PHYSICAL REVIEW E **105**, 044130 (2022)

Relationship between Schreiber's transfer entropy and Liang-Kleeman information flow from the perspective of stochastic thermodynamics

Hirohito Kiwata[*]

Division of Natural Science, Osaka Kyoiku University, Kashiwara, Osaka 582-8582, Japan

(Received 20 October 2020; revised 17 January 2022; accepted 5 April 2022; published 21 April 2022)

Schreiber's transfer entropy is an important index for investigating the causal relationship between random variables. The Liang-Kleeman information flow is another analysis to demonstrate the causality within dynamical systems. Horowitz's information flow is introduced through multicomponent stochastic thermodynamics. In this study, I elucidate the relationship between Schreiber's transfer entropy and the Liang-Kleeman information flow through Horowitz's information flow. I consider the case in which the system changes according to the stochastic differential equation. I find that the Liang-Kleeman and Horowitz information flows differ by a term derived from the stochastic fluctuation. I also show that Schreiber's transfer entropy is not less than Horowitz's information flow. This study helps unify various indexes that determine the causal relationship between variables.

DOI: 10.1103/PhysRevE.105.044130

I. INTRODUCTION

When one wants to investigate a complicated phenomenon but does not know how to apply the laws of physics to it, one typically observes the quantities that seem to be related to the phenomenon. When the system is in equilibrium, the existence of dependence among multiple observables can be determined by examining mutual information in terms of the information theory [1]. When the system is time varying, the measured quantities are obtained as time series data [2–4]. In some cases, information about the causal relationships among the measured quantities from the time series data is required. When studying aspects beyond the interdependency of x and y, it is important to know the causal relationship between x and y, where x is the cause of y. In other words, to determine if the variable x is required for the prediction model of the variable y, it is important to identify the causal relationship between the variables. Granger causality is a conventional method for testing the causal relationship between two quantities [5]. Schreiber invented transfer entropy to study the causal relationship between two random processes [6,7]. Transfer entropy has been used extensively to detect causal relationships, but it may not always successfully do so [8–10].

Liang and Kleeman considered a new approach of information flow to develop a method for constructing a causal relationship [11–18]. Their method is ingenious and divides the entropy change of a random variable x into a spontaneous change of x itself and the external influence from y. The external influence is evaluated using the difference between the entropy change and the spontaneous change. The external influence corresponds to the Liang-Kleeman information flow. Schreiber's transfer entropy is defined by the Kullback-Leibler divergence between two conditional probability distributions, $P(x_{t+dt}|x_t, y_t)$ and $P(x_{t+dt}|x_t)$, where dt represents the time increment. If x_{t+dt} does not depend on y_t, relation $P(x_{t+dt}|x_t, y_t) = P(x_{t+dt}|x_t)$ is satisfied; then, the Kullback-Leibler divergence between them becomes zero. Thus, based on their definitions, Schreiber's transfer entropy and the Liang-Kleeman information flow are considerably different, and their mutual relationship is not clear. Knowledge of the causal relationship between multiple indexes and when to use specific indexes is always valuable. Therefore, this study presents the relationship between Schreiber's transfer entropy and the Liang-Kleeman information flow.

A recent major trend in thermodynamics involves treating information thermodynamically [19]. When the second law of thermodynamics is formulated using stochastic thermodynamics, mutual information is added to the formula of the said law for macroscopic systems. In particular, when considering the thermodynamics of a system consisting of two variables, new terms are added to the equation of the second law of thermodynamics for the entropy changes of each system and the entropy changes from each environmental reservoir. The added terms concern the rates of changes in the mutual information between the two variables. These rates represent the information flow between the two systems. It is likely that the so-called information flow can be used for causal inference between the two variables. It has been shown that Schreiber's transfer entropy can be derived naturally by applying stochastic thermodynamics to the Bayesian network and formulating the second law of thermodynamics [20,21]. In light of the above, it seems worthwhile to consider the information flow from a thermodynamic perspective.

The importance of the Liang-Kleeman information flow in thermodynamics has been highlighted by Cafaro *et al.* [22]. They revealed the relationship between the Liang-Kleeman and Horowitz-Esposito notions of information flow [23–25]. The derivation of the Liang-Kleeman information flow was proven, first heuristically and then rigorously, but I believe that a thermodynamic interpretation strengthens its reliability. Furthermore, the thermodynamic interpretation of Schreiber's

[*]kiwata@cc.osaka-kyoiku.ac.jp

transfer entropy has been discussed by Prokopenko *et al.* [26]. Thus, the thermodynamic interpretations of both the Liang-Kleeman information flow and Schreiber's transfer entropy have been clarified; however, the thermodynamic relationship between the two has not been established. Therefore, I reexamine the relationship between the Liang-Kleeman and Horowitz-Esposito notions of information flow. Cafaro *et al.* targeted two systems interacting with each other and claimed that the two representations of the information flow between these systems are equal. In this study, I consider an arbitrary number of systems and investigate the relationship between these two notions of information flow. I elucidate the relationship between Schreiber's transfer entropy and Horowitz's information flow. As stated previously, establishing a direct relationship between them can be highly beneficial.

Physical systems are causal. Theoretically, one can consider systems that do not satisfy causality. Concretely, a coupled chaotic dynamic system [27–29] is a model that does not satisfy Granger's first principle of causality [30], but such a model will not be discussed in this study.

II. MODEL

I consider that the system changes according to a stochastic differential equation, which is given by [31,32]

$$d\vec{X}_t = \vec{F}(\vec{X}_t, t)dt + B(\vec{X}_t, t)\cdot d\vec{W}(t). \tag{1}$$

I assume an N-dimensional system. In Eq. (1), $\vec{X}_t$ and $\vec{F}$ are N-dimensional column vectors denoted by $\vec{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{N,t})^\top$ and $\vec{F}(\vec{X}, t) = (F_1(\vec{X}, t), F_2(\vec{X}, t), \ldots, F_N(\vec{X}, t))^\top$, respectively. The sign "$\top$" denotes the transpose. The quantity $B(\vec{X}, t)$ is an $N\times N$ matrix, and the sign "$\cdot$" represents the Itô product between a matrix and a column vector. In Itô's equation, the elements of the vector of the Wiener increment, that is, $dW_i(t),\ i\in\{1, 2, \ldots, N\}$, satisfy

$$dW_i(t)dW_j(t) = \delta_{ij}dt, \tag{2}$$

where δ_{ij} denotes the Kronecker delta. Because $\vec{X}$ at time t is represented by $\vec{X}_t$, the value of $\vec{X}$ after dt from t is obtained as

$$\vec{X}_{t+dt} = \vec{X}_t + d\vec{X}_t. \tag{3}$$

As $d\vec{W}(t)$ is a random variable, there exist various paths of $\vec{X}_t$. When the value $\vec{X}$ is measured at time t, the probability distribution $p_t(\vec{x})$ exists. This probability distribution is obtained as follows:

$$p_t(\vec{x}) = \langle\delta(\vec{x} - \vec{X}_t)\rangle, \tag{4}$$

where $\langle\cdots\rangle$ represents the average over various paths of $\vec{X}_t$, and $\delta(x)$ denotes Dirac's delta function. The probability distribution $p_t(x_k)$ is obtained by the following marginalization:

$$p_t(x_k) = \int d\vec{x}_{-k}\langle\delta(\vec{x} - \vec{X}_t)\rangle = \int\prod_{i\neq k}^{N}dx_i\left\langle\prod_{j=1}^{N}\delta(x_j - X_{j,t})\right\rangle$$
$$= \langle\delta(x_k - X_{k,t})\rangle, \tag{5}$$

where $\vec{x}_{-k}$ denotes all elements of $\vec{x}$ except for x_k.

To consider the time dependence of $p_t(\vec{x})$, I introduce

$$p_{t+dt}(\vec{x}) = \langle\delta(\vec{x} - \vec{X}_{t+dt})\rangle. \tag{6}$$

Substituting Eq. (3) in Eq. (6) and expanding it with respect to $d\vec{X}_t$, I obtain

$$\begin{aligned}p_{t+dt}(\vec{x}) &= \langle\delta(\vec{x} - \vec{X}_t - d\vec{X}_t)\rangle\\ &= \langle\delta(\vec{x} - \vec{X}_t)\rangle - \langle d\vec{X}_t^{\top}\nabla\delta(\vec{x} - \vec{X}_t)\rangle\\ &\quad + \tfrac{1}{2}\langle(d\vec{X}_t^{\top}\nabla)^2\delta(\vec{x} - \vec{X}_t)\rangle + O((d\vec{X}_t)^3),\end{aligned} \tag{7}$$

where $\nabla = (\partial_1, \partial_2, \ldots, \partial_N)^\top$. In addition, using Itô's lemma, Eq. (7) becomes

$$\begin{aligned}p_{t+dt}(\vec{x}) &= p_t(\vec{x}) - \nabla^\top[\vec{F}(\vec{x}, t)p_t(\vec{x})]dt\\ &\quad + \tfrac{1}{2}\nabla\nabla^\top : [B(\vec{x}, t)B(\vec{x}, t)^\top p_t(\vec{x})]dt + o(dt),\end{aligned} \tag{8}$$

where ":" denotes $A : B = \sum_{i,j}a_{ij}b_{ji}$. According to [23,33], I assume that the system is multipartite; thus, the terms containing $\partial_i\partial_j$ $(i\neq j)$ are disregarded. A detailed explanation of the disregard can be seen in Appendix A. In the multipartite case, Eq. (8) is written as

$$p_{t+dt}(\vec{x}) = p_t(\vec{x}) - \sum_{i=1}^{N}\partial_i J_i(\vec{x}, t)dt + o(dt), \tag{9}$$

where $J_i(\vec{x}, t)$ is defined by

$$J_i(\vec{x}, t) = F_i(\vec{x}, t)p_t(\vec{x}) - \tfrac{1}{2}\partial_i[g_{ii}(\vec{x}, t)p_t(\vec{x})], \tag{10}$$

and $g_{ij}(\vec{x}, t) = (B(\vec{x}, t)B(\vec{x}, t)^\top)_{ij}$. Equation (9) leads to the continuity equation of probability. As per the form of Eq. (9), $J_i(\vec{x}, t)$ represents the ith-direction current of the probabilities. In the multipartite system, the ith-direction current (10) contains only the ∂_i term. Because $g_{ij}(\vec{x}, t)$ is composed of $B(\vec{x}, t)$, $g_{ij}(\vec{x}, t)$ stems from the stochastic fluctuation. In the multipartite system, the ith-direction current is not influenced by the other direction elements. As the formula of the Liang-Kleeman information flow includes only the $g_{ii}(\vec{x}, t)$ term, the Liang-Kleeman information flow would have implicitly assumed that the system is multipartite.

