

Kernel canonical-correlation Granger causality for multiple time series

Guorong Wu, Xujun Duan, Wei Liao, Qing Gao, and HuaFu Chen*

Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology,
University of Electronic Science and Technology of China, Chengdu 610054, China

(Received 20 October 2010; revised manuscript received 1 February 2011; published 25 April 2011)

Canonical-correlation analysis as a multivariate statistical technique has been applied to multivariate Granger causality analysis to infer information flow in complex systems. It shows unique appeal and great superiority over the traditional vector autoregressive method, due to the simplified procedure that detects causal interaction between multiple time series, and the avoidance of potential model estimation problems. However, it is limited to the linear case. Here, we extend the framework of canonical correlation to include the estimation of multivariate nonlinear Granger causality for drawing inference about directed interaction. Its feasibility and effectiveness are verified on simulated data.

DOI: [10.1103/PhysRevE.83.041921](https://doi.org/10.1103/PhysRevE.83.041921)

PACS number(s): 87.10.Ca, 87.19.xm, 87.19.le, 87.85.Tu

I. INTRODUCTION

Modeling and estimating the interactions in dynamical systems is a key question in neuroscience, engineering, and other fields. An efficient and elegant method to understand this kind of system, rooted in graph theory and statistical physics is to draw a map representing the network of the system, on the scale of a subsystem [1–3]. The different scale structures may lead to different properties, and thus to different corresponding patterns of dynamic interaction [4,5].

Granger causality considers that a signal y_t causes another signal x_t , if x_t could be predicted better when taking into account the past of y_t than by not doing so, for multivariate Granger causality other relevant information being used in either case [6]. Multivariate kernel causality generalizes linear Granger causality to the nonlinear case [7]. It has been proved that multivariate Granger causality is an effective and powerful tool to detect the information transfer between the nodes of a complex system [6,8,9]. Many advanced and mathematically complex techniques have been proposed to statistically characterize Granger causality [10]. Cross correlation is a simple way to detect directional influence, but is limited to the linear and univariate case [11]. Canonical-correlation analysis, which is the generalization of the Pearson correlation (zero-lag cross correlation), is a multivariate statistics method for measuring associations between two sets of variables. It has been successfully applied in Granger causality analysis [12,13], whereas only limited to the linear case (although for Gaussian variables such as those simulated in [12], it has been proved that nonlinear analysis is not necessary since a Gaussian autoregressive (AR) process is always linear [14]). On the other hand, when we deal with brain imaging data, although linear relationship is enough for functional connectivity in resting-state functional magnetic resonance imaging [15], it might not be sufficient for detecting effective connectivity, and the extension to the nonlinear case could be in order. This paper presents a different canonical-correlation algorithm for multivariate nonlinear Granger causality analysis, and explores directionality in some chaotic systems.

II. KERNEL CANONICAL-CORRELATION GRANGER CAUSALITY

First we give the formal definition for Granger causality for multiple time series. Suppose that variables X , Y are wide-sense stationary multivariate stochastic processes, $\{\Omega_t\}$ is the information set containing all the relevant information available up to and including period t , and $\{\mathfrak{S}_t\}$ is the given set of information including at least $\{X_t, Y_t\}$. Y_t is said to cause X_{t+1} if $F(X_{t+1}|\Omega_t - Y_t) \neq F(X_{t+1}|\Omega_t)$ [16], where $F(X_{t+1}|J_t)$ is the conditional distribution function of X_{t+1} given J_t , $\Omega_t - Y_t$ denotes the information except the values taken by Y_t . In practice, Ω_t is impossible to be obtained so \mathfrak{S}_t is used in substitution. Accordingly, Y_t is said to be a *prima facie* cause of X_{t+1} with respect to \mathfrak{S}_t . In the applications, regression techniques and cross-correlation analysis are the main tools to identify Granger causality flows [11,17], and the standard test of Granger causality is based on the vector autoregressive (VAR) model [18,19].

