Information Transfer between Dynamical System Components

X. San Liang* and Richard Kleeman
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA
(Received 8 August 2005; published 8 December 2005)

We present a rigorous formalism of information transfer for systems with dynamics fully known. This follows from an accurate classification of the mechanisms for the entropy change of one component into a self-evolution plus a transfer from the other component. The formalism applies to both continuous flows and discrete maps. The resulting transfer measure possesses a property of asymmetry and is qualitatively consistent with the classical measures. It is further validated with the baker transformation and the Hénon map.

DOI: 10.1103/PhysRevLett.95.244101 PACS numbers: 05.45.−a, 02.50.−r, 89.70.+c, 89.75.−k

Information transfer between the components of a dynamical system has been of interest since decades ago [1,2]. It is usually investigated as a subfield of nonlinear signal coherence analysis in the context of some specific discipline [3], but its concept is also frequently seen in general physics literature. The well-studied baker transformation is such an example: It has been argued that information is transferred continually from the stretching direction to the folding direction [4,5], although the transfer is yet to be quantified. Recently, there comes much renewed interest with this problem, motivated mainly by the research of weather predictability and climate variability, where information transfer plays an important role [6].

So far, the formalisms of information transfer are mostly data-based, the widely used transfer measures including the time-delayed mutual information [1] and, in the framework of a Markov chain, the more sophisticated transfer entropy by Schreiber [2]. In this study, we will show that, when dynamics is fully known (as in the above physical problems), the measure of information transfer can be rigorously formulated.

We rely on a two-dimensional (2D) system to elucidate the fundamental idea. The formalism on a more generic basis for arbitrarily many dimensions will follow later [7]. We first derive the formula with a continuous flow, then extend it to deal with more complicated discrete mappings. The result is applied to study two classical problems where the physics of information transfer is qualitatively clear for validation.

Consider a 2D autonomous system:

$$\frac{d\underline{\mathbf{x}}}{dt} = \underline{\mathbf{F}}(\underline{\mathbf{x}}), \tag{1}$$

where $\underline{\mathbf{F}} = (F_1, F2)$, and $\underline{\mathbf{x}} = (x_1, x_2) \in \Omega$. In this study, the sample space Ω is assumed to be a direct product of Ω_1 and Ω_2. Let $\{\underline{\mathbf{X}}, t\}$ be a stochastic process, $\underline{\mathbf{X}} = (X_1, X_2)$ the random variables corresponding to (x_1, x_2), and $\rho = \rho(x_1, x_2, t)$ the probability density distribution at t. We need to find how the joint entropy of X_1 and X_2, which is defined as

$$H(t) = -\iint_\Omega \rho \log\rho \, dx_1 dx_2, \tag{2}$$

evolves with time. For notational simplicity, we will drop the Ω in the double integral, and all double integrations henceforth are understood to be over the entire sample space, unless otherwise indicated.

Associated with (1), there is a Liouville equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_1}(\rho F_1) + \frac{\partial}{\partial x_2}(\rho F_2) = 0. \tag{3}$$

Multiplication by $(1 + \log\rho)$ gives

$$\frac{\partial(\rho\log\rho)}{\partial t} + \underline{\mathbf{F}}\cdot\nabla(\rho\log\rho) + \rho(1+\log\rho)\nabla\cdot\underline{\mathbf{F}} = 0. \tag{4}$$

Integrating,

$$\frac{dH}{dt} - \iint \nabla\cdot(\rho\log\rho\underline{\mathbf{F}})dx_1 dx_2 - \iint \rho\nabla\cdot\underline{\mathbf{F}} dx_1 dx_2 = 0.$$

In practice (particularly for weather systems or climate models), ρ generally vanishes at boundaries; i.e., extreme events have zero probability. This assumption may also be weakened to include extreme events but with boundary fluxes balanced in the x_1 and x_2 directions, respectively. But, for practical purposes, we stick to the former assumption. The second term in the above equation thus vanishes. So $\frac{dH}{dt} - \iint \rho(x_1, x_2)\nabla\cdot\underline{\mathbf{F}} dx_1 dx_2 = 0$, or

$$\frac{dH}{dt} = E(\nabla\cdot\underline{\mathbf{F}}). \tag{5}$$

Equation (5) states that the time rate of change of H is due to the expectation of the divergence of $\underline{\mathbf{F}}$. In other words, the change of entropy is controlled totally by the contraction or expansion of the phase space. Soon we will see that the physics revealed by (5) is the key to the establishment of our transfer formalism.