III. DEFINITION OF INFORMATION FLOW IN STOCHASTIC THERMODYNAMICS

In this section, I briefly introduce the current knowledge on information flow in stochastic thermodynamics. To begin with, let us consider the case of a bipartite system, which is formulated by two random variables: $X_{1,t}$ and $X_{2,t}$. Mutual information is useful to examine the interdependence of these two random variables. The mutual information is defined as

$$I(X_{1,t}; X_{2,t}) = \int p_t(\vec{x})\ln\left[\frac{p_t(\vec{x})}{p_t(x_1)p_t(x_2)}\right]d\vec{x}. \tag{11}$$

The mutual information allows the recognition of the interdependence between them but not the causality. To recognize the causality between the two random variables, time-shifted mutual information is useful [34–36]. Using the time-shifted mutual information, the information flow from variable $X_{2,t}$ to $X_{1,t}$ is defined as [37]

$$T_{2\to1} = \lim_{dt\to0}\frac{1}{dt}[I(X_{1,t+dt}; X_{2,t}) - I(X_{1,t}; X_{2,t})]. \tag{12}$$

The time-shifted mutual information $I(X_{1,t+dt}; X_{2,t})$ represents the interdependence between $X_{1,t+dt}$ and $X_{2,t}$. The positive $T_{2\to1}$ shows that the interdependence increases as the time of $X_{1,t}$ changes from t to $t+dt$; this means that the effect of $X_{2,t}$ on $X_{1,t}$ increases with time for $X_{1,t}$. Thus, one can interpret the positive $T_{2\to1}$ as the information flow from $X_{2,t}$ to $X_{1,t}$. It is evident from Eq. (12) that the time derivative of the mutual information becomes

$$\frac{d}{dt}I(X_{1,t};X_{2,t}) = T_{1\to2} + T_{2\to1}. \tag{13}$$

Equation (13) shows that the time derivative of the mutual information is equal to the sum of the exchange of the information flow. Using $N=2$ in Eq. (9), one can derive another formula for information flow, namely,

$$T_{2\to1} = \int J_1(\vec{x},t)\partial_1 \ln[p_t(x_2|x_1)]d\vec{x}, \tag{14}$$

where $p_t(x_2|x_1)$ is a conditional probability distribution defined by $p_t(\vec{x})/p_t(x_1)$ [23,37].

The extension of the information flow to a multipartite system has been given by Horowitz [33]. Let us consider that a system varies according to the stochastic differential equation of the N-dimensional system. The system is divided into two parts: the kth variable $X_{k,t}$ and $\vec{X}_{-k,t} = \{X_{1,t},\dots,X_{k-1,t},X_{k+1,t},\dots,X_{N,t}\}$. The mutual information between $X_{k,t}$ and $\vec{X}_{-k,t}$ is defined as

$$I(X_{k,t};\vec{X}_{-k,t}) = \int p_t(\vec{x}) \ln\left[\frac{p_t(\vec{x})}{p_t(x_k)p_t(\vec{x}_{-k})}\right]d\vec{x}, \tag{15}$$

where $p_t(\vec{x}_{-k}) = \int p_t(\vec{x})dx_k$. The derivative of Eq. (15) with respect to t becomes

$$\frac{d}{dt}I(X_{k,t};\vec{X}_{-k,t}) = \int J_k(\vec{x},t)\partial_k \ln[p_t(\vec{x}_{-k}|x_k)]d\vec{x} + \sum_{l\neq k}\int J_l(\vec{x},t)\partial_l \ln[p_t(x_k|\vec{x}_{-k})]d\vec{x}. \tag{16}$$

Each term on the right-hand side of the equation can be interpreted as the information flow between $X_{k,t}$ and $\vec{X}_{-k,t}$. By analogy with Eq. (14), the first term on the right-hand side of Eq. (16) corresponds to the information flow from $\vec{X}_{-k,t}$ to $X_{k,t}$. The second term corresponds to the sum of the information flow from $X_{k,t}$ to $X_{l,t}$, that is,

$$\frac{d}{dt}I(X_{k,t};\vec{X}_{-k,t}) = \int J_k(\vec{x},t)\partial_k \ln[p_t(\vec{x}_{-k}|x_k)]d\vec{x} + \sum_{l\neq k} T_{k\to l}, \tag{17}$$

where $T_{k\to l}$ is defined as

$$T_{k\to l} = \int J_l(\vec{x},t)\partial_l \ln[p_t(x_k|\vec{x}_{-k})]d\vec{x}. \tag{18}$$

Equation (18) has been introduced by Horowitz [33], and henceforth, I call Eq. (18) Horowitz's information flow. The stochastic thermodynamic aspect of Horowitz's information flow has been discussed by Horowitz [33].

IV. RESULTS

A. Formulation of Horowitz's information flow in a multipartite system

In this section, I demonstrate the properties of Horowitz's information flow, namely, Eq. (18). Let us see another formulation of Horowitz's information flow that is useful for interpreting it. I consider the following difference between the two mutual information:

$$\begin{aligned} I(X_{k,t};\{X_{l,t+dt},\vec{X}_{-\{k,l\},t}\}) - I(X_{k,t};\{X_{l,t},\vec{X}_{-\{k,l\},t}\}) &= \int p(x'_l,\vec{x}_{-l}) \ln\left[\frac{p(x'_l,\vec{x}_{-l})}{p(x'_l,\vec{x}_{-\{k,l\}})p_t(x_k)}\right]dx'_l d\vec{x}_{-l} \\ &\quad - \int p_t(\vec{x}) \ln\left[\frac{p_t(\vec{x})}{p_t(\vec{x}_{-k})p_t(x_k)}\right]d\vec{x}, \end{aligned} \tag{19}$$

where $p(x'_l,\vec{x}_{-l})$ is a joint probability distribution defined as

$$p(x'_l,\vec{x}_{-l}) = \langle\delta(x'_l - X_{l,t+dt})\delta(\vec{x}_{-l}-\vec{X}_{-l,t})\rangle, \tag{20}$$

and $\vec{x}_{-\{k,l\}}$ represents all elements of vector $\vec{x}$, except for x_k and x_l. In Eq. (20), the variables with a prime represent random variables at time $t+dt$, and the variables without a prime represent those at time t. This probability distribution function is composed of random variables at different times; thus, the subscript t is not attached to p in Eq. (20). Each term in Eq. (19) consists of the mutual information between $X_{k,t}$ and $\vec{X}_{-k,t}$; moreover, the lth random variable of the first term is time shifted. The right-hand side of Eq. (19) becomes

$$\begin{aligned} I(X_{k,t};\ \{X_{l,t+dt},\vec{X}_{-\{k,l\},t}\}) - I(X_{k,t};\{X_{l,t},\vec{X}_{-\{k,l\},t}\}) &= \int p(x'_l,\vec{x}_{-l}) \ln\left[\frac{p(x'_l,\vec{x}_{-\{k,l\}}|x_k)}{p(x'_l,\vec{x}_{-\{k,l\}})}\right]dx'_l d\vec{x}_{-l} - \int p_t(\vec{x})\ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right]d\vec{x} \\ &= \underbrace{\int p(x'_l,\vec{x}_{-l}) \ln[p(x'_l,\vec{x}_{-\{k,l\}}|x_k)]dx'_l d\vec{x}_{-l}}_{\mathcal{A}} \\ &\quad - \underbrace{\int p(x'_l,\vec{x}_{-l}) \ln[p(x'_l,\vec{x}_{-\{k,l\}})]dx'_l d\vec{x}_{-l}}_{\mathcal{B}} \\ &\quad - \int p_t(\vec{x})\ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right]d\vec{x}. \end{aligned} \tag{21}$$

First, I estimate the term "$\mathcal{A}$" in Eq. (21). Substituting $X_{l,t+dt} = X_{l,t} + dX_{l,t}$ into Eq. (20), expanding it with respect to $dX_{l,t}$, and utilizing Itô's lemma, Eq. (20) becomes

$$\begin{aligned} p(x'_l,\vec{x}_{-l}) &= \langle\delta(x'_l - X_{l,t})\delta(\vec{x}_{-l}-\vec{X}_{-l,t})\rangle \\ &\quad - \partial'_l[F_l(x'_l,\vec{x}_{-l},t)\langle\delta(x'_l - X_{l,t})\delta(\vec{x}_{-l}-\vec{X}_{-l,t})\rangle]dt \\ &\quad + \tfrac{1}{2}\partial'^2_l[g_{ll}(x'_l,\vec{x}_{-l},t)\langle\delta(x'_l - X_{l,t}) \\ &\quad \times \delta(\vec{x}_{-l}-\vec{X}_{-l,t})\rangle]dt + o(dt) \\ &= p_t(x'_l,\vec{x}_{-l}) - \partial'_l J_l(x'_l,\vec{x}_{-l},t)dt + o(dt) \\ &= p_t(x'_l,\vec{x}_{-l}) - \bigcirc dt + o(dt). \end{aligned} \tag{22}$$

Comparing Eq. (22) with Eq. (9), the time is advanced from t to $t + dt$ with respect to $X_{l,t}$, leading to the appearance of the l-direction current in Eq. (22). To save space, I introduce the symbol "$\bigcirc$". Using Eq. (22), the conditional probability distribution $p(x_l', \vec{x}_{-\{k,l\}}|x_k)$ is derived as

$$p(x_l', \vec{x}_{-\{k,l\}}|x_k) = \frac{p(x_l', \vec{x}_{-l})}{p_t(x_k)} = \frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_k)} - \frac{\bigcirc}{p_t(x_k)}dt + o(dt). \tag{23}$$

Substituting Eqs. (22) and (23) into $\mathcal{A}$ and expanding the result with respect to dt up to the first-order term gives

$$\begin{aligned}
\mathcal{A} &= \int \{p_t(x_l', \vec{x}_{-l}) - \bigcirc dt + o(dt)\} \ln\left[\frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_k)} - \frac{\bigcirc}{p_t(x_k)}dt + o(dt)\right]dx_l' d\vec{x}_{-l} \\
&= \int \{p_t(x_l', \vec{x}_{-l}) - \bigcirc dt + o(dt)\}\left\{\ln\left[\frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_k)}\right] - \frac{\bigcirc}{p_t(x_l', \vec{x}_{-l})}dt + o(dt)\right\}dx_l' d\vec{x}_{-l} \\
&= \int p_t(x_l', \vec{x}_{-l}) \ln\left[\frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_k)}\right]dx_l' d\vec{x}_{-l} - \int \bigcirc dx_l' d\vec{x}_{-l} dt - \int \bigcirc \ln\left[\frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_k)}\right]dx_l' d\vec{x}_{-l} dt + o(dt) \\
&= \int p_t(\vec{x}) \ln\left[\frac{p_t(\vec{x})}{p_t(x_k)}\right]d\vec{x} - \int [\partial_l J_l(\vec{x}, t)] \ln\left[\frac{p_t(\vec{x})}{p_t(x_k)}\right]d\vec{x} dt + o(dt) \\
&= \int p_t(\vec{x}) \ln[p_t(\vec{x}_{-k}|x_k)]d\vec{x} - \int [\partial_l J_l(\vec{x}, t)] \ln[p_t(\vec{x}_{-k}|x_k)]d\vec{x} dt + o(dt).
\end{aligned} \tag{24}$$

The approximate expression $\ln(a + \Delta) \simeq \ln a + \Delta/a$, which holds when Δ is sufficiently small, is used during the transformation from the first line to the second one. In the transformation from the third line to the fourth one, I use the formula $\int \partial_l' J_l(x_l', \vec{x}_{-l}, t)dx_l' = 0$, which holds when $\lim_{x_l' \to \pm\infty} J_l(x_l', \vec{x}_{-l}, t) = 0$. Furthermore, I transform the integral variable from x_l' to x_l. Then, I estimate the term "$\mathcal{B}$" in Eq. (21). By integrating Eq. (22) with respect to x_k, I obtain

$$\begin{aligned}
p(x_l', \vec{x}_{-\{k,l\}}) &= \int p(x_l', \vec{x}_{-l})dx_k \\
&= p_t(x_l', \vec{x}_{-\{k,l\}}) - \int \partial_l' J_l(x_l', \vec{x}_{-\{k,l\}}, x_k'', t)dx_k'' dt + o(dt) \\
&= p_t(x_l', \vec{x}_{-\{k,l\}}) - \int \bigcirc dx_k'' dt + o(dt).
\end{aligned} \tag{25}$$

