-extended Granger causality (PEV-GC).

C. Computational properties

In this section, we describe the computational properties of this polynomial approach. Compared to existing nonlinear extensions of Granger causality this method has two advantages. First, the only parameter is the order R of the largest polynomial. It follows that the search for parameters is simplified in relation to the EGCI and NLGC, which both depend on the choice of at least two critical parameters. Second, due to its close relationship to the “classical” delayed embedding vector, the PEV is straightforward to implement and does not require any additional computational effort such as fuzzy c means clustering, evaluation of RBFs (as in NLGC), or search of local neighborhoods (as in the EGCI).

To compute the PEV of embedding dimension m and order R , the size of the nonlinear embedding vector is the sum of the number of different monomials of degree l , that is, $\sum_{l=1}^R \frac{1}{l!} m(m+1) \cdots (m+l-1)$, where m is the size of the original linear embedding vector. For example, in our experiments we used $(m, R) = (2, 2)$ or $(m, R) = (3, 3)$ which correspond to a five- and a 19-dimensional vector, respectively (the linear vector is two and three dimensional, respectively). It is clear that by using a polynomial embedding vector the computational cost due to the evaluation of the autoregressive equations increases. However, it is still reasonably small if compared to the additional operations required by the NLGC and EGCI.

IV. NUMERICAL EXPERIMENTS

To evaluate quantitatively the behavior of the PEV-GC and to illustrate its effectiveness, we conducted numerical experiments with two noiseless and two noisy model systems with built-in causality structure. In all experiments we compared the proposed measure with transfer entropy (TE), nonlinear Granger causality, and extended Granger causality. All simulations were done in MATLAB.

A. Noiseless polynomial chaotic map

We first studied a model system consisting of unidirectionally coupled noiseless chaotic (Hénon) maps [13]:

$$x_{t+2} = 1.4 - x_{t+1}^2 + 0.3x_t, \quad (26)$$

$$y_{t+2} = 1.4 - [ex_{t+1} + (1-e)y_{t+1}]y_{t+1} + 0.3y_t, \quad (27)$$

where the parameter e regulates the coupling strength between the two maps. We varied e from 0 to 1 with increments of 0.025. For each coupling and for sample sizes of $N=10^3$ and 10^4 , the time series x_t and y_t were generated (the initial 10^5 samples were discarded as transients). Each experiment was repeated five times.

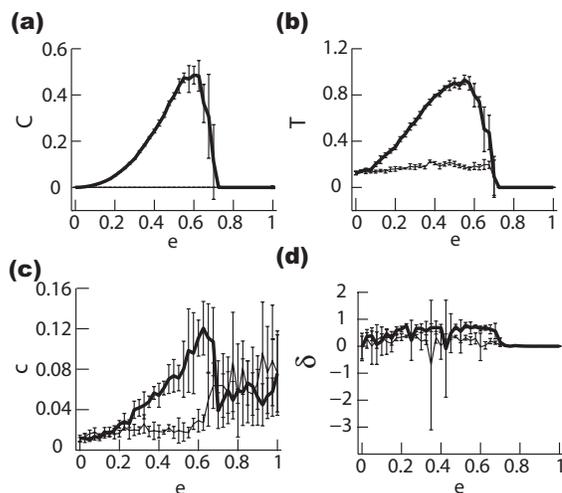


FIG. 1. Unidirectionally coupled Hénon maps ($N=10^3$; no noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger Causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table I for parameters.

By construction the two time series are unidirectionally coupled with time series x_t influencing time series y_t . It follows that the causal influence of x_t on y_t should increase if the strength of the coupling e increases. On the other hand, causal influence of y_t on x_t should be negligible because there is no interaction term in Eq. (26). From other studies (e.g., [3]), it is known that x_t and y_t will enter a (generally) synchronized state for $e > 0.7$. In that area, the two maps are indistinguishable from each other and the flow of information is zero (despite a nonzero coupling between the maps).

The results of our experiments are shown in Fig. 1 ($N=10^3$) and in Fig. 2 ($N=10^4$). The parameter settings used for the analysis are summarized in Table I. Those parameters are experimentally selected so as to give the best results. As evident from the graphs, PEV-GC gives noise-free and consistent results even for small sample sizes ($N=10^3$). We also observe that only PEV-GC successfully indicates absence of a causal influence of y_t on x_t ($C_{X \Rightarrow Y}=0$) at any e .

