Entropy **2013**, *15*, 327-360; doi:10.3390/e15010327

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Review

The Liang-Kleeman Information Flow: Theory and Applications

X. San Liang [1,2]

[1] School of Marine Sciences and School of Mathematics and Statistics, Nanjing University of Information Science and Technology (Nanjing Institute of Meteorology), 219 Ningliu Blvd, Nanjing 210044, China; E-Mail: sanliang@courant.nyu.edu; Tel.: +86-25-5869-5695; Fax: +86-25-5869-5698

[2] China Institute for Advanced Study, Central University of Finance and Economics, 39 South College Ave, Beijing 100081, China

Received: 17 October 2012; in revised form: 22 November 2012 / Accepted: 28 December 2012 / Published: 18 January 2013

Abstract: Information flow, or information transfer as it may be referred to, is a fundamental notion in general physics which has wide applications in scientific disciplines. Recently, a rigorous formalism has been established with respect to both deterministic and stochastic systems, with flow measures explicitly obtained. These measures possess some important properties, among which is flow or transfer asymmetry. The formalism has been validated and put to application with a variety of benchmark systems, such as the baker transformation, Hénon map, truncated Burgers-Hopf system, Langevin equation, *etc*. In the chaotic Burgers-Hopf system, all the transfers, save for one, are essentially zero, indicating that the processes underlying a dynamical phenomenon, albeit complex, could be simple. (Truth is simple.) In the Langevin equation case, it is found that there could be no information flowing from one certain time series to another series, though the two are highly correlated. Information flow/transfer provides a potential measure of the cause–effect relation between dynamical events, a relation usually hidden behind the correlation in a traditional sense.

Keywords: Liang-Kleeman information flow; causation; emergence; Frobenius-Perron operator; time series analysis; atmosphere-ocean science; El Niño; neuroscience; network dynamics; financial economics

---

1. Introduction

Information flow, or information transfer as it sometimes appears in the literature, refers to the transference of information between two entities in a dynamical system through some processes, with one entity being the source, and another the receiver. Its importance lies beyond its literal meaning in that it actually carries an implication of causation, uncertainty propagation, predictability transfer, *etc*., and, therefore, has applications in a wide variety of disciplines. In the following, we first give a brief demonstration of how it may be applied in different disciplines; the reader may skip this part and go directly to the last two paragraphs of this section.

According to how the source and receiver are chosen, information flow may appear in two types of form. The first is what one would envision in the usual sense, *i.e*., the transference between two parallel parties (for example, two chaotic circuits [1]), which are linked through some mechanism within a system. This is found in neuroscience (e.g., [2–4]), network dynamics (e.g., [5–7]), atmosphere–ocean science (e.g., [8–11]), financial economics (e.g., [12,13]), to name but a few. For instance, neuroscientists focus their studies on the brain and its impact on behavior and cognitive functions, which are associated with flows of information within the nervous system (e.g., [3]). This includes how information flows from one neuron to another neuron across the synapse, how dendrites bring information to the cell body, how axons take information away from the cell body, and so forth. Similar issues arise in computer and social networks, where the node–node interconnection, causal dependencies, and directedness of information flow, among others, are of concern [6,14,15]. In atmosphere–ocean science, the application is vast, albeit newly begun. An example is provided by the extensively studied El Niño phenomenon in the Pacific Ocean, which is well known through its linkage to global natural disasters, such as the floods in Ecuador and the droughts in Southeast Asia, southern Africa and northern Australia, to the death of birds and dolphins in Peru, to the increased number of storms over the Pacific, and to the famine and epidemic diseases in far-flung parts of the world [16–18]. A major focus in El Niño research is the predictability of the onset of the irregularly occurring event, in order to issue in-advance warning of potential hazardous impacts [19–21]. It has now become known that the variabilities in the Indian Ocean could affect the El Niño predictability (e.g., [22]). That is to say, at least a part of the uncertainty source for El Niño predictions is from the Indian Ocean. Therefore, to some extent, the El Niño predictability may also be posed as an information flow problem, *i.e*., a problem on how information flows from the Indian Ocean to the Pacific Ocean to make the El Niño more predictable or more uncertain.

Financial economics provides another field of application of information flow of the first type; this field has received enormous public attention since the recent global financial crisis triggered by the subprime mortgage meltdown. A conspicuous example is the cause–effect relation between the equity and options markets, which reflects the preference of traders in deciding where to place their trades. Usually, information is believed to flow unidirectionally from equity to options markets because informed traders prefer to trade in the options markets (e.g., [23]), but recent studies show that the flow may also exist in the opposite way: informed traders actually trade both stocks and "out-of-the-money" options, and hence the causal relation from stocks to options may reverse [12]. More (and perhaps the most important) applications are seen through predictability studies. For instance, the predictability of asset return characteristics is a continuing problem in financial economics, which is largely due to the

information flow in markets. Understanding the information flow helps to assess the relative impact from the markets and the diffusive innovation on financial management. Particularly, it helps the prediction of jump timing, a fundamental question in financial decision making, through determining information covariates that affect jump occurrence up to the intraday levels, hence providing empirical evidence in the equity markets, and pointing us to an efficient financial management [13].

The second type of information flow appears in a more abstract way. In this case, we have one dynamical event; the transference occurs between different levels, or sometimes scales, within the same event. Examples for this type are found in disciplines such as evolutionary biology [24–26], statistical physics [27,28], turbulence, *etc*., and are also seen in network dynamics. Consider the transitions in biological complexity. A reductionist, for example, views that the emergence of new, higher level entities can be traced back to lower level entities, and hence there is a "bottom-up" causation, *i.e*., an information flow from the lower levels to higher levels. Bottom-up causation lays the theoretical foundation for statistical mechanics, which explains macroscopic thermodynamic states from a point of view of molecular motions. On the other hand, "top-down" causation is also important [29,30]. In evolution (e.g., [31]), it has been shown that higher level processes may constrain and influence what happens at lower levels; particularly, in transiting complexity, there is a transition of information flow, from the bottom-up to top-down, leading to a radical change in the structure of causation (see, for example [32]). Similar to evolutionary biology, in network dynamics, some simple computer networks may experience a transition from a low traffic state to a high congestion state, beneath which is a flow of information from a bunch of almost independent entities to a collective pattern representing a higher level of organization (e.g., [33]). In the study of turbulence, the notoriously challenging problem in classical physics, it is of much interest to know how information flows over the spectrum to form patterns on different scales. This may help to better explain the cause of the observed higher moments of the statistics, such as excess kurtosis and skewness, of velocity components and velocity derivatives [34]. Generally, the flows/transfers are two-way, *i.e*., both from small scales to large scales, and from large scales to small scales, but the flow or transfer rates may be quite different.

Apart from the diverse real-world applications, information flow/transfer is important in that it offers a methodology for scientific research. In particular, it offers a new way of time series analysis [35–37]. Traditionally, correlation analysis is widely used for identifying the relation between two events represented by time series of measurements; an alternative approach is through mutual information analysis, which may be viewed as a type of nonlinear correlation analysis. But both correlation analysis and mutual information analysis put the two events on an equal stance. As a result, there is no way to pick out the cause and the effect. In econometrics, Granger causality [38] is usually employed to characterize the causal relation between time series, but the characterization is just in a qualitative sense; when two events are mutually causal, it is difficult to differentiate their relative strengths. The concept of information flow/transfer is expected to remedy this deficiency, with the mutual causal relation quantitatively expressed.

