Information flow and causality quantification in discrete and continuous stochastic systems

X. San Liang
China Institute for Advanced Study,
Central University of Finance and Economics, Beijing,
State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou,
and National University of Defense Technology, Changsha,
P. R. China

1. Introduction

Information flow, or information transfer as referred in the literature, is a fundamental physics concept that has applications in a wide variety of disciplines such as neuroscience (e.g., Pereda et al., 2005), atmosphere-ocean science (Kleeman, 2002; 2007; Tribbia, 2005), nonlinear time series analysis (e.g., Kantz & Schreiber, 2004; Abarbanel, 1996), economics, material science, to name several. In control theory, it helps to understand the information structure and hence characterize the cause-effect notion of causality in nonsequential stochastic control systems (e.g., Andersland & Teneketzis, 1992). Given the well-known importance, it has been an active arena of research for several decades (e.g.,Kaneko, 1986; Vastano & Swinney, 1988; Rosenblum et al., 1996; Arnhold et al., 1999; Schreiber, 2000; Kaiser & Schreiber, 2002). However, it was not until recently that the concept is formalized, on a rigorous mathematical and physical footing. In this chapter we will introduce the rigorous formalism initialized in Liang & Kleeman (2005) and established henceforth; we will particularly focus on the part of the studies by Liang (2008) and Liang & Kleeman (2007a,b) that is pertaining to the subjects of this book. For formalisms in a more generic setting or of broader interest the reader should consult and cite the original papers.

The concept of information flow/transfer was originally introduced to overcome the shortcoming of mutual information in reflecting the transfer asymmetry between the transmitter and the recipient. It is well known that mutual information tells the amount of information exchanged (cf. Cove & Thomas, 1991), but does not tell anything about the directionality of the exchange. This is the major thrust that motivates many studies in this field, among which are Vastano & Swinney (1988) and Schreiber (2000). Another thrust, which is also related to the above, is the concern over causality. Traditionally, causality, such as the Granger causality (Granger, 1969), is just a qualitative notion. While it is useful in identifying the causal relation between dynamical events, one would like to have a more accurate measure to quantify this relation. This would be of particular use in characterizing the intricate systems with two-way coupled events, as then we will be able to weigh the relative importance of one event over another. Information flow is expected to function as this quantitative measure.

The third thrust is out of the consideration from general physics. Information flow is a physical concept seen everywhere in our daily life experiences. The renowned baker transformation (cf. section 5 in this chapter), which mimics the kneading of a dough, is such an example. It has been argued intuitively that, as the transformation applies, information flows continuingly from the stretching direction to the folding direction, while no transfer is invoked the other way (e.g., Lasota & Mackey, 1994). Clearly the central issue here is how much the information is transferred between the two directions.

Historically information flow formalisms have been developed in different disciplines (particularly in neuroscience), usually in an empirical or half-empirical way within the context of the problems in question. These include the time-delayed information transfer (Vastano & Swinney, 1988) and the more sophisticated transfer entropy associated with a Markov chain (Schreiber, 2000). Others, though in different appearances, may nevertheless be viewed as the varieties of these two types. Recently, it was observed that even these two are essentially of the same like, in that both deal with the evolution of marginal entropies (Liang & Kleeman, 2005; 2007a). With this observation, Liang & Kleeman realized that actually this important concept can be rigorously formulated, and the corresponding formulas analytically derived rather than empirically proposed. The so-obtained transfer measure possesses nice properties as desired, and has been verified in different applications, with both benchmark systems and real world problems. The objective of this chapter is to give a concise introduction of this formalism. Coming up next is a setup of the mathematical framework, followed by two sections (§3 and §4) where the transfer measures for different systems are derived. In these sections, one will also see a very neat law about entropy production [cf. Eq. (18) in §3.1.2], paralleling the law of energy conservation, and the some properties of the resulting transfer measures (§4.3). Section 5 gives two applications, one about the afore-mentioned baker transformation, the other about a surprisingly interesting causality inference with two highly correlated time series. The final section (section 6) is a brief summary. Through the chapter only two-dimensional systems are considered; for high dimensional formalisms, see Liang & Kleeman (2007)a,b. As a convention in the literature, the terminologies "information flow" and "information transfer" will be used interchangeably throughout.

2. Mathematical formalism

Let Ω be the sample space and $\underline{x} \in \Omega$ the vector of state variables. For convenience, we follow the convention of notation in the physics literature, where random variables and deterministic variables are not distinguished. (In probability theory, they are usually distinguished with lower and upper cases like $\underline{x}$ and $\underline{X}$.) Consider a stochastic process of $\underline{x}$, which may take a continuous time form $\{\underline{x}(t), t \geq 0\}$ or a discrete time form $\{\underline{x}(\tau), \tau\}$, with τ being positive integers signifying discrete time steps. Throughout this chapter, unless otherwise indicated, we limit out discussion within two-dimensional (2D) systems $\underline{x} = (x_1, x_2)^T \in \Omega$ only. The stochastic dynamical systems we will be studying with are, in the discrete time case,

$$\underline{x}(\tau + 1) = \Phi(\underline{x}(\tau)) + \underline{\underline{B}}(\underline{x}, \tau)\underline{v} \tag{1}$$

and, in the continuous time case,

$$d\underline{x} = \underline{\underline{F}}(\underline{x}, t)dt + \underline{\underline{B}}(\underline{x}, t)d\underline{w}. \tag{2}$$

Here Φ is a 2-dimensional transformation

$$\Phi : \Omega \to \Omega, \qquad (x_1, x_2) \mapsto (\Phi_1(\underline{x}), \Phi_2(\underline{x})), \tag{3}$$

$\underline{\mathbf{F}}$ the vector field, $\underline{\mathbf{v}}$ the white noise, $\underline{\mathbf{w}}$ a standard Wiener process, and $\underline{\mathbf{B}}$ a 2×2 matrix of the perturbation amplitude. The sample space Ω is assumed to be a Cartesian product $\Omega_1 \times \Omega_2$. We therefore just need to examine how information is transferred between the two components, namely x_1 and x_2, of the system in question. Without loss of generality, it suffices to consider only the information transferred from x_2 to x_1, or $T_{2 \to 1}$ for short.

Associated with each state $\underline{\mathbf{x}} \in \Omega$ is a joint probability density function

$$\rho = \rho(\underline{\mathbf{x}}) = \rho(x_1, x_2) \in L^1(\Omega),$$

and two marginal densities

$$\rho_1(x_1) \quad = \quad \int_{\Omega_2} \rho(x_1, x_2) \, dx_2,$$

$$\rho_2(x_2) \quad = \quad \int_{\Omega_1} \rho(x_1, x_2) \, dx_1,$$

with which we have a joint (Shannon) entropy

$$H = - \iint_{\Omega} \rho(\underline{\mathbf{x}}) \log \rho(\underline{\mathbf{x}}) \, d\underline{\mathbf{x}}, \tag{4}$$

and marginal entropies

$$H_1 \quad = \quad - \int_{\Omega_1} \rho(x_1) \log \rho(x_1) \, dx_1, \tag{5}$$

$$H_2 \quad = \quad - \int_{\Omega_2} \rho(x_2) \log \rho(x_2) \, dx_2. \tag{6}$$

As $\underline{\mathbf{x}}$ evolves, the densities evolve subsequently. Specifically, corresponding to (2) there is a Fokker-Planck equation that governs the evolution of ρ; if $\underline{\mathbf{x}}$ moves on according to (1), the density is steered forward by the Frobenius-Perron operator (F-P operator henceforth). (Both the Fokker-Planck equation and the F-P operator will be introduced later.) Accordingly the entropies H, H_1, and H_2 also change with time. As reviewed in the introduction, the classical empirical/half-empirical information flow/transfer formalisms, though appearing in different forms, all essentially deal with the evolution of the marginal entropy of the receiving component, i.e., that of x_1 if $T_{2 \to 1}$ is considered. With this Liang & Kleeman (2005) noted that, by carefully classifying the mechanisms that govern the marginal entropy evolution, the concept of information transfer or information flow actually can be put on a rigorous footing. More specifically, the evolution of H_1 can be decomposed into two exclusive parts, according to their driving mechanisms: one is from x_2 only, another with the effect from x_2 excluded. The former, written $T_{2 \to 1}$, is the very information flow or information transfer from x_2 to x_1. Putting the latter as $\frac{dH_{1\slashed{2}}}{dt}$ for the continuous case, and $\Delta H_{1\slashed{2}}$ for the discrete case, we therefore have:

(1) For the discrete system (1), the information transferred from x_2 to x_1 is

$$T_{2 \to 1} = \Delta H_1 - \Delta H_{1\slashed{2}}; \tag{7}$$

(2) For the continuous system (2), the rate of information transferred from x_2 to x_1 is

$$T_{2 \to 1} = \frac{dH_1}{dt} - \frac{dH_{1\slashed{2}}}{dt}. \tag{8}$$

Likewise, the information flow from x_1 to x_2 can be defined. In the following we will be exploring how these are evaluated.

3. Deterministic systems with random inputs

We begin with the deterministic counterparts of (1) and (2), i.e.,

$$\underline{\mathbf{x}}(\tau + 1) = \Phi(\underline{\mathbf{x}}(\tau)), \tag{9}$$

and

$$\frac{d\underline{\mathbf{x}}}{dt} = \underline{\mathbf{F}}(\mathbf{x}, t), \tag{10}$$

respectively, with randomness limited within initial conditions, and then extend it to generic systems. This is not just because that (9) [resp. (10)] makes a special case of (1) [resp. (2)], but also because historically it is the idiosyncrasy of deterministic systems (Liang & Kleeman, 2005) that stimulates the rigorous formulation for this important physical notion, namely information flow or information transfer.

3.1 Entropy production

We first examine how entropy is produced with the systems (9) and (10). In this subsection, the system dimensionality is not limited to 2, but can be arbitrary.