If only one component is involved, entropy along that coordinate can also be defined but with the marginal distribution. Suppose we are interested in the entropy evolution of the first component. The marginal density is $\rho_1(x_1, t) = \int_{\Omega_2} \rho(x_1, x_2, t)dx_2$. The evolution equation of

ρ_1 is derived by integrating (3) with respect to x_2 over the subspace Ω_2:

$$\frac{\partial \rho_1}{\partial t} + \frac{\partial}{\partial x_1} \int_{\Omega_2} \rho F_1 dx_2 = 0. \tag{6}$$

The third term in the original equation has been integrated out with the compact support assumption for ρ. Following the same procedure as above, the entropy in direction 1 $H_1(t) = -\int_{\Omega_1} \rho_1 \log\rho_1 dx_1$ evolves as

$$\frac{dH_1}{dt} = \iint \left[\log\rho_1 \frac{\partial(\rho F_1)}{\partial x_1}\right] dx_1 dx_2. \tag{7}$$

Through integration by parts, this may also be written as

$$\frac{dH_1}{dt} = -\iint \rho \left[\frac{F_1}{\rho_1} \frac{\partial \rho_1}{\partial x_1}\right] dx_1 dx_2. \tag{8}$$

Equation (8) states how H_1 evolves with time. The evolutionary mechanism can be decomposed into two parts: One is from X_1 itself, which we write as dH_1^*/dt; another from X_2 through the coupling in the joint density distribution ρ. Clearly, the latter is the information transfer. We want to separate it out from the intertwined mechanism in the right-hand side of (8).

The separation is made through a heuristic reasoning based on what one observes with the joint entropy and (5): The time change of H depends only on $\nabla \cdot \underline{\mathbf{F}}$. One may argue that the entropy with X_1 changes only with $\partial F_1/\partial x_1$ if it evolves on its own. Paralleling (5) then should be an equation:

$$\frac{dH_1^*}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) = \iint \rho \frac{\partial F_1}{\partial x_1} dx_1 dx_2. \tag{9}$$

The rate of entropy transfer from X_2 to X_1 is then

$$\begin{aligned} T_{2\to 1} &= \frac{d}{dt}(H_1 - H_1^*) = -\iint \left(\frac{F_1}{\rho_1}\frac{\partial \rho_1}{\partial x_1} + \frac{\partial F_1}{\partial x_1}\right)\rho dx_1 dx_2 \\ &= -\iint \rho_{2|1}(x_2|x_1) \frac{\partial(F_1\rho_1)}{\partial x_1} dx_1 dx_2. \end{aligned} \tag{10}$$

Likewise, the transfer of entropy from x_1 to x_2 is

$$T_{1\to 2} = -\iint \rho_{1|2}(x_1|x_2) \frac{\partial(F_2\rho_2)}{\partial x_2} dx_1 dx_2. \tag{11}$$

As elaborated by Schreiber [2], a desired property about the information transfer is its asymmetry between the components. Particularly, in system (1), if F_1 has no dependence on x_2, then x_1 evolves on its own. That is to say, there should be no information transfer from random variable component X_2 to X_1, albeit the transfer in the other direction may be nonzero when F_2 depends on both x_1 and x_2. This is true with the transfers defined in (10) and (11). In fact, when $F_1 = F_1(x_1)$,

$$\begin{aligned} T_{2\to 1} &= -\int_{\Omega_1}\left(\int_{\Omega_2} \rho_{2|1}(x_2|x_1) dx_2\right) \frac{\partial(F_1\rho_1)}{\partial x_1} dx_1 \\ &= -\int_{\Omega_1} \frac{\partial(F_1\rho_1)}{\partial x_1} dx_1 = 0. \end{aligned}$$

It merits noting that, because of the asymmetry, information transfer is distinctly different from transfer of other quantities such as energy: Information does not have to be lost in one component in order for another component to receive it. As elucidated above, while X_1 gains information from X_2, X_2 might have nothing to do with X_1.

The foregoing formalism for continuous systems seems to be surprisingly simple. We now carry over the idea to discrete maps. Maps usually display more complex and more interesting features than continuous flows. In fact, the information transfers by previous researchers are all formulated in the context of discrete systems. We need to extend (10) and (11) to 2D maps, even for the purpose of making contacts with the classical formalisms. For simplicity, we first examine a map with invertible individual components and then generalize to more generic cases.