Causality implies directionality. Perhaps the most conspicuous observation on information flow/transfer is its asymmetry between the involved parties. A typical example is seen in our daily life when a baker is kneading a dough. As the baker stretches, cuts, and folds, he guides a unilateral flow of information from the horizontal to the vertical. That is to say, information goes only from the stretching

direction to the folding direction, not *vice versa*. The one-way information flow (in a conventional point of view) between the equity and options markets offers another good example. In other cases, such as in the aforementioned El Niño event, though the Indian and Pacific Oceans may interact with each other, *i.e.*, the flow route could be a two-way street, the flow rate generally differs from one direction to another direction. For all that account, transfer asymmetry makes a basic property of information flow; it is this property that distinguishes information flow from the traditional concepts such as mutual information.

As an aside, one should not confuse dynamics with causality, the important property reflected in the asymmetry of information flow. It is temptating to think that, for a system, when the dynamics are known, the causal relations are determined. While this might be the case for linear deterministic systems, in general, however, this need not be true. Nonlinearity may lead a deterministic system to chaos; the future may not be predictable after a certain period of time, even though the dynamics is explicitly given. The concept of emergence in complex systems offers another example. It has long been found that irregular motions according to some simple rules may result in the emergence of regular patterns (such as the inverse cascade in the planar turbulence in natural world [39,40]). Obviously, how this instantaneous flow of information from the low-level entities to high-level entities, *i.e.*, the patterns, cannot be simply explained by the rudimentary rules set *a priori*. In the language of complexity, emergence does not result from rules only (e.g., [41–43]); rather, as said by Corning (2002) [44], "Rules, or laws, have no causal efficacy; they do not in fact 'generate' anything... the underlying causal agencies must be separately specified."

Historically, quantification of information flow has been an enduring problem. The challenge lies in that this is a real physical notion, while the physical foundation is not as clear as those well-known physical laws. During the past decades, formalisms have been established empirically or half-empirically based on observations in the aforementioned diverse disciplines, among which are Vastano and Swinney's time-delayed mutual information [45], and Schreiber's transfer entropy [46,47]. Particularly, transfer entropy is established with an emphasis of the above transfer asymmetry between the source and receiver, so as to have the causal relation represented; it has been successfully applied in many real problem studies. These formalisms, when carefully analyzed, can be approximately understood as dealing with the change of marginal entropy in the Shannon sense, and how this change may be altered in the presence of information flow (see [48], section 4 for a detailed analysis). This motivates us to think about the possibility of a rigorous formalism when the dynamics of the system is known. As such, the underlying evolution of the joint probability density function (pdf) will also be given, for deterministic systems, by the Liouville equation or, for stochastic systems, by the Fokker-Planck equation (cf. §4 and §5 below). From the joint pdf, it is easy to obtain the marginal density, and hence the marginal entropy. One thus expects that the concept of information flow/transfer may be built on a rigorous footing when the dynamics are known, as is the case with many real world problems like those in atmosphere–ocean science. And, indeed, Liang and Kleeman (2005) [49] find that, for two-dimensional (2D) systems, there is a concise law on entropy evolution that makes the hypothesis come true. Since then, the formalism has been extended to systems in different forms and of arbitrary dimensionality, and has been applied with success in benchmark dynamical systems and more realistic problems. In the following sections, we will give a systematic introduction of the theories and a brief review of some of the important applications.

In the rest of this review, we first set up a theoretical framework, then illustrate through a simple case how a rigorous formalism can be achieved. Specifically, our goal is to compute within the framework, for a continuous-time system, the transference rate of information, and, for a discrete-time system or mapping, the amount of the transference upon each application of the mapping. To unify the terminology, we may simply use "information flow/transfer" to indicate either the "rate of information flow/transfer" or the "amount of information flow/transfer" wherever no ambiguity exists in the context. The next three sections are devoted to the derivations of the transference formulas for three different systems. Sections 3 and 4 are for deterministic systems, with randomness limited within initial conditions, where the former deals with discrete mappings and the latter with continuous flows. Section 5 discusses the case when stochasticity is taken in account. In the section that follows, four major applications are briefly reviewed. While these applications are important *per se*, some of them also provide validations for the formalism. Besides, they are also typical in terms of computation; different approaches (both analytical and computational) have been employed in computing the flow or transfer rates for these systems. We summarize in Section 7 the major results regarding the formulas and their corresponding properties, and give a brief discussion on the future research along this line. As a convention in the history of development, the terms "information flow" and "information transfer" will be used synonymously. Throughout this review, by entropy we always mean Shannon or absolute entropy, unless otherwise specified. Whenever a theorem is stated, generally only the result is given and interpreted; for detailed proofs, the reader is referred to the original papers.

2. Mathematical Formalism

2.1. Theoretical Framework

Consider a system with n state variables, x_1, x_2, ..., x_n, which we put together as a column vector $\boldsymbol{x} = (x_1, ..., x_n)^T$. Throughout this paper, $\boldsymbol{x}$ may be either deterministic or random, depending on the context where it appears. This is a notational convention adopted in the physics literature, where random and deterministic states for the same variable are not distinguished. (In probability theory, they are usually distinguished with lower and upper cases like $\boldsymbol{x}$ and $\boldsymbol{X}$.) Consider a sample space of $\boldsymbol{x}$, $\Omega \subset \mathbb{R}^n$. Defined on Ω is a joint probability density function (pdf) $\rho = \rho(\boldsymbol{x})$. For convenience, assume that ρ and its derivatives (up to an order as high as enough) are compactly supported. This makes sense, as in the real physical world, the probability of extreme events vanishes. Thus, without loss of generality, we may extend Ω to $\mathbb{R}^n$ and consider the problem on $\mathbb{R}^n$, giving a joint density in $L^1(\mathbb{R}^n)$ and n marginal densities $\rho_i \in L^1(\mathbb{R})$:

$$\rho_i(x_i) \quad = \quad \int_{\mathbb{R}^{n-1}} \rho(x_1, x_2, ..., x_n)\, dx_1 ... dx_{i-1} dx_{i+1} ... dx_n, \qquad i = 1, ...n$$

Correspondingly, we have an entropy functional of ρ (joint entropy) in the Shannon sense

$$H = -\int_{\mathbb{R}^n} \rho(\boldsymbol{x}) \log \rho(\boldsymbol{x})\, d\boldsymbol{x} \tag{1}$$

and n marginal entropies

$$H_i \quad = \quad -\int_{\mathbb{R}} \rho(x_i) \log \rho(x_i)\, dx_i, \qquad i = 1, ..., n \tag{2}$$