3.1.1 Entropy evolution with discrete systems

Let $\rho = \rho(\underline{\mathbf{x}})$ be the joint density of $\underline{\mathbf{x}}$ at step τ, with the dependence on τ suppressed for simplicity. Its evolution is governed by the Frobenius-Perron operator, or F-P operator as will be called,

$$\mathcal{P} : L^1(\Omega) \mapsto L^1(\Omega),$$

which is given by, in a loose sense,

$$\int_{\omega} \mathcal{P}\rho(\underline{\mathbf{x}}) \, d\underline{\mathbf{x}} = \int_{\Phi^{-1}(\omega)} \rho(\underline{\mathbf{x}}) \, d\underline{\mathbf{x}}, \tag{11}$$

for any $\omega \subset \Omega$. [A rigorous definition with measure theory can be seen in Lasota & Mackey (1994).] If Φ is nonsingular and invertible, the right hand side of (11) is

$$\int_{\Phi^{-1}(\omega)} \rho(\underline{\mathbf{x}}) \, d\underline{\mathbf{x}} \xupupdownarrow{\underline{\mathbf{y}} = \Phi(\underline{\mathbf{x}})} \int_{\omega} \rho\left[\Phi^{-1}(\underline{\mathbf{y}})\right] \left|J^{-1}\right| \, d\underline{\mathbf{y}},$$

where J is the Jacobian of Φ:

$$J = J(\underline{\mathbf{x}}) = \det\left[\frac{\partial\Phi(x_1, x_2)}{\partial(x_1, x_2)}\right].$$

and J^{-1} its inverse. So in this case $\mathcal{P}$ can be explicitly written out:

$$\mathcal{P}\rho(\underline{\mathbf{x}}) = \rho\left[\Phi^{-1}(\underline{\mathbf{x}})\right] \left|J^{-1}\right|. \tag{12}$$

With $\mathcal{P}$, the change of the joint entropy H from time step τ to step $\tau + 1$ is, by (4),

$$\Delta H \quad = \quad H(\tau + 1) - H(\tau)$$

$$= -\iint_\Omega \mathcal{P}\rho(\underline{\mathbf{x}}) \log \mathcal{P}\rho(\underline{\mathbf{x}}) \, d\underline{\mathbf{x}} + \iint_\Omega \rho(\underline{\mathbf{x}}) \log \rho(\underline{\mathbf{x}}) \, d\underline{\mathbf{x}}. \tag{13}$$

In the case of nonsingular and invertible Φ, the above can be evaluated:

$$
\begin{aligned}
\Delta H \quad &= -\iint_\Omega \rho\left[\Phi^{-1}(\underline{\mathbf{x}})\right] \left|J^{-1}\right| \cdot \log\left(\rho\left[\Phi^{-1}(\underline{\mathbf{x}})\right] \left|J^{-1}\right|\right) \, d\underline{\mathbf{x}} + \iint_\Omega \rho \log \rho \, d\underline{\mathbf{x}} \\
&\underset{=\!=\!=\!=\!=}{\overset{\underline{\mathbf{y}}=\Phi^{-1}(\underline{\mathbf{x}})}{}} -\iint_\Omega \rho(\underline{\mathbf{y}})\left[\log\rho(\underline{\mathbf{y}}) + \log\left|J^{-1}\right|\right] \, d\underline{\mathbf{y}} + \iint_\Omega \rho \log \rho \, d\underline{\mathbf{x}} \\
&= \iint_\Omega \rho(\underline{\mathbf{y}})\left|J\right| \, d\underline{\mathbf{y}}.
\end{aligned}
$$

We hence have the following theorem:

Theorem 3.1. *If the system (9) has a nonsingular and invertible mapping Φ, then the entropy change can be expressed as, in a concise form,*

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

where E is the mathematical expectation with respect to ρ.

Equation (14), which was established in Liang & Kleeman (2005), states that the entropy increase for a discrete system upon one application of an invertible transformation is simply the average logarithm of the rate of area change under the transformation. This extremely concise form of evolution gives us a hint on how the information flow concept may be easily obtained, as will be clear soon.

3.1.2 Entropy evolution with continuous systems

Now consider the continuous system (10). Here the dimensionality is not just limited to 2, but can be any positive integer n. First discretize it on the infinitesimal interval $[t, t + \Delta t]$:

$$\underline{\mathbf{x}}(t + \Delta t) = \underline{\mathbf{x}}(t) + \underline{\mathbf{F}}(\underline{\mathbf{x}}(t), t)\Delta t. \tag{15}$$

This equation defines a mapping $\Phi : \Omega \to \Omega$, $\underline{\mathbf{x}} \mapsto \underline{\mathbf{x}} + \underline{\mathbf{F}}(\underline{\mathbf{x}}, t)\Delta t$, with a Jacobian

$$
\begin{aligned}
J \quad &= \det\left[\frac{\partial\Phi(x_1, x_2, ..., x_n)}{\partial(x_1, x_2, ..., x_n)}\right] \\
&= \det\begin{bmatrix}
1 + \frac{\partial F_1}{\partial x_1}\Delta t & \frac{\partial F_1}{\partial x_2}\Delta t & ... & \frac{\partial F_1}{\partial x_n}\Delta t \\
\frac{\partial F_2}{\partial x_1}\Delta t & 1 + \frac{\partial F_2}{\partial x_2}\Delta t & ... & \frac{\partial F_2}{\partial x_n}\Delta t \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial F_n}{\partial x_1}\Delta t & \frac{\partial F_n}{\partial x_2}\Delta t & ... & 1 + \frac{\partial F_n}{\partial x_n}\Delta t
\end{bmatrix} \\
&= \Delta t \sum_i \frac{\partial F_i}{\partial x_i} + O(\Delta t^2).
\end{aligned} \tag{16}
$$

As $\Delta t \to 0$, it is easy to show that Φ is always nonsingular and invertible; in fact, $\Phi^{-1} : \Omega \to \Omega$ can be explicitly found:

$$\Phi^{-1}(\underline{\mathbf{x}}) = \underline{\mathbf{x}} - \underline{\mathbf{F}}(\underline{\mathbf{x}}, t)\Delta t + O(\Delta t^2). \tag{17}$$

So by (14), as $\Delta t \to 0$,

$$
\begin{aligned}
\frac{dH}{dt} &= \lim_{\Delta t \to 0} \frac{E \log |J|}{\Delta t} \\
&= E \lim_{\Delta t \to 0} \frac{1}{\Delta t} \log \left(1 + \Delta t \sum_i \frac{\partial F_i}{\partial x_i} + O(\Delta t^2) \right) \\
&= E \left(\sum_i \frac{\partial F_i}{\partial x_i} \right).
\end{aligned}
$$

This fulfills the proof of the following important theorem:

Theorem 3.2. *For the deterministic system (10), the entropy H evolves according to*

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

(18)

Like (14), Eq. (18) is also in an extremely concise form. It states that the time rate of change of H is totally controlled by the contraction or expansion of the phase space. This important theorem was established by Liang & Kleeman (2005), using the Liouville equation corresponding to (10). But the derivation therein requires some assumption (though very weak) at the boundaries, while here no assumption is invoked.

3.2 Information flow

The elegant formula (18) allows us to obtain with ease the information flow for the continuous system (10). Indeed, this is precisely what Liang & Kleeman (2005) did in establishing the first formalism in a rigorous sense. To be short, consider only the rate of information transfer from x_2 to x_1, namely $T_{2 \to 1}$, which is the difference between the rate of change of the marginal entropy $\frac{dH_1}{dt}$ and that with the effect from x_2 excluded, i.e., $\frac{dH_{1\cancel{2}}}{dt}$. In a 2D system, $\frac{dH_{1\cancel{2}}}{dt}$ is actually equivalent to the rate of H_1 evolution due to x_1 its own, denoted $\frac{dH_1^*}{dt}$. Observing the obvious additivity property of (18), Liang & Kleeman (2005) intuitively argued that

$$
\frac{dH_1^*}{dt} = E \left(\frac{\partial F_1}{\partial x_1} \right).
$$

We hence obtain the following theorem:

Theorem 3.3. *For the 2D system (10),*

$$
\frac{dH_{1\cancel{2}}}{dt} = E \left(\frac{\partial F_1}{\partial x_1} \right) = \iint_\Omega \rho \frac{\partial F_1}{\partial x_1} \, dx_1 dx_2.
$$

(19)

(The proof of this theorem is deferred to later in this subsection.) The information flow from x_2 to x_1 therefore follows easily from (8).

Theorem 3.4. *For the 2D system (10), the rate of information transferred from x_2 to x_1 is*

$$T_{2\to 1} = -\mathcal{E}_{2|1}\left(\frac{\partial (F_1 \rho_1)}{\partial x_1}\right), \tag{20}$$

where $\mathcal{E}$ is an integration operator defined with respect to the conditional density

$$\rho_{2|1}(x_1 | x_1) = \frac{\rho(x_1, x_2)}{\rho_1(x_1)} \tag{21}$$

such that, for any function $f = f(x_1, x_2)$,

$$\mathcal{E}_{2|1} f = \iint_\Omega \rho_{2|1}(x_2 | x_1) \cdot f(x_1, x_2)\, dx_1 dx_2. \tag{22}$$

Proof
Corresponding to (10) is the Liouville equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\mathbf{F}\rho) = 0 \tag{23}$$

that governs the density evolution. Integrating it with respect to x_2 over the subspace Ω_2,

$$\frac{\partial \rho_1}{\partial t} + \frac{\partial}{\partial x_1} \int_{\Omega_2} \rho F_1\, dx_2 = 0. \tag{24}$$

The other term is integrated out with the compact support assumption for ρ. Multiplication by $-(1 + \log \rho_1)$, followed by an integration over Ω_1, gives

$$
\begin{aligned}
\frac{dH_1}{dt} &= \iint_\Omega \left[\log \rho_1 \frac{\partial (\rho F_1)}{\partial x_1}\right] dx_1 dx_2 \\
&= -\iint_\Omega \rho \left[\frac{F_1}{\rho_1} \frac{\partial \rho_1}{\partial x_1}\right] dx_1 dx_2.
\end{aligned}
$$

In the second step integration by parts is used; also used is the compact support assumption for ρ. So

$$
\begin{aligned}
T_{2\to 1} &= \frac{dH_1}{dt} - \frac{dH_{1\!\backslash\!2}}{dt} \\
&= \frac{dH_1}{dt} - E\left(\frac{\partial F_1}{\partial x_1}\right) \\
&= -\iint_\Omega \left(\frac{F_1}{\rho_1} \frac{\partial \rho_1}{\partial x_1} + \frac{\partial F_1}{\partial x_1}\right) \rho\, dx_1 dx_2 \\
&= -\iint_\Omega \rho_{2|1}(x_2 | x_1) \frac{\partial (F_1 \rho_1)}{\partial x_1}\, dx_1 dx_2.
\end{aligned}
$$

Q.E.D.