For convenience, the integral variable is set to x_k'' to distinguish it from the other variables in the second line of Eq. (25). Substituting Eqs. (22) and (25) into $\mathcal{B}$, we arrive at

$$\begin{aligned}
\mathcal{B} &= \int \{p_t(x_l', \vec{x}_{-l}) - \bigcirc dt + o(dt)\} \ln\left[p_t(x_l', \vec{x}_{-\{k,l\}}) - \int \bigcirc dx_k'' dt + o(dt)\right]dx_l' d\vec{x}_{-l} \\
&= \int \{p_t(x_l', \vec{x}_{-l}) - \bigcirc dt + o(dt)\}\left\{\ln[p_t(x_l', \vec{x}_{-\{k,l\}})] - \frac{\int \bigcirc dx_k''}{p_t(x_l', \vec{x}_{-\{k,l\}})}dt + o(dt)\right\}dx_l' d\vec{x}_{-l} \\
&= \int p_t(x_l', \vec{x}_{-l}) \ln[p_t(x_l', \vec{x}_{-\{k,l\}})]dx_l' d\vec{x}_{-l} - \underline{\int \frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_l', \vec{x}_{-\{k,l\}})} \int \bigcirc dx_k'' dx_l' d\vec{x}_{-l} dt} \\
&\quad - \int \bigcirc \ln[p_t(x_l', \vec{x}_{-\{k,l\}})]dx_l' d\vec{x}_{-l} dt + o(dt).
\end{aligned} \tag{26}$$

Regarding the underlined term in Eq. (26), by integrating with respect to x_k, I obtain

$$\begin{aligned}
\int \frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_l', \vec{x}_{-\{k,l\}})} \int \bigcirc dx_k'' dx_l' d\vec{x}_{-l} dt &= \int \frac{p_t(x_l', \vec{x}_{-l})}{p_t(x_l', \vec{x}_{-\{k,l\}})} dx_k \int \bigcirc dx_k'' dx_l' d\vec{x}_{-\{k,l\}} dt \\
&= \int \bigcirc dx_k'' dx_l' d\vec{x}_{-\{k,l\}} dt \\
&= \int \partial_l' J_l(x_l', \vec{x}_{-\{k,l\}}, x_k'', t) dx_k'' dx_l' d\vec{x}_{-\{k,l\}} dt \\
&= 0.
\end{aligned} \tag{27}$$

In the last line, I assume that $J_l(x_l', \vec{x}_{-\{k,l\}}, x_k'', t)$ becomes 0 at $x_l' \to \pm\infty$. By transforming the integral variable from x_l' to x_l, formula $\mathcal{B}$ becomes

$$\mathcal{B} = \int p_t(\vec{x}) \ln[p_t(\vec{x}_{-k})] d\vec{x} - \int [\partial_l J_l(\vec{x}, t)] \ln[p_t(\vec{x}_{-k})] d\vec{x} dt + o(dt). \tag{28}$$

Substituting Eqs. (24) and (28) into Eq. (21), I obtain

$$\begin{aligned}
&I(X_{k,t}; \{X_{l,t+dt}, \vec{X}_{-\{k,l\},t}\}) - I(X_{k,t}; \{X_{l,t}, \vec{X}_{-\{k,l\},t}\}) \\
&= \underbrace{\int p_t(\vec{x}) \ln[p_t(\vec{x}_{-k}|x_k)] d\vec{x} - \int [\partial_l J_l(\vec{x}, t)] \ln[p_t(\vec{x}_{-k}|x_k)] d\vec{x} dt}_{\mathcal{A}} \\
&\quad \underbrace{- \int p_t(\vec{x}) \ln[p_t(\vec{x}_{-k})] d\vec{x} + \int [\partial_l J_l(\vec{x}, t)] \ln[p_t(\vec{x}_{-k})] d\vec{x} dt}_{\mathcal{B}} - \int p_t(\vec{x}) \ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right] d\vec{x} + o(dt) \\
&= -\int [\partial_l J_l(\vec{x}, t)] \ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right] d\vec{x} dt + o(dt) \\
&= \int J_l(\vec{x}, t) \partial_l \ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right] d\vec{x} dt + o(dt) \\
&= \int J_l(\vec{x}, t) \partial_l \ln[p_t(x_k|\vec{x}_{-k})] d\vec{x} dt + o(dt) \\
&= T_{k\to l} dt + o(dt).
\end{aligned} \tag{29}$$

As a result, I obtain the following formula of Horowitz's information flow:

$$T_{k\to l} = \lim_{dt\to 0} \frac{1}{dt} [I(X_{k,t}; \{X_{l,t+dt}, \vec{X}_{-\{k,l\},t}\}) - I(X_{k,t}; \{X_{l,t}, \vec{X}_{-\{k,l\},t}\})]. \tag{30}$$

This formula is useful in understanding Horowitz's information flow. Using the following identity of the mutual information,

$$I(X; \{Y, Z\}) = I(X; Z) + I(X; Y|Z), \tag{31}$$

Eq. (30) becomes

$$\begin{aligned}
T_{k\to l} &= \lim_{dt\to 0} \frac{1}{dt} [I(X_{k,t}; \vec{X}_{-\{k,l\},t}) + I(X_{k,t}; X_{l,t+dt}|\vec{X}_{-\{k,l\},t}) - I(X_{k,t}; \vec{X}_{-\{k,l\},t}) - I(X_{k,t}; X_{l,t}|\vec{X}_{-\{k,l\},t})] \\
&= \lim_{dt\to 0} \frac{1}{dt} [I(X_{k,t}; X_{l,t+dt}|\vec{X}_{-\{k,l\},t}) - I(X_{k,t}; X_{l,t}|\vec{X}_{-\{k,l\},t})].
\end{aligned} \tag{32}$$

Equation (32) corresponds to Eq. (12), and it can been seen that Horowitz's information flow is formulated by the conditional time-shifted mutual information. Furthermore, using the following identity of the conditional mutual information and the conditional Shannon entropy,

$$I(X; Y|Z) = S(Y|Z) - S(Y|\{X, Z\}), \tag{33}$$

Eq. (32) can be transformed into

$$\begin{aligned}
T_{k\to l} &= \lim_{dt\to 0} \frac{1}{dt} [S(X_{l,t+dt}|\vec{X}_{-\{k,l\},t}) - S(X_{l,t+dt}|\vec{X}_{-l,t}) - S(X_{l,t}|\vec{X}_{-\{k,l\},t}) + S(X_{l,t}|\vec{X}_{-l,t})] \\
&= \frac{d}{d\tau} S(X_{l,t+\tau}|\vec{X}_{-\{k,l\},t})\bigg|_{\tau=0} - \frac{d}{d\tau} S(X_{l,t+\tau}|\vec{X}_{-l,t})\bigg|_{\tau=0}.
\end{aligned} \tag{34}$$

Horowitz's information flow is represented by the difference of the conditional Shannon entropy. In the case of the bipartite system, Eq. (34) becomes

$$T_{2\to 1} = \frac{d}{d\tau} S(X_{1,t+\tau})\bigg|_{\tau=0} - \frac{d}{d\tau} S(X_{1,t+\tau}|X_{2,t})\bigg|_{\tau=0}, \tag{35}$$

and it reproduces the information flow of the bipartite system. The first term on the right-hand side of Eq. (35) represents the time derivative of the entropy of the marginal probability of $X_{1,t}$, and the second term represents the time derivative of the entropy of the conditional probability of $X_{1,t}$, given $X_{2,t}$. If I map the condition $X_{2,t}$ onto the frozen random variable $X_{2,t}$, then Eq. (35) is consistent with the Liang-Kleeman information flow. For the information flow of the multipartite system, namely, Eq. (34), each term consists of the time derivative of the entropy of the conditional probability of $X_{l,t}$, given other random variables. The difference between the two terms is that the first term does not include $X_{k,t}$ in the given random variables, unlike the second term. For the multipartite system, the Liang-Kleeman information flow is formulated in a form similar to that of Eq. (35). For the multipartite system, Horowitz's information flow does not accord with the Liang-Kleeman information flow.

B. Liang-Kleeman information flow and Horowitz's information flow

In this section, I demonstrate the relationship between the Liang-Kleeman information flow and Horowitz's information flow. First, let us consider the bipartite system. Substituting Eq. (10) in Eq. (14) and repeating the integration by parts, I derive

$$\begin{aligned}
T_{2\to 1} &= \int \left\{ F_1(\vec{x},t)p_t(\vec{x}) - \frac{1}{2}\partial_1[g_{11}(\vec{x},t)p_t(\vec{x})]\right\}\partial_1 \ln[p_t(x_2|x_1)]d\vec{x} \\
&= \int \left\{ F_1(\vec{x},t)p_t(\vec{x})\frac{1}{p_t(x_2|x_1)}\partial_1[p_t(x_2|x_1)] - \frac{1}{2}\partial_1[g_{11}(\vec{x},t)p_t(\vec{x})]\frac{1}{p_t(x_2|x_1)}\partial_1[p_t(x_2|x_1)]\right\}d\vec{x} \\
&= \int F_1(\vec{x},t)p_t(x_1)\partial_1[p_t(x_2|x_1)]d\vec{x} + \frac{1}{2}\int g_{11}(\vec{x},t)p_t(\vec{x})\partial_1\left\{\frac{1}{p_t(x_2|x_1)}\partial_1[p_t(x_2|x_1)]\right\}d\vec{x} \\
&= -\int p_t(x_2|x_1)\partial_1[F_1(\vec{x},t)p_t(x_1)]d\vec{x} - \frac{1}{2}\int g_{11}(\vec{x},t)p_t(\vec{x})\left\{\frac{1}{p_t(x_2|x_1)}\partial_1[p_t(x_2|x_1)]\right\}^2 d\vec{x} \\
&\quad + \frac{1}{2}\int g_{11}(\vec{x},t)p_t(\vec{x})\frac{1}{p_t(x_2|x_1)}\partial_1^2[p_t(x_2|x_1)]d\vec{x} \\
&= -\int p_t(x_2|x_1)\partial_1[F_1(\vec{x},t)p_t(x_1)]d\vec{x} - \frac{1}{2}\int g_{11}(\vec{x},t)p_t(\vec{x})\{\partial_1 \ln[p_t(x_2|x_1)]\}^2 d\vec{x} \\
&\quad + \frac{1}{2}\int g_{11}(\vec{x},t)p_t(x_1)\partial_1^2[p_t(x_2|x_1)]d\vec{x} \\
&= -\int p_t(x_2|x_1)\partial_1[F_1(\vec{x},t)p_t(x_1)]d\vec{x} + \frac{1}{2}\int p_t(x_2|x_1)\partial_1^2[g_{11}(\vec{x},t)p_t(x_1)]d\vec{x} \\
&\quad - \frac{1}{2}\int g_{11}(\vec{x},t)p_t(\vec{x})\{\partial_1 \ln[p_t(x_2|x_1)]\}^2 d\vec{x},
\end{aligned} \tag{36}$$

where the property of $p_t(\vec{x})$ becoming 0 at $|x_i| \to \infty$ is utilized during the integration by parts. The first and second terms in the last line of Eq. (36) are in accordance with the Liang-Kleeman information flow. However, an extra term appears in $T_{2\to 1}$. Because the extra term contains $g_{11}(\vec{x},t)$, it originates from the stochastic fluctuation.