B. Noisy polynomial chaotic map

In order to assess the robustness of PEV-GC against noise, in the second set of experiments we considered noise-affected unidirectionally coupled Hénon maps. After generating $10^5 + N$ samples (as in the previous experiment, the first 10^5 samples were discarded as transients), independent identically distributed white noise was added to the time series. We tested the causality measures for $N=10^4$ and noise levels of 10%, 30%, and 50%. Each experiment was conducted five times.

The results are presented in Fig. 3 (noise level 10%), Fig. 4 (noise level 30%), and Fig. 5 (noise level 50%). The pa-

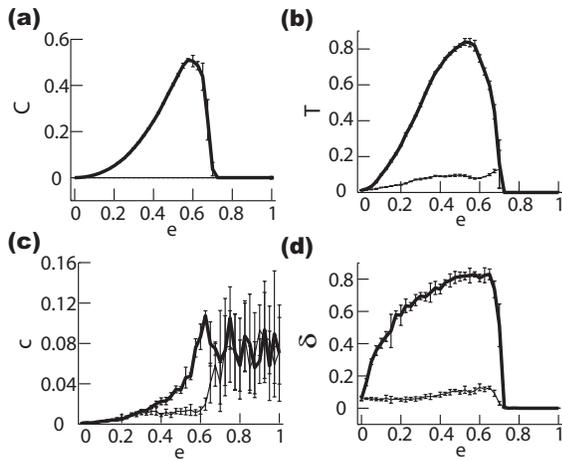


FIG. 2. Unidirectionally coupled Hénon maps ($N=10^4$, no noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table I for parameters.

parameter settings for each method are shown in Table I. We observe that the PEV-GC and TE methods are robust against the influence of fairly high levels of noise for all values of the coupling parameter e . PEV-GC performs better because it successfully shows that there is no information flow between the two maps. NLGC, on the other hand, manages to disclose the unidirectional causal dependency between the two maps

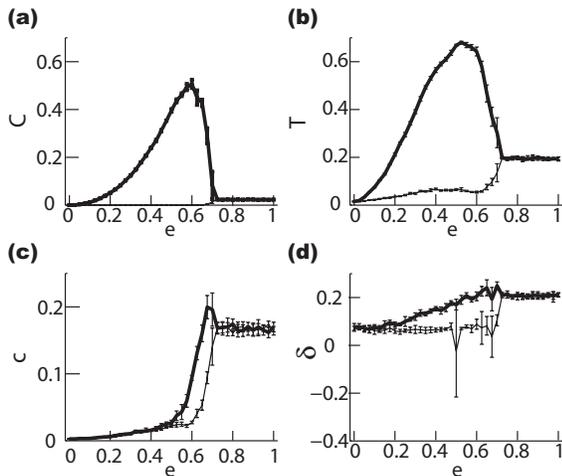


FIG. 3. Unidirectionally coupled Hénon maps ($N=10^4$; 10% noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table I for parameters.

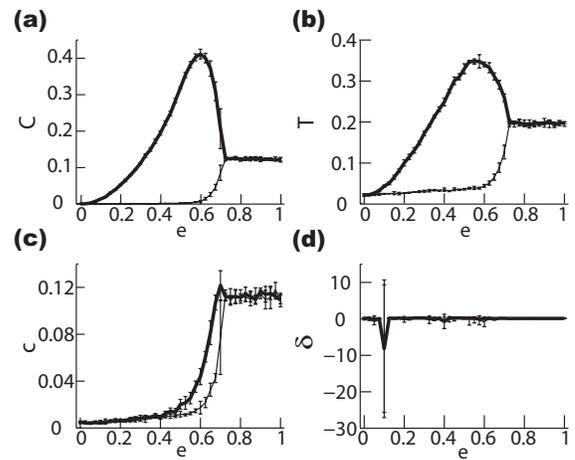


FIG. 4. Unidirectionally coupled Hénon map ($N=10^4$; 30% noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table I for parameters.

only for values of $e \in [0.5, 0.7]$. Finally, the EGCI, being sensitive to both sample size and noise level, fails to give a satisfying result. From these results, we can safely conclude that only PEV-GC and TE give consistent results for this noisy chaotic map data.