Consider an n-dimensional dynamical system, autonomous or nonautonomous,

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}, t) \tag{3}$$

where $\boldsymbol{F} = (F_1, F_2, ..., F_n)^T$ is the vector field. With random inputs at the initial stage, the system generates a continuous stochastic process $\{\boldsymbol{x}(t), t \geq 0\}$, which is what we are concerned with. In many cases, the process may not be continuous in time (such as that generated by the baker transformation, as mentioned in the introduction). We thence also need to consider a system in the discrete mapping form:

$$\boldsymbol{x}(\tau + 1) = \Phi(\boldsymbol{x}(\tau)) \tag{4}$$

with τ being positiver integers. Here Φ is an n-dimensional transformation

$$\Phi : \mathbb{R}^n \to \mathbb{R}^n, \qquad (x_1, x_2, ..., x_n) \mapsto (\Phi_1(\boldsymbol{x}), \Phi_2(\boldsymbol{x}), ..., \Phi_n(\boldsymbol{x})) \tag{5}$$

the counterpart of the vector field $\boldsymbol{F}$. Again, the system is assumed to be perfect, with randomness limited within the initial conditions. Cases with stochasticity due to model inaccuracies are deferred to Section 5. The stochastic process thus formed is in a discrete time form $\{\boldsymbol{x}(\tau), \tau\}$, with $\tau > 0$ signifying the time steps. Our formalism will be established henceforth within these frameworks.

2.2. Toward a Rigorous Formalism—A Heuristic Argument

First, let us look at the two-dimensional (2D) case originally studied by Liang and Kleeman [49]

$$\frac{dx_1}{dt} = F_1(x_1, x_2, t) \tag{6}$$

$$\frac{dx_2}{dt} = F_2(x_1, x_2, t) \tag{7}$$

This is a system of minimal dimensionality that admits information flow. Without loss of generality, examine only the flow/transfer from x_2 to x_1.

Under the vector field $\boldsymbol{F} = (F_1, F_2)^T$ $\boldsymbol{x}$ evolves with time; correspondingly its joint pdf $\rho(\boldsymbol{x})$ evolves, observing a Liouville equation [50]:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_1}(F_1 \rho) + \frac{\partial}{\partial x_2}(F_2 \rho) = 0 \tag{8}$$

As argued in the introduction, what matters here is the evolution of H_1 namely the marginal entropy of x_1. For this purpose, integrate (8) with respect to x_2 over $\mathbb{R}$ to get:

$$\frac{\partial \rho_1}{\partial t} + \frac{\partial}{\partial x_1} \int_{\mathbb{R}} F_1 \rho \, dx_2 = 0 \tag{9}$$

Other terms vanish, thanks to the compact support assumption for ρ. Multiplication of (9) by $-(1 + \log \rho_1)$ followed by an integration over $\mathbb{R}$ gives the tendency of H_1:

$$\frac{dH_1}{dt} = \int_{\mathbb{R}^2} \left[\log \rho_1 \frac{\partial(\rho F_1)}{\partial x_1} \right] dx_1 dx_2 = -E\left(\frac{F_1}{\rho_1} \frac{\partial \rho_1}{\partial x_1} \right) \tag{10}$$

where E stands for mathematical expectation with respect to ρ. In the derivation, integration by parts has been used, as well as the compact support assumption.

Now what is the rate of information flow from x_2 to x_1? In [49], Liang and Kleeman argue that, as the system steers a state forward, the marginal entropy of x_1 is replenished from two different sources: one is from x_1 itself, another from x_2. The latter is through the very mechanism namely information flow/transfer. If we write the former as dH_1^*/dt, and denote by $T_{2\to 1}$ the rate of information flow/transfer from x_2 to x_1 (T stands for "transfer"), this gives a decomposition of the marginal entropy increase according to the underlying mechanisms:

$$\frac{dH_1}{dt} = \frac{dH_1^*}{dt} + T_{2\to 1} \tag{11}$$

Here dH_1/dt is known from Equation (10). To find $T_{2\to 1}$, one may look for dH_1^*/dt instead. In [49], Liang and Kleeman find that this is indeed possible, based on a heuristic argument. To see this, multiply the Liouville Equation (8) by $-(1+\log\rho)$, then integrate over $\mathbb{R}^2$. This yields an equation governing the evolution of the joint entropy H which, after a series of manipulation, is reduced to

$$\frac{dH}{dt} = \int_{\mathbb{R}^2} \nabla\cdot(\rho\log\rho\boldsymbol{F})dx_1dx_2 + \int_{\mathbb{R}^2} \rho\nabla\cdot\boldsymbol{F}dx_1dx_2$$

where ∇ is the divergence operator. With the assumption of compact support, the first term on the right hand side goes to zero. Using E to indicate the operator of mathematical expectation, this becomes

$$\frac{dH}{dt} = E\left(\nabla\cdot\boldsymbol{F}\right) \tag{12}$$

That is to say, the time rate of change of H is precisely equal to the mathematical expectation of the divergence of the vector field. This remarkably concise result tells that, as a system moves on, the change of its joint entropy is totally controlled by the contraction or expansion of the phase space of the system. Later on, Liang and Kleeman show that this is actually a property holding for deterministic systems of arbitrary dimensionality, even without invoking the compact assumption [51]. Moreover, it has also been shown that, the local marginal entropy production observes a law in the similar form, if no remote effect is taken in account [52].

With Equation (12), Liang and Kleeman argue that, apart from the complicated relations, the rate of change of the marginal entropy H_1 due to x_1 only (*i.e.*, dH_1^*/dt as symbolized above), should be

$$\frac{dH_1^*}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) = \int_{\mathbb{R}^2} \rho\frac{\partial F_1}{\partial x_1}\,dx_1dx_2 \tag{13}$$

This heuristic reasoning makes the separation (11) possible. Hence the information flows from x_2 to x_1 at a rate of

$$\begin{aligned} T_{2\to 1} &= \frac{dH_1}{dt} - \frac{dH_1^*}{dt} = -E\left(\frac{F_1}{\rho_1}\frac{\partial\rho_1}{\partial x_1}\right) - E\left(\frac{\partial F_1}{\partial x_1}\right) \\ &= -E\left[\frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\right] \\ &= -\int_{\mathbb{R}^2} \rho_{2|1}(x_2|x_1)\frac{\partial(F_1\rho_1)}{\partial x_1}dx_1dx_2 \end{aligned} \tag{14}$$

where $\rho_{2|1}$ is the conditional pdf of x_2, given x_1. The rate of information flow from x_1 to x_2, written $T_{1\to 2}$, can be derived in the same way. This tight formalism (called "LK2005 formalism" henceforth), albeit based on heuristic reasoning, turns out to be very successful. The same strategy has been applied again in a similar study by Majda and Harlim [53]. We will have a chance to see these in Sections 4 and 6.

2.3. Mathematical Formalism

The success of the LK2005 formalism is remarkable. However, its utility is limited to systems of dimensionality 2. For an n-dimensional system with $n > 2$, the so-obtained Equation (14) is not the transfer from x_2 to x_1, but the cumulant transfer to x_1 from all other components x_2, x_3,..., x_n. Unless one can screen out from Equation (14) the part contributed from x_2, it seems that the formalism does not yield the desiderata for high-dimensional systems.