For the multipartite system, substituting Eq. (10) for Eq. (18) and repeating the integration by parts as is in the case of Eq. (36), I obtain

$$\begin{aligned}
T_{k\to l} = &-\int p_t(x_k|\underline{\vec{x}_{-k}})\partial_l[F_l(\vec{x},t)p_t(\vec{x}_{-k})]d\vec{x} + \frac{1}{2}\int p_t(x_k|\underline{\vec{x}_{-k}})\partial_l^2[g_{ll}(\vec{x},t)p_t(\vec{x}_{-k})]d\vec{x} \\
&- \frac{1}{2}\int g_{ll}(\vec{x},t)p_t(\vec{x})\{\partial_l \ln[p_t(x_k|\vec{x}_{-k})]\}^2 d\vec{x}.
\end{aligned} \tag{37}$$

The Liang-Kleeman information flow of the multipartite system is given by

$$T_{k\to l}^{(\mathrm{LK})} = -\int p_t(x_k|\underline{x_l})\partial_l[F_l(\vec{x},t)p_t(\vec{x}_{-k})]d\vec{x} + \frac{1}{2}\int p_t(x_k|\underline{x_l})\partial_l^2[g_{ll}(\vec{x},t)p_t(\vec{x}_{-k})]d\vec{x}. \tag{38}$$

The first two terms in Eq. (37) are identical to the Liang-Kleeman information flow, except for the underlined parts, which differ from each other. An extra term appears in Horowitz's information flow (i.e., compared to the expression for the Liang-Kleeman information flow). The same situation is observed in the case of the bipartite system.

The information flow of the bipartite system, defined by Eq. (14), satisfies Eq. (13). Horowitz's information flow satisfies Eq. (17). Equations (14) and (18) originate from the time derivative of the mutual information. Because the Liang-Kleeman information flow does not accord with either Eq. (14) or Eq. (18), the Liang-Kleeman information flow is incompatible with the time derivative of the mutual information, at least with respect to the terms generated by the stochastic fluctuations.

C. Schreiber's transfer entropy and Horowitz's information flow

In this section, I investigate the relationship between Schreiber's transfer entropy and Horowitz's information flow, as defined by Eq. (30). Schreiber's transfer entropy from k to l is defined as

$$T_{k\to l}^{(\mathrm{S})} = \int p(x_l', \vec{x}_{-k}, x_k) \ln\left[\frac{p(x_l'|\vec{x}_{-k}, x_k)}{p(x_l'|\vec{x}_{-k})}\right]dx_l' d\vec{x}_{-k} dx_k, \tag{39}$$

where $p(x'_l, \vec{x}_{-k}, x_k)$ is a joint probability distribution defined as

$$p(x'_l, \vec{x}_{-k}, x_k) = p(x'_l, \vec{x}) = \langle \delta(x'_l - X_{l,t+dt})\delta(\vec{x} - \vec{X}_t)\rangle. \tag{40}$$

Similar to the notation in Eq. (20), the variables with a prime represent random variables at time $t + dt$, and the variables without a prime represent those at time t in Eq. (40). This probability distribution function is composed of random variables at different times; thus, subscript t is not attached to p in Eq. (40). Equation (39) is transformed into

$$\begin{aligned}
T^{(\mathrm{S})}_{k\to l} &= \int p(x'_l, \vec{x}_{-k}, x_k) \ln\left[\frac{p(x'_l, \vec{x}_{-k}, x_k)}{p_t(\vec{x}_{-k}, x_k)} \frac{p_t(\vec{x}_{-k})}{p(x'_l, \vec{x}_{-k})}\right] dx'_l d\vec{x}_{-k} dx_k \\
&= \int p(x'_l, \vec{x}_{-k}, x_k) \ln\left[\frac{p(x'_l, \vec{x}_{-k}, x_k)}{p(x'_l, \vec{x}_{-k})} \frac{p_t(\vec{x}_{-k})}{p_t(\vec{x}_{-k}, x_k)} \frac{p_t(x_k)}{p_t(x_k)}\right] dx'_l d\vec{x}_{-k} dx_k \\
&= \int p(x'_l, \vec{x}_{-k}, x_k) \ln\left[\frac{p(x'_l, \vec{x}_{-k}|x_k)}{p(x'_l, \vec{x}_{-k})}\right] dx'_l d\vec{x}_{-k} dx_k - \int p_t(\vec{x}_{-k}, x_k) \ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right] d\vec{x}_{-k} dx_k \\
&= \int p(x'_l, \vec{x}_{-\{k,l\}}, x_l, x_k) \ln\left[\frac{p(x'_l, \vec{x}_{-\{k,l\}}, x_l|x_k)}{p(x'_l, \vec{x}_{-\{k,l\}}, x_l)}\right] dx'_l d\vec{x}_{-\{k,l\}} dx_l dx_k - \int p_t(\vec{x}) \ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right] d\vec{x}.
\end{aligned} \tag{41}$$

For the first term in the last line of Eq. (41), I select x_l from $\vec{x}_{-k}$, provided $k \neq l$. Applying the log-sum inequality for x_l to the first term of Eq. (41), I can extract x_l, and Eq. (41) becomes

$$\begin{aligned}
T^{(\mathrm{S})}_{k\to l} &\geqslant \int p(x'_l, \vec{x}_{-\{k,l\}}, x_k) \ln\left[\frac{p(x'_l, \vec{x}_{-\{k,l\}}|x_k)}{p(x'_l, \vec{x}_{-\{k,l\}})}\right] dx'_l d\vec{x}_{-\{k,l\}} dx_k - \int p_t(\vec{x}) \ln\left[\frac{p_t(\vec{x}_{-k}|x_k)}{p_t(\vec{x}_{-k})}\right] d\vec{x} \\
&= I(X_{k,t}; \{X_{l,t+dt}, \vec{X}_{-\{k,l\},t}\}) - I(X_{k,t}; \{X_{l,t}, \vec{X}_{-\{k,l\},t}\}) \\
&= T_{k\to l}dt + o(dt),
\end{aligned} \tag{42}$$

where I utilize Eq. (30) from the second line in the last line. Equation (42) shows that Schreiber's transfer entropy rate is an upper bound of Horowitz's information flow.

Let us discuss the case where the equality sign holds. From the property of the log-sum inequality, the equality sign of Eq. (42) holds if and only if $p(x'_l, \vec{x}_{-\{k,l\}}, x_l|x_k)/p(x'_l, \vec{x}_{-\{k,l\}}, x_l)$ are equal for all x_l. Introducing a constant c, which does not depend on x_l, the equality condition becomes $p(x'_l, \vec{x}_{-\{k,l\}}, x_l|x_k) = c\, p(x'_l, \vec{x}_{-\{k,l\}}, x_l)$. For the sake of explanation, let us assume that c does not depend on the other random variables. After marginalization, the equation becomes $p_t(x_l, x_k) = p_t(x_l)p_t(x_k)$. In this case, the independence of x_k from the other random variables is prescribed. In the general case for c, the probability distribution condition is more complicated. Horowitz's information flow contains a condition relating to the dynamics, but Schreiber's transfer entropy does not. I consider that this condition about dynamics is why Schreiber's transfer entropy is not lower than Horowitz's information flow.

D. Applications

1. Linear stochastic system

To examine the effectiveness of Horowitz's information flow, I apply it to a linear stochastic system. The linear stochastic system is prescribed by the following stochastic differential equation:

$$d\vec{X}_t = A\vec{X}_t dt + B \cdot d\vec{W}(t), \tag{43}$$

where A and B are $N \times N$ constant matrices. The linear stochastic system bears the following useful characteristic. When the initial distribution is a normal distribution, $p_t(\vec{x})$ is a normal distribution forever, namely,

$$p_t(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N \det\Sigma}} \exp\left[-\frac{1}{2}(\vec{x} - \vec{\mu})^\top \Sigma^{-1}(\vec{x} - \vec{\mu})\right]. \tag{44}$$

The mean vector $\vec{\mu}$ and covariance matrix Σ of the normal distribution change according to the following equations, respectively:

$$\frac{d\vec{\mu}}{dt} = A\vec{\mu}, \tag{45}$$

$$\frac{d\Sigma}{dt} = A\Sigma + \Sigma A^\top + BB^\top. \tag{46}$$

Substituting Eq. (44) with Eq. (37), I derive an analytic formula of Horowitz's information flow for the linear stochastic system. The formula is given by

$$T_{k\to l} = -a_{lk}\frac{C_{lk}}{C_{kk}} - \frac{1}{2}g_{ll}\frac{C_{lk}}{\det\Sigma}\frac{C_{lk}}{C_{kk}}, \tag{47}$$

where C_{ij} is an (ij) cofactor of the covariance matrix Σ. The difference is that the Liang-Kleeman information flow is represented by elements of the covariance matrix [17], while Horowitz's information flow is represented by the cofactor of the covariance matrix. The appearance of the cofactor in Eq. (47) originates from $p_t(x_k|\vec{x}_{-k})$ in Eq. (37). The second term on the right-hand side of Eq. (47) corresponds to the last term in Eq. (37) and stems from the stochastic fluctuation. The brief derivation of Eq. (47) can be found in Appendix B.

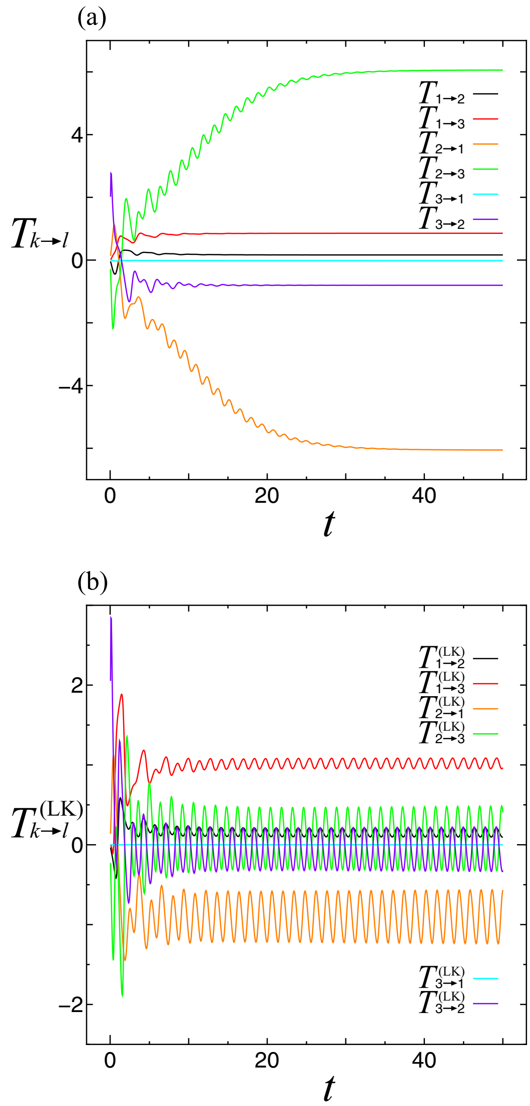

FIG. 1. Temporal changes in information flow: (a) Horowitz's information flow and (b) Liang-Kleeman information flow.

These two formulas of the information flow are compared using an example. The values of parameters are set to

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -5 \\ -1 & 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}. \tag{48}$$

Figure 1 shows time evolution of the information flow. The characteristic difference between the two types of information flow is that Horowitz's information flow [Fig. 1(a)] approaches a constant value after a transition period, while the Liang-Kleeman information flow [Fig. 1(b)] oscillates in time. Because of $a_{13} = 0$, there does not exist the information flow from 3 to 1. Figure 1 shows that the Liang-Kleeman information flow from 3 to 1 is zero; similar to Horowitz's information flow, $T_{3\to1}$ is almost zero compared to other Horowitz information flows. Each of Horowitz's information flows, in order of magnitude, is $T_{2\to3} > T_{1\to3} > T_{1\to2} > T_{3\to1} > T_{3\to2} > T_{2\to1}$. If we consider the center of the oscillation with respect to the Liang-Kleeman information flow as the value of the information flow and arrange each of the Liang-Kleeman information flows in order of magnitude, $T^{(LK)}_{1\to3} > T^{(LK)}_{1\to2} > T^{(LK)}_{2\to3} > T^{(LK)}_{3\to1} > T^{(LK)}_{3\to2} > T^{(LK)}_{2\to1}$ is obtained. Comparing these two types of information flow, the only difference is the position of $T^{(LK)}_{2\to3}$. Otherwise, the order is the same. For the Liang-Kleeman information flow, the amplitude of $T^{(LK)}_{2\to3}$ is the largest. Hence, it is questionable to use the center of the oscillation as the representative value of the information flow.