One may argue that, following the same way with (14), the information flow for the discrete system (9) can be obtained. Indeed this is true, but only in part, as the neat formula (14) requires that the mapping Φ and its components be nonsingular and invertible. Unfortunately,

for many important 2D mappings like the baker transformation we will be introducing in section 5, the requirements are generally not met. We therefore need to consider more generic situations.

By (7), we need to find ΔH_1 and $\Delta H_{1\slashed{2}}$ as the system (9) moves forward from step τ to $\tau + 1$. As in the continuous case, it is easy to obtain ΔH_1 from the given mapping Φ. The key is how to find $\Delta H_{1\slashed{2}}$, which is the entropy increase in direction 1 as the system goes from τ to $\tau + 1$ under Φ with x_2 frozen instantaneously at step τ, given $x_1(\tau)$. As $\Delta H_{1\slashed{2}} = H_{1\slashed{2}}(\tau + 1) - H_1(\tau)$, we are done if $H_{1\slashed{2}}(\tau + 1)$ is evaluated. This is the marginal entropy for the first component evolved from H_1 with contribution from x_2 excluded from τ to $\tau + 1$. Consider the quantity

$$f \equiv -\log \mathcal{P}_{1\slashed{2}}\rho_1(y_1), \tag{25}$$

where $y_1 = \Phi_1(\underline{x})$, and $\mathcal{P}_{1\slashed{2}}\rho_1(y_1)$ is the marginal density in direction 1 at step $\tau + 1$, as the density ρ_1 evolves from step τ to step $\tau + 1$ under the transformation:

$$\Phi_{\slashed{2}} : y_1 = \Phi_1(x_1, x_2) \tag{26}$$

i.e., the map Φ with x_2 frozen instantaneously at τ as a parameter. Note here we use $y_1 = \Phi_1(\underline{x})$ for the state of component 1 at step $\tau + 1$ (x_1 is for that at step τ); We do not use x_1 with some superscript or subscript in order to avoid any possible confusion in distinguishing the states of x_1 at these two time steps.

With our notation introduced above, $H_{1\slashed{2}}(\tau + 1)$ is the mathematical expectation of f. (Recall how Shannon entropy is defined.) In other words, it is equal to the integration of f times some probability density function over the corresponding sample space. The first density to be multiplied is $\mathcal{P}_{1\slashed{2}}\rho_1(y_1)$. But f also depends on x_2, we thence need another density for x_2. Recall that the freezing of x_2 is performed on interval $[\tau, \ \tau + 1]$, given all other components (here only x_1 in this 2D system) at step τ. What we need is therefore the conditional density of x_2 given x_1 at τ, i.e., $\rho(x_2|x_1)$. Put all these together, we therefore have the following result.

Proposition 3.1. *As the system (9) evolves from time step τ to time step $\tau + 1$, if x_2 is instantaneously frozen as a parameter at step τ, the marginal entropy of x_2 at step $\tau + 1$ is*

$$H_{1\slashed{2}}(\tau + 1) = -\iint_\Omega \mathcal{P}_{1\slashed{2}}\rho_1(y_1) \cdot \log \mathcal{P}_{1\slashed{2}}\rho_1(y_1) \cdot \rho(x_2|x_1) \, dy_1 dx_2, \tag{27}$$

where y_1 is given by (26).

Note here we do not do another averaging with respect to x_1, as x_1 is already embedded in y_1.

The information transferred from x_2 to x_1 is now easy to obtain. Since $H_1(\tau)$ is the same, the right hand side of (7) is simply the difference between

$$H_1(\tau + 1) = -\iint_\Omega (\mathcal{P}\rho)_1(y_1) \log(\mathcal{P}\rho)_1(y_1) \, dy_1, \tag{28}$$

where $(\mathcal{P}\rho)_1$ is the marginal density at step $\tau + 1$, and $H_{1\slashed{2}}(\tau + 1)$. We hence arrive at the following theorem on information flow.

Theorem 3.5. *For system (9), the information transferred from x_2 to x_1 is*

$$
\begin{aligned}
T_{2\to1} \quad = \quad & -\int_{\Omega_1} (\mathcal{P}\rho)_1(y_1) \cdot \log(\mathcal{P}\rho)_1(y_1)\, dy_1 \\
& +\int_{\Omega} \mathcal{P}_{1\not2}\rho_1(y_1) \cdot \log \mathcal{P}_{1\not2}\rho_1(y_1) \cdot \rho(x_2|x_1)\, dy_1 dx_2.
\end{aligned}
\tag{29}
$$

Likewise the information flow from x_1 to x_2 can be obtained. Note in arriving at (29) no issue about the invertibility of Φ or either of its components is ever invoked. But if invertibility is guaranteed, then the formula may be further simplified.

Corollary 3.1. *In the system (9), if the mapping Φ has a component Φ_1 that is invertible, then*

$$
\Delta H_{1\not2} = E \log |J_1|, \qquad where \quad J_1 = \frac{\partial \Phi_1(\mathbf{x})}{\partial x_1},
\tag{30}
$$

and hence

$$
T_{2\to1} = \Delta H_1 - E \log |J_1|.
\tag{31}
$$

Remark: This concise result is just one would expect by the similar heuristic argument in arriving at Theorem 3.3 and Theorem 3.4.

Proof
By (27),

$$
\begin{aligned}
\Delta H_{1\not2} \quad = \quad & -\iint_{\Omega_1 \times \Omega_2} \mathcal{P}_{1\not2}\rho_1(y_1) \cdot \log \mathcal{P}_{1\not2}\rho_1(y_1) \cdot \rho(x_2|x_1)\, dy_1 dx_2 \\
& +\int_{\Omega_1} \rho_1 \log \rho_1\, dx_1,
\end{aligned}
$$

When Φ_1 is invertible, $J_1 = \frac{\partial \Phi_1}{\partial x_1} \neq 0$, by (12),

$$
\begin{aligned}
\mathcal{P}_{1\not2}\rho_1(y_1) \quad = \quad & \rho\left[\Phi_1^{-1}(y_1, x_2)\right] \left|J_1^{-1}\right| \\
= \quad & \rho_1(x_1) \left|J_1^{-1}\right|.
\end{aligned}
\tag{32}
$$

So

$$
\begin{aligned}
\Delta H_{1\not2} \quad = \quad & -\iint \rho_1(x_1) \left|J_1^{-1}\right| \log \left(\rho_1(x_1) \left|J_1^{-1}\right|\right) \rho(x_2|x_1) |J_1|\, dx_1 dx_2 \\
& +\int \rho_1 \log \rho_1\, dx_1 \\
= \quad & -\iint \rho_1(x_1)\, \rho(x_2|x_1) \log \left|J_1^{-1}\right|\, dx_1 dx_2 \\
= \quad & \iint \rho(x_1, x_2) \log |J_1|\, dx_1 dx_2 \\
= \quad & E \log |J_1|,
\end{aligned}
\tag{33}
$$

and the second part follows subsequently.
Q.E.D.

We are now able to prove the first theorem of this subsection, namely Theorem 3.3, which originally was obtained by Liang & Kleeman (2005) through heuristic physical argument.

Proof of Theorem 3.3.
As before, look at an infinitesimal time interval $[t, t + \Delta t]$ and, for clarity, write the state variables at time t and $t + \Delta t$ as, respectively, $\underline{x}$ and $\underline{y}$. Discretization of (10) yields a mapping $\Phi = (\Phi_1, \Phi_2) : \Omega \to \Omega, \underline{x} = (x_1, x_2) \mapsto \underline{y} = (y_1, y_2)$,

$$\Phi : \begin{cases} y_1 = x_1 + \Delta t \cdot F_1(\underline{x}, t), \\ y_2 = x_2 + \Delta t \cdot F_2(\underline{x}, t). \end{cases} \tag{34}$$

As shown before, as $\Delta t \to 0$, Φ is nonsingular and always invertible, so are its components Φ_1 and Φ_2. Moreover, the Jacobian for Φ_1 is

$$J_1 = \frac{\partial y_1}{\partial x_1} = 1 + \Delta t \frac{\partial F_1}{\partial x_1} + O(\Delta t^2). \tag{35}$$

By Corollary 3.1, $\Delta H_{1\cancel{2}} = E \log |J_1|$. So

$$\begin{aligned} \frac{dH_{1\cancel{2}}}{dt} &= \lim_{\Delta t \to 0} \frac{\Delta H_{1\cancel{2}}}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E \left(\log \left| 1 + \Delta t \frac{\partial F_1}{\partial x_1} \right| + O(\Delta t^2) \right) \\ &= E \left(\frac{\partial F_1}{\partial x_1} \right). \end{aligned}$$

Q.E.D.

