Quantitative causality analysis with coarsely sampled time series

Preprint · March 2023

CITATIONS
0

READS
146

1 author:

X. San Liang
Fudan University
148 PUBLICATIONS **2,985** CITATIONS

SEE PROFILE

Quantitative causality analysis with coarsely sampled time series

X. San Liang

(Dated: March 8, 2023)

Abstract

The information flow-based quantitative causality analysis has been widely applied in different disciplines because of its origin from first principles, its concise form, and its computational efficiency. So far the algorithm for its estimation is based on differential dynamical systems, which, however, may make an issue for coarsely sampled time series. Here, we show that for linear systems, this is fine at least qualitatively; but for highly nonlinear systems, the bias increases significantly as the sampling frequency is reduced. This paper provides a partial solution to this problem, showing how causality analysis is assured faithful with coarsely sampled series when, of course, the statistics is sufficient. An explict and concise formula has been obtained, with only sample covariances involved. It has been successfully applied to a system comprising of a pair of coupled Rössler oscillators. Particularly remarkable is the success when the two oscillators are nearly synchronized.

PACS numbers:

Keywords: quantitative causality; information flow; coarsely sampled time series; synchronization; Rössler system; Lie group

I. INTRODUCTION

Causality analysis is an important problem in scientific research. Though traditionally formulated as a statistical problem in data science, computer science, among other disciplines, recently it has been found to be, within the framework of information flow/transfer, "a real notion in physics that can be derived *ab initio*"[1]. A comprehensive study with generic systems has been fulfilled recently, with explicit formulas attained in closed form; see [2] and [1]. These formulas have been validated with benchmark systems such as baker transformation, Hénon map, etc., and have been applied successfully to real world problems in the diverse disciplines such as global climate change (e.g., [3], [4], [5]), dynamic meteorology (e.g.,[6]), land-atmosphere interaction (e.g.,[7]), data-driven prediction (e.g., [8], [9]), near-wall turbulence (e.g., [10]), neuroscience (e.g., [11], [12]), financial analysis (e.g., [13], [14]), quantum information (e.g, [15]), to name several.

For the purpose of this study, we first give a brief introduction of the theory within the framework of a differential dynamical system. (Also available for discrete-time mappings, refer to Liang (2016).) Let

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\dot{\mathbf{w}}, \tag{1}$$

be a d-dimensional continuous-time stochastic system for $\mathbf{x} = (x_1, ..., x_d)$ (we do not distinguish notations for random and deterministic variables), where $\mathbf{F} = (F_1, ..., F_d)$ may be arbitrary nonlinear differentiable functions of $\mathbf{x}$ and t, $\mathbf{w}$ is a vector of white noises, and $\mathbf{B} = (b_{ij})$ is the matrix of perturbation amplitudes which may also be any differentiable functions of $\mathbf{x}$ and t. Liang (2016)[1] proves that the rate of information flowing from x_j to x_i (in nats per unit time) is

$$\begin{aligned} T_{j\to i} &= -E\left[\frac{1}{\rho_i}\int_{\mathbb{R}^{d-2}} \frac{\partial(F_i\rho_{ij})}{\partial x_i} d\mathbf{x}_{ij}\right] + \frac{1}{2}E\left[\frac{1}{\rho_i}\int_{\mathbb{R}^{d-2}} \frac{\partial^2(g_{ii}\rho_{ij})}{\partial x_i^2} d\mathbf{x}_{ij}\right], \\ &= -\int_{\mathbb{R}^d} \rho_{j|i}(x_j|x_i)\frac{\partial(F_i\rho_{ij})}{\partial x_i} d\mathbf{x} + \frac{1}{2}\int_{\mathbb{R}^d} \rho_{j|i}(x_j|x_i)\frac{\partial^2(g_{ii}\rho_{ij})}{\partial x_i^2} d\mathbf{x}, \end{aligned} \tag{2}$$

where $d\mathbf{x}_{ij}$ signifies $dx_1...dx_{i-1}dx_{i+1}...dx_{j-1}dx_{j+1}...dx_n$, E stands for mathematical expectation, $g_{ii} = \sum_{k=1}^n b_{ik}b_{ik}$, $\rho_i = \rho_i(x_i)$ is the marginal probability density function (pdf) of x_i, $\rho_{j|i}$ is the pdf of x_j conditioned on x_i, and $\rho_{ij} = \int_{\mathbb{R}} \rho(\mathbf{x})dx_j$. The algorithm for the information flow-based causal inference is as follows: If $T_{j\to i} = 0$, then x_j is not causal to x_i; otherwise it is causal, and the absolute value measures the magnitude of the causality from x_j to x_i. This is guaranteed by a property called "principle of nil causality." Another property regards the invariance upon coordinate transformation, indicating that the obtained information flow (IF) is an intrinsic property in nature[16]. Also established by Liang (2016)[1] is that, for a linear model, i.e., for $\mathbf{F}(\mathbf{x}, t) = \mathbf{A}\mathbf{x}$, $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are constant matrices in (1), then

$$T_{j\to i} = a_{ij}\frac{\sigma_{ij}}{\sigma_{ii}}$$

where σ_{ij} is the population covariance of x_i and x_j. By this, in the linear sense, causation implies correlation, but not vice versa. In an explicit expression, this corollary fixes the debate on causation vs. correlation ever since George Berkeley (1710)[17].

