This is the accepted manuscript made available via CHORUS. The article has been published as:

Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information

Songting Li, Yanyang Xiao, Douglas Zhou, and David Cai

Phys. Rev. E **97**, 052216 — Published 29 May 2018

DOI: 10.1103/PhysRevE.97.052216

Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information

Songting Li
Courant Institute of Mathematical Sciences, New York University, New York, NY, United States of America

Yanyang Xiao
Courant Institute of Mathematical Sciences, New York University, New York, NY, United States of America
NYUAD Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Douglas Zhou*
School of Mathematical Sciences, MOE-LSC, and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China

David Cai†
Courant Institute of Mathematical Sciences, New York University, New York, NY, United States of America
NYUAD Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
School of Mathematical Sciences, MOE-LSC, and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China

(Dated: May 10, 2018)

The Granger causality (GC) analysis has been extensively applied to infer causal interactions in dynamical systems arising from economy and finance, physics, bioinformatics, neuroscience, social science, and many other fields. In the presence of potential nonlinearity in these systems, the validity of the GC analysis in general is questionable. To illustrate this, here we first construct minimal nonlinear systems and show that the GC analysis fails to infer causal relations in these systems — it gives rise to all types of incorrect causal directions. In contrast, we show that the time-delayed mutual information (TDMI) analysis is able to successfully identify the direction of interactions underlying these nonlinear systems. We then apply both methods to neuroscience data collected from experiments and demonstrate that the TDMI analysis but not the GC analysis can identify the direction of interactions among neuronal signals. Our work exemplifies inference hazards in the GC analysis in nonlinear systems and suggests that the TDMI analysis can be an appropriate tool in such a case.

I. INTRODUCTION

Detecting causal interactions among units in a system is of great importance to understand the cooperative nature of the system. In general, it is methodologically challenging to identify causality only from the measurements of a system without any interventions [1]. A solution was proposed by Wiener based on the idea that the driver is always earlier than the recipient, and the time series of the driver should contain information of the recipient [2]. Therefore, one should expect to improve the prediction of the recipient's state by incorporating the historical information of the driver.

Wiener's principle has been later formalized by Granger in terms of linear regression known as Granger causality (GC) [3]. In particular, if the variance of the prediction error of signal X is reduced by including additionally the history of signal Y in the regressive model, then the causal relation from Y to X is inferred. The GC analysis has become increasingly popular recently with extensive applications in dynamical systems from economy and finance [4–7], physics [8, 9], bioinformatics [10, 11], neuroscience [12–15], social science [16, 17], and many other fields. In general, these systems could be highly nonlinear while the mathematical framework of the GC analysis is established for linear systems. Therefore, the validity of its application in these nonlinear systems is generally questionable.

To overcome the challenge of nonlinearity, another formalization of Wiener's principle has been proposed by Schreiber known as transfer entropy (TE) [18]. In contrast to GC that measures the improvement of signal prediction through variance, TE measures the reduction of signal uncertainty through probability distribution. The TE value from Y to X quantifies the amount of reduced uncertainty of X's state by incorporating the history of both X and Y compared with that only incorporating X's own history. A nonzero TE value infers the existence of causality. TE is an information-theoretic quantity, which makes it applicable in nonlinear systems. In particular, for linear Gaussian models, it has been proved that TE is equivalent to GC [19]. Therefore, TE is a generalization of GC for detecting causality in nonlinear systems. TE has been applied in many nonlinear systems such as neuroscience [20–22] and other fields [23–25]. Despite its applicability in nonlinear systems, in general, TE

* zdz@sjtu.edu.cn
† Dedicated to David Cai

requires long data to construct high-dimension probability distributions, and the dimension of probability distributions in TE is determined by the memory time of the signals. For a system with a long memory, the "curse of dimensionality" will present a great challenge in the application of TE. Variations of TE [20, 26] and related techniques [21] have been developed to improve the applicability of TE to systems with long memory, yet the issue of the reconstruction of high-dimensional probability distribution for a long-memory system remains to be fully resolved.

To alleviate the issue of dimensionality, an implementation of Wiener's principle with variations was introduced by Vastano and Swinney known as time-delayed mutual information (TDMI) [27]. In particular, it computes the mutual information between two signals with multiple time lags. Based on the principle that the driver who contains the information of the recipient is earlier than the recipient, the causal relation is inferred by the sign of the time lag when the value of the mutual information between two signals reaches its peak. Accordingly, this time lag can be interpreted as the time delay of information transport between the two signals. The TDMI analysis was initially proposed to investigate spatiotemporal information transport in physical systems [27], and later it has been applied to infer causal interactions in neuroscience [28–30] and other fields [31–34]. Accounting for the fact that mutual information is an information-theoretic quantity as TE, the application of the TDMI analysis thus is also valid for systems with substantial nonlinearity. For instance, the result of the TDMI analysis is shown to be invariant under any nonlinear invertible transformations of signals [35], i.e., independent of the method of signal measurement.

In this work, we compare the performance of the GC and TDMI analyses by constructing minimal examples of nonlinear systems as well as applying both methods in experimental neuroscience data. The result shows that TDMI can successfully identify the direction of interactions underlying these nonlinear systems while GC fails to identify them. The article is organized as follows. In Sec. *Methods*, we introduce the mathematical framework of GC and TDMI. In Sec. *Results*, we first construct five examples of nonlinear systems and demonstrate that the GC analysis can give rise to all possible types of incorrect causal directions. In contrast, the TDMI analysis is able to successfully identify the correct causal directions, suggesting the applicability of the TDMI analysis in nonlinear systems. We then apply the two methods in neuroscience data recorded in experiments and demonstrate that the TDMI analysis identifies a special group of neurons in rat hippocampus whose firing signal predominantly drives the theta-band (4-12 Hz) local field potential (LFP) signal while the GC analsis fails to label them. In Sec. *Discussion*, we discuss the pros and cons of the GC and TDMI analyses in practical applications.

II. METHODS

A. Granger causality

The mathematical framework of GC is established on linear regression. Given two stationary signals X and Y, one can measure their time traces denoted as $\{x_t\}$ and $\{y_t\}$ at each sampling point, and identify their causal relations using the regressive models described below.

On the one hand, the value of x_t can be predicted from the linear autoregression on its own history

$$x_t = \sum_{i=1}^{\infty} \hat{a}_i x_{t-i} + \hat{\varepsilon}_t,$$

where $\{\hat{a}_i\}$ are the autoregression coefficients, and $\{\hat{\varepsilon}_t\}$ are the corresponding prediction errors.