Consider the unrestricted multivariate autoregressive model:

$$X_t = [X_{t-\text{past}}, Y_{t-\text{past}}] \begin{bmatrix} A_1 \\ B_1 \end{bmatrix} + \epsilon_t, \quad (1)$$

where $X_{t-\text{past}}$ ($X_{t-\text{past}} = [X_{t-1}, \dots, X_{t-n}, \dots]$) and $Y_{t-\text{past}}$ are the lag distributions of X_t and Y_t , respectively. The widely accepted test of Granger causality is based on the prediction error of the VAR model. When the vector Y causes X in the Wiener-Granger sense, the coefficients B_1 are jointly significantly different from zero. As we know, correlation may be interpreted as the square root of R^2 of a linear regression model, where R^2 represents the variance of the response variable which may be explained by the regressors [13,20], so the correlation is associated with predictive power. By some extensions [21], it could be proved that the partial canonical correlation between X_t and $Y_{t-\text{past}}$ identify the Granger causality from Y to X .

In the next, we consider problems in the sample space. Suppose we have $N + m + 1$ realization of the q_1 -dimensional and q_2 -dimensional stochastic variables X_t and Y_t . We used the following shorthand notations: $\xi_i = (X_i, \dots, X_{i+N-1})^T \in \mathfrak{R}^{N \times q_1}$, $\zeta_i = (Y_i, \dots, Y_{i+N-1})^T \in \mathfrak{R}^{N \times q_2}$, $\xi = (\xi_1, \dots, \xi_m) \in \mathfrak{R}^{N \times mq_1}$, $x = \xi_{m+1} \in \mathfrak{R}^{N \times q_1}$, $\zeta = (\zeta_1, \dots, \zeta_m) \in \mathfrak{R}^{N \times mq_2}$,

*Author to whom correspondence should be addressed: chenhf@uestc.edu.cn

$\boldsymbol{\gamma} = [\boldsymbol{\xi}, \boldsymbol{\zeta}] \in \mathfrak{R}^{N \times m(q_1 + q_2)}$. Let $\mathbf{K}_1 = \boldsymbol{\xi}\boldsymbol{\xi}^T$, $\mathbf{K}_2 = \boldsymbol{\gamma}\boldsymbol{\gamma}^T$. In order to center data in the feature space, we replace \mathbf{K}_i by $\mathbf{K}_i - m^{-1}\mathbf{j}\mathbf{j}'\mathbf{K}_i - m^{-1}\mathbf{K}_i\mathbf{j}\mathbf{j}' + m^{-2}(\mathbf{j}'\mathbf{K}_i\mathbf{j})\mathbf{j}\mathbf{j}'$ ($i = 1, 2$), where $\mathbf{j} = (1, \dots, 1)^T \in \mathfrak{R}^{N \times 1}$ [22]. In the following we assume that the data in the feature space has been centered. The matrices \mathbf{K}_i are always singular, as the effective ranks of the matrix are much lower than their sizes because of the small sample size and high-dimensional case [23]. Here we convert the high-dimensional feature space to reduced dimensional subspace, in the premise of not losing any valid information of primitive samples [23]. Let $\mathbf{P} = \sum_{i=1}^m \mathbf{v}_i \mathbf{v}_i^T$, where \mathbf{v}_i are the orthogonal eigenvectors of \mathbf{K}_1 with nonzero eigenvalue. Moreover, $\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{x}$ is the prediction error vector of linear regression of \mathbf{x} versus $\boldsymbol{\xi}$, the prediction error defined as $\boldsymbol{\varepsilon}_x = |\mathbf{y}|_F^2$ (where $\|\cdot\|_F$ is the Frobenius norm). Using both $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ to predict \mathbf{x} , the prediction error is now $\boldsymbol{\varepsilon}_{xy}$. Let $\{\boldsymbol{\tau}\}_\alpha$ consist of standard orthogonal eigenvectors with nonvanishing eigenvalue of the matrix $\mathbf{K} = (\mathbf{I} - \mathbf{P})\mathbf{K}_2(\mathbf{I} - \mathbf{P})$ [24]. Without loss of generality we assume that each $\boldsymbol{\tau}_\alpha$ has zero mean. We have $\boldsymbol{\varepsilon}_{xy} = \boldsymbol{\varepsilon}_x - \sum_i r_i^2$ (where r_i is the Pearson's correlation coefficient of each column of \mathbf{x} and $\boldsymbol{\tau}_i$), and the variance of residual-based linear Granger causality index reads [24]