In the case with only d time series $x_1, x_2, ..., x_d$, the quantitative causality, i.e., the IF, between them can be estimated using maximum likelihood estimation (see [18] and [19]). Under the assumption of a linear system with additive noises, the maximum likelihood estimator (mle) of (2) for $T_{j\to i}$ is[19]

$$\hat{T}_{j\to i} = \frac{1}{\det \mathbf{C}} \cdot \sum_{\nu=1}^{d} \Delta_{j\nu} C_{\nu,di} \cdot \frac{C_{ij}}{C_{ii}}, \tag{3}$$

where C_{ij} is the sample covariance between x_i and x_j, Δ_{ij} the cofactors of the matrix $\mathbf{C} = (C_{ij})$, and $C_{i,dj}$ the sample covariance between x_i and a series derived from x_j using the Euler forward differencing scheme: $\dot{x}_{j,n} = (x_{j,n+k} - x_{j,n})/(k\Delta t)$, with $k \geq 1$ some integer. Eq. (3) is rather concise in form, involving only the common statistics, i.e., sample covariances. The transparent formula makes causality analysis, which otherwise would be complicated, very easy and computationally efficient. Note, however, that Eq. (3) cannot replace (2); it is just the maximum likelihood estimator (mle) of the latter. Statistical significance tests can be performed for the estimators. This is done with the aid of a Fisher information matrix. See Liang (2014)[18] and Liang (2021)[19] for details.

Originally the formalism is established in the light of a differential system; in other words, it is with infinitesimal time increments. (The formalism with discrete mappings has also been established by Liang (2016)[1], but still there has no estimation with it.) One would naturally ask a question about the applicability in the case of coarsely sampled time series. Indeed, it is not unusual that the given series may be coarsely sampled because of the limited observations. As will be seen in the following section this may make a problem for nonlinear systems if the sample interval is large. This paper henceforth attempts to address this issue in the original linear framework. In the following we first check the applicability of (3) for series from a linear system and a highly nonlinear system (section II), with a variety of sampling intervals. A new approach is presented in section III, which is then utilized to redo the causal inferences in section II. Some remaining issues are discussed in section V.

II. THE ISSUE WITH COARSELY SAMPLED SERIES

A. Time series from linear systems

We first test the applicability of (3), as the sampling interval increases, with a well-studied linear system whose IF rates have been found half-analytically. This is the validation example in [18]:

$$\frac{dx_1}{dt} = -x_1 + 0.5x_2 + 0.1\dot{w}_1 \tag{4a}$$

$$\frac{dx_2}{dt} = -x_2 + 0.1\dot{w}_2 \tag{4b}$$

where $\dot{w}_i$, $i = 1, 2$ are independent white noises. It has been shown that, the rates of information flow per unit time, $T_{2\to1} \to 0.11$ as $t \to \infty$, and $T_{1\to2} = 0$ for all t, reflecting accurately the one-way causality from x_2 to x_1. Now, using the same sample path as that in Liang (2014),[21] we re-sample the series with low frequencies to obtain new series. Shown in Figure 1 is part of the sample path, with triangles marking the sampling points.

The computed IFs for different sampling intervals are listed below:

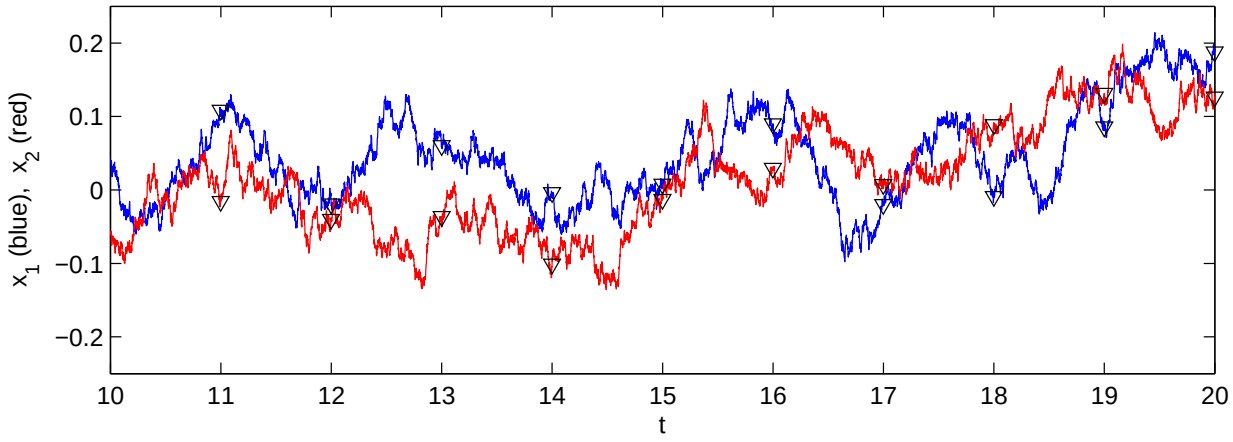

FIG. 1: A segment of the sample path generated by Eq. (4), with a time step $\Delta t = 0.001$. The time spans from 0 through 100. As we are only interested in the limit IF as $t \to \infty$, the first 5000 steps, i.e., those for $t \leq 5$, are discarded.