On the other hand, the value of x_t can also be predicted from the linear regression on both the history of X and Y,

$$x_t = \sum_{i=1}^{\infty} a_i x_{t-i} + \sum_{j=1}^{\infty} b_j y_{t-j} + \varepsilon_t,$$

where $\{a_i\}$, $\{b_j\}$ are the joint regression coefficients, and $\{\varepsilon_t\}$ are the corresponding prediction errors. If Y causally drives X, then the prediction of x_t is expected to be improved by incorporating the history of Y in addition to its own history. In other words, the variance of the prediction error ε_t should be smaller than that of $\hat{\varepsilon}_t$. Inspired by this, the GC value from Y to X is defined as

$$F_{y\to x} = \log \frac{\mathrm{var}(\hat{\varepsilon}_t)}{\mathrm{var}(\varepsilon_t)}.$$

If the GC value vanishes, then $\mathrm{var}(\hat{\varepsilon}_t) = \mathrm{var}(\varepsilon_t)$ and $\{b_j\} = 0$, which indicates that the change of Y will not affect the future of X. Hence Y does not drive X. Otherwise, Y drives X.

In this work, we perform the GC analysis in our constructed examples by following the standard procedure introduced in Ref. [36]. The regression orders of the two signals are determined using Akaike information criterion [39] by following Refs. [36–38]. The significance test is achieved by taking into account the fact that the GC value for a pair of independent signals is asymptotically χ^2 distributed as the data length approaches infinity, and the significance threshold of a nonzero GC value is determined by setting a significance level $\alpha = 0.001$. In the example of neuroscience data, we apply the GC analysis in the frequency domain [36] with the order and significance threshold determined in the identical way as that in the time-domain GC analysis.

B. Time-delayed mutual information

TDMI is an information-theoretic approach for detecting causal interactions. In general, the quantity of mutual information characterizes the common information

shared between two signals. Given two stationary signals X and Y with their time traces $\{x_t\}$ and $\{y_t\}$, respectively, the mutual information between them is defined as

$$I(X,Y)=\sum_{x_t}\sum_{y_t}p(x_t,y_t)\log\frac{p(x_t,y_t)}{p(x_t)p(y_t)},$$

where $p(x_t,y_t)$ is the joint probability distribution of $X=x_t$ and $Y=y_t$, $p(x_t)$ and $p(y_t)$ are their marginal probability distributions. In particular, $I(X,Y)=0$ is equivalent to $p(x_t,y_t)=p(x_t)p(y_t)$, which indicates that signals X and Y are independent if they do not share information.

Mutual information is symmetric, i.e., $I(X,Y)=I(Y,X)$. Therefore, it cannot be applied directly to infer the direction of interactions between two signals. To overcome this limitation, one can introduce a time-lag parameter to capture the delay of information transfer between the two signals. TDMI as a function of time-lag τ is defined as

$$I(X,Y,\tau)=\sum_{x_t}\sum_{y_{t-\tau}}p(x_t,y_{t-\tau})\log\frac{p(x_t,y_{t-\tau})}{p(x_t)p(y_{t-\tau})},$$

where $p(x_t,y_{t-\tau})$ is the joint probability distribution of $X=x_t$ and $Y=y_{t-\tau}$. A nonzero amplitude of the mutual information as a function of τ indicates the existence of interactions between two signals; the sign of the time-lag τ where $I(X,Y,\tau)$ reaches its peak magnitude is used to infer the information flow direction that can be further interpreted as the causal direction of interaction. A negative τ indicates that X shares a maximum amount of information with the future of Y, thus X drives Y. A positive τ indicates that X shares a maximum amount of information with the past of Y, thus X is driven by Y.

In this work, we perform the TDMI analysis in our examples by following the standard procedure introduced in Ref. [29]. In particular, the probability distribution $p(x_t,y_{t-\tau})$ in TDMI is reconstructed by viewing the time traces $\{(x_t,y_{t-\tau})\}$ as a realization of a two-dimensional stationary process governed by the joint probability $p(x_t,y_{t-\tau})$, and the mutual information at a given time lag is calculated via adaptive partitioning of the sample space [40]. To perform the significance test, we randomly shuffle the time series of the two signals separately using Matlab inline function *randperm.m* and calculate the mutual information between the shuffled pair of signals. After shuffling data and calculating mutual information for 100 times, the significance threshold is set as the largest value of the mutual information between shuffled signals. The level of the significance threshold is not shown in the figures in Sec. *Results* because it is too small compared with the peaks of any TDMI curve.

The TDMI code is available at `http://ins.sjtu.edu.cn/people/zdz/code/GC_vs_TDMI.RAR`.

TABLE I. Summary of constructed linear systems in each example.

Example	linear system
A	$\begin{cases} x_t=\varepsilon_t \\ y_t=-0.1x_{t-1}+\eta_t \end{cases}$
B	$\begin{cases} x_t=-0.3x_{t-1}+\varepsilon_t \\ y_t=0.3y_{t-1}-0.9x_{t-1}+\eta_t \end{cases}$
C	$\begin{cases} x_t=-\sum_{k=1}^{8}c_kx_{t-k}+\varepsilon_t \\ y_t=-\sum_{k=1}^{8}c_ky_{t-k}+100\sum_{k=1}^{9}c_{k-1}x_{t-k}+\eta_t \end{cases}$
D	$\begin{cases} x_t=-0.1y_{t-1}+\varepsilon_t \\ y_t=-0.1x_{t-1}+\eta_t \end{cases}$
E	$\begin{cases} x_t=-\sum_{k=1}^{8}c_kx_{t-k}+0.5\sum_{k=1}^{9}c_{k-1}y_{t-k}+\varepsilon_t \\ y_t=-\sum_{k=1}^{8}c_ky_{t-k}+0.5\sum_{k=1}^{9}c_{k-1}x_{t-k}+\eta_t \end{cases}$

III. RESULTS

In this Section, we first construct five minimal examples to study whether GC and TDMI can capture the causal interactions in these systems. In each example, we start with a two-dimensional linear dynamical system describing the dynamics of signals X and Y, and then construct the corresponding nonlinear system by performing a static nonlinear transform to each data points of signal X generated from the linear system. We show that GC may easily fail to detect causal interactions in these examples, while TDMI can successfully detect them. For the ease of reading, all the results have been summarized in TABLE I and TABLE II.

We then investigate the performance of the two methods when applied to neuroscience data measured in rat hippocampus, and show that TDMI is able to identify nonlinear interactions among neuronal signals while GC cannot. In particular, TDMI identifies a special group of neurons in rat hippocampus whose firing signal predominantly drives the theta-band LFP signal. These neurons have been discovered in mouse hippocampus in a previous study [29], and in this work their existence in rat hippocampus has also been confirmed.