4. Stochastic systems

With the information flow for deterministic systems derived, we now take into account stochasticity and re-consider the problem. We first consider discrete systems in the form of (1), then continuous systems (2).

4.1 Discrete stochastic systems
As our convention, write $\underline{x}(\tau + 1)$ as $\underline{y}$ to avoid confusion. Eq. (1) then defines a mapping sending $\underline{x}$ to $\underline{y}$:

$$\underline{y} = \Phi(\underline{x}) + \underline{B}(\underline{x})\underline{v}, \tag{36}$$

where $\underline{v}$ is a vector of white noise defined on $\mathbb{R}^2$, $\underline{B} = (b_{ij})$ is a matrix of the perturbation amplitude, and the dependence on τ in the terms is suppressed for notation simplicity. Corresponding to this mapping is a Markov operator $\mathcal{P} : L^1(\Omega) \to L^1(\Omega)$, similar to the F-P operator for the system (9), that sends $\rho(\underline{x}(\tau))$ to $\rho(\underline{x}(\tau + 1))$ or $\rho(\underline{y})$. To find $\mathcal{P}$, we need just find $\rho(\underline{y})$, given $\rho(\underline{x})$, Φ, $\underline{B}$, and $\rho(\underline{v})$ which is also written as $\rho_v(\underline{v})$ for clarity. For convenience, $\underline{B}$ is assumed to be nonsingular.

Let Π be a transformation of $(\underline{\mathbf{x}}, \underline{\mathbf{v}})$ into $(\underline{\mathbf{z}}, \underline{\mathbf{y}})$ such that

$$\Pi : \begin{cases} \underline{\mathbf{z}} = \underline{\mathbf{x}}, \\ \underline{\mathbf{y}} = \Phi(\underline{\mathbf{x}}) + \underline{\underline{\mathbf{B}}}(\underline{\mathbf{x}})\underline{\mathbf{v}}. \end{cases} \tag{37}$$

Its Jacobian is

$$J = \det \left[\frac{\partial(\underline{\mathbf{z}}, \underline{\mathbf{y}})}{\partial(\underline{\mathbf{x}}, \underline{\mathbf{v}})} \right] = \det \left[\begin{array}{cc} \underline{\underline{\mathbf{I}}}_2 & \underline{\underline{\mathbf{0}}}_2 \\ \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} & \underline{\underline{\mathbf{B}}} \end{array} \right] = \det \underline{\underline{\mathbf{B}}}, \tag{38}$$

where $\underline{\underline{\mathbf{I}}}_2$ and $\underline{\underline{\mathbf{0}}}_2$ are 2×2 identity and zero matrices, respectively. Given that $\underline{\underline{\mathbf{B}}}$ is nonsingular, $\det \underline{\underline{\mathbf{B}}}$ is nonzero, and hence Π is invertible:

$$\Pi^{-1} : \begin{cases} \underline{\mathbf{x}} = \underline{\mathbf{z}}, \\ \underline{\mathbf{v}} = \underline{\underline{\mathbf{B}}}^{-1}(\underline{\mathbf{z}}) \left(\underline{\mathbf{y}} - \Phi(\underline{\mathbf{z}}) \right). \end{cases} \tag{39}$$

We now look at how the joint distribution of $(\underline{\mathbf{z}}, \underline{\mathbf{y}})$ is expressed in terms of $(\underline{\mathbf{x}}, \underline{\mathbf{v}})$.

For any $\omega_x \in \Omega$, $\omega_v \in \mathbb{R}^2$,

$$\begin{aligned} \iiiint_{\omega_x \times \omega_v} \rho_{z,y}(\underline{\mathbf{z}}, \underline{\mathbf{y}}) \, d\underline{\mathbf{z}} d\underline{\mathbf{y}} &= \iiiint_{\Pi^{-1}(\omega_x \times \omega_v)} \rho_{x,v}(\underline{\mathbf{x}}, \underline{\mathbf{v}}) \, d\underline{\mathbf{x}} d\underline{\mathbf{v}} \\ &= \iiiint_{\omega_x \times \omega_v} \rho_{x,v} \left(\Pi^{-1}(\underline{\mathbf{z}}, \underline{\mathbf{y}}) \right) \cdot \left| J^{-1} \right| \, d\underline{\mathbf{z}} d\underline{\mathbf{y}}. \end{aligned} \tag{40}$$

As ω_x and ω_v are arbitrarily chosen, the integrands must be equal, and hence

$$\begin{aligned} \rho_{z,y}(\underline{\mathbf{z}}, \underline{\mathbf{y}}) &= \rho_{x,v} \left(\Pi^{-1}(\underline{\mathbf{z}}, \underline{\mathbf{y}}) \right) \cdot \left| J^{-1} \right| \\ &= \rho_{x,v} \left[\underline{\mathbf{z}}, \underline{\underline{\mathbf{B}}}^{-1}(\underline{\mathbf{z}})(\underline{\mathbf{y}} - \Phi(\underline{\mathbf{z}})) \right] \cdot \left| J^{-1} \right| \\ &= \rho_x(\underline{\mathbf{z}}) \cdot \rho_v \left[\underline{\underline{\mathbf{B}}}^{-1}(\underline{\mathbf{z}})(\underline{\mathbf{y}} - \Phi(\underline{\mathbf{z}})) \right] \cdot \left[\det \underline{\underline{\mathbf{B}}}(\underline{\mathbf{z}}) \right]^{-1}. \end{aligned}$$

In the last step, the fact that $\underline{\mathbf{x}}$ and $\underline{\mathbf{v}}$ are independent has been used. Integrate $\underline{\mathbf{z}}$ out and we obtain

$$\begin{aligned} \rho_y(\underline{\mathbf{y}}) &= \iint_\Omega \rho_{z,y}(\underline{\mathbf{z}}, \underline{\mathbf{y}}) \, d\underline{\mathbf{z}} \\ &= \iint_\Omega \rho_x(\underline{\mathbf{z}}) \cdot \rho_v \left[\underline{\underline{\mathbf{B}}}^{-1}(\underline{\mathbf{z}})(\underline{\mathbf{y}} - \Phi(\underline{\mathbf{z}})) \right] \cdot \left[\det \underline{\underline{\mathbf{B}}}(\underline{\mathbf{z}}) \right]^{-1} \, d\underline{\mathbf{z}}. \end{aligned}$$

This equation defines a Markov operator $\mathcal{P}$ (corresponding to the F-P operator in the deterministic case) for system (1):

$$\mathcal{P}\rho(\underline{\mathbf{x}}) = \iint_\Omega \rho(\underline{\mathbf{z}}) \cdot \rho_v \left[\underline{\underline{\mathbf{B}}}^{-1}(\underline{\mathbf{z}})(\underline{\mathbf{y}} - \Phi(\underline{\mathbf{z}})) \right] \cdot \left[\det \underline{\underline{\mathbf{B}}}(\underline{\mathbf{z}}) \right]^{-1} \, d\underline{\mathbf{z}}. \tag{41}$$

In this case ρ_v is a Gaussian distribution with zero mean and an identity covariance matrix, and hence $\mathcal{P}$ can be computed. With it one may calculate the marginal density $(\mathcal{P}\rho)_1$ and hence the marginal entropy at time step $\tau + 1$:

$$H_1(\tau + 1) = - \int_{\Omega_1} (\mathcal{P}\rho)_1(y_1) \cdot \log(\mathcal{P}\rho)_1(y_1) \, dy_1. \tag{42}$$

Next look at $H_{1\hat{2}}(\tau + 1)$. Freezing x_2 at step τ modifies the dynamics to

$$\Phi_{\hat{2}}: \qquad y_1 = \Phi_1(x_1, x_2) + b_{11}v_1 + b_{12}v_2. \qquad (43)$$

Here we distinguish several cases: (1) If $b_{11} = b_{12} = 0$, then this degenerates into a deterministic system, and the Markov operator is the F-P operator as we derived before; (2) if either of the last two terms vanishes, then follow the same procedure as above and a modified Markov operator $\mathcal{P}_{1\hat{2}}$ is obtained; (3) if b_{11} and b_{12} have no dependence on $\underline{x}_1$, then $b_{11}v_1 + b_{12}v_2 \sim N(0, b_{11}^2 + b_{12}^2)$ can be combined to be one random variable with known distribution, and, again, the above procedure applies, and $\mathcal{P}_{1\hat{2}}$ follows accordingly; (4) if neither of the perturbations are zero, then we need to do a transformation from (x_1, v_1, v_2) to (z_1, z_2, y_1) with some random variables z_1 and z_2 as simple as possible. The so-obtained joint density of (z_1, z_2, y_1) is then integrated over the sample spaces of z_1 and z_2, and the resulting marginal entropy is the desired $\mathcal{P}_{1\hat{2}}$. So anyway $\mathcal{P}_{1\hat{2}}$ can be computed, giving

$$H_{1\hat{2}}(\tau + 1) = -\iint_\Omega \mathcal{P}_{1\hat{2}}\rho_1(y_1) \cdot \log \mathcal{P}_{1\hat{2}}\rho_1(y_1) \cdot \rho(x_2|x_1) \, dy_1 dx_2$$

by Proposition 3.1. This subtracted from $H_1(\tau + 1)$ results the information transferred from x_2 to x_1:

$$T_{2\to 1} = H_1(\tau + 1) - H_{1\hat{2}}(\tau + 1). \qquad (44)$$

In principle, following the above procedure all the information flows between the system components can be computed. But more often than not this turn out to be very tedious and difficult. In practice, we would like to suggest different approaches, depending on the problem itself.