S.I. (# of pts)	1	10	50	100	300	500
$\hat{T}_{2\to1}$	0.110±0.051	0.113±0.051	0.099±0.050	0.090±0.048	0.055±0.036	0.055±0.034
$\hat{T}_{1\to2}$	-0.002±0.056	-0.001±0.055	-0.015±0.054	-0.011±0.053	-0.008±0.038	-0.015±0.034

Also computed are the confidence intervals at a level of 90% (at a significance level of 0.1). First, the estimators $\hat{T}_{2\to1}$ for all the SIs here are significantly distinct from zero, while those the other way around, $\hat{T}_{1\to2}$, are not significant at a level of 90%. So the causality in a qualitative sense has been faithfully recovered even with very low sampling frequencies (large SI). (In fact, even with SI=1000 the result is still correct; we do not consider cases beyond SI=500 since the sample size is too small for SI¿500, resulting in insufficient statistics.)

Since this example actually has a half-analytical solution ($T_{2\to1} \approx 0.11$, $T_{1\to2} = 0$), we have more to say about the computed results. Generally, the result of $\hat{T}_{1\to2}$ looks satisfactory. For $\hat{T}_{2\to1}$, it is rather accurate for SI ≤ 100. Beyond 100, it is not accurate any more.

B. Time series from synchronized chaotic oscillators

The following example is from the synchronization problem as examined by Palus et al. (2018). The system is composed of two Rössler oscillators, $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$, where

$$\frac{dx_1}{dt} = -\omega_1 x_2 - x_3, \tag{5a}$$

$$\frac{dx_2}{dt} = \omega_1 x_1 + 0.15 x_2, \tag{5b}$$

$$\frac{dx_3}{dt} = 0.2 + x_3(x_1 - 10), \tag{5c}$$

is the master system, and

$$\frac{dy_1}{dt} = -\omega_2 y_2 - y_3 + \varepsilon(x_1 - y_1), \tag{6a}$$

$$\frac{dy_2}{dt} = \omega_2 y_1 + 0.15 y_2, \tag{6b}$$

$$\frac{dy_3}{dt} = 0.2 + y_3(y_1 - 10), \tag{6c}$$

is the driven one. Following Palus et al. (2018)[20], choose $\omega_1 = 1.015$ and $\omega_2 = 0.985$. Using the Runge-Kutta scheme and choosing a time step $\Delta t = 0.001$, the coupled 6-dimensional system can be solved rather accurately with different ε. Figure 2 plots the solutions of x_1 and y_1 when the coupling strength $\varepsilon = 0.11$ (upper panel) and $\varepsilon = 0.15$ (lower panel). As shown in the latter case, the two subsystems become synchronized if $\varepsilon \geq 0.15$. Again, we

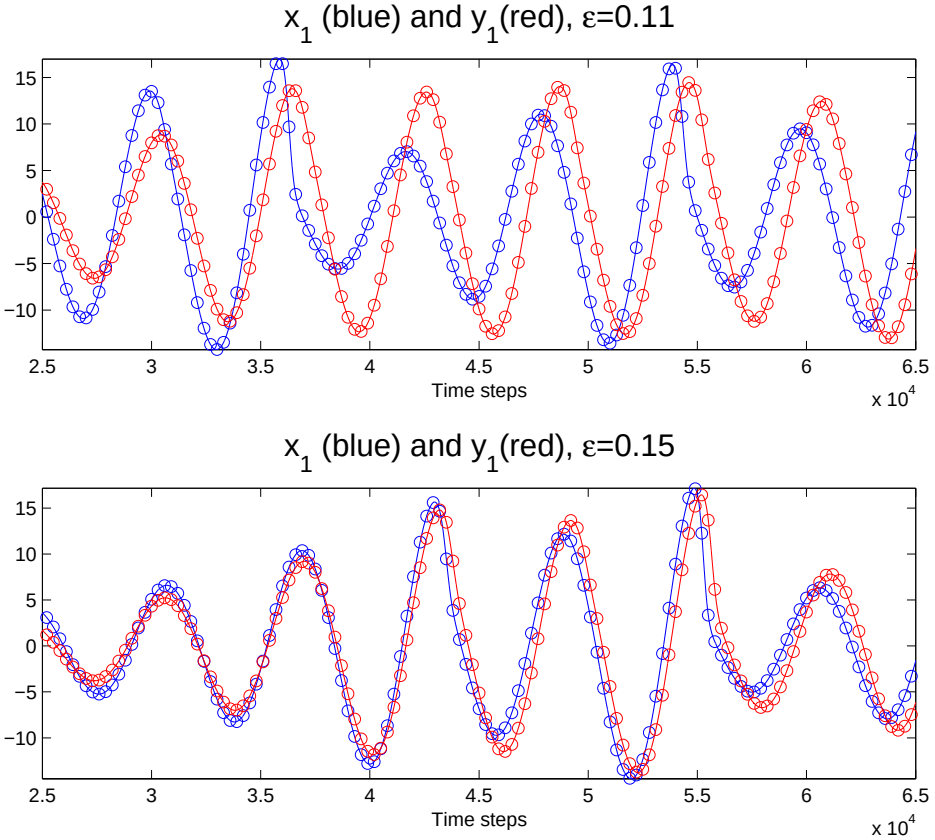

FIG. 2: Part of the time series of x_1 and y_1 of the coupled Rössler systems for $\varepsilon = 0.11$ (top) and $\varepsilon = 0.15$ (bottom). The circles indicate the sampling points (every 300 steps here, corresponding to 18 points in each period). The two oscillators become synchronized as $\varepsilon \geq 0.15$.

choose to study the problem for $t \in [0, 100]$ (10^5 time steps in total).