Consider a transformation Φ: $\Omega \to \Omega$, which is made up of two components Φ_1: $\Omega \mapsto \Omega_1$, Φ_2: $\Omega \mapsto \Omega_2$. The evolution of its density is steered by the Frobenius-Perron operator (*F-P* operator hereafter) $\mathcal{P}$: $L^1(\Omega) \to L^1(\Omega)$, where the sample space Ω is, again, a shorthand for $\Omega_1 \times \Omega_2$. Given a density $\rho = \rho(x_1, x_2)$, $\mathcal{P}$ is defined, in a loose sense (see [4] for a rigorous treatment using the measure theory) such that

$$\iint_\omega \mathcal{P}\rho(x_1, x_2) dx_1 dx_2 = \iint_{\Phi^{-1}(\omega)} \rho(x_1, x_2) dx_1 dx_2, \tag{12}$$

where ω is any subset of Ω. For an invertible transformation Φ, $\mathcal{P}$ can be expressed explicitly as

$$\mathcal{P}\,\rho(x_1, x_2) = \rho[\Phi^{-1}(x_1, x_2)]|J^{-1}|, \tag{13}$$

where $J^{-1} = J^{-1}(x_1, x_2) = \det[\partial(\Phi^{-1}(x_1, x_2))/\partial(x_1, x_2)]$ is the determinant of the Jacobian matrix for the inverse transformation of Φ [4].

Now compute the entropy increase after applying once an invertible mapping Φ. From Eq. (13),

$$\begin{aligned} \Delta H &= -\iint \mathcal{P}\rho \log\mathcal{P}\rho dx_1 dx_2 + \iint \rho \log\rho dx_1 dx_2 \\ &= -\iint \rho(\Phi^{-1}(x_1, x_2))|J^{-1}| \log[\Phi^{-1}(x_1, x_2)|J^{-1}|] dx_1 dx_2 + \iint \rho \log\rho dx_1 dx_2 \\ &= -\iint \rho(u_1, u_2)|J^{-1}|[\log\rho(u_1, u_2) + \log|J^{-1}|]|J| du_1, du_2 + \iint \rho \log\rho dx_1 dx_2 \\ &= -\iint \rho(x_1, x_2) \log|J^{-1}| dx_1 dx_2. \end{aligned}$$

In the derivation, we have made a transformation of variables from (x_1, x_2) to $(u_1, u_2) = \Phi^{-1}(x_1, x_2)$ and use the fact that $\Phi^{-1}(\Omega) = \Omega$. The above equality can be concisely rewritten as

$$\Delta H = E \log|J|. \tag{14}$$

Notice that $|J|$ is the rate of area change of the transformation Φ. Equation (14), hence, simply states that the change of entropy is the average of the logarithm of area change of the phase space.

Equation (14) is consistent with its continuous counterpart (5). If the transformation Φ is replaced by an infinitesimal operator, the two are identical. An obvious observation about (14) is, for an invertible measure-preserving transformation ($|J| = 1$), entropy stays invariant.

We proceed to investigate the entropy transfer. Consider the transfer from X_2 to X_1 first. The entropy of X_1 increases as

$$\Delta H_1 = -\int_{\Omega_1}\left(\int_{\Omega_2} \mathcal{P}\rho dx_2\right)\log\left(\int_{\Omega_2} \mathcal{P}\rho dx_2\right)dx_1 + \int_{\Omega_1} \rho_1 \log\rho_1 dx_1,$$

where ρ_1 is the marginal distribution of X_1. This increase is due to two mechanisms: One is by X_1 itself; another is the transfer from X_2. We denote these two by ΔH_1^* and $T_{2\to1}$, respectively. In the case when Φ_1, the first component of $\Phi(x_1, x_2)$, is invertible, (14) implies that ΔH_1^* is due to the expansion or contraction of the phase space in dimension x_1, while keeping dimension x_2 unchanged. That is to say,

$$\Delta H_1^* = E \log|J_1|, \tag{15}$$

where $J_1 = \partial\Phi_1/\partial x_1$. Following the same argument with the continuous system, the transfer of entropy from X_2 to X_1, $T_{2\to1}$, is simply the difference between ΔH_1 and ΔH_1^*. Likewise, the transfer from the opposite direction may also be calculated.

The above formulation works for the case when both Φ_1 and Φ_2 are invertible. Maps, however, are much more complex than the discretized form of (1). In this context, even though a 2D system is invertible, its individual components may not be so. Well-studied examples include the baker transformation and the Hénon map. We need to extend the foregoing formalism to allow for noninvertibility.