$$\delta(\mathbf{Y} \rightarrow \mathbf{X}) = 1 - \boldsymbol{\varepsilon}_{xy} \cdot \boldsymbol{\varepsilon}_x^{-1} = \|\mathbf{y}\|_F^{-2} \cdot \sum_i r_i^2. \quad (2)$$

Furthermore, the basis $\{\boldsymbol{\tau}\}_\alpha$ is a maximal linearly independent set of the column space of $\boldsymbol{\zeta}$ partialing out the effect of $\boldsymbol{\xi}$ (the sample representation of $\mathbf{Y}_{t-\text{past}}$ partialing out the effect of $\mathbf{X}_{t-\text{past}}$), and \mathbf{y} is the sample representation of \mathbf{X}_t partialing out the effect of $\mathbf{X}_{t-\text{past}}$. Then a correlation-based linear Granger causality index can be defined as

$$CCA(\mathbf{Y} \rightarrow \mathbf{X}) = \text{canoncorr}(\mathbf{y}, \boldsymbol{\tau}) \quad (3)$$

where *canoncorr* is a defined MATLAB function, which denotes canonical-correlation analysis (CCA), i.e., $\text{canoncorr}(\mathbf{y}, \boldsymbol{\tau}) = \text{argmax}_{\alpha, \beta} \rho = \alpha' \mathbf{y}' \boldsymbol{\tau} \beta \cdot \sqrt{\alpha' \mathbf{y}' \mathbf{y} \alpha \beta' \boldsymbol{\tau}' \boldsymbol{\tau} \beta}$, which could be simply solved by linear algebra techniques. Kernel CCA (KCCA) uses the ‘‘kernel trick’’ to project the data into a higher dimensional feature space [22], which provides a convenient way for generalization of the linear CCA (LCCA). As described for the linear case, the kernel Granger causality [24] could be implemented along the same lines, but \mathbf{K}_i ($i = 1, 2$) should be replaced by the corresponding Gram matrix $\mathbf{K}_1 = \kappa(\boldsymbol{\gamma}, \boldsymbol{\gamma})$, $\mathbf{K}_2 = \kappa(\boldsymbol{\xi}, \boldsymbol{\xi})$. However, bivariate

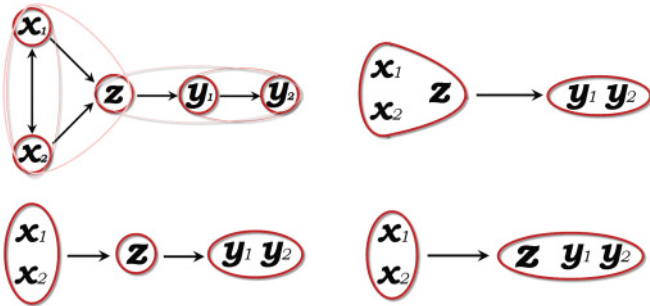


FIG. 1. (Color online) Schematic plot for the simulation random vector model. The upper-left frame demonstrates the selective causal relationship of the variables in Eq. (4). The remains denote the decomposed frames of the upper-left frame.