For each ε we generate six time series of 10^5 steps, and evaluate the IFs according to Eq. (3) ($k = 1$ is chosen). The IFs as functions of ε are then obtained, and plotted in Fig. 3a, which accurately tells that the master is $\mathbf{x}$, and $\mathbf{y}$ is the slave. An appealing

observation is that this causality inference even works when the two oscillators are nearly synchronized as $\varepsilon > 0.15$, demonstrating the power of this rigorously formulated causality analysis.

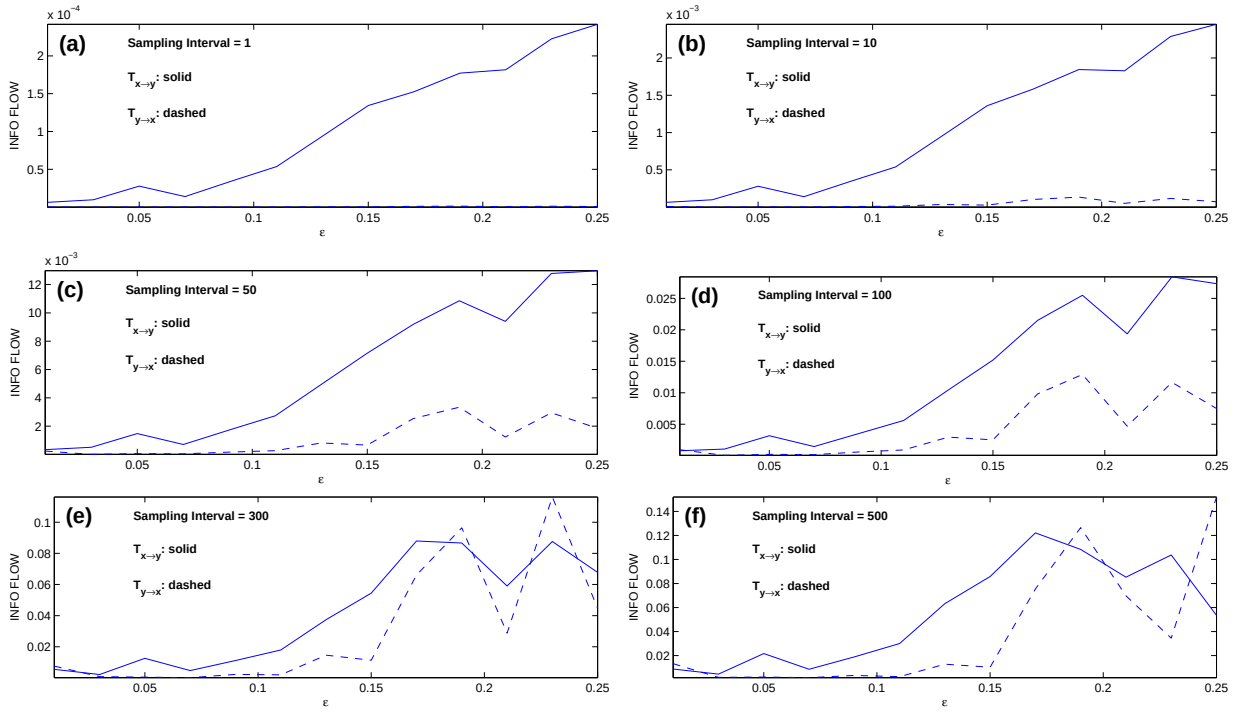

FIG. 3: Absolute information flow rates $\left|\hat{T}_{x\to y}\right|$ and $\left|\hat{T}_{y\to x}\right|$ (dashed) between the two Rössler oscillators as functions of the coupling coefficient ε with different sampling intervals. Units are in nats per unit time. By the preset causality, the dashed line should coincide with the abscissa.

We subsample the series every SI steps, SI=10, 50, 100, 300, 500, and redo the computation using the same scheme. The resulting IF rates are shown in Figs. 3b, c, d, e, f, respectively. By the preset causality, the dashed line should be the zero-line. Clearly, the causal inference works well for SI$\leq$ 10. The computed IF becomes biased for SI$\geq$ 50, and the bias grows significantly as SI increases. If we focus on $\varepsilon <= 0.15$, i.e., when the systems are not synchronized (see Palus et al., 2018), the causal inference still functions fine for SI$\leq$50. If the synchronized cases are taken into account ($\varepsilon > 0.15$), then the inferences in the cases for $50 \leq$ SI ≤ 100 are much biased, and those for SI exceeding 300, a case corresponding to an approximate sampling frequency of 20 per period, are not correct any more.

III. APPROACHING TO A PARTIAL SOLUTION

As shown above, if the sampling frequency of the time series is low, the resulting linear IF for nonlinear series may be biased. Indeed, in the case with high nonlinearity, the linear assumption is always easy to be blamed. While theoretically it is not a problem (causality is guaranteed as proved in a theorem), we agree that, before a fully nonlinear algorithm is developed, this will be a continuing issue. What we want to show here is, how much room there is for improvement. So far, the algorithm documented in [18], and later in [19], is

based on the Bernstein-Euler differencing scheme, which is, of course, very rudimentary due to the first order differencing. If a time series is coarsely sampled, the error could be large.