A. unidirection misinferred as no interaction

In this example, we start with the following linear dynamical system,

$$\begin{cases} x_t=\varepsilon_t \\ y_t=-0.1x_{t-1}+\eta_t \end{cases}$$

TABLE II. Summary of GC and TDMI results in each example.

Example	Threshold	$F_{x\to y}$	$F_{y\to x}$	GC direction	Nonlinearity	$F_{\tilde{x}\to y}$	$F_{y\to\tilde{x}}$	GC direction	TDMI direction
A	1.1×10^{-6}	1.0×10^{-2}	6.0×10^{-8}	$X\to Y$	$\tilde{x}_t=x_t^2$	1.0×10^{-8}	2.8×10^{-7}	independent	$X(\tilde{X})\to Y$
B	1.9×10^{-6}	6.2×10^{-1}	7.5×10^{-7}	$X\to Y$	$\tilde{x}_t=[\frac{x_t+\lvert x_t\rvert}{2}]^5$	5.7×10^{-2}	3.0×10^{-4}	$\tilde{X}\leftrightarrow Y$	$X(\tilde{X})\to Y$
C	5.1×10^{-5}	4.2×10^{-1}	1.8×10^{-5}	$X\to Y$	$\tilde{x}_t=x_t^5$	3.6×10^{-5}	2.3×10^{-2}	$\tilde{X}\leftarrow Y$	$X(\tilde{X})\to Y$
D	1.4×10^{-6}	9.8×10^{-3}	9.9×10^{-3}	$X\leftrightarrow Y$	$\tilde{x}_t=x_t^2$	1.0×10^{-8}	3.1×10^{-7}	independent	$X(\tilde{X})\leftrightarrow Y$
E	4.5×10^{-5}	2.2×10^{-1}	2.2×10^{-1}	$X\leftrightarrow Y$	$\tilde{x}_t=\tanh(10x_t)$	2.1×10^{-5}	1.1×10^{-1}	$\tilde{X}\leftarrow Y$	$X(\tilde{X})\leftrightarrow Y$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent and identically distributed (i.i.d) standard Gaussian random variables. A realization of this linear system generates the time series $\{x_t\}$ and $\{y_t\}$ with data length 10^7. By performing the GC analysis to $\{x_t\}$ and $\{y_t\}$, the GC value from X to Y is $F_{x\to y}\approx 1.0\times10^{-2}$ and that from Y to X is $F_{y\to x}\approx 6.0\times10^{-8}$. Accounting for the fact that the significance threshold for non-zero GC value is $F_{thr}\approx 1.1\times10^{-6}$ (See Sec. *Methods* for details), the GC analysis identifies the direction of causal interaction from X to Y but not from Y to X, which is consistent with the underlying dynamics of the linear system.

By performing the quadratic transform to signal X, we can obtain a new signal $\tilde{X}$ with its realization as $\tilde{x}_t=x_t^2$. The GC analysis between $\{\tilde{x}_t\}$ and $\{y_t\}$ gives that $F_{\tilde{x}\to y}\approx 1.0\times10^{-8}$ and $F_{y\to\tilde{x}}\approx 2.8\times10^{-7}$. Both of the GC values are below the significance threshold $F_{thr}\approx 1.1\times10^{-6}$, indicating that there is no interaction between the signals $\tilde{X}$ and Y. This result given by GC is obviously incorrect because $\tilde{X}$ and Y are closely linked by X rather than being independent of each other.

In contrast, by performing the TDMI analysis, the causal direction of interactions in both the linear and nonlinear systems can be correctly identified. From Fig. 1, both the TDMI curves reach a significant large peak at the negative time lag $\tau=-1$, corresponding to the delay time of the information transfer from $X(\tilde{X})$ to Y. This indicates that the direction of causal interaction is always from $X(\tilde{X})$ to Y in both the linear and nonlinear systems, not being disturbed by the nonlinear transform.

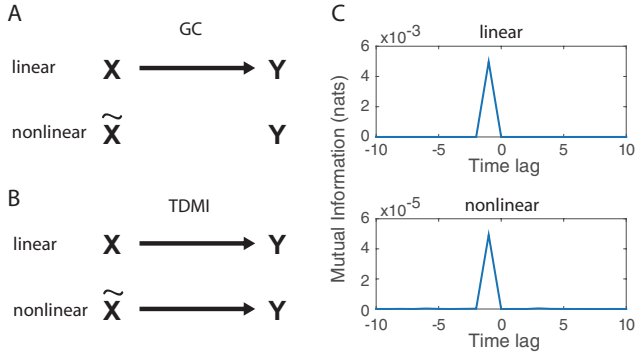

FIG. 1. (A) Causal direction identified by GC. (B) Causal direction identified by TDMI. (C) TDMI curve as a function of time lag. The upper and lower parts in each panel correspond to the linear and nonlinear systems in Example A, respectively.

B. unidirection misinferred as bidirection

In this example, we start with the following linear dynamical system,

$$\begin{cases} x_t=-0.3x_{t-1}+\varepsilon_t \\ y_t=0.3y_{t-1}-0.9x_{t-1}+\eta_t \end{cases}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are i.i.d standard Gaussian random variables. A realization of this linear system generates the time series $\{x_t\}$ and $\{y_t\}$ with data length 10^7. By performing the GC analysis to $\{x_t\}$ and $\{y_t\}$, the GC value from X to Y is $F_{x\to y}\approx 6.2\times10^{-1}$ and that from Y to X is $F_{y\to x}\approx 7.5\times10^{-7}$. Accounting for the fact that the significance threshold for non-zero GC value is $F_{thr}\approx 1.9\times10^{-6}$, the GC analysis identifies the direction of causal interaction from X to Y but not from Y to X, which is consistent with the underlying dynamics of the linear system.

By performing the nonlinear transform $f(x)=[(x+|x|)/2]^5$ to signal X, we can obtain a new signal $\tilde{X}$ with its realization as $\tilde{x}_t=[(x_t+|x_t|)/2]^5$. The GC analysis between $\{\tilde{x}_t\}$ and $\{y_t\}$ gives that $F_{\tilde{x}\to y}\approx 5.7\times10^{-2}$ and $F_{y\to\tilde{x}}\approx 3.0\times10^{-4}$. Both of the GC values are above the significance threshold $F_{thr}\approx 1.9\times10^{-6}$, indicating that there are bidirectional interactions between the signals $\tilde{X}$ and Y. This result given by GC is obviously incorrect because by no means will Y affect $\tilde{X}$.