To overcome the difficulty, Liang and Kleeman [48,51] observe that, the key part in Equation (14) namely dH_1^*/dt actually can be alternatively interpreted, for a 2D system, as the evolution of H_1 with the effect of x_2 excluded. In other words, it is the tendency of H_1 with x_2 frozen instantaneously at time t. To avoid confusing with dH_1^*/dt, denote it as $dH_{1\not{2}}/dt$, with the subscript $\not{2}$ signifying that the effect of x_2 is removed. In this way dH_1/dt is decomposed into two disjoint parts: $T_{2\to 1}$ namely the rate of information flow and $dH_{1\not{2}}/dt$. The flow is then the difference between dH_1/dt and $dH_{1\not{2}}/dt$:

$$T_{2\to 1} = \frac{dH_1}{dt} - \frac{dH_{1\not{2}}}{dt} \tag{15}$$

For 2D systems, this is just a restatement of Equation (14) in another set of symbols; but for systems with dimensionality higher than 2, they are quite different. Since the above partitioning does not have any restraints on n, Equation (15) is applicable to systems of arbitrary dimensionality.

In the same spirit, we can formulate the information transfer for discrete systems in the form of Equation (4). As $\boldsymbol{x}$ is mapped forth under the transformation Φ from time step τ to $\tau+1$, correspondingly its density ρ is steered forward by an operator termed after Georg Frobenius and Oskar Perron, which we will introduce later. Accordingly the entropies H, H_1, and H_2 also change with time. On the interval $[\tau, \tau+1]$, let H_1 be incremented by ΔH_1 from τ to $\tau+1$. By the foregoing argument, the evolution of H_1 can be decomposed into two exclusive parts according to their driving mechanisms, *i.e.*, the information flow from x_2, $T_{2\to 1}$, and the evolution with the effect of x_2 excluded, written as $\Delta H_{1\not{2}}$. We therefore obtain the discrete counterpart of Equation (15):

$$T_{2\to 1} = \Delta H_1 - \Delta H_{1\not{2}} \tag{16}$$

Equations (15) and (16) give the rates of information flow/transfer from component x_2 to component x_1 for systems (3) and (4), respectively. One may switch the corresponding indices to obtain the flow between any component pair x_i and x_j, $i \neq j$. In the following two sections we will be exploring how these equations are evaluated.

3. Discrete Systems

3.1. Frobenius-Perron Operator

For discrete systems in the form of Equation (4), as $\boldsymbol{x}$ is carried forth under the transformation Φ, there is another transformation, called Frobenius–Perron operator $\mathcal{P}$ (F-P operator hereafter), steering $\rho(\boldsymbol{x})$, *i.e.*, the pdf of $\boldsymbol{x}$, to $\mathcal{P}\rho$ (see a schematic in Figure 1). The F-P operator governs the evolution of the density of $\boldsymbol{x}$.

A rigorous definition requires some ingredients of measure theory which is beyond the scope this review, and the reader may consult with the reference [50]. Loosely speaking, given a transformation $\Phi : \Omega \to \Omega$ (in this review, $\Omega = \mathbb{R}^n$), $\boldsymbol{x} \mapsto \Phi\boldsymbol{x}$, it is a mapping $\mathcal{P} : L^1(\Omega^n) \to L^1(\Omega^n)$, $\rho \mapsto \mathcal{P}\rho$, such that

$$\int_{\omega} \mathcal{P}\rho(\boldsymbol{x})d\boldsymbol{x} = \int_{\Phi^{-1}(\omega)} \rho(\boldsymbol{x})d\boldsymbol{x} \tag{17}$$

for any $\omega \subset \Omega$. If Φ is nonsingular and invertible, it actually can be explicitly evaluated. Making transformation $\boldsymbol{y} = \Phi(\boldsymbol{x})$, the right hand side is, in this case,

$$\int_{\Phi^{-1}(\omega)} \rho(\boldsymbol{x})\ d\boldsymbol{x} = \int_{\omega} \rho\left[\Phi^{-1}(\boldsymbol{y})\right] \cdot \left|J^{-1}\right|\ d\boldsymbol{y}$$

where J is the Jacobian of Φ:

$$J = \det\left[\frac{\partial(y_1, y_2, ..., y_n)}{\partial(x_1, x_2, ..., x_n)}\right]$$

and J^{-1} its inverse. Since ω is arbitrarily chosen, we have

$$\mathcal{P}\rho(\boldsymbol{x}) = \rho\left[\Phi^{-1}(\boldsymbol{x})\right] \cdot \left|J^{-1}\right| \tag{18}$$

If no nonsingularity is assumed for the transformation Φ, but the sample space Ω is in a Cartesian product form, as is for this review, the F-P operator can also be evaluated, though not in an explicit form. Consider a domain

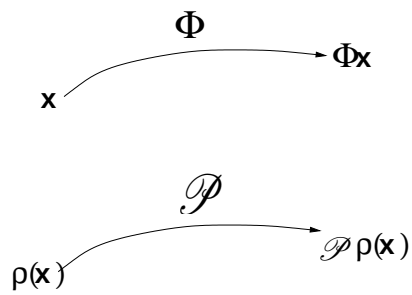
$$\omega = [a_1, x_1] \times [a_2, x_2] \times ... \times [a_n, x_n]$$

where $\boldsymbol{a} = (a_1, ..., a_n)$ is some constant point (usually can be set to be the origin). Let the counterimage of ω be $\Phi^{-1}(\omega)$, then it has been proved (c.f. [50]) that

$$\mathcal{P}\rho(\boldsymbol{x}) = \frac{\partial^n}{\partial x_n ... \partial x_2 \partial x_1} \int_{\Phi^{-1}(\omega)} \rho(\xi_1, \xi_2, ..., \xi_n)\ d\xi_1 d\xi_2 ... d\xi_n$$

In this review, we consider a sample space $\mathbb{R}^n$, so essentially all the F-P operators can be calculated this way.

Figure 1. Illustration of the Frobenius-Perron operator $\mathcal{P}$, which takes $\rho(\boldsymbol{x})$ to $\mathcal{P}\rho(\boldsymbol{x})$ as Φ takes $\boldsymbol{x}$ to $\Phi\boldsymbol{x}$.



3.2. Information Flow

The F-P operator $\mathcal{P}$ allows for an evaluation of the change of entropy as the system evolves forth. By the formalism (16) , we need to examine how the marginal entropy changes on a time interval $[\tau,\ \tau+1]$. Without loss of generality, consider only the flow from x_2 to x_1. First look at increase of H_1. Let ρ be the joint density at step τ, then the joint density at step $\tau+1$ is $\mathcal{P}\rho$, and hence

$$\begin{aligned}\Delta H_1 &= H_1(\tau+1) - H_1(\tau) \\ &= -\int_{\mathbb{R}} (\mathcal{P}\rho)_1(y_1)\cdot\log(\mathcal{P}\rho)_1(y_1)\,dy_1 + \int_{\mathbb{R}} \rho_1(x_1)\cdot\log\rho_1(x_1)\,dx_1 \end{aligned} \tag{19}$$

Here $(\mathcal{P}\rho)_1$ means the marginal density of x_1 at $\tau+1$; it is equal to $\mathcal{P}\rho$ with all components of $\boldsymbol{x}$ but x_1 being integrated out. The independent variables with respect to which the integrations are taken are dummy; but for the sake of clarity, we use different notations, *i.e.*, x and y, for them at time step τ and $\tau+1$, respectively.