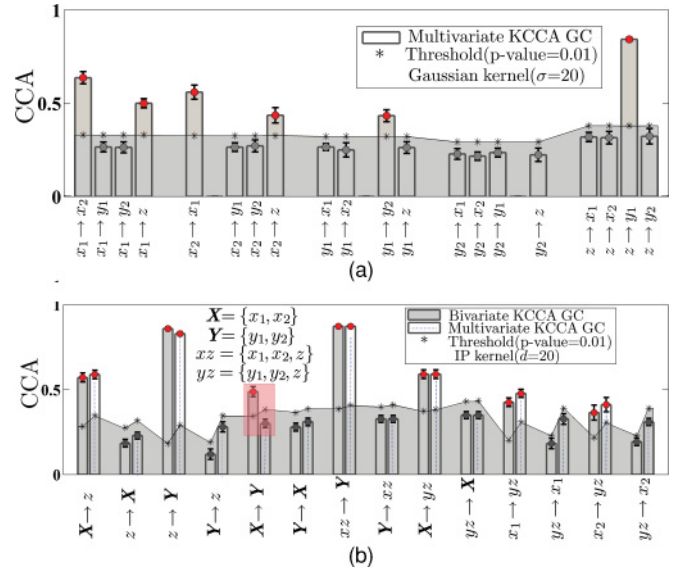


FIG. 2. (Color online) Simulation results of bivariate and multivariate KCCA Granger causality analyses for random vector model. IP kernel with degree of 2 and Gaussian with $\sigma = 20$ were considered. (a) the analysis result of multivariate KCCA Granger causality for all pairs of maps. (b) the analysis result of bivariate and multivariate KCCA Granger causality among combinations.

Granger causality (here information set $\mathfrak{S}_t = \{\mathbf{X}_t, \mathbf{Y}_t\}$) is a *prima facie* causality, which may vary when to \mathfrak{S}_t we add additional information which could influence \mathbf{X}_t or \mathbf{Y}_t or both. Geweke proposed that the conditional Granger causality would overcome the spurious causalities [25], caused by delay or indirect influence through a third series. Accordingly, we plan here to extend the bivariate KCCA Granger causality to the multivariate case. First, let $\mathbf{x} = (\mathbf{X}_{m+1}, \dots, \mathbf{X}_{m+N})^T$, then replace \mathbf{X}_t by $(\mathbf{X}_t^T, \mathbf{Z}_t^T)^T$, with multivariate \mathbf{Z} consisting of other measured variables; the remaining step is same as the bivariate case.

In the following, we consider two kinds of kernels: the inhomogeneous polynomial (IP) and the Gaussian kernel. The IP kernel of degree d is $\kappa_d(\mathbf{X}, \mathbf{Y}) = (\mathbf{I} + \mathbf{X}^T \mathbf{Y})^d$, and the Gaussian kernel is defined as $\kappa_\sigma(\mathbf{X}, \mathbf{Y}) = \exp[-(\sqrt{2}\sigma)^{-2} \|\mathbf{X} - \mathbf{Y}\|^2]$.

III. APPLICATION TO SIMULATIONS

A. Toy model

In order to verify the feasibility and effectiveness of the proposed method, we applied it on a random vector model first, which associated the following equations:

$$\begin{aligned} x_1(t) &= -0.8x_1(t-1) + 0.25\sqrt{2}x_2(t-2) + 0.2\tau_1(t), \\ x_2(t) &= 0.75x_1(t-1)[1 - x_2(t-2)] + 0.3\tau_2(t), \\ y_1(t) &= -0.4e^{-y_1^2(t-1)} + 0.95z(t-1)^2 + 0.2\tau_3(t), \\ y_2(t) &= -0.75y_1(t-1)^2 + 0.5y_2(t-1) + 0.4\tau_4(t), \\ z(t) &= 0.3 \tan[x_1(t-1)] - 0.8 \cos[x_2(t-1)] + 0.2\tau_5(t), \end{aligned} \quad (4)$$

where the τ 's are zero-mean uncorrelated Gaussian white noise processes with unit variances. Assuming no prior knowledge of the above equation, the model is simulated to generate a data set

of 500 realizations. Granger causality analysis was performed on the simulated data, and the lag order was determined by the Bayesian information criterion (BIC). We define the group set as the following combination: $\mathbf{X} = \{x_1, x_2\}$, $\mathbf{Y} = \{y_1, y_2\}$, $xz = \{x_1, x_2, z\}$, $yz = \{y_1, y_2, z\}$. Since there are many ways in which x_1, x_2, y_1, y_2 , and z can be combined, we have chosen some selective architectures (see Fig. 1). We repeated the simulation 50 times with random τ 's to generate a null distribution. The result is displayed in Fig. 2. We found that both the bivariate and multivariate KCCA Granger causality could detect the correct causality flow well, but the bivariate method failed to distinguish the indirect influences $\mathbf{X} \rightarrow \mathbf{Y}$, which were actually mediated by z , while multivariate KCCA Granger causality recognizes it to be nonsignificant. Besides, the causal influence $\mathbf{X} \rightarrow yz$ is stronger than x_1 and x_2 causal effect on yz , respectively. It indicates that the group effect is different from the individual effect [26,27].