4.2 Continuous stochastic systems

For the continuous system (2), there is a Fokker-Planck equation governing the density evolution:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(F_1\rho)}{\partial x_1} + \frac{\partial(F_2\rho)}{\partial x_2} = \frac{1}{2}\sum_{i,j=1}^{2}\frac{\partial^2(g_{ij}\rho)}{\partial x_i \partial x_j}, \qquad (45)$$

where

$$g_{ij} = g_{ji} = \sum_{k=1}^{2} b_{ik}b_{jk}, \qquad ij = 1, 2, \qquad (46)$$

and b_{ij} are the entries of the perturbation matrix $\underline{\underline{B}}$. From this it is easy to obtain the evolution of all the entropies, and H_1 in particular.

Proposition 4.1. *For system (2), the marginal entropy of x_1 evolves according to*

$$\frac{dH_1}{dt} = -E\left(F_1\frac{\partial \log \rho_1}{\partial x_1}\right) - \frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right), \qquad (47)$$

where E stands for expectation with respect to ρ.

Proof.
Integrate (45) with respect to x_2 over Ω_2 to get

$$\frac{\partial \rho_1}{\partial t} + \int_{\Omega_2} \frac{\partial (F_1 \rho)}{\partial x_1} dx_2 = \frac{1}{2} \int_{\Omega_2} \frac{\partial^2 (g_{11} \rho)}{\partial x_1^2} dx_2. \tag{48}$$

Here we have done integration by parts, and applied the compact support assumption for ρ and its derivatives. For simplicity, hereafter we will suppress the integral domain Ω, unless otherwise noted. Multiplication of (48) by $-(1 + \log \rho_1)$, followed by an integration with respect to x_1 over Ω_1, yields

$$\frac{dH_1}{dt} - \iint \log \rho_1 \frac{\partial (F_1 \rho)}{\partial x_1} dx_1 dx_2 = -\frac{1}{2} \iint \log \rho_1 \frac{\partial^2 (g_{11} \rho)}{\partial x_1^2} dx_1 dx_2.$$

Integrate by parts again, and (47) follows. Q.E.D.

As before, the key part is the evaluation of $\frac{dH_{1\bar{2}}}{dt}$. The result is summarized in the following theorem:

Proposition 4.2. *For the system (2), the time rate of change of the marginal entropy of x_1 with x_2 frozen instantaneously is*

$$\frac{dH_{1\bar{2}}}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) - \frac{1}{2} E\left(g_{11} \frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) - \frac{1}{2} E\left(\frac{1}{\rho_1} \frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2}\right). \tag{49}$$

Proof.
Examine a small time interval $[t, t + \Delta t]$. We are going to prove the proposition by taking the limit:

$$\frac{dH_{1\bar{2}}}{dt} = \lim_{\Delta t \to 0} \frac{H_{1\bar{2}}(t + \Delta t) - H_1(t)}{\Delta t},$$

which boils down to the derivation of $H_{1\bar{2}}(t + \Delta t)$, namely the marginal entropy of x_1 at time $t + \Delta t$ as x_2 frozen as a parameter instantaneously at t. In principle this may be obtained using the strategy in the preceding subsection, but the evaluation of the convolution proves to be very difficult. To avoid the difficulty, Liang (2008) took a different approach, which we will follow hereafter.

In the stochastic system (2), the state $\underline{\mathbf{x}} = (x_1, x_2)^T$ is carried forth as time goes on. When time reaches t, freeze x_2 instantaneously and see how the state may evolves thenceforth until $t + \Delta t$. For convenience, denote by $x_{1\bar{2}}$ the first component of $\underline{\mathbf{x}}$ with x_2 frozen as a parameter. The system (2) is then modified to

$$dx_{1\bar{2}} = F_1(x_{1\bar{2}}, x_2, t) dt + \sum_k b_{1k} dw_k, \qquad \text{on } [t, t + \Delta t], \tag{50}$$

$$x_{1\bar{2}} = x_1 \qquad \text{at time } t. \tag{51}$$

Just as (45), correspondingly there is a modified Fokker-Planck equation for the density of $x_{1\bar{2}}$, written $\rho_{1\bar{2}}$:

$$\frac{\partial \rho_{1\bar{2}}}{\partial t} + \frac{\partial (F_1 \rho_{1\bar{2}})}{\partial x_1} = \frac{1}{2} \frac{\partial^2 (g_{11} \rho_{1\bar{2}})}{\partial x_1^2}, \qquad t \in [t, t + \Delta t] \tag{52}$$

$$\rho_{1\bar{2}} = \rho_1 \qquad \text{at } t. \tag{53}$$

Here g_{11} is the same as that defined in (46), i.e., $g_{11} = \sum_k b_{1k}^2$. Eq. (52) divided by $\rho_{1\check{2}}$ yields

$$\frac{\partial f_t}{\partial t} + \frac{1}{\rho_{1\check{2}}} \frac{\partial F_1 \rho_{1\check{2}}}{\partial x_1} = \frac{1}{\rho_{1\check{2}}} \frac{\partial^2 g_{11} \rho_{1\check{2}}}{\partial x_1^2},$$

where f_t is a function of x_1,

$$f_t(x_1) = \log \rho_{1\check{2}}(t, x_1). \tag{54}$$

We are doing this in the hope of obtaining an evolution law for $H_{1\check{2}}$, as by the definition of Shannon entropy we will just need to consider how the expectation of $-f_t(x_1)$ evolves. Discretizing,

$$f_{t+\Delta t}(x_1) = f_t(x_1) - \frac{\Delta t}{\rho_1} \frac{\partial (F_1 \rho_1)}{\partial x_1} + \frac{\Delta t}{2\rho_1} \frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2} + O(\Delta t^2),$$

where the fact $\rho_{1\check{2}} = \rho_1$ at time t has been used. For simplicity, the arguments have been suppressed for functions evaluated at $x_1(t)$, and this convention will be kept throughout this subsection. So

$$f_{t+\Delta t}(x_{1\check{2}}(t+\Delta t)) = f_t(x_{1\check{2}}(t+\Delta t)) - \frac{\Delta t}{\rho_1} \frac{\partial (F_1 \rho_1)}{\partial x_1} + \frac{\Delta t}{2\rho_1} \frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2} + O(\Delta t^2).$$

Using the Euler-Bernstein approximation (e.g., Lasota & Mackey, 1994) of (50), the $x_{1\check{2}}(t+\Delta t)$ in the argument of f_t on the right hand side can be expanded as

$$x_{1\check{2}}(t+\Delta t) = x_1(t) + F_1 \Delta t + \sum_k b_{1k} \Delta w_k + O(\Delta t^2).$$

And hence

$$
\begin{aligned}
& f_{t+\Delta t}(x_{1\check{2}}(t+\Delta t)) \\
&= f_t\left(x_1 + F_1 \Delta t + \sum_k b_{1k} \Delta w_k\right) - \frac{\Delta t}{\rho_1} \frac{\partial (F_1 \rho_1)}{\partial x_1} + \frac{\Delta t}{2\rho_1} \frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2} + O(\Delta t^2) \\
&= f_t(x_1) + \frac{\partial f_t}{\partial x_1}\left(F_1 \Delta t + \sum_k b_{1k} \Delta w_k\right) + \frac{1}{2} \frac{\partial^2 f_t}{\partial x_1^2}\left(F_1 \Delta t + \sum_k b_{1k} \Delta w_k\right)^2 \\
&\quad - \frac{\Delta t}{\rho_1} \frac{\partial (F_1 \rho_1)}{\partial x_1} + \frac{\Delta t}{2\rho_1} \frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2} + O(\Delta t^2),
\end{aligned}
\tag{55}
$$

where Taylor series expansion has been performed. Take expectations on both sides with respect to their respective random variables. Recalling how density evolution is defined, these expectations are equal (see Lasota & Mackey, 1994). Thus the left hand side results in $-H_{1\check{2}}(t+\Delta t)$, and the first term on the right hand side is $-H_1(t)$. Notice that for a Wiener process w_k, $\Delta w_k \sim N(0, \Delta t)$, that is to say,

$$E\Delta w_k = 0, \qquad E(\Delta w_k)^2 = \Delta t;$$

also notice that Δw_k are independent of (x_1, x_2). So

$$E\left(\frac{\partial f_1}{\partial x_1} \sum_k b_{1k} \Delta w_k\right) = E\left(\frac{\partial f_1}{\partial x_1}\right) \sum_k b_{1k} E\Delta w_k = 0.$$

Hence the second term on the right hand side is

$$\Delta t \cdot E\left(F_1 \frac{\partial f_t}{\partial x_1}\right).$$

For the same reason, the third term after expansion leaves only one sub-term of order Δt:

$$\frac{1}{2}E\left[\frac{\partial^2 f_t}{\partial x_1^2}\sum_k b_{1k}\Delta w_k \sum_j b_{1j}\Delta w_j\right]$$
$$= \frac{1}{2}E\left[\frac{\partial^2 f_t}{\partial x_1^2}\left(\sum_k b_{1k}^2(\Delta w_k)^2 + \sum_{k\neq j}b_{1k}b_{1j}\Delta w_k\Delta w_j\right)\right].$$

Using the independence between the perturbations, the summation over $k\neq j$ inside the parentheses must vanish upon applying expectation. The first summation is equal to $g_{11}\Delta t$, by the definition of g_{ij} and the fact $E(\Delta w_k)^2 = \Delta t$. So the whole term is

$$\frac{\Delta t}{2}E\left[g_{11}\frac{\partial^2 f_t}{\partial x_1^2}\right].$$

These, plus the fact that
$$f_t = \log \rho_{12}(t; x_1) = \log \rho_1,$$
all put together, (55) followed by an expectation on both sides yields

$$H_{12}(t+\Delta t) = H_1(t) - \Delta t \cdot E\left(F_1\frac{\partial \log \rho_1}{\partial x_1}\right) - \frac{\Delta t}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right)$$
$$+\Delta t \cdot E\left(\frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\right) - \frac{\Delta t}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2(g_{11}\rho_1)}{\partial x_1^2}\right) + O(\Delta t^2).$$

The second and fourth terms on the right hand side can be combined to give

$$\Delta t \cdot E\left(-F_1\frac{\partial \log \rho_1}{\partial x_1} + \frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\right) = \Delta t \cdot E\left(\frac{\partial F_1}{\partial x_1}\right).$$

So

$$\frac{dH_{12}}{dt} = \lim_{\Delta t \to 0}\frac{H_{12}(t+\Delta t) - H_1(t)}{\Delta t}$$
$$= E\left(\frac{\partial F_1}{\partial x_1}\right) - \frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) - \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2(g_{11}\rho_1)}{\partial x_1^2}\right).$$

Q.E.D.