The key to the formalism (16) is the finding of

$$\Delta H_{1\cancel{2}} = H_{1\cancel{2}}(\tau+1) - H_1(\tau) \tag{20}$$

namely the increment of the marginal entropy of x_1 on $[\tau,\tau+1]$ with the contribution from x_2 excluded. Here the system in question is no longer Equation (4), but a system with a mapping modified from Φ:

$$\Phi_{\cancel{2}}: \begin{cases} y_1 = & \Phi_1(x_1,x_2,x_3,...,x_n) \\ y_3 = & \Phi_3(x_1,x_2,x_3,...,x_n) \\ \vdots & \qquad\vdots \\ y_n = & \Phi_n(x_1,x_2,x_3,...,x_n) \end{cases} \tag{21}$$

with x_2 frozen instantaneously at τ as a parameter. Again, we use $x_i = x_i(\tau)$, $y_i = \Phi(\boldsymbol{x}(\tau)) = x_i(\tau+1)$, $i = 1,...,n$, to indicate the state variables at steps τ and $\tau+1$, respectively, to avoid any possible confusion. In the mean time, the dependence on τ and $\tau+1$ are suppressed for notational economy. Corresponding to the modified transformation $\Phi_{\cancel{2}}$ is a modified F-P operator, written $\mathcal{P}_{\cancel{2}}$. To find $H_{1\cancel{2}}(\tau+1)$, examine the quantity $h = -\log(\mathcal{P}_{\cancel{2}}\rho)_1(y_1)$, where the subscript 1 indicates that this is a marginal density of the first component, and the dependence on y_1 tells that this is evaluated at step $\tau+1$. Recall how Shannon entropy is defined: $H_{1\cancel{2}}(\tau+1)$ is essentially the mathematical expectation,

or "average" in loose language, of h. More specifically, it is h multiplied with some pdf followed by an integration over $\mathbb{R}^n$, *i.e.*, the corresponding sample space. The pdf is composed of several different factors. The first is, of course, $(\mathcal{P}_{\not 2}\rho)_1(y_1)$. But h, as well as $(\mathcal{P}_{\not 2}\rho)_1$, also has dependence on x_2, which is embedded within the subscript $\not 2$. Recall how x_2 is treated during $[\tau, \tau+1]$: It is frozen at step τ and kept on as a parameter, given all other components at τ. Therefore, the second part of the density is $\rho(x_2|x_1, x_3, ..., x_n)$, *i.e.*, the conditional density of x_2 on $x_1, x_3, ..., x_n$. (Note again that x_i means variables at time step τ.) This factor introduces extra dependencies: x_3, x_4, ..., x_n (that of x_1 is embedded in y_1), which must also be averaged out, so the third factor of the density is $\rho_{3...n}(x_3, ..., x_n)$ namely the joint density of $(x_3, x_4, ..., x_n)$. Put all these together,

$$\begin{aligned} H_{1\not 2}(\tau+1) &= -\int_{\mathbb{R}^n} (\mathcal{P}_{\not 2}\rho)_1(y_1) \cdot \log(\mathcal{P}_{\not 2}\rho)_1(y_1) \cdot \rho(x_2|x_1, x_3, ..., x_n) \\ &\quad \cdot \rho_{3...n}(x_3, ..., x_n)\, dy_1 dx_2 dx_3 ... dx_n \end{aligned} \tag{22}$$

Subtraction of $H_{1\not 2}(\tau+1) - H_1(\tau)$ from Equation (19) gives, eventually, the rate of information flow/transfer from x_2 to x_1:

$$\begin{aligned} T_{2\to 1} &= -\int_{\mathbb{R}} (\mathcal{P}\rho)_1(y_1) \cdot \log(\mathcal{P}\rho)_1(y_1)\, dy_1 \\ &\quad + \int_{\mathbb{R}^n} (\mathcal{P}_{\not 2}\rho)_1(y_1) \cdot \log(\mathcal{P}_{\not 2}\rho)_1(y_1) \cdot \rho(x_2|x_1, x_3, ..., x_n) \cdot \\ &\quad \rho_{3...n}(x_3, ..., x_n)\, dy_1 dx_2 dx_3 ... dx_n \end{aligned} \tag{23}$$

Notice that the conditional density of x_2 is on x_1, not on y_1. (x_1 and y_1 are the same state variable evaluated at different time steps, and are connected via $y_1 = \Phi_1(x_1, x_2, ..., x_n)$.

Likewise, it is easy to obtain the information flow between any pair of components. If, for example, we are concerned with the flow from x_j to x_i ($i, j = 1, 2, ..., n$, $i \neq j$), replacement of the indices 1 and 2 in Equation (23) respectively with i and j gives

$$\begin{aligned} T_{j\to i} &= -\int_{\mathbb{R}} (\mathcal{P}\rho)_i(y_i) \cdot \log(\mathcal{P}\rho)_i(y_i)\, dy_i \\ &\quad + \int_{\mathbb{R}^n} (\mathcal{P}_{\not j}\rho)_i(y_i) \cdot \log(\mathcal{P}_{\not j}\rho)_1(y_i) \cdot \rho(x_j \mid x_1, x_2, ..., x_{j-1}, x_{j+1}, ..., x_n) \cdot \\ &\quad \rho_{\not i \not j}\, dx_1 dx_2 ... dx_{i-1} dy_i dx_{i+1} ... dx_n \end{aligned} \tag{24}$$

Here the subscript $\not j$ of $\mathcal{P}$ means the F-P operator with the effect of the jth component excluded through freezing it instantaneously as a parameter. We have also abused the notation a little bit for the density function to indicate the marginalization of that component. That is to say,

$$\rho_{\not i} = \rho_{\not i}(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) = \int_{\mathbb{R}} \rho(\boldsymbol{x})\, dx_i \tag{25}$$

and $\rho_{\not i \not j}$ is the density after being marginalized twice, with respect to x_i and x_j. To avoid this potential notation complexity, alternatively, one may reorganize the order of the components of the vector $\boldsymbol{x} = (x_1, ..., x_n)^T$ such that the pair appears in the first two slots, and modify the mapping Φ accordingly. In this case, the flow/transfer is precisely the same in form as Equation (23). Equations (23) and (24) can be evaluated explicitly for systems that are definitely specified. In the following sections we will see several concrete examples.

3.3. Properties

The information flow obtained in Equations (23) or (24) has some nice properties. The first is a concretization of the transfer asymmetry emphasized by Schreiber [47] (as mentioned in the introduction), and the second a special property for 2D systems.