For the normal distribution, the analytic formula of Schreiber's transfer entropy has already been derived in [38–40]. Schreiber's transfer entropy rate of the linear stochastic model was evaluated with the parameters given by Eqs. (48); however, Schreiber's transfer entropy rate violently oscillates in time. I consider that the failure of the evaluation of Schreiber's transfer entropy rate is due to the nonstationarity of the system. On employing the above-mentioned parameter values, the mean vector $\vec{\mu}$ and covariance matrix Σ are time varying; thus, the system is nonstationary.

2. Stochastic gradient system

The stochastic gradient system can be considered as another application of Horowitz's information flow [41,42]. Here, the drift vector $\vec{F}(\vec{X}, t)$ in Eq. (1) is given by $\vec{F}(\vec{X}, t) = -\nabla V(\vec{X})$, where V is a potential function. For simplicity, let $B(\vec{X}, t)$ be a constant times the identity matrix, i.e., $b \times I$, and then one obtains $p(\vec{x}) \propto \exp[-2V(\vec{x})/b^2]$ as the stationary probability distribution. To compare Horowitz's information flow with the results in [17,18], I adopted the following three-dimensional case as a potential function:

$$V(\vec{x}) = \frac{1}{2}\left(x_1^2x_2^2 + x_2^2x_3^2 + x_1^2 + x_2^2 + x_3^2\right). \tag{49}$$

By differentiating the potential function, the drift vector is obtained as

$$F_1 = -x_1x_2^2 - x_1, \tag{50}$$

$$F_2 = -x_2x_1^2 - x_2x_3^2 - x_2, \tag{51}$$

$$F_3 = -x_3x_2^2 - x_3. \tag{52}$$

As is clear from the formulas, the equations are nonlinear. Considering the symmetry of the system, the following relationships of the information flow are obtained:

$$T_{1\to2} = T_{3\to2}, \quad T_{2\to1} = T_{2\to3}, \quad T_{3\to1} = T_{1\to3}. \tag{53}$$

Substituting the stationary probability distribution, i.e., Eq. (49), in Eq. (37), Horowitz's information flow is derived. The brief derivation of Horowitz's information flow is available in Appendix C. It is worth noting that analytically $T_{3\to1} = 0$. As the first element of the drift vector [i.e., Eq. (50)] does not depend on x_3, the variable x_3 does not cause x_1 immediately. Hence, one can observe the principle of nil causality [17], not only in the Liang-Kleeman information flow but also in Horowitz's information flow. Regarding the other Horowitz information flow, $T_{1\to2}$ and $T_{2\to1}$, the formulas are available in Appendix C. These two can be obtained by performing numerical integration. Figure 2(a) shows Horowitz's information flow of the stochastic gradient system. It can be observed that, as b increases, Horowitz's information flow also increases except for $T_{3\to1}$.

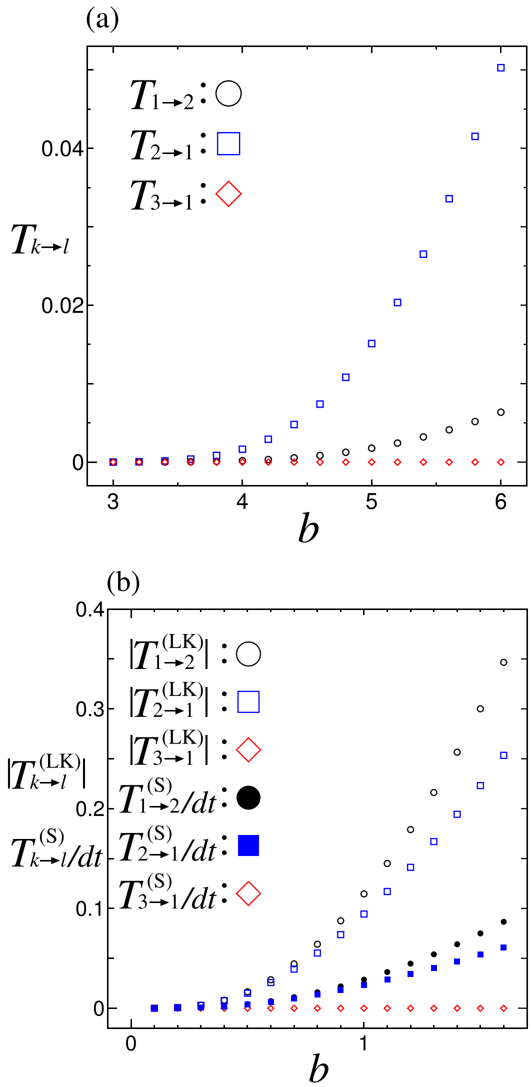

FIG. 2. Information flow and Schreiber's transfer entropy rate for a stochastic gradient system as a function of b: (a) Horowitz's information flow; (b) Liang-Kleeman information flow and Schreiber's transfer entropy rate. $T^{(S)}_{k\to l}/dt$ implies $\lim_{dt\to 0} T^{(S)}_{k\to l}/dt$.

The Liang-Kleeman information flow of the stochastic gradient system is shown in Fig. 2(b). Instead of performing all the integrals numerically, as in [18], I performed the calculations analytically as far as possible to estimate the Liang-Kleeman information flow. I analytically demonstrate $T^{(LK)}_{3\to1} = 0$. The detailed calculation is presented in Appendix D. With regard to $|T^{(LK)}_{1\to2}|$ and $|T^{(LK)}_{2\to1}|$, it is evident from Fig. 2(b) that, as b increases, these two information flows increase with a smaller value of b, as compared to the case of Horowitz's information flow.

Schreiber's transfer entropy rate is also depicted in Fig. 2(b). The formulas for Schreiber's transfer entropy rate are presented in Appendix E. Schreiber's transfer entropy is a quantity with the order of magnitude of dt. I compare the other information flows with Schreiber's transfer entropy divided by dt and taken to the $dt \to 0$ limit. With an increase in b, as is evident from the figure, Schreiber's transfer entropy rate becomes nonzero for small values of b, as compared to the case of Horowitz's information flow. Thus, it can be confirmed that Eq. (42) is satisfied.

V. SUMMARY AND DISCUSSION

The relationship between Schreiber's transfer entropy and the Liang-Kleeman information flow, which are indicators of causality between variables, was clarified through Horowitz's information flow. I initially examined Horowitz's information flow derived from the time derivative of the mutual information [i.e., Eq. (18)]. I derived an alternative formula for Horowitz's information flow, which helped the interpretation. The alternative formula demonstrates that Horowitz's information flow is a reasonable extension from the bipartite system to a multipartite one. As for the deterministic term of the stochastic differential equation, Horowitz's information flow is in accordance with the Liang-Kleeman information flow. However, considering the stochastic fluctuation, I recognized the disagreement between the two formulas. Thus, I considered that the Liang-Kleeman information flow is not compatible with the time derivative of the mutual information.

I also examined the relationship between Schreiber's transfer entropy and Horowitz's information flow. Schreiber's transfer entropy is defined by the Kullback-Leibler divergence of the conditional probability distributions of a random variable at time $t + dt$, given the random variables at time t. I demonstrated the inequality between Schreiber's transfer entropy rate and Horowitz's information flow. The appearance of the rate can be attributed to the Kullback-Leibler divergence of the conditional probabilities of minute time changes. Schreiber's transfer entropy rate is the upper limit of Horowitz's information flow.

In [33], Horowitz demonstrated inequality between Schreiber's transfer entropy rate and the information flow. In this inequality, Schreiber's transfer entropy is defined by that from X_k to $\vec{X}_{-k}$, and the information flow is represented by $\sum_{l\neq k} T_{k\to l}$. In this study, I divided a collection of random variables and derived the formula from one random variable for another random variable. On this point, the inequality I proved and that of Horowitz are very different from each other.

I formulated Horowitz's information flow for a linear stochastic system. It was found that determining whether the time derivative of x_l is independent of x_k is possible not only by evaluating the Liang-Kleeman information flow but also via Horowitz's information flow. In this case, it was found that the Liang-Kleeman information flow is exactly zero, whereas the Horowitz information flow slightly deviates from zero due to the stochastic fluctuation. Although the Liang-Kleeman information flow is expressed using covariance, Horowitz's information flow is expressed using the cofactor of the covariance matrix. Given that the partial correlation is formulated by the inverse of the correlation matrix, Horowitz's information flow is expressed using partial correlation. This is the difference between the two information flows. Although I calculated Schreiber's transfer entropy rate for the linear stochastic system, a stable value of Schreiber's transfer entropy rate could not be obtained because the system is not in the steady state at the calculated parameter values.

The Liang-Kleeman and Horowitz information flows are dynamic quantities because they are derived from the time derivative of entropy or mutual information, and Schreiber's transfer entropy is a static quantity because it is defined by conditional mutual information. Dynamic quantities are considered to be robust with respect to nonstationarity. For this reason, in the linear stochastic system case studied here, the Liang-Kleeman and Horowitz information flows could be evaluated, but Schreiber's transfer entropy could not be evaluated. I consider that further research is needed on this.

I also examined Horowitz's information flow for a stochastic gradient system. For the considered model, it was analytically shown that, when a variable x does not directly depend on a variable y, Horowitz's information flow, $T_{y\to x}$, is zero. Therefore, both the Liang-Kleeman and Horowitz information flows exhibit the principle of nil causality. Unlike the considered linear stochastic system, which is in a nonstationary state, the stochastic gradient system is in a stationary state; thus, Schreiber's transfer entropy rate stabilizes and can be determined. For the considered stochastic gradient system, it was analytically shown that Horowitz's information flow and Schreiber's transfer entropy rate satisfy the inequality.

Let us make a comparison of the three methods: Schreiber's transfer entropy rate, the Liang-Kleeman information flow, and Horowitz's information flow. As for any of the three methods, if the evolution of a variable x does not depend on a variable y, each quantity from y to x becomes zero. Whereas Schreiber's transfer entropy rate is numerically and analytically cumbersome to estimate, Schreiber's transfer entropy is considered versatile because it is equivalent to the Granger causality for linear Gaussian stochastic process. The Liang-Kleeman and Horowitz information flows have similar derivation processes. Because Horowitz's information flow is derived from the time derivative of mutual information, I consider that Horowitz's information flow is the most reasonable in terms of derivation. Considering the stochastic gradient system results, Horowitz's information flow is less sensitive than the other two, because it cannot be greater than zero without a larger stochastic fluctuation than the other two.