In contrast, by performing the TDMI analysis, the causal direction of interactions in both the linear and nonlinear systems can be correctly identified. From Fig. 2, both the TDMI curves reach a significant large peak at the negative time lag $\tau=-1$. This indicates that the direction of causal interaction is always from $X(\tilde{X})$ to Y in both the linear and nonlinear systems, not being disturbed by the nonlinear transform.

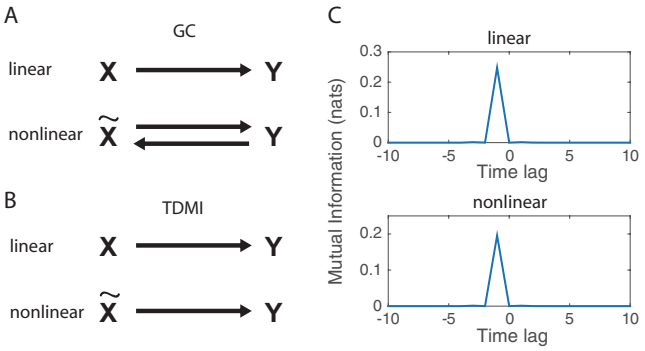

FIG. 2. (A) Causal direction identified by GC. (B) Causal direction identified by TDMI. (C) TDMI curve as a function of time lag. The upper and lower parts in each panel correspond to the linear and nonlinear systems in Example B, respectively.

C. unidirection misinferred as reversed unidirection

In this example, we start with the following linear dynamical system,

$$\begin{cases} x_t = -\sum_{k=1}^{8} c_k x_{t-k} + \varepsilon_t \\ y_t = -\sum_{k=1}^{8} c_k y_{t-k} + 100\sum_{k=1}^{9} c_{k-1} x_{t-k} + \eta_t \end{cases}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are Gaussian random variables with zero mean and covariance as $\mathrm{var}(\varepsilon_t, \eta_t) = \begin{pmatrix} 5.16\times10^{-6} & 0 \\ 0 & 1 \end{pmatrix}$, and the coefficients $\{c_k\}$ is given by the following polynomial

$$\sum_{k=0}^{8} c_k z^k = [(1 - re^{-2\pi if} z)(1 - re^{2\pi if} z)]^4$$

with $f = 0.1$ and $r = 0.8$.

A realization of this linear system generates the time series $\{x_t\}$ and $\{y_t\}$ with data length 10^6. By performing the GC analysis to $\{x_t\}$ and $\{y_t\}$, the GC value from X to Y is $F_{x\to y} \approx 4.2\times10^{-1}$ and that from Y to X is $F_{y\to x} \approx 1.8\times10^{-5}$. Accounting for the fact that the significance threshold for non-zero GC value is $F_{thr} \approx 5.1\times10^{-5}$, the GC analysis identifies the direction of causal interaction from X to Y but not from Y to X, which is consistent with the underlying dynamics of the linear system.

By performing the nonlinear quintic transform $f(x) = x^5$ to signal X, we can obtain a new signal $\tilde{X}$ with its realization as $\tilde{x}_t = x_t^5$. The GC analysis between $\{\tilde{x}_t\}$ and $\{y_t\}$ gives that $F_{\tilde{x}\to y} \approx 3.6\times10^{-5}$ and $F_{y\to\tilde{x}} \approx 2.3\times10^{-2}$. Therefore, the GC value $F_{\tilde{x}\to y}$ is above the significance threshold $F_{thr} \approx 5.1\times10^{-5}$ while the GC value $F_{y\to\tilde{x}}$ is below the significance threshold. This indicates that, after the nonlinear transform, the causal direction changes from that X drives Y to that Y drives $\tilde{X}$, which is obviously inconsistent with the dynamics.

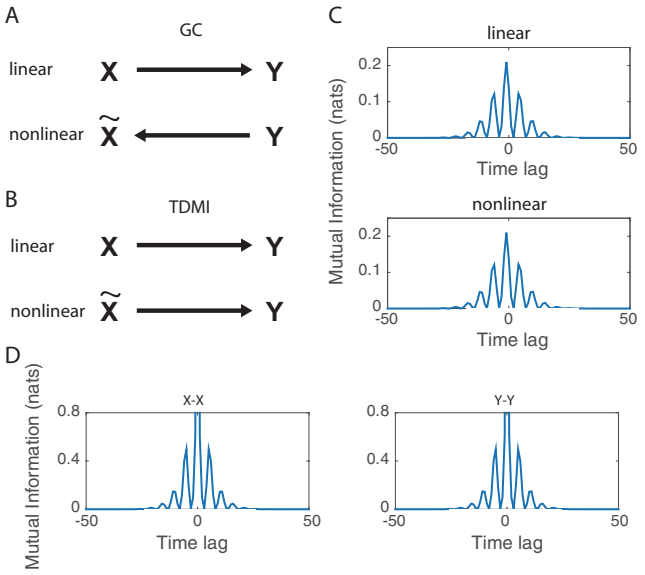

FIG. 3. (A) Causal direction identified by GC. (B) Causal direction identified by TDMI. (C) TDMI curve as a function of time lag. The upper and lower parts in each panel correspond to the linear and nonlinear systems in Example C, respectively. (D) TDMI curve between X and itself (left) and that between Y and itself (right). The global peak of the TDMI curves at zero time lag is truncated in order to better visualize the local peaks of small amplitude.

In contrast, by performing the TDMI analysis, the causal direction of interactions in both the linear and nonlinear systems can be correctly identified. From Fig. 3, both the TDMI curves reach a significant non-zero global peak at a negative time lag, indicating that the direction of causal interaction is always from $X(\tilde{X})$ to Y in both the linear and nonlinear systems, not being disturbed by the nonlinear transform.

We note that there are also several local peaks at some positive time lags. To interpret these peaks, we calculate the TDMI between X and itself, and that between Y and itself. As shown in Fig. 3D, both of the curves show the highly similar decaying profile and identical number of oscillatory periods to the TDMI feature in Fig. 3C, indicating that these positive peaks in Fig. 3C are probably induced by the memory of the signal itself rather than the causal information flow from Y to X. The interpretation of local peaks in TDMI curves will be further discussed in Sec. *Discussion*.