A theorem as established by Liang (2008)[2] reads that, if the noise is additive in Eq. (1), i.e., if $\mathbf{B}$ is a constant matrix, then the noise itself does not appear in the formula of $T_{j\to i}$. So under the additive noise assumption, we can estimate the IF within the framework of a deterministic system. In this case, note that the linear equation set actually can be solved for an interval $[t, t+\Delta t]$, no matter how large Δt is. This gives a hint to the solution of the low sampling frequency problem.

Consider

$$\frac{d\mathbf{x}}{dt} = \mathbf{f} + \mathbf{A}\mathbf{x}, \tag{7}$$

where $\mathbf{A} = (a_{ij})$ is a $d \times d$ matrix. Let us assume that $\mathbf{f} = \mathbf{0}$, since the time series can always be pre-treated by removing the linear trend, and it has been proved that this removal does not alter the IF rates. In this case, on the interval $[t, t+\Delta t]$, we actually have a mapping $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ that takes the state $\mathbf{x}(t)$ to the state $\mathbf{x}(t+\Delta t)$ at $t+\Delta t$, with the propagating operator:

$$\Phi = e^{\mathbf{A}\Delta t} = e^{\begin{bmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \dots & a_{dd} \end{bmatrix}\Delta t} \equiv \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1d} \\ \vdots & \ddots & \vdots \\ \alpha_{d1} & \dots & \alpha_{dd} \end{bmatrix}. \tag{8}$$

It is not easy to estimate a_{ij}, but it is easy to estimate α_{ij} instead, by observing the relation

$$\begin{bmatrix} \alpha_{11} & \dots & \alpha_{1d} \\ \vdots & \ddots & \vdots \\ \alpha_{d1} & \dots & \alpha_{dd} \end{bmatrix} \mathbf{x}(n) = \mathbf{x}(n+1), \qquad n = 0, 1, 2, ..., N. \tag{9}$$

This written in a matrix form is

$$\begin{bmatrix} x_1(0) & \dots & x_d(0) \\ \vdots & \ddots & \vdots \\ x_1(N-1) & \dots & x_d(N-1) \end{bmatrix} \begin{bmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{id} \end{bmatrix} = \begin{bmatrix} x_i(1) \\ \vdots \\ x_i(N) \end{bmatrix}$$

for $i = 1, .., d$. Averaging all the rows of the algebraic equation set, and subtracting the mean from each row, we get

$$\begin{bmatrix} x_1(0) - \bar{x}_1 & \dots & x_d(0) - \bar{x}_d \\ \vdots & \ddots & \vdots \\ x_1(N-1) - \bar{x}_1 & \dots & x_d(N-1) - \bar{x}_d \end{bmatrix} \begin{bmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{id} \end{bmatrix} = \begin{bmatrix} x_i(1) - \bar{x}_{i+} \\ \vdots \\ x_i(N) - \bar{x}_{i+} \end{bmatrix},$$

where $\bar{x}_i = \frac{1}{N}\sum_{n=0}^{N-1} x_i(n)$, $\bar{x}_{i+} = \frac{1}{N}\sum_{n=1}^{N} x_i(n)$, i.e., the series $\{x_{i+}(n)\}$ is the series $\{x_i(n)\}$ advanced by one step. Let i run through $\{1, 2, ..., d\}$. We have the following d overdetermined equation sets:

$$\begin{bmatrix} x_1(0) - \bar{x}_1 & \dots & x_d(0) - \bar{x}_d \\ \vdots & \ddots & \vdots \\ x_1(N-1) - \bar{x}_1 & \dots & x_d(N-1) - \bar{x}_d \end{bmatrix} \begin{bmatrix} \alpha_{11} & \dots & \alpha_{d1} \\ \vdots & \ddots & \vdots \\ \alpha_{1d} & \vdots & \alpha_{dd} \end{bmatrix} = \begin{bmatrix} x_1(1) - \bar{x}_{1+} & \vdots & x_d(1) - \bar{x}_{d+} \\ \vdots & \ddots & \vdots \\ x_1(N) - \bar{x}_{1+} & \vdots & x_d(N) - \bar{x}_{d+} \end{bmatrix} \tag{10}$$

Denote by $\mathbf{\Lambda}$ the matrix (α_{ij}), then the matrix of unknowns in the above equation sets is $\mathbf{\Lambda}^T$. Left multiplication by

$$\begin{bmatrix} x_1(0)-\bar{x}_1 & \dots & x_d(0)-\bar{x}_d \\ \vdots & \ddots & \vdots \\ x_1(N-1)-\bar{x}_1 & \dots & x_d(N-1)-\bar{x}_d \end{bmatrix}^T$$

on both sides yields d $d\times d$ equation sets:

$$\mathbf{C}\mathbf{\Lambda}^T = \tilde{\mathbf{C}}, \tag{11}$$

where $\mathbf{C} = (C_{ij})$ is the sample covariance matrix of $\mathbf{x}$, and $\tilde{\mathbf{C}} = (C_{i,j+})$, and $C_{i,j+}$ is the sample covariance between x_i and x_{j+}, i.e., x_j advanced by one time step. The least square solutions of the overdetermined sets (10) are the solutions of (11):

$$\mathbf{\Lambda}^T = \mathbf{C}^{-1}\tilde{\mathbf{C}},$$

and hence

$$\mathbf{\Lambda} = (\mathbf{C}^{-1}\tilde{\mathbf{C}})^T = \tilde{\mathbf{C}}^T\mathbf{C}^{-1}.$$

The estimator of $\mathbf{A}$ is, therefore,

$$\hat{\mathbf{A}} = \frac{1}{\Delta t}\log\left(\tilde{\mathbf{C}}^T\mathbf{C}^{-1}\right). \tag{12}$$

(Caution should be used in case of singularity. The irrelevant imaginary part also should be discarded.)