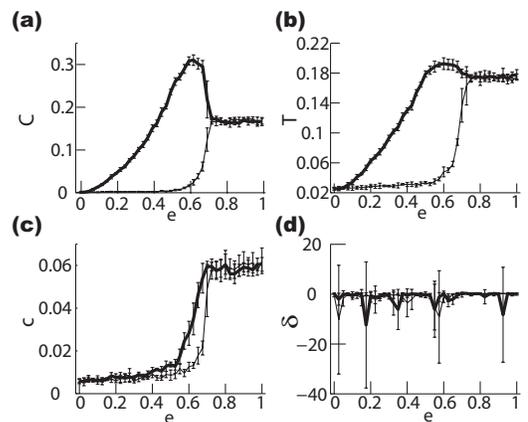


FIG. 5. Unidirectionally coupled Hénon maps ($N=10^4$; 50% noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table I for parameters.

C. Noiseless sine-based chaotic map

In the third numerical experiment, we applied all four measures to coupled chaotic maps featuring nonpolynomial components:

$$x_{t+2} = 1.7 - 0.4x_{t+1}^2 + 1.1 \sin(x_t), \quad (28)$$

$$y_{t+2} = 1.7 - 0.4y_{t+1}^2 + 1.1 \sin(y_t - ex_t), \quad (29)$$

where the parameter $e \in [0, 1]$ denotes the coupling strength of the causal interaction from X to Y . The embedding of the coupling parameter in the phase term of the sine function makes the detection of the causal relationship more challenging. The parameter e was incremented by 0.025 steps, and for each e , we collected $N=10^3$ and 10^4 samples after discarding 10^5 samples as transients. For each experiment, we performed five runs.

By construction the two time series were unidirectionally coupled with time series x_t influencing time series y_t . It follows that x_t causally affects y_t (the strength of the causal interaction should increase if the coupling parameter e increases). On the other hand, the causal influence of y_t on x_t should be negligible due to the absence of an interaction term in Eq. (28).

The results are presented in Fig. 6 ($N=10^3$) and in Fig. 7 ($N=10^4$). The parameters used for the analysis are summarized in Table II. Those parameters are experimentally selected so as to give the best results. In this experiment, all methods except the EGCI fail to measure the unidirectional causal structure in the weak coupling condition ($e < 0.5$). Another rather surprising result is that TE is unstable in the weak coupling area. This result differs from the previous two experiments, in which TE performed quite well.

D. Noisy sine-based chaotic map

In the fourth experiment, we studied a noisy sine-based map. As in the case of the second experiment, independent identically distributed white noise was added to the extracted bivariate time series. We tested the causality measures for $N=10^4$ and noise levels of 10%, 20%, and 30%; each experiment was conducted five times.

The results of our numerical experiments are presented in Fig. 8 (noise level 10%), Fig. 9 (noise level 20%), and Fig. 10 (noise level 30%). The parameters for each method are presented in Table II. As evident from the graphs, PEV-GC is robust against noise, and the estimated coupling strengths are

TABLE I. Parameters for experiments on unidirectionally coupled Hénon maps. If there are two values, they correspond to the cases of $N=10^3$, 10^4 , respectively. The embedding dimension is $m=2$ except for TE.

Method	Parameters
Transfer entropy (TE)	$r=8, m=1$
Extended GC index (EGCI)	$L=50, d=0.5, 0.3$
Nonlinear GC (NLGC)	$P=50, \sigma=0.05$
PEV-GC	$R=2$

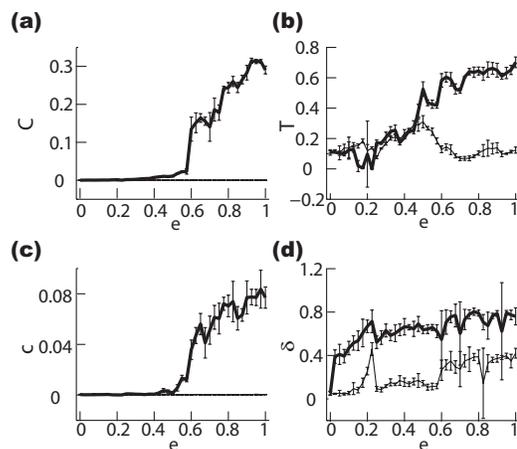


FIG. 6. Unidirectionally coupled sine-based maps ($N=10^3$; no noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table II for parameters.