D. bidirection misinferred as no interaction

In this example, we start with the following linear dynamical system,

$$\begin{cases} x_t = -0.1 y_{t-1} + \varepsilon_t \\ y_t = -0.1 x_{t-1} + \eta_t \end{cases}$$

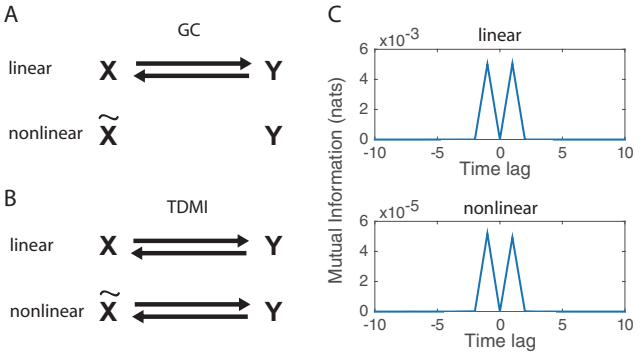

FIG. 4. (A) Causal direction identified by GC. (B) Causal direction identified by TDMI. (C) TDMI curve as a function of time lag. The upper and lower parts in each panel correspond to the linear and nonlinear systems in Example D, respectively.

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are i.i.d standard Gaussian random variables. A realization of this linear system generates the time series $\{x_t\}$ and $\{y_t\}$ with data length 10^7. By performing the GC analysis to $\{x_t\}$ and $\{y_t\}$, the GC value from X to Y is $F_{x\to y} \approx 9.8 \times 10^{-3}$ and that from Y to X is $F_{y\to x} \approx 9.9 \times 10^{-3}$. Accounting for the fact that the significance threshold for non-zero GC value is $F_{thr} \approx 1.4\times10^{-6}$, the GC analysis identifies bidirectional causal interactions between X and Y, which is consistent with the underlying dynamics of the linear system.

By performing the quadratic transform to signal X, we can obtain a new signal $\tilde{X}$ with its realization as $\tilde{x}_t = x_t^2$. The GC analysis between $\{\tilde{x}_t\}$ and $\{y_t\}$ gives that $F_{\tilde{x}\to y} \approx 1.0 \times 10^{-8}$ and $F_{y\to\tilde{x}} \approx 3.1 \times 10^{-7}$. Both of the GC values are below the significance threshold $F_{thr} \approx 1.4\times10^{-6}$, indicating that there is no interaction between the signals $\tilde{X}$ and Y. This result given by GC is obviously incorrect because $\tilde{X}$ and Y are closely linked by X rather than being independent of each other.

In contrast, by performing the TDMI analysis, the causal direction of interactions in both the linear and nonlinear systems can be correctly identified. From Fig. 4, both the TDMI curves reach two significant large peaks, one at positive lag $\tau = 1$ and the other at negative time lag $\tau = -1$. This indicates that the direction of causal interactions between $X(\tilde{X})$ and Y are always bidirectional in both the linear and nonlinear systems, not being disturbed by the nonlinear transform.

E. bidirection misinferred as unidirection

In this example, we start with the following linear dynamical system,

$$\begin{cases} x_t = -\sum_{k=1}^{8} c_k x_{t-k} + 0.5\sum_{k=1}^{9} c_{k-1} y_{t-k} + \varepsilon_t \\ y_t = -\sum_{k=1}^{8} c_k y_{t-k} + 0.5\sum_{k=1}^{9} c_{k-1} x_{t-k} + \eta_t \end{cases}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are the same Gaussian random variables defined in Example C, and the coefficients $\{c_k\}$ are given by the same polynomial in Example C.

A realization of this linear system generates the time series $\{x_t\}$ and $\{y_t\}$ with data length 10^6. By performing the GC analysis to $\{x_t\}$ and $\{y_t\}$, the GC value from X to Y is $F_{x\to y} \approx 2.2 \times 10^{-1}$ and that from Y to X is $F_{y\to x} \approx 2.2 \times 10^{-1}$. Accounting for the fact that the significance threshold for non-zero GC value is $F_{thr} \approx 4.5\times10^{-5}$, the GC analysis identifies bidirectional causal interactions between X and Y, which is consistent with the underlying dynamics of the linear system.

By performing the nonlinear hyperbolic tangent transform $f(x) = \tanh(10x)$ to signal X, we can obtain a new signal $\tilde{X}$ with its realization as $\tilde{x}_t = \tanh(10x_t)$. The GC analysis between $\{\tilde{x}_t\}$ and $\{y_t\}$ gives that $F_{\tilde{x}\to y} \approx 2.1 \times 10^{-5}$ and $F_{y\to\tilde{x}} \approx 1.1 \times 10^{-1}$. Therefore, the GC value $F_{y\to\tilde{x}}$ is above the significance threshold $F_{thr} \approx 4.5 \times 10^{-5}$ while the GC value $F_{\tilde{x}\to y}$ is below the significance threshold. This indicates that, after the nonlinear transform, the causal direction changes from bidirectional to unidirectional, i.e., from Y to $\tilde{X}$, which is obviously incorrect because $\tilde{X}$ also drives Y.

In contrast, by performing the TDMI analysis, the causal direction of interactions in both the linear and nonlinear systems can be correctly identified. From Fig. 5, the TDMI curves have global peaks at both positive and negative time lags. This indicates that the directions of causal interaction are always bidirectional in both the linear and nonlinear systems, not being disturbed by the nonlinear transform.

F. interactions between neuronal signals

From the above examples, we have shown that TDMI rather than GC is able to successfully identify causal interactions in our constructed minimal nonlinear systems. Yet the applicability of TDMI in high-dimensional, complex, real systems remains to be evaluated. To illustrate this, here we apply TDMI to detect causal interactions between neuronal signals using data recorded in neuroscience experiments [41, 42] and contrast its performance with GC.

In the experiments [41, 42], multichannel extracellular recordings were performed in hippocampal CA1 area of three rats during open field tasks. Spike signals of

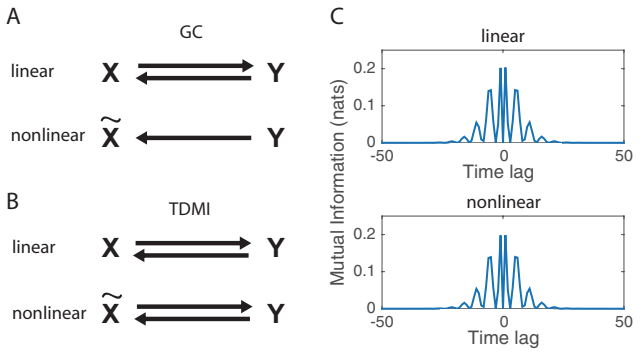

FIG. 5. (A) Causal direction identified by GC. (B) Causal direction identified by TDMI. (C) TDMI curve as a function of time lag. The upper and lower parts in each panel correspond to the linear and nonlinear systems in Example E, respectively. The mutual information at zero time lag is set as zero since it contains little information of the causal direction.

each neuron and LFP signals reflecting neuronal population activity were extracted from the raw data by bandpass filtering and spike sorting. Examples of the spike and LFP signals are shown in Fig. 6A and Fig. 6B, respectively. The duration of each recording session lasted from 17 minutes to 1 hour 46 minutes, providing us with more than 10^6 data points to perform the TDMI analysis. Additional experimental details can be found in Refs. [41, 42].