B. Rössler attractors

As another simulated example, we apply the proposed method to the Rössler system. The Rössler equations are given by

$$\begin{aligned} \dot{x}_1 &= -y_1 - z_1, \\ \dot{y}_1 &= x_1 + ay_1, \\ \dot{z}_1 &= b + z_1(x_1 - c), \\ \dot{x}_2 &= -y_2 - z_2 + dx_1, \\ \dot{y}_2 &= x_2 + ay_2, \\ \dot{z}_2 &= b + z_2(x_2 - c) \end{aligned}$$

with $b = 2, c = 4$, and $d = 0.05$; let $\mathbf{x} = \{x_1, x_2\}$, $\mathbf{y} = \{y_1, y_2\}$, and $\mathbf{z} = \{z_1, z_2\}$. For $a = 0.398$ this system exhibits broadband chaos, and for $a = 0.3909$ it exhibits period 6 behavior [28].

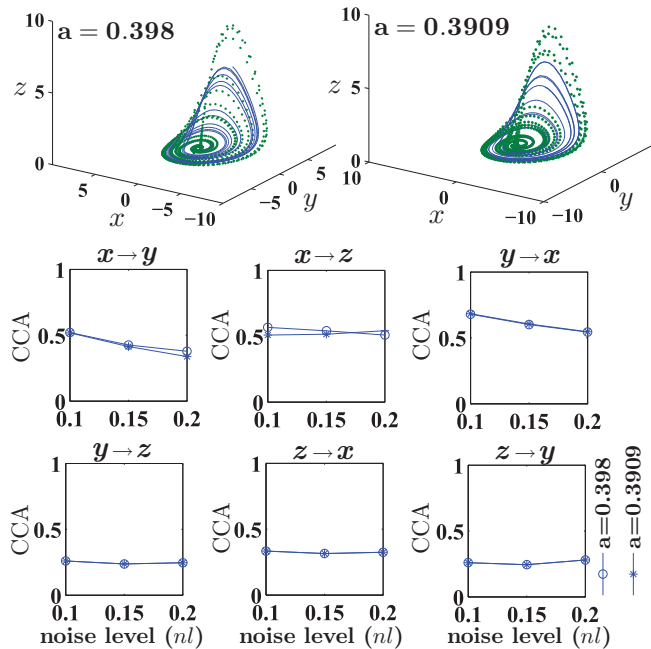


FIG. 3. (Color online) Plots of the trajectory Rössler system for values $b = 2, c = 4$, varied a , and analysis result of multivariate KCCA Granger causality between all pairs of maps in $\{x, y, z\}$. Here $a = 0.398, 0.3909, L = 870$, IP kernel with $d = 2, m = 2$, is used.

For these values of a the data were simulated using the Runge-Kutta method (we used *ode45* function in MATLAB) with L integration steps. Multivariate Gaussian white noise is added to the system, and the noise level nl is defined as the noise standard deviation divided by the standard deviation of the noise-free time series. Here we control $nl = 0.1, 0.15, 0.2, L = 870$. We evaluated multivariate KCCA Granger causality using IP kernel $d = 2$, and fixed lag order 2. The results are displayed in Fig. 3. It reveals that $z \rightarrow y, y \rightarrow z$ are not significant for the above two values of a under 1% significance level, but are unstable or even false to recognize $y \rightarrow z$ as a spurious causal connection for larger L (such as 1000). The result is affected by the noise level; as denoted in [24,29], the detecting ability of causal relation is associated with synchronized state, which is always affected by noise.