not even affected by heavy noise (30%). On the other hand, a deterioration of performance is notable in the TE and EGCI. Especially for the TE, the results are far from the theoretical predictions; for weaker couplings, the estimated coupling direction is even inverted. NLGC displays a strange behavior for values of e close to 0.6 where it suddenly falls and rises again.

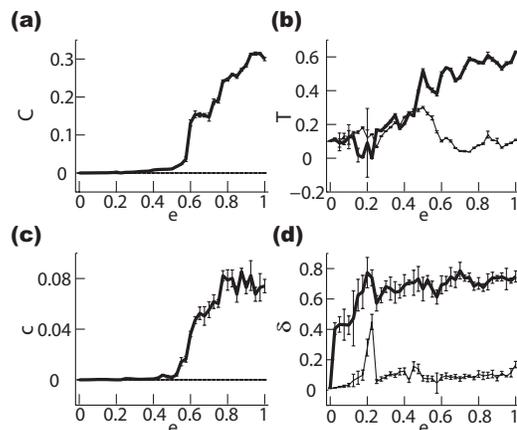


FIG. 7. Unidirectionally coupled sine-based maps ($N=10^4$, no noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table II for parameters.

TABLE II. Parameters for experiments on unidirectionally coupled sine-based maps. If there are two values, they correspond to the cases of $N=10^3$, 10^4 , respectively. The embedding dimension is $m=3$ except for TE.

Method	Parameters
Transfer entropy (TE)	$r=8, m=1$
Extended GC index (EGCI)	$L=50, d=0.8, 0.6$
Nonlinear GC (NLGC)	$P=50, \sigma=1.0$
PEV-GC	$R=3$

E. Exponential function coupling data

So far, we have examined the performances of causality measures on several chaotic maps. As you can see, PEV-GC shows preferable results against the others.

In this experiment, we use the following dynamic models for data generation:

$$x_{t+1} = 0.125x_t + \frac{25x_t}{4(x_t^2 + 1)} + 2 \cos(1.2t) + N(0, 1), \quad (30)$$

$$y_{t+1} = 0.1y_t^2 - e[f(x_t)^2 - 0.3] + N(0, 1), \quad (31)$$

where

$$f(x_t) = \frac{1.0}{1.0 + \exp(-x_t)} \quad (32)$$

and $N(0, 1)$ is a random value from a normal distribution of zero mean and unit variance. One big difference from previous chaotic maps is that X drives Y through an exponential (sigmoid) function. Apparently, exponential bases of NLGC will suit for estimating the $X \Rightarrow Y$ causal couplings. Note also

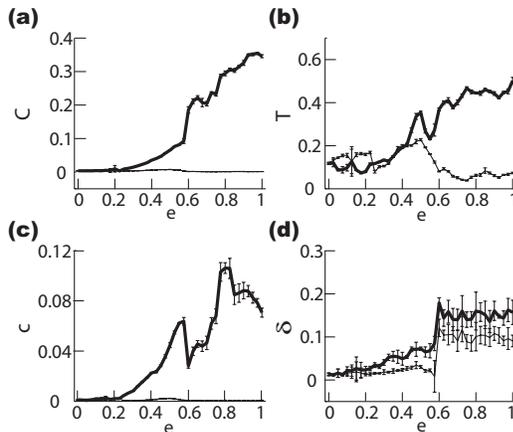


FIG. 8. Unidirectionally coupled sine-based maps ($N=10^4$; 10% noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table II for parameters.