We first perform power spectrum analysis to the LFP and spike signals and find that both signals have substantially strong energy at the theta band (4-12 Hz), as shown in Fig. 6C. This fact inspires us to study causal interactions between the spike activity of each neuron and the theta-band LFP signals, which could provide us with insights into the generating mechanisms underlying the theta-band wave. For instance, it is yet unclear whether the theta wave is generated by the majority of the neurons or only a small group of neurons in the network.

By performing the TDMI analysis between the spike signal of each selected neuron (Fig. 6A) and the corresponding theta wave filtered from the LFP signal (Fig. 6D), we discover that only 10 out of ~ 130 recorded neurons share substantial information with the theta wave, thus interact strongly with the theta wave. In addition, only two out of the ten neurons possess negative time lags at the global peak of the TDMI curve through the full recording session (Fig. 6E), indicating that their spike activity consistently drives the theta wave. The two neurons are thereafter named as "theta-driving neurons". The result indicates that only a small group of neurons participate in the generation of the theta wave.

As shown in Fig. 6E, in addition to the global peak, the TDMI curve has a decaying oscillatory pattern with several local peaks at which the corresponding time lags are positive. To interpret these peaks, we perform the TDMI analysis between the theta wave and itself. As shown in Fig. 6F, we find that the self-TDMI curve exhibits similar decaying profile and identical number of oscillatory periods as the TDMI curve has in Fig. 6E. Therefore, the local peaks are likely induced by the memory of the the theta wave signal rather than the causal drive from the theta wave to the spike activity of the theta-driving neuron.

To further investigate whether the interaction between the theta-driving neuron and the theta wave is predominant compared with other frequency bands of waves, we calculate the TDMI between the theta-driving neuron's firing activity and waves of different frequency bands filtered from the LFP signal including the delta wave (1-4 Hz), the beta wave (12-30 Hz), the gamma wave (30-100 Hz), and the ripple wave (100-250 Hz). Examples of these filtered waves are shown in Fig. 6D. As shown in Fig. 6E, except for the case of the theta wave, the amount of information shared between the theta-driving neuron and the other waves is almost negligible. Therefore, the theta-driving neuron interacts and drives the theta wave predominantly.

For the same group of theta-driving neurons, the GC analysis draws the opposite conclusion from the TDMI analysis. As shown in Fig. 6G, after applying the GC analysis between the spike signal and the original LFP signal in the frequency domain, the interaction between them is inferred to be bidirectional, and the interaction from LFP to spike is even stronger than that from spike to LFP by comparing their corresponding amplitudes of GC values as a function of frequency. In addition, as shown in Fig. 6G, the interaction between the theta-driving neuron and the LFP signal exists across a wide range of frequency from delta band to gamma band, which is inconsistent with the TDMI analysis showing that the theta-driving neuron predominantly drives the theta-band wave.

The feature of the theta-driving neuron in *rat* hippocampus discovered here is very similar to the feature of neurons discovered in *mouse* hippocampus in a previous experiment [29, 43]. As shown in Fig. 6D, the theta-driving neuron fires action potentials that are phase-locked to the theta wave, and it fires preferentially in the ascending phase of the theta wave. The theta-driving neuron has a relatively high firing rate ~ 40 Hz contrast to the remaining neurons ~ 1 Hz recorded in the experiments. The total number of theta-driving neurons are less than 5% in the recorded population. All these features have been observed in mouse theta-driving neurons. Therefore, although the result of the TDMI analysis cannot be thoroughly verified since the underlying detailed dynamics is unknown, the consistent results of the TDMI analysis from the two independent experiments [29, 41, 42] validates TDMI to a certain extent. To fully validate the result of TDMI, one needs to rely on further experiments, e.g., those using optogenetic tools to control the activity of neurons [44]. To be specific, the inferred direction from the theta-driving neuron to the theta wave by TDMI can be thoroughly validated if the theta wave appears (disappears) when the activity of the

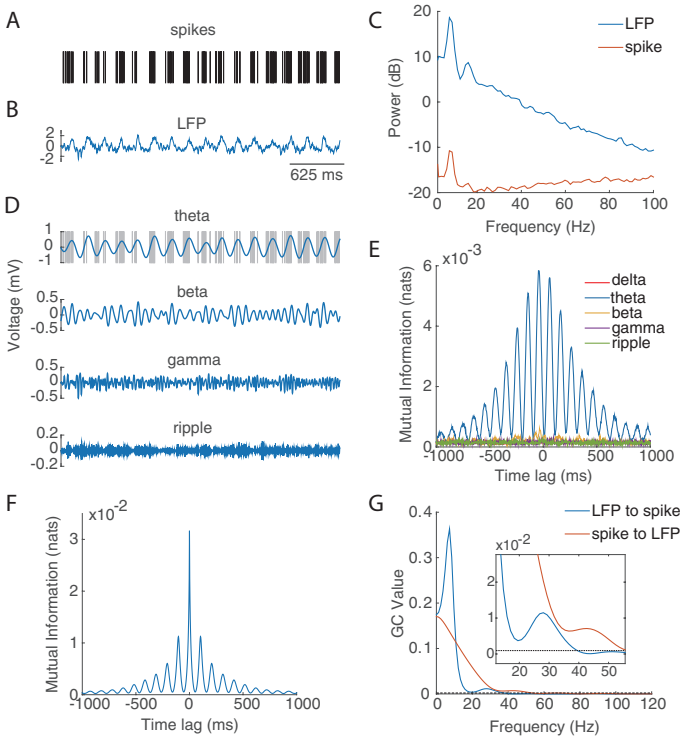

FIG. 6. (A) the spike signal of a theta-driving neuron. (B) the local field potential signal. (C) The power spectrum density of the spike and LFP signals. (D) the corresponding waves of various frequencies filtered from the LFP signal in *B*. The grey bars in the background of the theta wave is the spike train of the theta-driving neuron in *A*. (E) The TDMI curve between the theta-driving neuron's spike and the waves of various frequencies. (F) The self-TDMI curve for the theta wave. (G) The GC curves between the spike and LFP signals in both directions in the frequency domain. Dashed line is the significance threshold. Inset is the zoom in of the regime close to the significance threshold.

theta-driving neuron is evoked (inhibited).