Once getting A, hence the coefficients (a_{ij}), we substitute a_{ij} for the whole part

$$\frac{1}{\det \mathbf{C}}\sum_{k=1}^{d}\Delta_{jk}C_{k,di}$$

in Eq. (3), i.e., multiply a_{ij} by C_{ij}/C_{ii} to arrive at the desideratum, $\hat{T}_{j\to i}$. If we denote by $[\mathbf{A}]_{ij}$ the extraction of the $(i,j)^{th}$ entry of the matrix $\mathbf{A}$, this is

$$\hat{T}_{j\to i} = \frac{1}{\Delta t}\left[\log(\tilde{\mathbf{C}}^T\mathbf{C}^{-1})\right]_{ij}\cdot\frac{C_{ij}}{C_{ii}}. \tag{13}$$

(Note here log is the matrix logarithm. In matlab, the function is logm.)

IV. THE COARSELY SAMPLED SERIES PROBLEMS REVISITED

As demonstrated above, for the series generated from linear systems, the estimation of the IF is fine qualitatively. We here, nevertheless, want to see how the new scheme may have the results improved. Shown below is a recalculation of the estimates. Since this case has a rather accurate result (≈ 0.11 nats per unit time), we can see that the result is accurate enough for all the SIs here.

Sampling Interval	1	10	50	100	300	500
$\hat{T}_{2\to1}$	0.114	0.118	0.109	0.106	0.082	0.098
$\hat{T}_{1\to2}$	0.007	0.008	−0.007	−0.002	−0.002	−0.015

The new scheme for the estimation is particularly for the nonlinear case. For the pair of Rössler oscillators, the computed results are plotted in Fig. 4. Compared to Fig. 3, now the performance has been much improved. For the cases with SI$\leq$ 100 (Figs. 4a-d), the results are rather accurate for all the coupling strengths ε considered (both synchronized and non-synchronized). For the case SI=300, which corresponds to a sampling frequency of 20 points per period, the one-way causality is accurately recovered for the nonsynchronized cases ($\varepsilon \leq 0.15$). But beyond that $\varepsilon > 0.15$, the inference fails. Particularly, when SI=500 (Fig. 4f), the result is even worse that its counterpart with the traditional scheme as plotted in Fig. 3f. This, of course, may be due to the resulting small sample size, which causes singularity to the matrix logarithm.

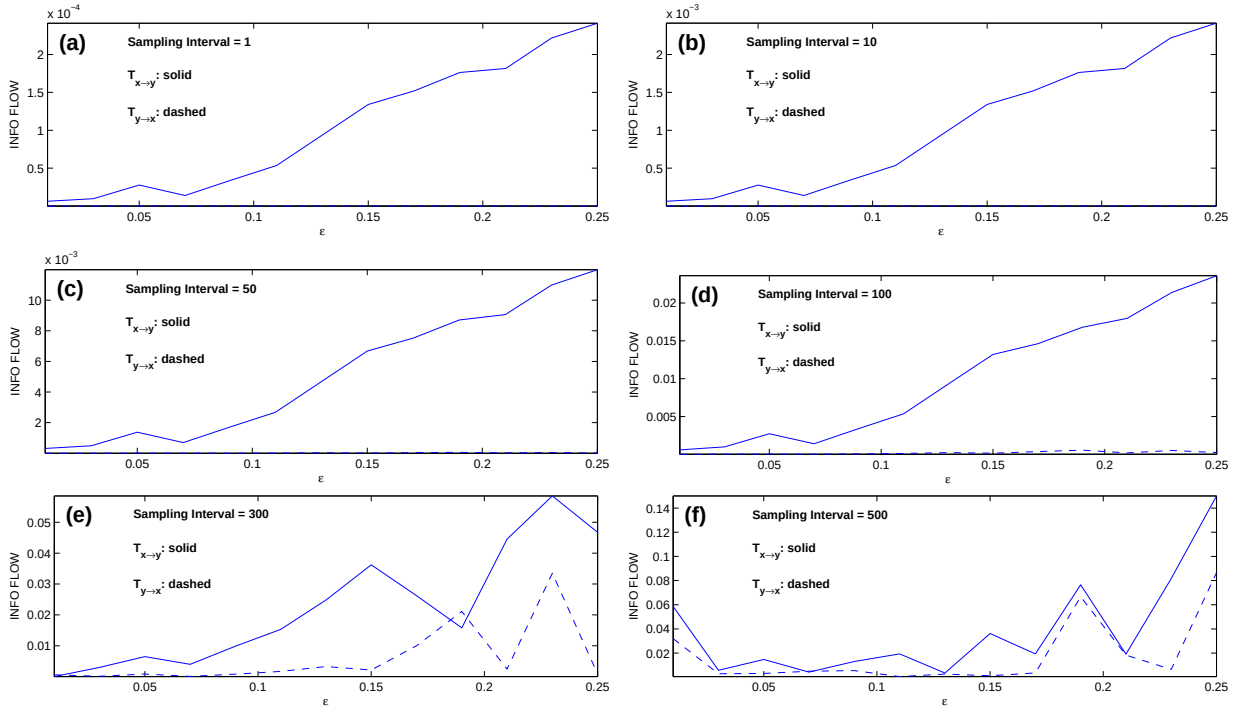

FIG. 4: As Fig. 3, but the information flow rates are computed with the new scheme.