With $\frac{dH_{12}}{dt}$ evaluated, now it is easy to obtain $T_{2\to 1}$, namely, the information flow from x_2 to x_1.

Theorem 4.1. *For the system (2), the time rate of information transferred from x_2 to x_1 is*

$$T_{2 \to 1} = -\mathcal{E}_{2|1} \left(\frac{\partial (F_1 \rho_1)}{\partial x_1} \right) + \frac{1}{2} \mathcal{E}_{2|1} \left(\frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2} \right), \tag{56}$$

where $\mathcal{E}_{2|1}$ is the integration operator defined in Theorem 3.4.

Proof.
Subtracting (49) from (47), one obtains

$$
\begin{aligned}
T_{2 \to 1} &= -E \left(F_1 \frac{\partial \log \rho_1}{\partial x_1} \right) - E \left(\frac{\partial F_1}{\partial x_1} \right) + \frac{1}{2} E \left(\frac{1}{\rho_1} \frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2} \right) \\
&= -E \left(\frac{1}{\rho_1} \frac{\partial (F_1 \rho_1)}{\partial x_1} \right) + \frac{1}{2} E \left(\frac{1}{\rho_1} \frac{\partial^2 (g_{11} \rho_1)}{\partial x_1^2} \right),
\end{aligned}
\tag{57}
$$

where E is the expectation with respect to $\rho(x_1, x_2)$. Notice that the conditional density of x_2 given x_1 is

$$\rho_{2|1}(x_2 | x_1) = \frac{\rho(x_1, x_2)}{\rho_1(x_1)}.$$

The operator

$$E \left(\frac{1}{\rho_1} \cdot \right) = \iint_{\Omega} \left(\frac{\rho}{\rho_1} \cdot \right) d\underline{\mathbf{x}}$$

is then simply the integration operator $\mathcal{E}_{2|1}$ as defined before in Theorem 3.4. The result thus follows.

Notice that in (56), the first term on the right hand side is precisely that in (20) i.e., the result of Liang & Kleeman (2005) based on intuitive argument for deterministic systems. This derivation supplies an alternative proof of the argument, and hence Theorem 3.4.

Above is the information flow from x_2 to x_1. Likewise, the flow from x_1 to x_2 can be derived. It is

$$T_{1 \to 2} = -\mathcal{E}_{1|2} \left(\frac{\partial (F_2 \rho_2)}{\partial x_2} \right) + \frac{1}{2} \mathcal{E}_{1|2} \left(\frac{\partial^2 (g_{22} \rho_2)}{\partial x_2^2} \right), \tag{58}$$

where $\rho_2 = \int \rho \, dx_1$ is the marginal density of x_2, and $\mathcal{E}_{1|2}$ is the operator such that, for any function $f \in L^1(\Omega)$, $\mathcal{E}_{1|2} f = \iint_{\Omega} \rho_{1|2}(x_1 | x_2) f(\mathbf{x}) \, d\underline{\mathbf{x}}$.

4.3 Properties
The above-derived information flow between system components possesses a very important property, namely the property of transfer directionality or asymmetry as emphasized in Schreiber (2000). One may have observed that the transfer in one direction need not imply anything about the transfer in the other direction, in contrast to the traditional correlation analysis or mutual information analysis. Particularly, in the extreme case that one component evolves independently from the other, the observation is concretized in the following theorem.

Theorem 4.2. (Causality)
If the evolution of x_1 is independent of x_2, then $T_{2 \to 1} = 0$.

Proof.

This property holds for formalisms with all the systems, but we here just prove with the continuous case. For the discrete system, the proof is lengthy, and the reader is referred to Liang & Kleeman (2007a) for details.

In (56), if $F_1 = F_1(x_1)$, and g_{11} is independent of x_2, integration can be taken for $\rho_{2|1}$ with respect to x_2 inside the double integrals, which gives

$$\int_{\Omega_2} \rho_{2|1}(x_2|x_1)\, dx_2 = 1.$$

The right hand side hence becomes

$$-\int_{\Omega_1} \frac{\partial (F_1 \rho_1)}{\partial x_1}\, dx_1 + \int_{\Omega_1} \frac{\partial (g_{11}\rho_1)}{\partial x_2}\, dx_1.$$

By the compact support assumption, these integrations both vanish, leaving a zero $T_{2\to 1}$.

Alternatively, if neither F_1 nor g_{11} has dependency on x_2, the integrals in (48) can be taken within the integrands, making ρ into ρ_1. This way the whole equation becomes a 1D Fokker-Planck equation for ρ_1, and hence x_1 is totally decoupled from the system, behaving like an independent variable. By intuition there should be no information flowing from x_2. Q.E.D.

This theorem shows that, between two evolving state variables x_1 and x_2, evaluation of $T_{2\to 1}$ and $T_{1\to 2}$ is able to tell which one causes which one and, in a quantitative way, tell how important one is to the other. Our information analysis thus gives a quantitative measure of the causality between two dynamical events. For this reason, this property is also referred to as the property of causality.

Another property holds only for the continuous system (2). Observe that the two terms of (56), the first is the same in form as that in (20), i.e., the corresponding deterministic system. Stochasticity contributes from the second term. An interesting observation is that:

Theorem 4.3. *Given a stochastic system component, if the stochastic perturbation is independent of another component, then the information transfer from the latter is the same in form as that for the corresponding deterministic system.*

Proof.

It suffices to consider only component x_1. If the stochastic perturbation $g_{11} = \sum_k b_{1k}^2$ is independent of x_2, then

$$\mathcal{E}_{2|1}\left(\frac{\partial^2 (g_{11}\rho_1)}{\partial x_1^2}\right) = \int \frac{\partial^2 (g_{11}\rho_1)}{\partial x_1^2}\, dx_1 = 0.$$

Here we have used the fact $\int \rho_{2|1} dx_2 = 1$. In this case, (56) and (20) have precisely the same form. Q.E.D.

This property is also very interesting since a great deal of noise in real systems appear to be additive; in other words, b_{ij}, and hence g_{ij}, are often constants. By the theorem these

stochastic systems thus function like deterministic in terms of information flow. Of course, the similarity is just in form; they are different in reality. The "deterministic" part of (56) (i.e., the first term) actually need not be deterministic, for stochasticity contributes to the state evolution and hence is embedded in the marginal density. As an illustration of the difference, the differential entropy for deterministic systems may go to minus infinity, e.g., in the case of the attractor of a fixed point or limit cycle, while this does not make an issue for stochastic systems (Ruelle, 1997).

5. Applications

The information flow formalism has been verified with benchmark problems, and applied to the study of several important dynamical system problems. Particularly, in Liang & Kleeman (2007a) we computed the transfers within a Hénon map, and obtained a result unique to our formalism just as one may expect on physical ground. In this section, we present two of these applications/verifications, echoing the challenges initially posed in the introduction.

5.1 Baker transformation

The baker transformation is a 2D mapping $\Phi : \Omega \to \Omega$, $\Omega = [0,1] \times [0,1]$, that mimics the kneading of a dough; it is given by

$$\Phi(x_1, x_2) = \begin{cases} (2x_1, \frac{x_2}{2}) & 0 \le x_1 \le \frac{1}{2}, \ 0 \le x_2 \le 1 \\ (2x_1 - 1, \frac{1}{2}x_2 + \frac{1}{2}) & \frac{1}{2} < x_1 \le 1, \ 0 \le x_2 \le 1 \end{cases}. \tag{59}$$

As introduced in the beginning, physicists have observed and intuitively argued that, upon applying the transformation, information flows continuingly from the stretching direction (here x_1) to the folding direction (x_2), while no transfer occurs the other way (see Lasota & Mackey, 1994). However, until Liang & Kleeman (2007a), this important physical phenomenon had not ever been quantitatively studied. In the following, we give a brief presentation of the Liang & Kleeman result.

To start, first look at the F-P operator. It is easy to check that the baker transformation is invertible, and measure preserving (the Jacobian $J = 1$), so by Eq. (14) its joint entropy stays unchanged. (But one of its components is not; see below.) The inverse map is given by

$$\Phi^{-1}(x_1, x_2) = \begin{cases} (\frac{x_1}{2}, 2x_2) & 0 \le x_2 \le \frac{1}{2}, \ 0 \le x_1 \le 1 \\ (\frac{x_1+1}{2}, 2x_2 - 1) & \frac{1}{2} \le x_2 \le 1, \ 0 \le x_1 \le 1 \end{cases} \tag{60}$$

Using Φ^{-1}, we can find the counterimage of $[0, x_1] \times [0, x_2]$ to be

1) $0 \le x_2 < \frac{1}{2}$,

$$\Phi^{-1}([0, x_1] \times [0, x_2]) = [0, \frac{x_1}{2}] \times [0, 2x_2]; \tag{61}$$

2) $\frac{1}{2} \le x_2 \le 1$,

$$\begin{aligned} \Phi^{-1}([0, x_1] \times [0, x_2]) &= \Phi^{-1}\left([0, x_1] \times [0, \frac{1}{2}]\right) \cup \Phi^{-1}\left([0, x_1] \times [\frac{1}{2}, x_2]\right) \\ &= [0, \frac{x_1}{2}] \times [0, 1] \cup [\frac{1}{2}, \frac{x_1+1}{2}] \times [0, 2x_2 - 1]. \end{aligned} \tag{62}$$

The F-P operator $\mathcal{P}$ is thus (cf. Lasota & Mackey, 1994)

$$\mathcal{P}\rho(x_1, x_2) = \frac{\partial^2}{\partial x_2 \partial x_1} \iint_{\Phi^{-1}([0,x_1]\times[0,x_2])} \rho(s,t) \, ds dt,$$

which, after a series of transformations, leads to

$$\mathcal{P}\rho(x_1, x_2) = \begin{cases} \rho(\frac{x_1}{2}, 2x_2), & 0 \le x_2 < \frac{1}{2}, \\ \rho(\frac{1+x_1}{2}, 2x_2 - 1), & \frac{1}{2} \le x_2 \le 1. \end{cases} \tag{63}$$

We now prove the following important result:

Theorem 5.1. *For the baker transformation (59),*

   (a) $T_{2\to1} = 0$,

   (b) $T_{1\to2} > 0$,

at any time steps.