APPENDIX A: EXPLANATION OF DISREGARD OF TERMS CONTAINING $\partial_i\partial_j$ ($i \neq j$)

In this Appendix, I demonstrate that terms containing $\partial_i\partial_j$ ($i \neq j$) do not appear in Eq. (8) for the multipartite case. When $(B(\vec{X}_t, t)B(\vec{X}_t, t)^\top)_{ij} = 0$ ($i \neq j$), terms containing $\partial_i\partial_j$ do not appear in Eq. (8). Let us consider the ith element of the stochastic fluctuation term of Eq. (1), that is, $\sum_k B_{ik}(\vec{X}_t, t)dW_k(t)$. The product of the ith and jth elements of the stochastic fluctuation term becomes

$$\begin{aligned}\sum_k B_{ik}(\vec{X}_t, t)dW_k(t) \times \sum_l B_{jl}(\vec{X}_t, t)dW_l(t)\\ = \sum_k B_{ik}(\vec{X}_t, t)B_{jk}(\vec{X}_t, t)dt\\ = (B(\vec{X}_t, t)B(\vec{X}_t, t)^\top)_{ij}dt,\end{aligned} \tag{A1}$$

where Eq. (2) is used in the transformation from the first to the second line. Since the different components of the stochastic fluctuation term are not correlated with each other in the multipartite case, Eq. (A1) shows that $(B(\vec{X}_t, t)B(\vec{X}_t, t)^\top)_{ij} = 0$ ($i \neq j$). This is why the $\partial_i\partial_j$ terms do not appear in the multipartite case.

APPENDIX B: DERIVATION OF EQ. (47)

To ensure clarity, I will proceed with the derivation using a concrete example. Let us consider the case of $N = 3$ and derive the formula of $T_{2\to 1}$. The marginal probability distribution $p_t(\vec{x}_{-2})$ is given by

$$\begin{aligned}p_t(\vec{x}_{-2}) &= \int p_t(\vec{x})dx_2 = p_t(x_1, x_3)\\ &= \frac{1}{\sqrt{(2\pi)^2M_{22}}}\exp\left\{-\frac{1}{2M_{22}}[\Sigma_{33}(x_1-\mu_1)^2 + \Sigma_{11}(x_3-\mu_3)^2 - 2\Sigma_{13}(x_1-\mu_1)(x_3-\mu_3)]\right\},\end{aligned} \tag{B1}$$

where M_{ij} is a minor of the covariance matrix Σ. The minor of the covariance matrix and the cofactor are associated by the equation $C_{ij} = (-1)^{i+j}M_{ij}$. Given that $F_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3$, $\partial_1[F_1p_t(\vec{x}_{-2})]$ becomes

$$\partial_1[F_1p_t(\vec{x}_{-2})] = a_{11}p_t(\vec{x}_{-2}) + (a_{11}x_1 + a_{12}x_2 + a_{13}x_3)\left\{-\frac{1}{M_{22}}[\Sigma_{33}(x_1-\mu_1) - \Sigma_{13}(x_3-\mu_3)]\right\}p_t(\vec{x}_{-2}). \tag{B2}$$

By substituting Eq. (B2) for the first term of the right-hand side of Eq. (37), I obtain

$$\begin{aligned}-\int p_t(x_2|\vec{x}_{-2})\partial_1[F_1p_t(\vec{x}_{-2})]d\vec{x} &= -\int p_t(\vec{x})\left\{a_{11} - \frac{1}{M_{22}}(a_{11}x_1 + a_{12}x_2 + a_{13}x_3)[\Sigma_{33}(x_1-\mu_1) - \Sigma_{13}(x_3-\mu_3)]\right\}d\vec{x}\\ &= -\int p_t(\vec{x})\left\{a_{11} - \frac{1}{M_{22}}[a_{11}(x_1-\mu_1) + a_{11}\mu_1][\Sigma_{33}(x_1-\mu_1) - \Sigma_{13}(x_3-\mu_3)]\right.\\ &\quad - \frac{1}{M_{22}}[a_{12}(x_2-\mu_2) + a_{12}\mu_2][\Sigma_{33}(x_1-\mu_1) - \Sigma_{13}(x_3-\mu_3)]\\ &\quad \left.- \frac{1}{M_{22}}[a_{13}(x_3-\mu_3) + a_{13}\mu_3][\Sigma_{33}(x_1-\mu_1) - \Sigma_{13}(x_3-\mu_3)]\right\}d\vec{x}\end{aligned}$$

$$
\begin{aligned}
&= -a_{11} + \frac{1}{M_{22}}a_{11}(\Sigma_{11}\Sigma_{33} - \Sigma_{13}\Sigma_{13}) + \frac{1}{M_{22}}a_{12}(\Sigma_{12}\Sigma_{33} - \Sigma_{23}\Sigma_{13}) \\
&\quad + \frac{1}{M_{22}}a_{13}(\Sigma_{13}\Sigma_{33} - \Sigma_{13}\Sigma_{33}) \\
&= a_{12}\frac{M_{12}}{M_{22}} = -a_{12}\frac{C_{12}}{C_{22}}.
\end{aligned} \tag{B3}
$$

As g_{11} is assumed to be constant, the second term of Eq. (37) becomes zero as follows:

$$
\frac{1}{2}\int p_t(x_2|\vec{x}_{-2})\partial_1^2[g_{11}p_t(\vec{x}_{-2})]d\vec{x} = \frac{1}{2}\int \partial_1^2[g_{11}p_t(\vec{x}_{-2})]d\vec{x}_{-2} = \frac{1}{2}\int \partial_1[g_{11}p_t(\vec{x}_{-2})]_{x_1=-\infty}^{x_1=\infty}d\vec{x}_{-\{1,2\}} = 0. \tag{B4}
$$

The conditional probability $p_t(x_2|\vec{x}_{-2})$ becomes

$$
\begin{aligned}
p_t(x_2|\vec{x}_{-2}) &= \frac{p_t(\vec{x})}{p_t(\vec{x}_{-2})} \\
&= \sqrt{\frac{M_{22}}{2\pi\det\Sigma}}\exp\left\{-\frac{1}{2}\frac{1}{\det\Sigma}\left[(x_1-\mu_1)^2\left(C_{11} - \frac{\det\Sigma}{M_{22}}\Sigma_{33}\right) + (x_2-\mu_2)^2C_{22} + (x_3-\mu_3)^2\left(C_{33} - \frac{\det\Sigma}{M_{22}}\Sigma_{11}\right)\right.\right. \\
&\quad \left.\left. + 2(x_1-\mu_1)(x_2-\mu_2)C_{12} + 2(x_1-\mu_1)(x_3-\mu_3)\left(C_{13} + \frac{\det\Sigma}{M_{22}}\Sigma_{13}\right) + 2(x_2-\mu_2)(x_3-\mu_3)C_{23}\right]\right\} \\
&= \sqrt{\frac{M_{22}}{2\pi\det\Sigma}}\exp\left\{-\frac{1}{2}\frac{1}{\det\Sigma}\left[(x_1-\mu_1)^2\frac{M_{12}M_{12}}{M_{22}} + (x_2-\mu_2)^2M_{22} + (x_3-\mu_3)^2\frac{M_{23}M_{23}}{M_{22}}\right.\right. \\
&\quad \left.\left. - 2(x_1-\mu_1)(x_2-\mu_2)M_{12} + 2(x_1-\mu_1)(x_3-\mu_3)\frac{M_{12}M_{23}}{M_{22}} - 2(x_2-\mu_2)(x_3-\mu_3)M_{23}\right]\right\} \\
&= \sqrt{\frac{M_{22}}{2\pi\det\Sigma}}\exp\left\{-\frac{1}{2}\frac{M_{22}}{\det\Sigma}\left[(x_1-\mu_1)\frac{M_{12}}{M_{22}} - (x_2-\mu_2) + (x_3-\mu_3)\frac{M_{23}}{M_{22}}\right]^2\right\}.
\end{aligned} \tag{B5}
$$

Using Eq. (B5), I derive

$$
\partial_1 \ln[p_t(x_2|\vec{x}_{-2})] = -\frac{M_{12}}{\det\Sigma}\left[(x_1-\mu_1)\frac{M_{12}}{M_{22}} - (x_2-\mu_2) + (x_3-\mu_3)\frac{M_{23}}{M_{22}}\right]. \tag{B6}
$$

The third term of Eq. (37) becomes

$$
\begin{aligned}
&-\frac{1}{2}\int g_{11}p_t(\vec{x})\Big\{\partial_1 \ln[p_t(x_2|\vec{x}_{-2})]\Big\}^2 d\vec{x} \\
&= -\frac{1}{2}\int g_{11}p_t(\vec{x})\left(\frac{M_{12}}{\det\Sigma}\right)^2\left[(x_1-\mu_1)\frac{M_{12}}{M_{22}} - (x_2-\mu_2) + (x_3-\mu_3)\frac{M_{23}}{M_{22}}\right]^2 d\vec{x} \\
&= -\frac{1}{2}g_{11}\left(\frac{M_{12}}{\det\Sigma}\right)^2\left(\Sigma_{22} + \Sigma_{11}\frac{M_{12}}{M_{22}}\frac{M_{12}}{M_{22}} + \Sigma_{33}\frac{M_{23}}{M_{22}}\frac{M_{23}}{M_{22}} - 2\Sigma_{12}\frac{M_{12}}{M_{22}} + 2\Sigma_{13}\frac{M_{12}}{M_{22}}\frac{M_{23}}{M_{22}} - 2\Sigma_{23}\frac{M_{23}}{M_{22}}\right) \\
&= -\frac{1}{2}g_{11}\left(\frac{M_{12}}{\det\Sigma}\right)^2\left(\underbrace{\Sigma_{22}}_{\mathcal{A}} + \underbrace{\Sigma_{11}\frac{M_{12}}{M_{22}}\frac{M_{12}}{M_{22}}}_{\mathcal{B}} + \underbrace{\Sigma_{33}\frac{M_{23}}{M_{22}}\frac{M_{23}}{M_{22}}}_{\mathcal{C}} - \underbrace{\Sigma_{12}\frac{M_{12}}{M_{22}}}_{\mathcal{A}} - \underbrace{\Sigma_{12}\frac{M_{12}}{M_{22}}}_{\mathcal{B}}\right. \\
&\quad \left. + \underbrace{\Sigma_{13}\frac{M_{12}}{M_{22}}\frac{M_{23}}{M_{22}}}_{\mathcal{B}} + \underbrace{\Sigma_{13}\frac{M_{12}}{M_{22}}\frac{M_{23}}{M_{22}}}_{\mathcal{C}} - \underbrace{\Sigma_{23}\frac{M_{23}}{M_{22}}}_{\mathcal{A}} - \underbrace{\Sigma_{23}\frac{M_{23}}{M_{22}}}_{\mathcal{C}}\right).
\end{aligned} \tag{B7}
$$

Summarizing the terms with the same symbol, I obtain the following equations:

$$
\mathcal{A} = \Sigma_{22} - \Sigma_{12}\frac{M_{12}}{M_{22}} - \Sigma_{23}\frac{M_{23}}{M_{22}} = \frac{\Sigma_{22}M_{22} - \Sigma_{12}M_{12} - \Sigma_{23}M_{23}}{M_{22}} = \frac{\Sigma_{12}C_{12} + \Sigma_{22}C_{22} + \Sigma_{32}C_{32}}{M_{22}} = \frac{\det\Sigma}{M_{22}}, \tag{B8}
$$

$$
\mathcal{B} = \Sigma_{11}\frac{M_{12}}{M_{22}}\frac{M_{12}}{M_{22}} - \Sigma_{12}\frac{M_{12}}{M_{22}} + \Sigma_{13}\frac{M_{12}}{M_{22}}\frac{M_{23}}{M_{22}} = \frac{\Sigma_{11}M_{12} - \Sigma_{12}M_{22} + \Sigma_{13}M_{23}}{M_{22}}\frac{M_{12}}{M_{22}}
$$

$$= \frac{-\Sigma_{11}C_{21} - \Sigma_{12}C_{22} - \Sigma_{13}C_{23}}{M_{22}} \frac{M_{12}}{M_{22}} = 0, \tag{B9}$$

$$\mathcal{C} = \Sigma_{33}\frac{M_{23}}{M_{22}}\frac{M_{23}}{M_{22}} + \Sigma_{13}\frac{M_{12}}{M_{22}}\frac{M_{23}}{M_{22}} - \Sigma_{23}\frac{M_{23}}{M_{22}} = \frac{\Sigma_{33}M_{23} + \Sigma_{13}M_{12} - \Sigma_{23}M_{22}}{M_{22}}\frac{M_{23}}{M_{22}}$$

$$= \frac{-\Sigma_{13}C_{12} - \Sigma_{23}C_{22} - \Sigma_{33}C_{32}}{M_{22}} \frac{M_{23}}{M_{22}} = 0, \tag{B10}$$

where I quoted a formula of the cofactor expansion of the determinant. As a result, Eq. (B7) becomes

$$-\frac{1}{2}\int g_{11}p_t(\vec{x})\{\partial_1 \ln[p_t(x_2|\vec{x}_{-2})]\}^2 d\vec{x} = -\frac{1}{2}g_{11}\left(\frac{M_{12}}{\det\Sigma}\right)^2\frac{\det\Sigma}{M_{22}} = -\frac{1}{2}g_{11}\frac{C_{12}}{\det\Sigma}\frac{C_{12}}{C_{22}}. \tag{B11}$$

Combining them, I obtain

$$T_{2\to 1} = -a_{12}\frac{C_{12}}{C_{22}} - \frac{1}{2}g_{11}\frac{C_{12}}{\det\Sigma}\frac{C_{12}}{C_{22}}. \tag{B12}$$

Generalizing this result, I obtain Eq. (47).