V. DISCUSSION

The maximum likelihood estimator of the information flow (IF), Eq. (3), provides a very easy way to causal inference. Theoretically it is based on a linear assumption, but practically it has shown tremendous success with series generated from highly nonlinear systems; anyway, linearization piecewise in time proves to be an efficient asymptote to an otherwise nonlinear system. In reality, series may be coarsely sampled; the time resolution may be low. An issue thus arises, as this formalism is theoretically on the basis of infinitesimal time

increments. In this case, as we have shown, it still works for linear systems in a qualitative sense; but for a highly nonlinear system composed of two Rössler oscillators, the bias becomes more and more significant as the sampling frequency is reduced.

A new scheme has been proposed to address this problem and provide a partial solution. Due to the nice property of IF, as proved in [2], that additive noises do not alter the IF flow in form, it is reasonable to directly estimate the IF without paying attention to the stochasticity. Instead of estimating through the differential equations using the Euler-Bernstein differencing, we choose to consider the integral form on the finite time interval, i.e., to estimate the Lie group members. In doing this, the original formula (3), which is rewritten here for easy reference,

$$\hat{T}_{j\to i} = \frac{1}{\det \mathbf{C}} \cdot \sum_{\nu=1}^{d} \Delta_{j\nu} C_{\nu,di} \cdot \frac{C_{ij}}{C_{ii}},$$

is replaced by (13),

$$\hat{T}_{j\to i} = \frac{1}{\Delta t} \left[\log(\tilde{\mathbf{C}}^T \mathbf{C}^{-1})\right]_{ij} \cdot \frac{C_{ij}}{C_{ii}},$$

where $\tilde{\mathbf{C}} = (C_{i,j+})$, and $C_{i,j+}$ is the sample covariance between x_i and x_{j+}, i.e., x_j advanced by one time step. Note here log is the matrix logarithm; in MATLAB, the function is logm. This way, it shows that the preset causality within the coupled system of chaotic oscillators has been rather accurately reproduced even when the sampling interval is large (sampling frequency is low).

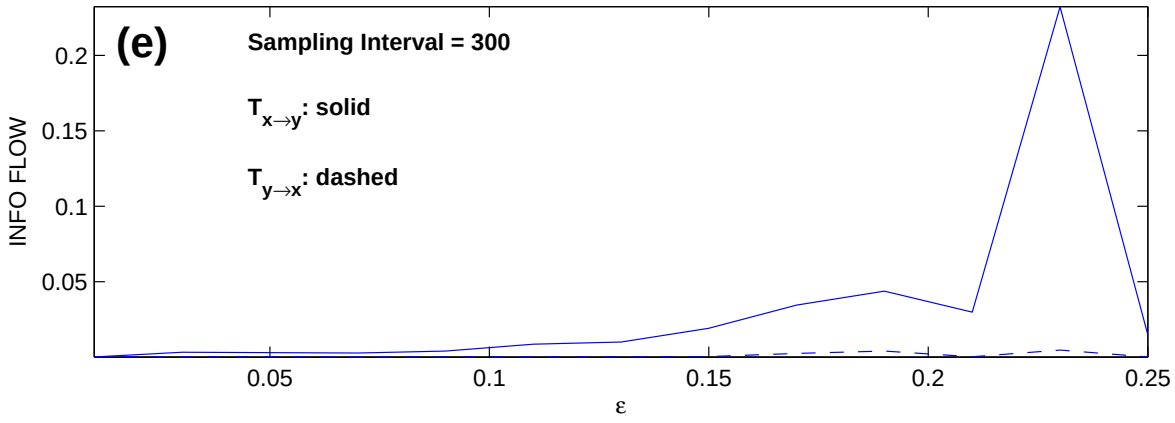

FIG. 5: As Fig. 4e, but the covariances are estimated using the residuals of x_i relative to the model result, instead of x_i themselves. Here SI=300 approximately corresponds to a sampling frequency of 18 points each period.

There is still much room for improvement for the above approach. For example, the estimation of the covariances in the quotient $\frac{\sigma_{ij}}{\sigma_{ii}}$ is by replacing the population covariances with sample covariances, while the sample is formed from the time series. While this is satisfactory for stochastic systems under the ergodic assumption, this may not be good for deterministic chaos, such as the Rössler oscillators case here. The reason is obvious: The time mean of the series in Fig. 2 is zero, but one can imagine that the ensemble mean of

all the possible paths is by no means zero; rather, it should be a function of time (just like the series itself), which may be close to the asymptotic linear system solution. So it makes more sense to treat the linear system solution as the mean. As such, we have attempted to improve the estimation by replacing the covariances of $\mathbf{x}$ with those of $\mathbf{x} - \overline{\mathbf{x}}$, where $\overline{\mathbf{x}}$ stands for the resulting linear system solution; With this we get another causal inference result for SI=300; the resulting IFs are plotted in Fig. 5. As one can see, the result looks rather accurate, just as expected, in contrast to Fig. 4e.