C. Coupled chaotic maps

Finally, let us consider a coupled map lattice of 20 nodes, with equations, for $i = 1, \dots, 20$,

$$\begin{aligned} x_{j,t} &= \left(1 - \sum_{i=1}^n c_{ji}\right) (1 - 1.8x_{j,t-1}^2) \\ &+ \sum_{i=1}^n c_{ji} (1 - 1.8x_{i,t-1}^2) + 0.01\tau_{j,t}, \end{aligned}$$

$c_{j,i}$ represents the coupling $i \rightarrow j$. We consider a network consisting of two community structures (C1 and C2), which displays a small-world architecture in the intracommunity. The connectivity scheme is displayed in Fig. 4 (above). We set the value of c_{ji} equal to 0.01, 0.03, 0.05, and generate three numbers of realization: $L = 1500, 2000$, and 2500. Then we evaluated multivariate KCCA Granger causality between

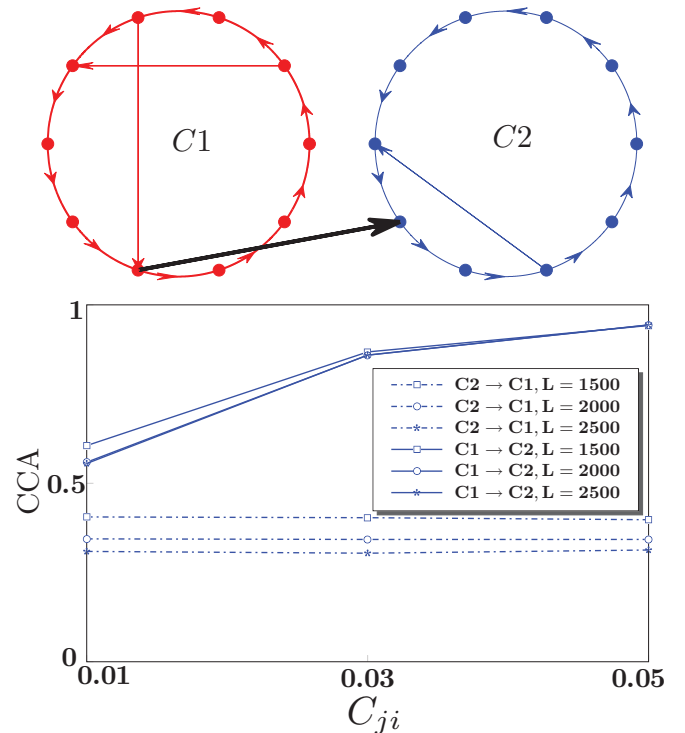


FIG. 4. (Color online) Left: schematic plot for the two community interaction. Right: the analysis result of multivariate KCCA Granger causality with Gaussian kernel, $m = 1$.

the two communities using a Gaussian kernel with adaptive width $\sigma = \max a_{ij}$ [for $\kappa_\sigma(\mathbf{X}, \mathbf{Y})$, where matrix $(a_{ij}) = |(\mathbf{X} - \mathbf{Y})^T(\mathbf{X} - \mathbf{Y})|$], and fixed lag order 1. The analysis result is displayed in Fig. 4 (below). We found a significant causal influence from $C1$ to $C2$.

IV. CONCLUSIONS

In summary, we present a canonical-correlation algorithm for multivariate nonlinear Granger causality analysis based on reproducing kernel Hilbert spaces. Dimension reduction of kernel matrix is added to the algorithm in order to overcome the common problem of the redundancy in the kernel matrix, which enhances its robustness. We expect the

proposed approach to be a simply statistical method to assess the causal relations and could lead to deeper understanding of the underlying information flow under a different scale level in the complex systems, which could be paralleled to extend to other corresponding domains.

ACKNOWLEDGMENTS

The authors thank Daniele Marinazzo (Université Paris Descartes, France; Ghent University, Belgium) for his useful comments on this study. This work was supported by the Natural Science Foundation of China (Grants No. 61035006 and No. 90820006).