APPENDIX C: HOROWITZ'S INFORMATION FLOW IN THE CASE OF A STOCHASTIC GRADIENT SYSTEM

I estimated Horowitz's information flow of a stochastic gradient system. For clarity, I proceeded to the calculation precisely. By using Eq. (37), $T_{3\to 1}$ becomes

$$T_{3\to 1} = -\int p(x_3|\vec{x}_{-3})\partial_1[F_1(\vec{x})p(\vec{x}_{-3})]d\vec{x} + \frac{1}{2}\int p(x_3|\vec{x}_{-3})\partial_1^2[g_{11}p(\vec{x}_{-3})]d\vec{x} - \frac{1}{2}\int g_{11}p(\vec{x})\{\partial_1 \ln[p(x_3|\vec{x}_{-3})]\}^2 d\vec{x}. \tag{C1}$$

The stationary probability distribution is given by

$$p(\vec{x}) = \frac{1}{Z}\exp\left[-\frac{2V(\vec{x})}{b^2}\right] = \frac{1}{Z}\exp\left[-\frac{\left(x_1^2x_2^2 + x_2^2x_3^2 + x_1^2 + x_2^2 + x_3^2\right)}{b^2}\right], \tag{C2}$$

where Z is a normalization constant and the subscript t of the probability distribution is omitted because of stationarity. To evaluate Eq. (C1), the marginal probability distribution is required. The marginal probability distribution, $p(\vec{x}_{-3})$, is obtained as

$$p(\vec{x}_{-3}) = p(x_1, x_2) = \int_{-\infty}^{\infty} p(\vec{x})dx_3 = \frac{1}{Z}\sqrt{\frac{\pi b^2}{x_2^2+1}}\exp\left[-\frac{\left(x_1^2x_2^2 + x_1^2 + x_2^2\right)}{b^2}\right]. \tag{C3}$$

The component of the first term in Eq. (C1) is transformed into

$$\begin{aligned}\partial_1[F_1(\vec{x})p(\vec{x}_{-3})] &= \partial_1\left[-x_1\left(x_2^2+1\right)p(\vec{x}_{-3})\right]\\ &= -\left(x_2^2+1\right)p(\vec{x}_{-3}) - x_1\left(x_2^2+1\right)\partial_1 p(\vec{x}_{-3})\\ &= -\left(x_2^2+1\right)p(\vec{x}_{-3}) - x_1\left(x_2^2+1\right)\left[-\frac{\left(2x_1x_2^2+2x_1\right)}{b^2}\right]p(\vec{x}_{-3})\\ &= \left[-\left(x_2^2+1\right) + \frac{2}{b^2}x_1^2\left(x_2^2+1\right)^2\right]p(\vec{x}_{-3}).\end{aligned} \tag{C4}$$

Substituting Eq. (C4) in the first term in Eq. (C1), the first term becomes

$$\begin{aligned}(\text{the first term}) &= -\int p(x_3|\vec{x}_{-3})\left[-\left(x_2^2+1\right) + \frac{2}{b^2}x_1^2\left(x_2^2+1\right)^2\right]p(\vec{x}_{-3})d\vec{x}\\ &= -\int\left[-\left(x_2^2+1\right) + \frac{2}{b^2}x_1^2\left(x_2^2+1\right)^2\right]p(\vec{x}_{-3})dx_1dx_2\\ &= -\int\left[-\left(x_2^2+1\right) + \frac{2}{b^2}\frac{b^2}{2\left(x_2^2+1\right)}\left(x_2^2+1\right)^2\right]p(x_2)dx_2\\ &= 0.\end{aligned} \tag{C5}$$

From the first line to the second line, the integration with respect to x_3 is performed. From the second line to the third line, the integration with respect to x_1 is performed, and then x_1^2 is replaced by $b^2/2(x_2^2+1)$. The second term in Eq. (C1) is easily shown to be zero in the same manner as Eq. (B4). Dividing Eq. (C2) by Eq. (C3), I derived the following conditional probability distribution:

$$p(x_3|\vec{x}_{-3}) = \frac{p(\vec{x})}{p(\vec{x}_{-3})} = \sqrt{\frac{x_2^2+1}{\pi b^2}}\exp\left[-\frac{\left(x_2^2x_3^2+x_3^2\right)}{b^2}\right]. \tag{C6}$$

Quoting Eq. (C6), the component of the third term in Eq. (C1) becomes

$$\partial_1 \ln[p(x_3|\vec{x}_{-3})] = 0. \tag{C7}$$

Therefore, the third term in Eq. (C1) also becomes zero. Considering that each term is zero, I concluded that $T_{3\to1}$ is zero.

Let us discuss Eq. (C7). Equation (C7) leads to the relationship $p(x_3|\vec{x}_{-3}) = p(x_3|x_1, x_2) = p(x_3|x_2)$, indicating that x_1 and x_3 are conditionally independent. Conversely, if the variables are conditionally independent, the third term in Eq. (C1) is considered to be zero.

Similarly, the other Horowitz information flow is derived as

$$T_{1\to2} = \int\left[\frac{b^2\left(-3x_2^2+1\right)}{\left(x_2^2+1\right)^2} + \frac{-3x_2^2+1}{x_2^2+1} - \frac{2}{b^2}x_2^2\right]p(x_2)dx_2, \tag{C8}$$

$$T_{2\to1} = \int\left[-\frac{3b^2x_1^2}{2\left(x_1^2+x_3^2+1\right)^2} + \frac{b^2-4x_1^2}{2\left(x_1^2+x_3^2+1\right)} + 1 - \frac{2}{b^2}x_1^2\right]p(\vec{x}_{-2})dx_1dx_3, \tag{C9}$$

where the marginal probability distributions are given by

$$p(x_2) = \int_{-\infty}^{\infty} p(\vec{x})dx_1dx_3 = \frac{1}{Z}\frac{\pi b^2}{x_2^2+1}\exp\left[-\frac{x_2^2}{b^2}\right], \tag{C10}$$

$$p(\vec{x}_{-2}) = p(x_1, x_3) = \int_{-\infty}^{\infty} p(\vec{x})dx_2 = \frac{1}{Z}\sqrt{\frac{\pi b^2}{x_1^2+x_3^2+1}}\exp\left[-\frac{\left(x_1^2+x_3^2\right)}{b^2}\right]. \tag{C11}$$

Numerical integration was performed to estimate Eqs. (C8) and (C9) for various values of b.

APPENDIX D: LIANG-KLEEMAN INFORMATION FLOW IN THE CASE OF A STOCHASTIC GRADIENT SYSTEM

In this Appendix, I demonstrate the Liang-Kleeman information flow of the stochastic gradient system. First, I evaluate $T_{3\to1}^{(\mathrm{LK})}$, which is defined as

$$T_{3\to1}^{(\mathrm{LK})} = -\int p(x_3|x_1)\partial_1[F_1(\vec{x})p(\vec{x}_{-3})]d\vec{x} + \frac{1}{2}\int p(x_3|x_1)\partial_1^2[g_{11}p(\vec{x}_{-3})]d\vec{x}. \tag{D1}$$

Based on Eq. (C4), the first term of Eq. (D1) becomes

$$\begin{aligned}
(\text{the first term}) &= -\int p(x_3|x_1)\left[-\left(x_2^2+1\right) + \frac{2}{b^2}x_1^2\left(x_2^2+1\right)^2\right]p(\vec{x}_{-3})d\vec{x}\\
&= -\int\left[-\left(x_2^2+1\right) + \frac{2}{b^2}x_1^2\left(x_2^2+1\right)^2\right]p(\vec{x}_{-3})dx_1dx_2\\
&= -\int\left[-\left(x_2^2+1\right) + \frac{2}{b^2}\frac{b^2}{2\left(x_2^2+1\right)}\left(x_2^2+1\right)^2\right]p(x_2)dx_2\\
&= 0.
\end{aligned} \tag{D2}$$

The second term of Eq. (D1) becomes zero in the same manner as Eq. (B4). As a result, it can be analytically shown that $T_{3\to1}^{(\mathrm{LK})}$ is zero. For $T_{1\to2}^{(\mathrm{LK})}$, the derivation is more complicated than that for $T_{3\to1}^{(\mathrm{LK})}$. The Liang-Kleeman information flow $T_{1\to2}^{(\mathrm{LK})}$ is defined by

$$T_{1\to2}^{(\mathrm{LK})} = -\int p(x_1|x_2)\partial_2[F_2(\vec{x})p(\vec{x}_{-1})]d\vec{x} + \frac{1}{2}\int p(x_1|x_2)\partial_2^2[g_{22}p(\vec{x}_{-1})]d\vec{x}, \tag{D3}$$

where $p(x_1|x_2)$ and $p(\vec{x}_{-1})$ are respectively given by

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \sqrt{\frac{x_2^2+1}{\pi b^2}}\exp\left[-\frac{\left(x_1^2x_2^2+x_1^2\right)}{b^2}\right], \tag{D4}$$

$$p(\vec{x}_{-1}) = p(x_2, x_3) = \int_{-\infty}^{\infty} p(\vec{x})dx_1 = \frac{1}{Z}\sqrt{\frac{\pi b^2}{x_2^2+1}}\exp\left[-\frac{\left(x_2^2x_3^2+x_2^2+x_3^2\right)}{b^2}\right]. \tag{D5}$$