The central point of the extension is with the starred entropy increase. Taking direction 1 as an example, what is needed is to generalize (15) so that noninvertibility is permissible. We claim the generalized version should be

$$\Delta H_1^* = \int \rho_1(x_1)\log\rho_1(x_1)dx_1 - \iint \mathcal{P}_1\rho_1(\Phi_1(x_1, x_2))\log\mathcal{P}_1\rho_1(\Phi_1(x_1, x_2))\rho(x_2|x_1)|J_1|dx_1dx_2, \tag{16}$$

where $\mathcal{P}_1$ is the *F-P* operator when x_2 is fixed; i.e., x_2 appears in $\mathcal{P}_1$ as a parameter. Equation (16) is just a restatement of our previous argument that ΔH_1^* be the entropy increase in X_1 from one time step to the next step with x_2 frozen as a parameter, given X_1 at the present step. It is easy to prove that [7] (a) when Φ_1 is invertible, (16) reduces to (15), namely, $E\log|J_1|$, and (b) when Φ_1 is independent of x_2, $\Delta H_1^* = \Delta H_1$. Property (a) means that (16) is indeed an extension of our previous formalism, while (b) is just the asymmetry property as expected. For all that account, the entropy transfer from X_2 to X_1 is, in a unified form,

$$T_{2\to1} = -\int_{\Omega_1}\left(\int_{\Omega_2} \mathcal{P}\rho dx_2\right)\log\left(\int_{\Omega_2} \mathcal{P}\rho dx_2\right)dx_1 + \iint \mathcal{P}_1\rho_1(\Phi_1(x_1,x_2))\log\mathcal{P}_1\rho_1(\Phi_1(x_1,x_2))\rho(x_2|x_1))|J_1|dx_1dx_2. \tag{17a}$$

Likewise, the transfer from X_1 to X_2 is

$$T_{1\to2} = -\int_{\Omega_2}\left(\int_{\Omega_1} \mathcal{P}\rho dx_1\right)\log\left(\int_{\Omega_1} \mathcal{P}\rho dx_1\right)dx_2 + \iint \mathcal{P}_2\rho_2(\Phi_2(x_1,x_2))\log\mathcal{P}_2\rho_2(\Phi_2(x_1,x_2))\rho(x_1|x_2))|J_2|dx_1dx_2, \tag{17b}$$

where $\mathcal{P}_2$ is the *F-P* operator corresponding to transformation Φ_2 with x_1 fixed as a parameter.

The information transfers given by (17a) and (17b) are physically consistent with the previous formalisms. In particular, it is consistent with Schreiber's transfer entropy [2]. The transfer entropy is a Kullback entropy-like quantity, which measures the incorrectness when the probability mass function of X_1 at time step n conditioned on the measurements at previous time steps is taken as the probability of X_1 given the measurements of both X_1 and X_2 at their previous time steps. The essence of this philosophy is reflected in our formalism. In the case of a Markov chain of order one, Schreiber's formula for the transfer entropy from X_2 to X_1 is, at time step n,

$$T_{2\to1} = \sum P(x_1^{n+1}, x_1^n, x_2^n)\log\frac{P(x_1^{n+1}|x_1^n, x_2^n)}{P(x_1^{n+1}|x_1^n)}, \tag{18}$$

which is equal to the difference between $A = -\sum P(x_1^{n+1}, x_1^n)\log P(x_1^{n+1}|x_1^n)$ and $B = -\sum P(x_1^{n+1}, x_1^n, x_2^n)\log P(x_1^{n+1}|x_1^n, x_2^n)$. These two terms correspond, respectively, to the entropy increases ΔH_1

and ΔH_1^* in our formalism upon one transformation: The transition probability reflects the dynamics; when we freeze a direction of the phase space to extract information transfer, it essentially produces the conditional probability given the component for that direction. In this sense, our formalism is physically consistent with the Schreiber formula with a Markov chain of order one.

However, our formalism differs quantitatively from (18). The major difference lies in that A and B are not strictly in a form of entropy increase, while entropy increase forms the building blocks for our formalism. This difference might lead to different results with the same problem, as is clear in the following Hénon map application.