-
- [1] A. L. Barabási and R. E. Crandall, *Am. J. Phys.* **71**, 409 (2003).
 [2] M. Chavez, M. Valencia, V. Navarro, V. Latora, and J. Martinerie, *Phys. Rev. Lett.* **104**, 118701 (2010).
 [3] W. Liao *et al.*, *NeuroImage* **54**, 2683 (2011).
 [4] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
 [5] D. Fraiman, P. Balenzuela, J. Foss, and D. R. Chialvo, *Phys. Rev. E* **79**, 061922 (2009).
 [6] J. Geweke, *J. Am. Stat. Assoc.* **79**, 907 (1984).
 [7] D. Marinazzo, M. Pellicoro, and S. Stramaglia, *Phys. Rev. E* **77**, 56215 (2008).
 [8] C. Ladrone, S. Guo, K. Kendrick, and J. Feng, *PloS one* **4**, e6899 (2009).
 [9] S. Bressler and A. Seth, *Neuroimage* (2010) (in press).
 [10] B. Gourévitch, R. L. Bouquin-Jeannès, and G. Faucon, *Biol. Cybern.* **95**, 349 (2006).
 [11] C. A. Sims, *Am. Econ. Rev.* **62**, 540–552 (1972).
 [12] P. Otter, *Econ. Lett.* **35**, 187 (1991).
 [13] J. R. Sato *et al.*, *Neuroimage* **52**, 1444 (2010).
 [14] L. Barnett, A. B. Barrett, and A. K. Seth, *Phys. Rev. Lett.* **103**, 238701 (2009).
 [15] J. Hlinka *et al.*, *Neuroimage* **54**, 2218 (2011).
 [16] C. Granger, *J. Econometrics* **39**, 199 (1988).
 [17] D. Pierce, and L. Haugh, *J. Econometrics* **5**, 265 (1977).
 [18] C. W. J. Granger, *Econometrica: J. Econometric Soc.* **37**, 424 (1969).
 [19] H. Lutkepohl, *New Introduction to Multiple Time Series Analysis* (Springer, New York, 2005).
 [20] D. Pierce, *J. Am. Stat. Assoc.* **74**, 901 (1979).
 [21] Partial out the effect of $X_{t-\text{past}}$. $B_1 = (\hat{Y}_{t-\text{past}}^T \hat{Y}_{t-\text{past}})^{-1} \hat{Y}_{t-\text{past}}^T X_t$, $\hat{Y}_{t-\text{past}} = M \hat{Y}_{t-\text{past}}$, $\hat{X}_t = M X_t$, $M = I - X_{t-\text{past}} (X_{t-\text{past}}^T X_{t-\text{past}})^{-1} X_{t-\text{past}}^T$. For the convenience of analysis, Let \hat{X}_t and $\hat{Y}_{t-\text{past}}$ be zero-mean and unit variances in each column. LCCA between \hat{X}_t and $\hat{Y}_{t-\text{past}}$ is to seek a pair of projective vectors α and β , and make the correlation coefficient $\rho = (\beta^T \beta)^{-1} \beta^T B_1 \alpha$ maximum. Furthermore, $\rho \neq 0 \Leftrightarrow B_1 \neq 0$, which is due to the property of CCA.
 [22] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis* (Cambridge University Press, Cambridge, England, 2004).
 [23] Q. Sun *et al.*, *Pattern Recognition* **38**, 2437 (2005).
 [24] D. Marinazzo, M. Pellicoro, and S. Stramaglia, *Phys. Rev. Lett.* **100**, 144103 (2008).
 [25] M. Ding, G. Chen, and S. L. Bressler, in *Handbook of Time Series Analysis*, edited by B. Schelter, M. Winterhalder, and J. Timmer (Wiley-VCH, New York, 2006).
 [26] L. Angelini *et al.*, *Phys. Rev. E* **81**, 037201 (2010).
 [27] D. Marinazzo *et al.*, *Phys. Lett. A* **374**, 4040 (2010).
 [28] M. Small, D. Yu, and R. Harrison, *Phys. Rev. Lett.* **87**, 188101 (2001).
 [29] M. Paluš *et al.*, *Phys. Rev. E* **63**, 46211 (2001).