IV. DISCUSSIONS

In this work, we have first constructed five examples of nonlinear systems in which the GC analysis fails to infer the causal interactions and gives rise to all possible types of incorrect directions. The inference failure of GC for causal interactions in nonlinear systems can be understood as follows. In our previous work, the GC value is shown to be well approximated by the summation of squared correlation coefficients [45, 46]. Because the correlation coefficients are not invariant under nonlinear transforms, the GC analysis in general fails to capture the causal information flow in nonlinear systems. In contrast, TDMI is invariant under invertible static nonlinear transforms [35] and is capable of identifying the correct causal directions in nonlinear systems consistent with the underlying dynamics.

The mathematical framework of GC is built on the linear regressive model. However, in practice, the GC analysis has been applied in nonlinear systems arising from a large number of fields. This may lead to inconsistent results with the underlying dynamics. The inferred causality is generally interpreted as effective causality, which is acceptable to be different from the underlying dynamics. However, through our examples, we demonstrate that the meaning of such effective causality could be confusing and unclear. For example, when applying the GC analysis in a nonlinear system, a pure recipient signal could be inferred as the effective cause of a pure driver signal.

Yet it is important to point out that the GC analysis could still be valid when being applied to certain particular nonlinear systems. For example, in neuroscience, despite the fact that the underlying neuronal dynamics is nonlinear, the application of the GC analysis to neuronal voltage signals under certain conditions has been shown to be able to successfully reconstruct the anatomical connectivity of neuronal networks [45, 46]. In addition, the reason why the GC analysis is applicable in this case has been fully elucidated by analyzing the linear subthreshold and nonlinear firing-reset structures of neuronal dynamics [45, 46]. In contrast, it has also been shown that, the results from the GC analysis can be difficult to interpret without examining the component behaviors of the system model [47]. Therefore, the GC analysis is not applicable to a nonlinear system unless one has sufficient *a priori* knowledge of the system.

Our work is partly similar to a recent study [47] that GC can fail to identify causal direction practically. However, our work emphasizes on different aspects from the above study. In Ref. [47], the failure of GC is due to the computational approach of Granger-Geweke causality even when the underlying model is a linear system. Granger-Geweke causality estimates can be severely biased or of high variance. The issues of bias and high variance demonstrated in Ref. [47] can be resolved using different computational approaches such as the state space approach [48], as pointed out by other studies [49, 50]. In contrast, our work demonstrates that GC in general fails to capture the causal interactions in the presence of nonlinearity, which cannot be simply resolved by using those computational approaches, e.g., the state space approach. To capture the causal relations in systems with nonlinearity, we have introduced the method of TDMI.

TDMI is an information-theoretic approach, therefore, it can be applied to nonlinear systems to infer causality. In addition, the result of the TDMI analysis is invariant under any nonlinear invertible transformations of signals [35], thus it is independent of the method of the signal measurement. This fact can be observed in Example C and Example E, in which the nonlinear transform is monotonic and the TDMI curves corresponding to both the linear and nonlinear systems have exactly the same profile.

The expression of TDMI can also be casted as

$$I(X,Y,\tau)=\sum_{x_t}\sum_{y_{t-\tau}}p(x_t,y_{t-\tau})\log\frac{p(x_t|y_{t-\tau})}{p(x_t)}.$$

It measures the difference between $p(x_t|y_{t-\tau})$ and $p(x_t)$ that quantifies the reduction of the uncertainty of X's future by incorporating Y's history. Different from GC, TDMI does not exclude the effect induced by its own history. Therefore, except for the global peaks, it is possible that the non-vanishing local peaks of the TDMI curve is attributed to the history of the signal itself rather than the causal information flow from another signal. Therefore, the physical interpretation of the non-vanishing local peaks in general requires one to further investigate the underlying structure of dynamics in nonlinear systems. For instance, in Example C, if one has the prior knowledge that the interaction in the system is unidirectional, then the direction of interaction can be easily identified as from X to Y by the global peak. Furthermore, if one has the prior knowledge that the directed interaction between any two signals in the system, if exists, possesses the identical form, then the bidirectional (unidirectional) interaction can be easily identified from the symmetric (asymmetric) profile of TDMI. In certain cases, the underlying structure can also be revealed by computing the TDMI between the signal and itself, as shown in our Example C.

In addition to TDMI, TE is another measure of information transport from signal Y to signal X with the exclusion of the influence of the X's own history [18]. It has further been shown that TE is equivalent to GC for linear Gaussian models [19]. The value of TE from signal Y to signal X is defined as

$$TE_{Y\to X}=\sum_{x_t}\sum_{x_t^-}\sum_{y_t^-}p(x_t,x_t^-,y_t^-)\log\frac{p(x_t|x_t^-,y_t^-)}{p(x_t|x_t^-)},$$

where $x_t^-=\{x_{t-1},x_{t-2},...,x_{t-m}\}$, $y_t^-=\{y_{t-1},y_{t-2},...,y_{t-n}\}$ are the history of x_t and y_t, respectively. In principle, TE can eliminate the effect of the history of the two signals while TDMI cannot. In practice, however, TE requires much longer data length to construct the high-dimension probability density distribution, i.e., $p(x_t,x_t^-,y_t^-)$. The dimension of probability distributions in TE is determined by the number of time lags between X and Y and that between Y and its own history, whereas the dimension of probability distribution in TDMI is only two. For instance, in Examples C, E, and F, the dimension of the conditional probability distributions is greater than ten. If the value of the continuous signals X and Y are discretized into 10^2 levels respectively, then the probability sample space will be greater than 10^{20}, which makes TE difficult to implement. Therefore, TDMI is a more practicable approach than TE for causal inference in nonlinear systems by using relatively short length of observational data.

The TDMI analysis is applicable to systems with dynamic nonlinearity. In our examples, we use static nonlinearity because it commonly appears in many engineering and biological systems such as artificial and biological neural networks. In addition to systems with static nonlinearity, TDMI also works for certain systems with dynamic nonlinearity especially when the absolute value of the coefficient decreases rapidly as the time lag increases. For instance, it can be shown that TDMI is able to successfully identify the directions of interactions in each of the five constructed examples when the corresponding static nonlinearity $F(x_t)$ is replaced by the dynamic nonlinearity $F(c_tx_t+c_{t-1}x_{t-1}+c_{t-2}x_{t-2})$, where $c_t=0.85, c_{t-1}=0.1$, and $c_{t-2}=0.05$.