Proof.
(a) With (63), we know that, upon one transformation, the marginal density of x_1 increases from

$$\rho_1 = \int_0^1 \rho(x_1, x_2) \, dx_2$$

to

$$\begin{aligned} \int_0^1 \mathcal{P}\rho(x_1, x_2) \, dx_2 &= \int_0^{1/2} \rho(\frac{x_1}{2}, 2x_2) \, dx_2 + \int_{1/2}^1 \rho(\frac{x_1+1}{2}, 2x_2 - 1) \, dx_2 \\ &= \frac{1}{2} \int_0^1 \left[\rho\left(\frac{x_1}{2}, x_2\right) + \rho\left(\frac{x_1+1}{2}, x_2\right) \right] \, dx_2 \\ &= \frac{1}{2} \left[\rho_1\left(\frac{x_1}{2}\right) + \rho_1\left(\frac{x_1+1}{2}\right) \right]. \end{aligned} \tag{64}$$

Note that the (59) as a whole is invertible. Its x_1 direction, however, is not. Consider x_1 only, the transformation reduces to a dyadic mapping, $\Phi_1 : [0,1] \to [0,1]$, $\Phi_1(x_1) = 2x_1 \pmod 1$. It is easy to obtain

$$\Phi_1^{-1}([0, x_1]) = [0, \frac{x_1}{2}] \cup [\frac{1}{2}, \frac{1+x_1}{2}]$$

for $x_1 < 1$. So it has an F-P operator

$$\begin{aligned} (\mathcal{P}\rho)_{12}(x_1) &= \frac{\partial}{\partial x_1} \int_{\Phi_1^{-1}([0,x_1])} \rho_1(s) \, ds \\ &= \frac{\partial}{\partial x_1} \int_0^{x_1/2} \rho_1(s) \, ds + \frac{\partial}{\partial x_1} \int_{1/2}^{(1+x_1)/2} \rho_1(s) \, ds \\ &= \frac{1}{2} \left[\rho_1\left(\frac{x_1}{2}\right) + \rho_1\left(\frac{1+x_1}{2}\right) \right]. \end{aligned}$$

This is exactly the same as (64), implying that

$$T_{2\to1} = 0, \tag{65}$$

which is just as expected.

(b) To compute the transfer in the opposite direction, first compute the marginal distribution

$$
\int_0^1 \mathcal{P}\rho(x_1, x_2)\, dx_1 = \begin{cases} \int_0^1 \rho\left(\frac{x_1}{2}, 2x_2\right)\, dx_1, & 0 \le x_2 < \frac{1}{2}; \\ \int_0^1 \rho\left(\frac{x_1+1}{2}, 2x_2 - 1\right)\, dx_1, & \frac{1}{2} \le x_2 \le 1. \end{cases} \tag{66}
$$

This substituted in

$$
\begin{aligned}
\Delta H_2 \;=\; & -\int_0^1 \int_0^1 \mathcal{P}\rho(x_1, x_2) \cdot \left[\log\left(\int_0^1 \mathcal{P}\rho(\lambda, x_2)d\lambda\right)\right] dx_1 dx_2 \\
& + \int_0^1 \int_0^1 \rho(x_1, x_2) \cdot \left[\log\left(\int_0^1 \rho(\lambda, x_2)d\lambda\right)\right] dx_1 dx_2,
\end{aligned} \tag{67}
$$

after a series of transformation of variables, gives

$$
\Delta H_2 = -\log 2 + (I + \mathbb{I}), \tag{68}
$$

where

$$
I \;=\; \int_0^1 \int_0^{1/2} \rho(x_1, x_2) \cdot \left[\log \frac{\int_0^1 \rho(\lambda, x_2)d\lambda}{\int_0^{1/2} \rho(\lambda, x_2)d\lambda}\right] dx_1 dx_2, \tag{69}
$$

$$
\mathbb{I} \;=\; \int_0^1 \int_{1/2}^1 \rho(x_1, x_2) \cdot \left[\log \frac{\int_0^1 \rho(\lambda, x_2)d\lambda}{\int_{1/2}^1 \rho(\lambda, x_2)d\lambda}\right] dx_1 dx_2. \tag{70}
$$

Note both I and $\mathbb{I}$ are nonnegative, because $\rho(x_1, x_2) \ge 0$ and

$$
\int_0^1 \rho(x_1, x_2)\, dx_1 \ge \int_0^{1/2} \rho(x_1, x_2)\, dx_1 \tag{71}
$$

$$
\int_0^1 \rho(x_1, x_2)\, dx_1 \ge \int_{1/2}^1 \rho(x_1, x_2)\, dx_1. \tag{72}
$$

Moreover, the two equalities cannot hold simultaneously, otherwise ρ will be zero, contradicting to the fact that it is a density distribution. So $I + \mathbb{I}$ is strictly positive.

On the other hand, in the folding or x_2 direction the transformation is always invertible, and the Jacobian $J_2 = \frac{1}{2}$. By Corollary 3.1,

$$
\Delta H_{2\backslash} = E \log \frac{1}{2} = -\log 2. \tag{73}
$$

So,

$$
T_{1 \to 2} = \Delta H_2 - \Delta H_{2\backslash} = I + \mathbb{I} > 0. \tag{74}
$$

Q.E.D.

In plain language, Eqs. (74) and (65) tell that there is always information flowing from x_1 or the stretching direction to x_2 or the folding direction ($T_{1 \to 2} > 0$), while no transfer occurs the other way ($T_{2 \to 1} = 0$). Illustrated in Fig. 1 is such a scenario, which has been intuitively argued in physics. Our formalism thus yields a result just as one may have expected on physical grounds.
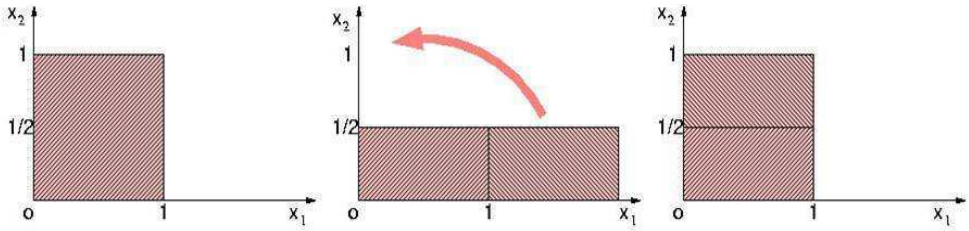
Fig. 1. Illustration of the baker transformation, and the associated information flow (middle) between the components.

5.2 Langevin equation

The formulas (56) and (58) with the stochastic system (2) are expected to be applicable in a wide variety of fields. To help further understand them, Liang (2008) examined a 2D linear system which hereafter we will be using:

$$d\underline{x} = \underline{\underline{A}}\,\underline{x}dt + \underline{\underline{B}}d\underline{w}, \tag{75}$$

where $\underline{w}$ is a Wiener process, and $\underline{\underline{A}} = (a_{ij})$ and $\underline{\underline{B}} = (b_{ij})$ are constant matrices. For convenience, suppose that initially $\underline{x}$ is Gaussian:

$$\underline{x} \sim N(\underline{\mu}, \underline{\underline{C}}).$$

Then it is Gaussian all the time because the system is linear (cf. Gardiner, 1985). Write the mean and covariance as

$$\underline{\mu}(t) = \left(\begin{array}{c} \mu_1(t) \\ \mu_2(t) \end{array} \right), \qquad \underline{\underline{C}}(t) = \left(\begin{array}{cc} c_{11}(t) & c_{12}(t) \\ c_{21}(t) & c_{22}(t) \end{array} \right).$$

It is easy to find the equations according to which they evolve:

$$\frac{d\underline{\mu}}{dt} = \underline{\underline{A}}\,\underline{\mu}, \tag{76a}$$

$$\frac{d\underline{\underline{C}}}{dt} = \underline{\underline{A}}\,\underline{\underline{C}} + \underline{\underline{C}}\,\underline{\underline{A}}^T + \underline{\underline{B}}\,\underline{\underline{B}}^T. \tag{76b}$$

($\underline{\underline{B}}\,\underline{\underline{B}}^T$ is the matrix (g_{ij}) we have seen before.) Solve them for $\underline{\mu}$ and $\underline{\underline{C}}$, and we obtain the probability density distribution at any time:

$$\rho(\underline{x}) = \frac{1}{2\pi(\det\underline{\underline{C}})^{1/2}}e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T\underline{\underline{C}}^{-1}(\underline{x}-\underline{\mu})}. \tag{77}$$

Substitute this into (56) and (58), and the transfers $T_{2\rightarrow1}$ and $T_{2\rightarrow2}$ are obtained accordingly.