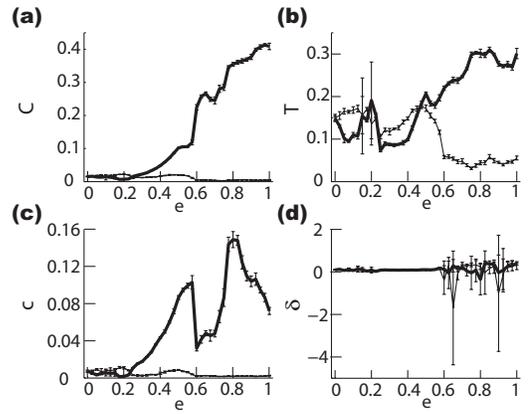


FIG. 9. Unidirectionally coupled sine-based maps ($N=10^4$; 20% noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table II for parameters.

that the noise term is incorporated in each dynamic model.

The results are presented in Fig. 11 for $N=10^4$. The parameters for each method are presented in Table III. Those parameters are experimentally selected so as to give the best results. The NLGC presents excellent estimation results as predicted, reflecting the coupling structure of the data. Though the bases are not polynomials, the PEV-GC still performs quite well.

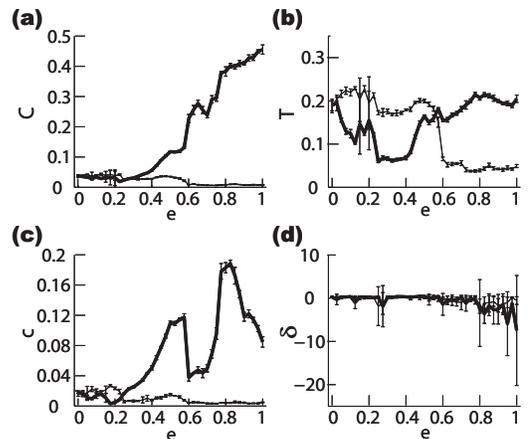


FIG. 10. Unidirectionally coupled sine-based maps ($N=10^4$; 30% noise). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table II for parameters.

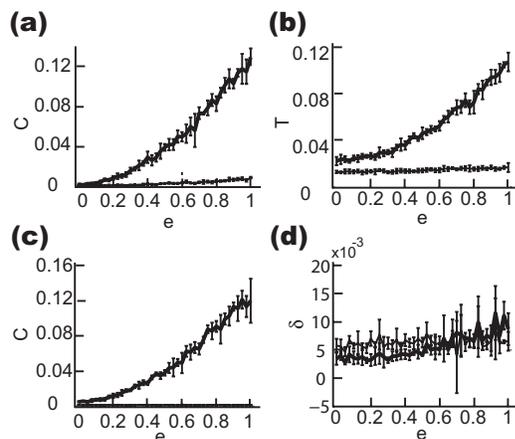


FIG. 11. Unidirectional coupling data with exponential function ($N=10^4$). (a) PEV-extended Granger causality. (b) Transfer entropy. (c) Nonlinear Granger causality. (d) Extended Granger causality index. The horizontal axis denotes the strength of coupling and the vertical axis denotes the causality measure. Thick lines represent $X \Rightarrow Y$ causality and thin lines represent $Y \Rightarrow X$ causality. The plots are averages over five experimental runs. The error bars denote standard deviations of causality measures for five runs. See Table III for parameters.

V. DISCUSSION

As the experimental results show, the PEV-GC successfully detects the underlying causal relationship in many cases, regardless of the number of samples and the heaviness of noise contamination. One possible reason for this is that the PEV may be interpreted as a truncated Taylor expansion of order R . This view is interesting because any elementary and well-behaved function can be expanded into polynomials by Taylor expansion. This indicates that the PEV is potentially applicable to many different kinds of bivariate time series.

It is also appealing for users that the PEV is easy to implement and has only one control parameter (R), compared to other alternatives presented in this paper. Since tuning of multiple parameters is a difficult task, the smaller number of parameters is attractive in practice.

We also point out a substantial difference between the polynomial embedding vector and the “kernel method” used in nonlinear Granger causality [1]. Kernel methods are rather popular nonlinear techniques in statistical pattern recognition (e.g., [14]). They rely on the use of “kernel functions,” de-

signed as nonlinear distance functions (e.g., Gaussian, sigmoid) between samples. One characteristic of kernel techniques is that the size of the nonlinear (embedding) vector is very dependent on the number of samples, not on the dimensionality of each sample. On the other hand, the size of the PEV is basically dependent on the dimension of a sample (embedding dimension), not on the sample size.