The TDMI analysis is applicable in high dimensional complex systems. In our Example F, by applying TDMI to neural data measured in experiment, we have confirmed the existence of the theta-driving neuron in rat hippocampus that has been reported in mouse hippocampus previously [29, 43]. In addition, the TDMI analysis can be successfully applied to neuronal spike signals to reconstruct the connectivity matrix of a neuronal network, in which the accuracy could be above 95% in certain dynamical regimes [51].

ACKNOWLEDGMENTS

This work was supported by NYU Abu Dhabi Institute G1301 (S.L., Y.X., D.Z., and D.C.), NSFC-11671259, NSFC-11722107, NSFC-91630208, Shanghai Rising-Star Program-15QA1402600 (D.Z.), NSFC 31571071, NSF DMS-1009575 (D.C.), Shanghai 14JC1403800, Shanghai 15JC1400104, SJTU-UM Collaborative Research Program (D.Z. and D.C.).

[1] J. Pearl, Econometric Theory **19**, 46 (2003).
[2] E. F. Beckenbach and R. Weller, *Modern Mathematics for the Engineer: First Series* (Courier Corporation, 2013).
[3] C. W. Granger, Econometrica: Journal of the Econometric Society , 424 (1969).
[4] S. Z. Chiou-Wei, C.-F. Chen, and Z. Zhu, Energy Economics **30**, 3063 (2008).
[5] K. Saidi and M. B. Mbarek, Progress in Nuclear Energy **88**, 364 (2016).
[6] S. Hossain, European Scientific Journal, ESJ **8** (2014).
[7] R. Horváth, J. Seidler, and L. Weill, Journal of Financial Services Research **45**, 341 (2014).
[8] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Physics Letters A **324**, 26 (2004).
[9] T. J. Mosedale, D. B. Stephenson, M. Collins, and T. C. Mills, Journal of climate **19**, 1182 (2006).

[10] N. D. Mukhopadhyay and S. Chatterjee, Bioinformatics **23**, 442 (2006).
[11] R. Nagarajan and M. Upreti, Statistical applications in genetics and molecular biology **9** (2010).
[12] A. K. Seth, A. B. Barrett, and L. Barnett, Journal of Neuroscience **35**, 3293 (2015).
[13] M. Ding, Y. Chen, and S. L. Bressler, Handbook of time series analysis: recent theoretical developments and applications **437** (2006).
[14] S. L. Bressler and A. K. Seth, Neuroimage **58**, 323 (2011).
[15] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, Proceedings of the National Academy of Sciences of the United States of America **101**, 9849 (2004).
[16] J. R. Freeman, American Journal of Political Science , 327 (1983).
[17] L. Keele, American Journal of Political Science **51**, 241 (2007).
[18] T. Schreiber, Physical review letters **85**, 461 (2000).
[19] L. Barnett, A. B. Barrett, and A. K. Seth, Physical review letters **103**, 238701 (2009).
[20] M. Staniek and K. Lehnertz, Physical Review Letters **100**, 158101 (2008).
[21] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, Journal of computational neuroscience **30**, 45 (2011).
[22] M. Wibral, R. Vicente, M. Lindner, *et al.*, *Transfer entropy in neuroscience* (Springer-Verlag: Berlin, Germany, 2014).
[23] A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks, Computational Economics **47**, 341 (2016).
[24] L. Sandoval, Entropy **16**, 4443 (2014).
[25] T. Dimpfl and F. J. Peter, Studies in Nonlinear Dynamics and Econometrics **17**, 85 (2013).
[26] H. Dickten and K. Lehnertz, Physical Review E **90**, 062706 (2014).
[27] J. A. Vastano and H. L. Swinney, Physical Review Letters **60**, 1773 (1988).
[28] W. Endo, F. P. Santos, D. Simpson, C. D. Maciel, and P. L. Newland, Journal of computational neuroscience **38**, 427 (2015).
[29] S. Li, J. Xu, G. Chen, L. Lin, D. Zhou, and D. Cai, Scientific Reports **7** (2017).
[30] A. Destexhe, Physical Review E **50**, 1594 (1994).
[31] M.-C. Ho and F.-C. Shin, Physical Review E **67**, 056214 (2003).
[32] J. Nichols, Proceedings of the Royal Society of London B: Biological Sciences **272**, 871 (2005).
[33] J. Nichols, M. Seaver, and S. Trickey, Journal of Sound and Vibration **297**, 1 (2006).
[34] C. Kirst, M. Timme, and D. Battaglia, Nature communications **7** (2016).
[35] A. Kraskov, H. Stögbauer, and P. Grassberger, Physical review E **69**, 066138 (2004).
[36] J. Geweke, Journal of the American statistical association **77**, 304 (1982).
[37] L. Barnett and A. K. Seth, Journal of neuroscience methods **223**, 50 (2014).
[38] M. Ding, Y. Chen, and S. L. Bressler, Handbook of time series analysis: recent theoretical developments and applications **437** (2006).
[39] H. Akaike, Psychometrika **52**, 317 (1987).
[40] G. A. Darbellay and I. Vajda, IEEE Transactions on Information Theory **45**, 1315 (1999).
[41] K. Mizuseki, A. Sirota, E. Pastalkova, and G. Buzsáki, Neuron **64**, 267 (2009).
[42] K. Mizuseki, A. Sirota, E. Pastalkova, and G. Buzsáki, http://dx.doi.org/10.6080/K0Z60KZ9.
[43] L. Zhang, G. Chen, R. Niu, W. Wei, X. Ma, J. Xu, J. Wang, Z. Wang, and L. Lin, Hippocampus **22**, 1781 (2012).
[44] G. Chen, Y. Zhang, X. Li, X. Zhao, Q. Ye, Y. Lin, H. W. Tao, M. J. Rasch, and X. Zhang, Neuron **96**, 1403 (2017).
[45] D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, and D. Cai, PloS one **9**, e87636 (2014).
[46] D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, and D. Cai, Physical review letters **111**, 054102 (2013).
[47] P. A. Stokes and P. L. Purdon, Proceedings of the National Academy of Sciences **114**, E7063 (2017).
[48] L. Barnett and A. K. Seth, Physical Review E **91**, 040101 (2015).
[49] L. Faes, S. Stramaglia, and D. Marinazzo, F1000Research **6** (2017).
[50] L. Barnett, A. B. Barrett, and A. K. Seth, arXiv preprint arXiv:1708.08001 (2017).
[51] S. Li, G. Li, Z. Tian, and D. Zhou, In preparation.