Theorem 3.1 *For the system Equation (4), if Φ_i is independent of x_j, then $T_{j\to i} = 0$ (in the mean time, $T_{i\to j}$ need not be zero).*

The proof is rather technically involved; the reader is referred to [48] for details. This theorem states that, if the evolution of x_i has nothing to do with x_j, then there will be no information flowing from x_j to x_i. This is in agreement with observations, and with what one would argue on physical grounds. On the other hand, the vanishing $T_{j\to i}$ yields no clue on $T_{i\to j}$, *i.e.*, the flow from x_i to x_j need not be zero in the mean time, unless Φ_j does not rely on x_i. This is indicative of a very important physical fact: information flow between a component pair is not symmetric, in contrast to the notion of mutual information ever existing in information theory. As emphasized by Schreiber [47], a faithful formalism must be able to recover this asymmetry. The theorem shows that our formalism yields precisely what is expected. Since transfer asymmetry is a reflection of causality, the above theorem is also referred to as *property of causality* by Liang and Kleeman [48].

Theorem 3.2 *For the system Equation (4), if $n = 2$ and Φ_1 is invertible, then $T_{2\to 1} = \Delta H_1 - E\log|J_1|$, where $J_1 = \partial\Phi_1/\partial x_1$.*

A brief proof will help to gain better understanding of the theorem. If $n = 2$, the modified system has a mapping $\Phi_{\mathfrak{x}}$ which is simply Φ_1 with x_2 as a parameter. Equation (22) is thus reduced to

$$H_{1\mathfrak{x}}(\tau+1) \;=\; -\int_{\mathbb{R}^2} (\mathcal{P}_{\mathfrak{x}}\rho)_1(y_1)\cdot\log(\mathcal{P}_{\mathfrak{x}}\rho)_1(y_1)\cdot\rho(x_2|x_1)\,dy_1dx_2$$

where $y_1 = \Phi_1(x_1, x_2)$, and $(\mathcal{P}_{\mathfrak{x}}\rho)_1$ the marginal density of x_1 evolving from $\rho_{\mathfrak{x}} = \rho_1$ upon one transformation of $\Phi_{\mathfrak{x}} = \Phi_1$. By assumption Φ_1 is invertible, that is to say, $J_1 = \frac{\partial\Phi_1}{\partial x_1} \neq 0$. The F-P operator hence can be explicitly written out:

$$\begin{aligned}(\mathcal{P}_{\mathfrak{x}}\rho)_1(y_1) &= \rho\left[\Phi_1^{-1}(y_1,x_2)\right]\cdot\left|J_1^{-1}\right| \\ &= \rho_1(x_1)\,\left|J_1^{-1}\right| \end{aligned} \tag{26}$$

So

$$\begin{aligned}\Delta H_{1\mathfrak{x}} &= H_{1\mathfrak{x}}(\tau+1) - H_1(\tau) \\ &= -\int_{\mathbb{R}^2}\rho_1(x_1)\left|J_1^{-1}\right|\,\log\left(\rho_1(x_1)\left|J_1^{-1}\right|\right)\,\rho(x_2|x_1)\,|J_1|\,dx_1dx_2 + \int_{\mathbb{R}}\rho_1\log\rho_1\,dx_1 \\ &= -\int_{\mathbb{R}^2}\rho_1(x_1)\,\rho(x_2|x_1)\,\log\left|J_1^{-1}\right|\,dx_1dx_2 \\ &= \int_{\mathbb{R}^2}\rho(x_1,x_2)\log|J_1|\,dx_1dx_2 \\ &= E\log|J_1| \end{aligned} \tag{27}$$

The conclusion follows subsequently from Equation (16).

The above theorem actually states another interesting fact that parallels what we introduced previously in §2.2 via heuristic reasoning. To see this, reconsider the mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{x} \mapsto \boldsymbol{x}$. Let Φ be nonsingular and invertible. By Equation (18), the F-P operator of the joint pdf ρ can be explicitly evaluated. Accordingly, the entropy increases, as time moves from step τ to step $\tau + 1$, by

$$\begin{aligned}\Delta H &= -\int_{\mathbb{R}^n} \mathcal{P}\rho(\boldsymbol{x}) \log \mathcal{P}\rho(\boldsymbol{x})\, d\boldsymbol{x} + \int_{\mathbb{R}^n} \rho(\boldsymbol{x}) \log \rho(\boldsymbol{x})\, d\boldsymbol{x} \\ &= -\int_{\mathbb{R}^n} \rho\left[\Phi^{-1}(\boldsymbol{x})\right] \left|J^{-1}\right| \log \rho\left[\Phi^{-1}(\boldsymbol{x})\right] \left|J^{-1}\right|\, d\boldsymbol{x} + \int_{\mathbb{R}^n} \rho(\boldsymbol{x}) \log \rho(\boldsymbol{x})\, d\boldsymbol{x}\end{aligned}$$

After some manipulation (see [48] for details), this is reduced to

$$\Delta H = E \log |J| \tag{28}$$

This is the discrete counterpart of Equation (12), yet another remarkably concise formula. Now, if the system in question is 2-dimensional, then, as argued in §2.2, the information flow from x_2 to x_1 should be $\Delta H_1 - \Delta H_1^*$, with ΔH_1^* being the marginal entropy increase due to x_1 itself. Furthermore, if Φ_1 is nonsingular and invertible, then Equation (28) tells us it must be that

$$\Delta H_1^* = E \log |J_1|$$

and this is precisely what Theorem 3.2 reads.

4. Continuous Systems

For continuous systems in the form of Equation (3), we may take advantage of what we already have from the previous section to obtain the information flow. Without loss of generality, consider only the flow/transfer from x_2 to x_1, $T_{2\to1}$. We adopt the following strategy to fulfill the task:

- Discretize the continuous system in time on $[t, t + \Delta t]$, and construct a mapping Φ to take $\boldsymbol{x}(t)$ to $\boldsymbol{x}(t + \Delta t)$;
- Freeze x_2 in Φ throughout $[t, t + \Delta t]$ to obtain a modified mapping $\Phi_{\wp}$;
- Compute the marginal entropy change ΔH_1 as Φ steers the system from t to $t + \Delta t$;
- Derive the marginal entropy change $\Delta H_{1\wp}$ as $\Phi_{\wp}$ steers the modified system from t to $t + \Delta t$;
- Take the limit
$$T_{2\to1} = \lim_{\Delta t \to 0} \frac{\Delta H_1 - \Delta H_{1\wp}}{\Delta t}$$
to arrive at the desiderata.