For the same reason as $T^{(\mathrm{LK})}_{3\to 1}$, the second term of Eq. (D3) becomes zero. Substituting Eqs. (51), (D4), and (D5) into Eq. (D3) yields

$$\begin{aligned} T^{(\mathrm{LK})}_{1\to 2} &= \int p(x_1|x_2)\partial_2\big[\big(x_2x_1^2+x_2x_3^2+x_2\big)p(\vec{x}_{-1})\big]d\vec{x} \\ &= \int p(x_1|x_2)\big[\big(x_1^2+x_3^2+1\big)p(\vec{x}_{-1})+\big(x_2x_1^2+x_2x_3^2+x_2\big)\partial_2 p(\vec{x}_{-1})\big]d\vec{x} \\ &= \int p(x_1|x_2)\bigg\{\big(x_1^2+x_3^2+1\big)p(\vec{x}_{-1})+\big(x_2x_1^2+x_2x_3^2+x_2\big)\Big[-\frac{x_2}{x_2^2+1}-\frac{2\big(x_3^2+1\big)}{b^2}x_2\Big]p(\vec{x}_{-1})\bigg\}d\vec{x} \\ &= \int p(x_1|x_2)\big(x_1^2+x_3^2+1\big)\bigg[1-\frac{x_2^2}{x_2^2+1}-\frac{2\big(x_3^2+1\big)}{b^2}x_2^2\bigg]p(\vec{x}_{-1})d\vec{x} \\ &= \int \big(x_1^2+x_3^2+1\big)\bigg[\frac{1}{x_2^2+1}-\frac{2\big(x_3^2+1\big)}{b^2}x_2^2\bigg]p(\vec{x})d\vec{x}, \end{aligned} \tag{D6}$$

where the relationship $p(x_1|x_2)p(\vec{x}_{-1})=p(\vec{x})$, which is apparent from Eqs. (D4) and (D5), is employed. Integrating over x_1 and x_3, I can replace x_1^2 and x_3^2 with $b^2/2(x_2^2+1)$ and x_3^4 with $3b^4/4(x_2^2+1)^2$. After tedious calculations, the following is obtained:

$$T^{(\mathrm{LK})}_{1\to 2} = \int \bigg[b^2\frac{\big(2x_2^2-1\big)}{\big(x_2^2+1\big)^2}+\frac{3x_2^2-1}{x_2^2+1}+\frac{2}{b^2}x_2^2\bigg]p(x_2)dx_2, \tag{D7}$$

where $p(x_2)$ is given by Eq. (C10). As we cannot evaluate the analytical calculations any further, we need to evaluate Eq. (D7) numerically. For $T^{(\mathrm{LK})}_{2\to 1}$, the Liang-Kleeman information flow is defined by

$$T^{(\mathrm{LK})}_{2\to 1} = -\int p(x_2|x_1)\partial_1[F_1(\vec{x})p(\vec{x}_{-2})]d\vec{x}+\frac{1}{2}\int p(x_2|x_1)\partial_1^2[g_{11}p(\vec{x}_{-2})]d\vec{x}. \tag{D8}$$

Utilizing Eqs. (50) and (C11), I obtain

$$T^{(\mathrm{LK})}_{2\to 1} = \int p(x_2|x_1)\big(x_2^2+1\big)\bigg(-\frac{x_3^2+1}{x_1^2+x_3^2+1}+\frac{2}{b^2}x_1^2\bigg)p(\vec{x}_{-2})d\vec{x}, \tag{D9}$$

and evaluate it numerically to obtain Fig. 2(b).

APPENDIX E: SCHREIBER'S TRANSFER ENTROPY RATE IN THE CASE OF A STOCHASTIC GRADIENT SYSTEM

Estimating Schreiber's transfer entropy rate is more complicated than those of the two information flows. The first step is to calculate the transition probability distributions [43]. By expanding Eq. (40) with respect to $dX_{l,t}$, the distribution can be transformed as follows:

$$\begin{aligned} p(x_l',\vec{x}) &= \langle\delta(x_l'-X_{l,t+dt})\delta(\vec{x}-\vec{X}_t)\rangle \\ &= \langle\delta(x_l'-X_{l,t}-dX_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle \\ &= \langle\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle-\langle dX_{l,t}\partial_l'\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle+\tfrac{1}{2}\langle(dX_{l,t}\partial_l')^2\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle+O((dX_{l,t})^3) \\ &= \langle\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle-\langle F_l(\vec{X}_t)\partial_l'\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle dt+\tfrac{1}{2}\langle b^2\partial_l'^2\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle dt+o(dt) \\ &= \langle\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle-\langle\partial_l'F_l(\vec{x}')\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle dt+\tfrac{1}{2}\langle b^2\partial_l'^2\delta(x_l'-X_{l,t})\delta(\vec{x}-\vec{X}_t)\rangle dt+o(dt) \\ &= \delta(x_l'-x_l)\langle\delta(\vec{x}-\vec{X}_t)\rangle-\partial_l'[F_l(\vec{x}')\delta(x_l'-x_l)]\langle\delta(\vec{x}-\vec{X}_t)\rangle dt+\tfrac{1}{2}b^2\partial_l'^2\delta(x_l'-x_l)\langle\delta(\vec{x}-\vec{X}_t)\rangle dt+o(dt) \\ &= \delta(x_l'-x_l)p_t(\vec{x})-\partial_l'[F_l(\vec{x}')\delta(x_l'-x_l)]p_t(\vec{x})dt+\tfrac{1}{2}b^2\partial_l'^2\delta(x_l'-x_l)p_t(\vec{x})dt+o(dt). \end{aligned} \tag{E1}$$

Hence, the transition probability distribution is given by

$$p(x_l'|\vec{x}) = \frac{p(x_l',\vec{x})}{p_t(\vec{x})} = \delta(x_l'-x_l)-\partial_l'[F_l(\vec{x}')\delta(x_l'-x_l)]dt+\frac{1}{2}b^2\partial_l'^2\delta(x_l'-x_l)dt+o(dt). \tag{E2}$$

By performing the integration using the Fourier integral representation of the Dirac delta function, for a small dt, I obtain

$$p(x_l'|\vec{x}) = \frac{1}{\sqrt{2\pi b^2 dt}}\exp\bigg\{-\frac{[x_l'-x_l-F_l(\vec{x})dt]^2}{2b^2dt}\bigg\}. \tag{E3}$$

Using Eq. (50), the transition probability distribution to x_1' becomes

$$p(x_1'|\vec{x}) = \frac{1}{\sqrt{2\pi b^2 dt}} \exp\left\{-\frac{[x_1' - x_1 + x_1(x_2^2+1)dt]^2}{2b^2 dt}\right\}. \tag{E4}$$

As the right-hand side of Eq. (E4) does not depend on x_3, $p(x_1'|\vec{x}) = p(x_1'|\vec{x}_{-3})$ is satisfied. Thus, Schreiber's transfer entropy from x_3 to x_1 becomes

$$T^{(\mathrm{S})}_{3\to 1} = \int p(x_1', \vec{x}) \ln\left[\frac{p(x_1'|\vec{x})}{p(x_1'|\vec{x}_{-3})}\right] dx_1' d\vec{x} = 0. \tag{E5}$$

For Schreiber's transfer entropy from x_1 to x_2 and from x_2 to x_1, the analytic derivation is complicated and protracted; hence, I only present the results:

$$T^{(\mathrm{S})}_{1\to 2} = \int_{-\infty}^{\infty} \frac{b^2}{4} \frac{x_2^2}{(x_2^2+1)^2} p(x_2) dx_2 dt + o(dt), \tag{E6}$$

$$T^{(\mathrm{S})}_{2\to 1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{b^2}{4} \frac{x_1^2}{(x_1^2+x_3^2+1)^2} p(x_1, x_3) dx_1 dx_3 dt + o(dt). \tag{E7}$$

Schreiber's transfer entropy rate in Fig. 2(b) is estimated via the numerical integration of Eqs. (E6) and (E7).

[1] T. M. Cover and J. A. Thomas, Elem. Inf. Theory **1**, 279 (1991).
[2] A. Dionisio, R. Menezes, and D. A. Mendes, Physica A **344**, 326 (2004).
[3] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).
[4] A. Papana and D. Kugiumtzis, Int. J. Bifurcation Chaos Appl. Sci. Eng. **19**, 4197 (2009).
[5] C. W. Granger, Econometrica: J. Econometric Soc., **37**, 424 (1969).
[6] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[7] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, *An Introduction to Transfer Entropy* (Springer, Berlin, 2016).
[8] D. A. Smirnov, Phys. Rev. E **87**, 042917 (2013).
[9] J. Sun and E. M. Bollt, Physica D **267**, 49 (2014).
[10] R. G. James, N. Barnett, and J. P. Crutchfield, Phys. Rev. Lett. **116**, 238701 (2016).
[11] X. S. Liang and R. Kleeman, Phys. Rev. Lett. **95**, 244101 (2005).
[12] X. S. Liang and R. Kleeman, Physica D **231**, 1 (2007).
[13] X. S. Liang and R. Kleeman, Physica D **227**, 173 (2007).
[14] X. S. Liang, Phys. Rev. E **78**, 031113 (2008).
[15] X. S. Liang, Entropy **15**, 327 (2013).
[16] X. S. Liang, Phys. Rev. E **90**, 052150 (2014).
[17] X. S. Liang, Phys. Rev. E **94**, 052201 (2016).
[18] X. S. Liang, Chaos **28**, 075311 (2018).
[19] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, Nat. Phys. **11**, 131 (2015).
[20] S. Ito and T. Sagawa, Phys. Rev. Lett. **111**, 180603 (2013).
[21] S. Ito and T. Sagawa, Math. Found. Appl. Graph Entropy **3**, 63 (2016).
[22] C. Cafaro, S. A. Ali, and A. Giffin, Phys. Rev. E **93**, 022114 (2016).
[23] J. M. Horowitz and M. Esposito, Phys. Rev. X **4**, 031015 (2014).
[24] M. Esposito and C. Van den Broeck, Phys. Rev. E **82**, 011143 (2010).
[25] C. Van den Broeck and M. Esposito, Phys. Rev. E **82**, 011144 (2010).
[26] M. Prokopenko, J. T. Lizier, and D. C. Price, Entropy **15**, 524 (2013).
[27] M. Paluš and M. Vejmelka, Phys. Rev. E **75**, 056211 (2007).
[28] M. Paluš, A. Krakovská, J. Jakubík, and M. Chvosteková, Chaos **28**, 075307 (2018).
[29] A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, and M. Paluš, Phys. Rev. E **97**, 042207 (2018).
[30] C. W. Granger, Am. Econ. Rev. **94**, 421 (2004).
[31] C. W. Gardiner *et al.*, *Handbook of Stochastic Methods* (Springer, Berlin, 1985), Vol. 3.
[32] K. Sekimoto, *Stochastic Energetics* (Springer, Berlin, 2010), Vol. 799.
[33] J. M. Horowitz, J. Stat. Mech.: Theory Exp. (2015) P03006.
[34] P. Martien, S. Pope, P. Scott, and R. Shaw, Phys. Lett. A **110**, 399 (1985).
[35] K. Kaneko, Physica D **23**, 436 (1986).
[36] J. A. Vastano and H. L. Swinney, Phys. Rev. Lett. **60**, 1773 (1988).
[37] A. E. Allahverdyan, D. Janzing, and G. Mahler, J. Stat. Mech.: Theory Exp. (2009) P09011.
[38] J. Sun, D. Taylor, and E. M. Bollt, SIAM J. Appl. Dyn. Syst. **14**, 73 (2015).
[39] M. Baldovin, F. Cecconi, and A. Vulpiani, Phys. Rev. Research **2**, 043436 (2020).
[40] C. Sarra, M. Baldovin, and A. Vulpiani, Phys. Rev. E **104**, 024116 (2021).
[41] T. Chen, E. Fox, and C. Guestrin, in *International Conference on Machine Learning* (PMLR, Bejing, China, 2014), pp. 1683–1691.
[42] Y.-A. Ma, T. Chen, and E. Fox, in *Advances in Neural Information Processing Systems*, edited by C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., 2015), Vol. 28.
[43] H. Risken, in *The Fokker-Planck Equation* (Springer, Berlin, 1996), pp. 63–95.