This point makes a difference between the two techniques in computational costs. We can see this by comparing two research tasks: image processing and time series analysis. In image processing, an image has very high dimensionality (e.g., even a very small 64×64 pixel image has 4096 pixels) and the number of samples (images) is rather small (in many cases, some hundreds). Clearly, this is not an ideal setting for the PEV—small number of samples for high-dimensional data. Next, let us consider time series analysis. Time series analysis techniques need to handle large sample pools ($N > 10^3$) for which, generally speaking, the embedding dimension is not too large (in many cases, $m < 10$). In this case, kernel techniques may face a computational problem—many samples for low-dimensional data. The number of samples can be reduced by introducing a heuristic trick such as clustering preprocessing in nonlinear Granger causality [1], but it may lead to a deterioration of performance unless the heuristics properly reflect the nature of the data. Therefore, in time series analysis we have two reasons to use the PEV: one is computational advantage and the other is that the PEV does not need heuristics which may deteriorate the performance. As the results show, the PEV-GC performs efficiently in the low-dimensional data set. However, we should note that this may not be true for highly nonlinear or high-dimensional data.

The main rationale behind using a nonlinear embedding vector in combination with Granger causality was to extend the traditional linear analysis and take into account also nonlinear structure. In principle it should be possible to extend other measures of causality also, such as transfer entropy, using the polynomial embedding vector. One caveat is that its formulation could become rather complex. Specifically, in the case of transfer entropy, we would need to estimate the probability density function of the time series formed by the embedding vectors. If the PEV is used, then such estimation is rather difficult and requires a substantial computational effort and a very large data pool (e.g., for $m=3$, $R=3$ the PEV is a 19-dimensional vector for which we would have to estimate probability distribution functions with up to 19 dimensions).

VI. CONCLUSION

In this paper, we survey several nonlinear extensions of Granger causality and assess their performances through numerical experiments using a low-dimensional and low-order nonlinearity data set. In particular, we focus on the extension of the time-delayed embedding vector (the polynomial embedding vector) which in combination with Granger causality leads to a computationally efficient nonlinear extension of Granger causality. We studied the effectiveness of this type of extension in numerical experiments with some artificial

TABLE III. Parameters for experiments on data unidirectionally coupled through exponential function. The embedding dimension is $m=3$ except for TE.

Method	Parameters
Transfer entropy (TE)	$r=8$, $m=1$
Extended GC index (EGCI)	$L=100$, $d=2.0$
Nonlinear GC (NLGC)	$P=100$, $\sigma=1.0$
PEV-GC	$R=3$

chaotic maps. Intensive experiments concerning the type of the maps, the number of samples, and the heaviness of noise contamination were conducted, and the results showed that the polynomial embedding vector based technique is superior or comparable to other alternative measures in quantifying

the asymmetric causal interdependence between bivariate time series data, at least in low-complexity systems. In future work, we will apply the proposed PEV-GC to data gathered from real world systems such as climate data, neurophysiological signals, and sensory and motor variables of robots.

-
- [1] N. Ancona, D. Marinazzo, and S. Stramaglia, *Phys. Rev. E* **70**, 056221 (2004).
- [2] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, *Physica D* **134**, 419 (1999).
- [3] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
- [4] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, *Phys. Lett. A* **324**, 26 (2004).
- [5] U. Feldmann and J. Bhattacharya, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **14**, 505 (2004).
- [6] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **17**, 903 (2007).
- [7] C. W. J. Granger, *Econometrica* **37**, 424 (1969).
- [8] I. I. Mokhov and D. A. Smirnov, *Geophys. Res. Lett.* **33**, L03708 (2006).
- [9] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, *Phys. Rev. Lett.* **45**, 712 (1980).
- [10] M. Takens, in *Dynamical Systems and Turbulence*, edited by D. Rand and L. Young, *Lecture Notes in Mathematics* Vol. 898 (Springer Verlag, Berlin, 1981), pp. 366–381.
- [11] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, *Phys. Rev. A* **45**, 3403 (1992).
- [12] A. M. Fraser and H. L. Swinney, *Phys. Rev. A* **33**, 1134 (1986).
- [13] M. Henon, *Commun. Math. Phys.* **50**, 69 (1976).
- [14] V. N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, Berlin, 1995).