4.1. Discretization of the Continuous System

As the first step, construct out of Equation (3) an n-dimensional discrete system, which steers $\boldsymbol{x}(t) = (x_1, x_2, ..., x_n)$ to $\boldsymbol{x}(t + \Delta t)$. To avoid any confusion that may arise, $\boldsymbol{x}(t + \Delta t)$ will be

denoted as $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ hereafter. Discretization of Equation (3) results in a mapping, to the first order of Δt, $\Phi = (\Phi_1, \Phi_2, ..., \Phi_n)$: $\mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{x} \mapsto \boldsymbol{y}$:

$$\Phi : \begin{cases} y_1 = & x_1 + \Delta t \cdot F_1(\boldsymbol{x}) \\ y_2 = & x_2 + \Delta t \cdot F_2(\boldsymbol{x}) \\ \vdots & \vdots \\ y_n = & x_n + \Delta t \cdot F_n(\boldsymbol{x}) \end{cases} \tag{29}$$

Clearly, this mapping is always invertible so long as Δt is small enough. In fact, we have

$$\Phi^{-1} : \begin{cases} x_1 = & y_1 - \Delta t \cdot F_1(\boldsymbol{y}) + O(\Delta t^2) \\ x_2 = & y_2 - \Delta t \cdot F_2(\boldsymbol{y}) + O(\Delta t^2) \\ \vdots & \vdots \\ x_n = & y_n - \Delta t \cdot F_n(\boldsymbol{y}) + O(\Delta t^2) \end{cases} \tag{30}$$

to the first order of Δt. Furthermore, its Jacobian J is

$$\begin{aligned} J &= \det \left[\frac{\partial(y_1, y_2, ..., y_n)}{\partial(x_1, x_2, ..., x_n)} \right] \\ &= \prod_i \left(1 + \Delta t \frac{\partial F_i}{\partial x_i} \right) + O(\Delta t^2) \\ &= 1 + \Delta t \sum_{i=1}^{n} \frac{\partial F_i}{\partial x_i} + O(\Delta t^2) \end{aligned} \tag{31}$$

Likewise, it is easy to get

$$\begin{aligned} J^{-1} = &= \det \left[\frac{\partial(x_1, x_2, ..., x_n)}{\partial(y_1, y_2, ..., y_n)} \right] \\ &= 1 - \Delta t \sum_{i=1}^{n} \frac{\partial F_i}{\partial x_i} + O(\Delta t^2) \end{aligned} \tag{32}$$

This makes it possible to evaluate the F-P operator associated with Φ. By Equation (18),

$$\begin{aligned} \mathcal{P}\rho(y_1, ..., y_n) &= \rho\left(\Phi^{-1}(y_1, ...y_n)\right) \left|J^{-1}\right| \\ &= \rho(x_1, x_2, ..., x_n) \cdot |1 - \Delta t \nabla \cdot \boldsymbol{F}| + O(\Delta t^2) \end{aligned} \tag{33}$$

Here $\nabla \cdot \boldsymbol{F} = \sum_i \frac{\partial F_i}{\partial x_1}$; we have suppressed its dependence on $\boldsymbol{x}$ to simplify the notation.

As an aside, the explicit evaluation (31), and subsequently (32) and (33), actually can be utilized to arrive at the important entropy evolution law (12) without invoking any assumptions. To see this, recall that $\Delta H = E \log |J|$ by Equation (28). Let Δt go to zero to get

$$\frac{dH}{dt} = \lim_{\Delta t \to 0} \frac{\Delta H}{\Delta t} = E \lim_{\Delta t \to 0} \frac{1}{\Delta t} \log \left(1 + \Delta t \nabla \cdot \boldsymbol{F} + O(\Delta t^2)\right)$$

which is the very result $E(\nabla \cdot \boldsymbol{F})$, just as one may expect.

4.2. Information Flow

To compute the information flow $T_{2\to 1}$, we need to know dH_1/dt and $dH_{1\cancel{2}}/dt$. The former is easy to find from the Liouville equation associated with Equation (3), *i.e.*,

$$\frac{\partial \rho}{\partial t} + \frac{\partial (F_1 \rho)}{\partial x_1} + \frac{\partial (F_2 \rho)}{\partial x_2} + ... + \frac{\partial (F_n \rho)}{\partial x_n} = 0 \tag{34}$$

following the same derivation as that in §2.2:

$$\frac{dH_1}{dt} = \int_{\mathbb{R}^n} \log \rho_1 \frac{\partial (F_1 \rho)}{\partial x_1} \, d\boldsymbol{x} \tag{35}$$

The challenge lies in the evaluation of $dH_{1\cancel{2}}/dt$. We summarize the result in the following proposition:

Proposition 4.1 *For the dynamical system (3), the rate of change of the marginal entropy of x_1 with the effect of x_2 instantaneously excluded is:*

$$\begin{aligned} \frac{dH_{1\cancel{2}}}{dt} &= \int_{\mathbb{R}^n} (1 + \log \rho_1) \cdot \frac{\partial (F_1 \rho_{\cancel{2}})}{\partial x_1} \cdot \Theta_{2|1} \, d\mathbf{x} + \\ & \int_{\mathbb{R}^n} \rho_1 \log \rho_1 \cdot F_1 \cdot \frac{\partial (\rho / \rho_{\cancel{2}})}{\partial x_1} \cdot \rho_{\cancel{1}\cancel{2}} \, d\mathbf{x} \end{aligned} \tag{36}$$

where

$$\theta_{2|1} = \theta_{2|1}(x_1, x_2, x_3, ..., x_n) = \frac{\rho}{\rho_{\cancel{2}}} \rho_{\cancel{1}\cancel{2}} \tag{37}$$

$$\Theta_{2|1} = \int_{\Omega_{\mathbb{R}^{n-2}}} \theta_{2|1}(\mathbf{x}) \, dx_3 ... dx_n \tag{38}$$

and $\rho_{\cancel{2}} = \int_{\mathbb{R}} \rho \, dx_2$, $\rho_{\cancel{1}\cancel{2}} = \int_{\mathbb{R}^2} \rho \, dx_1 dx_2$ are the densities after marginalized with x_2 and (x_1, x_2), respectively.

The proof is rather technically involved; for details, see [51], section 5.

With the above result, subtract $dH_{1\cancel{2}}/dt$ from dH_1/dt and one obtains the flow rate from x_2 to x_1. Likewise, the information flow between any component pair (x_i, x_j), $i, j = 1, 2, ..., n$; $i \neq j$, can be obtained henceforth.

Theorem 4.1 *For the dynamical system (3), the rate of information flow from x_j to x_i is*

$$\begin{aligned} T_{j\to i} &= \int_{\Omega} (1 + \log \rho_i) \left(\frac{\partial (F_i \rho)}{\partial x_i} - \frac{\partial (F_i \rho_{\cancel{j}})}{\partial x_i} \cdot \Theta_{j|i} \right) \, d\mathbf{x} \\ &+ \int_{\Omega} \frac{\partial (F_i \rho_i \log \rho_i)}{\partial x_i} \cdot \theta_{j|i} \, d\mathbf{x} \end{aligned} \tag{39}$$

where

$$\theta_{j|i} = \theta_{j|i}(\mathbf{x}) = \frac{\rho}{\rho_{\cancel{j}}} \rho_{\cancel{i}\cancel{j}} \tag{40}$$

$$\rho_{\cancel{i}} = \int_{\mathbb{R}} \rho(\mathbf{x}) \, dx_i \tag{41}$$

$$\rho_{\cancel{i}\cancel{j}} = \int_{\mathbb{R}^2} \rho(\mathbf{x}) \, dx_i dx_j \tag{42}$$

$$\Theta_{j|i} = \Theta_{j|i}(x_i, x_j) = \int_{\mathbb{R}^{n-2}} \theta_{j|i}(\mathbf{x}) \prod_{\nu \neq i,j} dx_\nu \tag{43}$$

In this formula, $\Theta_{j|i}$ reminds one of the conditional density x_j on x_i, and, if $n = 2$, it is indeed so. We may therefore call it the "generalized conditional density" of x_j on x_i.