As an example, let $\underline{\underline{A}} = \left[\begin{array}{cc} -0.5 & 0.1 \\ a_{21} & -0.5 \end{array} \right]$, $\underline{\underline{B}} = \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right]$. It is easy to show that both the eigenvalues of $\underline{\underline{A}}$ are negative; the system is hence stable and has an equilibrium solution:

$$\underline{\mu}(\infty) = \left(\begin{array}{c} 0 \\ 0 \end{array} \right), \qquad \underline{\underline{C}}(\infty) = \left(\begin{array}{cc} 2.44 & 2.20 \\ 2.20 & 2.00 \end{array} \right),$$
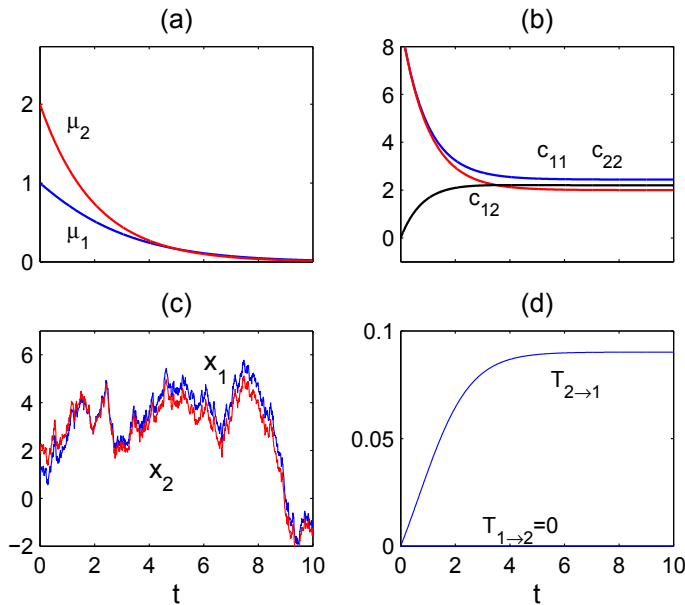
Fig. 2. A solution of (76) and the corresponding information transfers with the matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ and initial condition as shown in the text. (a) μ_1 and μ_2; (b) c_{11}, $c_{12} = c_{21}$, c_{22}; (c) a sample path starting from (1,2); (d) the computed information transfers $T_{2 \to 1}$ (upper) and $T_{1 \to 2} = 0$.

no matter how the system is initialized. Figs. 2a,b give the time evolutions of $\underline{\mu}$ and $\underline{\underline{C}}$ with initial conditions $\underline{\mu}(0) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, and $\underline{\underline{C}}(0) = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$; For reference, in Fig. 2c we also plot a sample path starting from $\underline{x}(0) = \underline{\mu}(0)$. Clearly, though initially x_1 (red line) and x_2 (blue line) they differ by a significant value, soon they begin to merge and thenafter almost follow the same path. To analyze the information transfer, observe that in this case the vector field component

$$F_2 = -0.5x_2,$$

has no dependence on x_1; furthermore,

$$g_{ij} = \sum_k b_{ik} b_{jk}$$

are all constants. So by Theorem 4.2, the information transferred from x_1 to x_2 should vanish at all times:

$$T_{1 \to 2} = 0.$$

This assertion is confirmed by the computed result. In Fig. 2d, $T_{1 \to 2}$ is zero through time. The other transfer, $T_{2 \to 1}$, increases monotonically and eventually approaches to a limit.

Comparing Figs. 2c and 2d one may have more to talk about. Obviously the typical sample paths of x_1 and x_2 in the former are highly correlated—-In fact they are almost the same. This

is in drastic contrast to the zero information flow from x_1 to x_2, namely $T_{1\rightarrow2}$, in the latter. The moral here is, even though $x_1(t)$ and $x_2(t)$ are highly correlated, the evolution of x_2 has nothing to do with x_1. To x_1, x_2 is causal, while to x_2, x_1 is not. Through this simple example one sees how information transfer extends the traditional notion of correlation analysis and/or mutual information analysis by including causality.

6. Summary

The past few years have seen a major advance in the formulation of information flow or information transfer, a fundamental general physics and dynamical system concept which has important applications in different disciplines. This advance, beginning with an elegant formula obtained by Liang & Kleeman (2005) for the law of entropy production

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

for system (10), has led to important scientific discoveries in the applied fields such as atmospheric science and oceanography. In this chapter, a concise introduction of the systematic research has been given within the framework of 2D dynamical systems. The resulting transfer is measured by the rate of entropy transferred from one component to another. The measure possesses a property of transfer asymmetry and, if the stochastic perturbation to the receiving component does not rely on the giving component, has a form same as that for the corresponding deterministic system. Explicit formulas, i.e., (56) and (58), have been obtained for generic stochastic systems (2), which we here write down again for easy reference:

$$T_{2\rightarrow1} = -E\left[\frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\right] + \frac{1}{2}E\left[\frac{1}{\rho_1}\frac{\partial^2(g_{11}\rho_1)}{\partial x_1^2}\right],$$

$$T_{1\rightarrow2} = -E\left[\frac{1}{\rho_2}\frac{\partial(F_2\rho_2)}{\partial x_2}\right] + \frac{1}{2}E\left[\frac{1}{\rho_2}\frac{\partial^2(g_{22}\rho_2)}{\partial x_2^2}\right],$$

where E stands for the mathematical expectation, and $g_{ij} = \sum_{k=1}^{2} b_{ik}b_{jk}$, $i = 1, 2$.

We have applied the results to examine the information flow within the baker transformation and a linear system. In the former, it is proved that there is always information flowing from the stretching direction to the folding direction, while no information is transferred the other way. In the latter, one sees that correlation does not necessarily mean causality; for two highly correlated time series, the one-way information transfer could be zero. Information flow analysis thus extends the traditional notion of correlation analysis with causality quantitatively represented, and this quantification is firmly based on a rigorous mathematical and physical footing.

7. References

[1] Abarbanel, H.D.I. (1996). *Analysis of Observed Chaotic Data*, Springer, New York.
[2] Andersland, M.S., and Teneketzis, D. (1992). Information structures, causality, and non-sequential stochastic control I: Design-independent properties, *SIAM J. Control and Optimization*, Vol. 30 (No. 6): 1447-1475.

[3] Arnhold, J., et al (1999). A robust method for detecting interdependences: application to intracranially recorded EEG, *Physica D*, Vol. 134: 419-430.

[4] Cove, T. M., & Thomas, J.A. (1991). *Elements of Information Theory*, Wiley, New York.

[5] Gardiner, C.W. (1985). *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Springer-Verlag, New York.

[6] Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, Vol. 37: 424-438.

[7] Kaiser, A. & Schreiber, T. (2002). Information transfer in continuous processes, *Physica D*, Vol. 166: 43-62.

[8] Kaneko, K. (1986). Lyapunov analysis and information flow in coupled map lattices, *Physica D*, Vol. 23: 436-447.

[9] Kantz, H. & Schreiber, Thomas (2004). *Nonlinear Time Series Analysis*, Cambridge University Press.

[10] Kleeman, Richard (2002). Measuring dynamical prediction utility using relative entropy, *J. Atmos. Sci.*, Vol. 59:2057-2072.

[11] Kleeman, Richard (2007). Information flow in ensemble weather predictions, *J. Atmos. Sci.*, Vol. 64(3): 1005-1016.

[12] Lasota, A. & Mackey, M.C. (1994). *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, Springer, New York.

[13] Liang, X. San (2008). Information flow within stochastic dynamical systems, *Phys. Rev. E*, Vol. 78: 031113.

[14] Liang, X. San. Local predictability and information flow in complex dynamical systems, *Physica D* (in press).

[15] Liang, X. San & Kleeman, Richard (2005). Information transfer between dynamical system components, *Phys. Rev. Lett.*, 95, (No. 24): 244101.

[16] Liang, X. San & Kleeman, Richard (2007a). A rigorous formalism of information transfer between dynamical system components. I. Discrete mapping *Physica D*, Vol 231: 1-9.

[17] Liang, X. San, & Kleeman, Richard (2007b). A rigorous formalism of information transfer between dynamical system components. II. Continuous flow, *Physica D*, Vol. 227: 173-182.

[18] Majda, A.J. & Harlim, J. (2007). Information flow between subspaces of complex dynamical systems, *Proc. Nat'l Acad. Sci.*, Vol. 104: 9558-9563.

[19] Pereda, E., Quiroga, R.Q. & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals, *Prog. Neurobiol.*, Vol. 77(1-2): 1-37.

[20] Rosenblum, M.G., Pikovsky, A.S. & Kurths, J. (1996). Phase synchronization of chaotic oscillators, *Phys. Rev. Lett.*, Vol. 76:1804-1807.

[21] Ruelle, D. (1997). Positivity of entropy production in the presence of a random thermostat, *J. Stat. Phys.*, Vol. 86(Nos. 5-6):935-951.

[22] Schreiber, Thomas (2000). Measuring information transfer, *Phys. Rev. Lett.*, Vol. 85(2):461.

[23] Tribbia, J.J. (2005). Waves, information and local predictability, *Workshop on Mathematical Issues and Challenges in Data Assimilation for Geophysical Systems: Interdisciplinary perspectives*, IPAM, UCLA, February 22-25, 2005.

[24] Vastano, J.A. & Swinney, H.L. (1988). Information transport in spatiotemporal systems, *Phys. Rev. Lett.*, Vol. 60:1773-1776.

Stochastic Control

Edited by Chris Myers

ISBN 978-953-307-121-3

Hard cover, 650 pages

Publisher Sciyo

Published online 17, August, 2010

Published in print edition August, 2010

Uncertainty presents significant challenges in the reasoning about and controlling of complex dynamical systems. To address this challenge, numerous researchers are developing improved methods for stochastic analysis. This book presents a diverse collection of some of the latest research in this important area. In particular, this book gives an overview of some of the theoretical methods and tools for stochastic analysis, and it presents the applications of these methods to problems in systems theory, science, and economics.

INTECH
open science | open minds