We, however, do not claim that we have solved the problem. What we want to show here is, how much room there is for improvement within a linear framework. Indeed, in the case with high nonlinearity, the linear assumption is always easy to be blamed. While theoretically it is not a problem (causality is guaranteed as proved in a theorem; see [1] and other references), it is believed that, before a fully nonlinear algorithm is developed, this will be a continuing issue.

Code availability

The codes are available, and will be updated, at www.ncoads.org/article/show/67.aspx.

Acknowledgments

Many thanks for Milan Paluš's question, which motivated this research.

---

[1] Liang, X.S. (2016). Information flow and causality as rigorous notions ab initio. *Phys. Rev. E*, 94, 052201.

[2] Liang, X.S. Information flow within stochastic dynamical systems. *Phys. Rev. E* **2008**, *78*, 031113.

[3] Stips, A.; Macias, D.; Coughlan, C.; Garcia-Gorriz, E.; Liang, X.S. On the causal structure between CO_2 and global temperature. *Sci. Rep.* **2016**, *6*, 21691.

[4] Vannitsem, S.; Dalaiden, Q.; Goosse, H. (2019) Testing for dynamical dependence—Application to the surface mass balance over Antarctica. *Geophys. Res. Lett.*, doi:10.1029/2019GL084329.

[5] Doquier, D., S. Vannitsem, F. Ragone, K. Wyser, and X.S. Liang (2022). Causal links between Arctic sea ice and its potential drivers based on the rate of information transfer. *Geophys. Res. Lett.*. DOI: 10.1029/2021GL095892.

[6] Liang, X.S. (2019). A study of the cross-scale causation and information flow in a stormy model mid-latitude atmosphere. *Entropy*, 21, 149.

[7] Hagan, D.F.T.; Wang, G.; Liang, X.S.; Dolman, H.A.J. (2019) A time-varying causality formalism based on the Liang-Kleeman information flow for analyzing directed interactions in nonstationary climate systems. *J. Clim.*, *32*, 7521–7537.

[8] Bai, C., R. Zhang, and Coauthors (2018). Forecasting the tropical cyclone genesis over the Northwest Pacific through identifying the causal factors in cyclone-climate interactions. *J. Atmos. Ocean. Tech.*, 35, 247-259.

[9] Liang, X.S., F. Xu, Y. Rong, R. Zhang, X. Tang, and F. Zhang (2021). El Niño Modoki can be mostly predicted more than 10 years ahead of time. *Sci. Rep.*, 11, 17860.

[10] Liang, X.S., A. Lozano-Durán (2016). A preliminary study of the causal structure in fully developed near-wall turbulence. *Proceedings of the Summer Program 2016*, 233-242. Center for Turbulence Research, Stanford University, CA, USA.

[11] Hristopulos, D.T., A. Babul, S. Babul, L.R. Brucar, N. Virji-Babul (2019). Disrupted information flow in resting-state in adolescents with sports related concussion. *Front. Hum. Neurosci.*, *13*, 419. doi:10.3389/fnhum.2019.00419.

[12] Cong, J., and coauthors (2023). Altered default mode network causal connectivity patterns in autism spectrum disorder revealed by Liang information flow analysis. *Human Brain Mapping*, DOI:10.1002/hbm.26209.

[13] Lu, X., K. Liu, and coauthors (2022). The dynamic causality in sporadic bursts between CO_2 emission allowance prices and clean energy index. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-022-21316-5.

[14] Lu, X.F., K. Liu, H. Cui, J. Li (2023). Study on the causality between international crude oil futures market and Chinese stock index future market. *Chinese Journal of Management Science*. http://doi.org/10.16381/j.cnki.issn1003-207x.2021.1128.

[15] Yi, B., S. Bose (2022). Quantum Liang information flow as causation quantifier. *Phys. Rev. Lett.* **129**, 020501.

[16] Liang, X.S. (2018). Information flow with respect to relative entropy. *Chaos*, *28*, 075311.

[17] Berkeley, G. *A Treatise Concerning the Principles of Human Knowledge*; Aaron Rhames: Dublin, Ireland, 1710.

[18] Liang, X.S. (2014). Unraveling the cause-effect relation between time series. *Phys. Rev. E*, 90, 052150.

[19] Liang, X.S. (2021). Normalized multivariate time series causality analysis and causal graph reconstruction. *Entropy*, 23, 679.

[20] Paluš, M., A. Krakovská, J. Jakubfk, and M. Chvosteková (2018). Causality, dynamical systems and the arrow of time. *Chaos*, **28**, 075307.

[21] Note, because of the pseudorandom number generator, the generated sample path using the normal differencing scheme may not be satisfactory. To see whether the obtained sample path is correct, one may check the resulting covariances, which can be rather accurately obtained by solving a deterministic ODE. Here the data for the generated sample path can be downloaded from http://www.ncoads.org/article/show/68.aspx under the item PRE_2014.dat.