A Hénon map sets a good example for the 2D information transfer. Defined as $\Phi = (\Phi_1, \Phi_2)\colon \mathbb{R}^2 \to \mathbb{R}^2$,

$$(x_1, x_2) \mapsto (1 + x_2 - a x_1^2, b x_1),$$

with $a > 0$ and $b > 0$ two parameters; X_2 depends solely on X_1. One therefore expects simple physics existing for $T_{1\to 2}$. ($T_{2\to 1}$ is much more complex, and we will consider it in Ref. [7].) Specifically, a pure transfer from X_1 is expected to account for all the entropy change in X_2, and, hence, $T_{1\to 2}$ should be equal to the change of entropy in terms of the marginal distribution of X_2. This is indeed the case by (17), which results in

$$T_{1\to 2} = \log b + H_1. \tag{19}$$

Equation (19) reconfirms the fact that the entropy transferred to X_2 is all that X_1 possesses, plus the part due to the expansion/contraction of phase space (with a factor b). Particularly, when $b = 1$, $T_{1\to 2} = H_1$. This simple result is just what one may expect from the transformation $\Phi_2(x_1, x_2) = b x_1$. Among the transfer measures we know, (17) is the only one that results in a transfer as consistent with the dynamics of Hénon map as (19) is.

The baker transformation is another example whose information transfer is qualitatively clear in physics, as mentioned in the beginning of this Letter. It is defined as a mapping

$$\Phi(x_1, x_2) = \begin{cases} (2x_1, \frac{x_2}{2}) & 0 \le x_1 \le \frac{1}{2}, 0 \le x_2 \le 1, \\ (2x_1 - 1, \frac{x_2}{2} + \frac{1}{2}) & \frac{1}{2} < x_1 \le 1, 0 \le x_2 \le 1, \end{cases}$$

mimicking a kneading dough. By computing Φ^{-1} and $\mathcal{P}\rho$ and using the formulas (17a) and (17b), it is straightforward to obtain [7]

$$T_{2\to 1} = 0, \qquad T_{1\to 2} > 0.$$

This is to say, X_1 (stretching) always loses information to X_2 (folding), while no transfer is invoked in the other way. This is just as expected with the common physical intuition [4,5].

We have investigated the information transfer between components in a 2D system, both in the form of a continuous flow and with a discrete mapping. The resulting entropy transfer possesses a property of asymmetry, which requires that the transfer to random variable X_1 from variable X_2 goes to zero if X_1 evolves independently of X_2, while in the same time the opposite does not need to be so. This formalism has been compared to Schreiber's transfer entropy and applied to several physical problems. Applications to continuous dynamical systems are deferred to Ref. [7], as the minimal dimensionality required for chaos is 3 [5]. Applications to discrete systems include one with the Hénon map and one with the baker transformation. For the former, it gives a simple and physically clear transfer from the quadratic component to the linear component; for the latter, we found that there is always information flowing from the stretching coordinate to the folding coordinate, while no transfer occurs in the opposite direction. The whole idea can be extended to dynamical systems with many dimensions. We will present that in the follow-ups to this Letter.

X. S. L. has benefited from Andrew Majda's lectures on information theory. This work was supported by NSF under CMG Grant No. 0417728 to Courant Institute of Mathematical Sciences.

*Electronic address: sanliang@cims.nyu.edu

[1] K. Kaneko, Physica (Amsterdam) **23D**, 436 (1986); J. A. Vastano and H. L. Swinney, Phys. Rev. Lett. **60**, 1773 (1988); T. Schreiber, J. Phys. A **23**, L393 (1990); M. G. Rosenblum *et al.*, Phys. Rev. Lett. **76**, 1804 (1996); H. D. I. Abarbanel, N. Masuda, M. I. Rabinovich, and E. Tumer, Phys. Lett. A **281**, 368 (2001).
[2] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[3] References in different contexts can be found in H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England, 2004); H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996).
[4] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics* (Springer, New York, 1994).
[5] E. Ott, T. Sauer, and J. A. Yorke, *Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems* (Wiley-Interscience, New York, 1994); A. J. Lichtenberger and M. A. Lieberman, *Regular and Chaotic Dynamics* (Springer, New York, 1992).
[6] For example, R. Kleeman, J. Atmos. Sci. **59**, 2057 (2002); R. Kleeman, A. Majda, and I. Timofeyev, Proc. Natl. Acad. Sci. U.S.A. **99**, 15 291 (2002); R. Kleeman and A. Majda, J. Atmos. Sci. **62**, 2864 (2005); R. Abram, A. Majdaov, and R. Kleeman, J. Atmos. Sci. **62**, 65 (2005).
[7] X. S. Liang and R. Kleeman (to be published).