4.3. Properties

Recall that, as we argue in §2.2 based on the entropy evolution law (12), the time rate of change of the marginal entropy of a component, say x_1, due to its own reason, is $dH_1^*/dt = E(\partial F_1/\partial x_1)$. Since for a 2D system, dH_1^*/dt is precisely $dH_{1\wr 2}/dt$, we expect that the above formalism (36) or (39) verifies this result.

Theorem 4.2 *If the system (3) has a dimensionality 2, then*

$$\frac{dH_{1\wr 2}}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) \tag{44}$$

and hence the rate of information flow from x_2 to x_1 is

$$T_{2\to 1} = -E\left[\frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\right] \tag{45}$$

What makes a 2D system so special is that, when $n = 2$, $\rho_{\wr 2} = \rho_1$, and $\Theta_{2|1}$ is just the conditional distribution of x_2 given x_1, $\rho/\rho_1 = \rho(x_2|x_1)$. Equation (36) can thereby be greatly simplified:

$$\begin{aligned}\frac{dH_{1\wr 2}}{dt} &= \int_{\mathbb{R}^n}(1+\log\rho_1)\frac{\partial F_1\rho_1}{\partial x_1}\cdot\frac{\rho}{\rho_1}\,d\boldsymbol{x} + \int_{\mathbb{R}^n}\rho_1\log\rho_1\cdot F_1\cdot\frac{\partial\rho(x_2|x_1)}{\partial x_1}\,d\boldsymbol{x}\\ &= \int_{\mathbb{R}^n}\frac{\partial(F_1\rho_1)}{\partial x_1}\frac{\rho}{\rho_1}\,d\boldsymbol{x} + \int_{\mathbb{R}^n}\log\rho_1\cdot\frac{\partial(F_1\rho)}{\partial x_1}\,d\boldsymbol{x}\\ &= \int_{\mathbb{R}^n}\rho\left(\frac{\partial F_1}{\partial x_1}\right)d\boldsymbol{x} = E\left(\frac{\partial F_1}{\partial x_1}\right)\end{aligned} \tag{46}$$

Subtract this from what has been obtained above for dH_1/dt, and we get an information flow just as that in Equation (14) via heuristic argument.

As in the discrete case, one important property that $T_{j\to i}$ must possess is transfer asymmetry, which has been emphasized previously, particularly by Schreiber [47]. The following is a concretization of the argument.

Theorem 4.3 **(Causality)** *For system (3), if F_i is independent of x_j, then $T_{j\to i} = 0$; in the mean time, $T_{i\to j}$ need not vanish, unless F_j has no dependence on x_i.*

Look at the right-hand side of the formula (39). Given that $(1+\log\rho_i)$ and $\rho_{\wr i}$, as well as F_i (by assumption), are independent of x_j, the integration with respect to x_j can be taken within the multiple integrals. Consider the second integral first. All the variables except $\theta_{j|i}$ have dependence on x_j. But $\int\theta_{j|i}dx_j = 1$, so the whole term is equal to $\int_{\mathbb{R}^{n-1}}\frac{\partial(F_i\rho_i\log\rho_i)}{\partial x_i}dx_1...dx_{j-1}dx_{j+1}...dx_n$ which vanishes by the assumption of compact support. For the first integral, move the integration with respect to x_j into the parentheses, as the factor outside has nothing to do with x_j. This integration yields

$$\begin{aligned}&\int_{\mathbb{R}}\frac{\partial(F_i\rho)}{\partial x_i}dx_j - \int_{\mathbb{R}}\frac{\partial(F_i\rho_{\wr i})}{\partial x_i}\cdot\Theta_{j|i}dx_j\\ &= \int_{\mathbb{R}^{n-1}}\left[\frac{\partial}{\partial x_i}\left(F_i\int\rho dx_j\right) - \frac{\partial}{\partial x_i}(F_i\rho_{\wr i})\cdot\int\Theta_{j|i}dx_j\right]dx_1...dx_{j-1}dx_{j+1}...dx_n\\ &= 0\end{aligned}$$

because $\int \rho dx_j = \rho_{\not{x}}$ and $\int \Theta_{j|i} dx_j = 1$. For all that account, both the two integrals on the right-hand side of Equation (39) vanish, leaving a zero flow of information from x_j to x_i. Notice that this vanishing $T_{j\to i}$ gives no hint on the flow in the opposite direction. In other words, this kind of flow or transfer is not symmetric, reflecting the causal relation between the component pair. As Theorem 3.1 is for discrete systems, Theorem 4.3 is the *property of causality* for continuous systems.

5. Stochastic Systems

So far, all the systems considered are deterministic. In this section we turn to systems with stochasticity included. Consider the stochastic counterpart of Equation (3)

$$d\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, t)dt + \boldsymbol{B}(\boldsymbol{x}, t)d\boldsymbol{w} \tag{47}$$

where $\boldsymbol{w}$ is a vector of standard Wiener processes, and $\boldsymbol{B} = (b_{ij})$ the matrix of perturbation amplitudes. In this section, we limit our discussion to 2D systems, and hence have only two flows/transfers to discuss. Without loss of generality, consider only $T_{2\to 1}$, *i.e.*, the rate of flow/transfer from x_2 to x_1.

As before, we first need to find the time rate of change of H_1, the marginal entropy of x_1. This can be easily derived from the density evolution equation corresponding to Equation (47), *i.e.*, the Fokker-Planck equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (F_1 \rho)}{\partial x_1} + \frac{\partial (F_2 \rho)}{\partial x_2} = \frac{1}{2} \sum_{i,j=1}^{2} \frac{\partial^2 (g_{ij}\rho)}{\partial x_i \partial x_j} \tag{48}$$

where $g_{ij} = g_{ji} = \sum_{k=1}^{2} b_{ik} b_{jk}$, $i, j = 1, 2$. This integrated over $\mathbb{R}$ with respect to x_2 gives the evolution of ρ_1:

$$\frac{\partial \rho_1}{\partial t} + \int_{\mathbb{R}} \frac{\partial (F_1 \rho)}{\partial x_1} dx_2 = \frac{1}{2} \int_{\mathbb{R}} \frac{\partial^2 (g_{11}\rho)}{\partial x_1^2} dx_2 \tag{49}$$

Multiply (49) by $-(1 + \log \rho_1)$, and integrate with respect to x_1 over $\mathbb{R}$. After some manipulation, one obtains, using the compact support assumption,

$$\frac{dH_1}{dt} = -E\left(F_1 \frac{\partial \log \rho_1}{\partial x_1}\right) - \frac{1}{2} E\left(g_{11} \frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) \tag{50}$$

where E is the mathematical expectation with respect to ρ.

Again, the key to the formalism is the finding of $dH_{1\not{2}}/dt$. For stochastic systems, this could be a challenging task. The major challenge is that we cannot obtain an F-P operator as nice as that in the previous section for the map resulting from discretization. In early days, Majda and Harlim [53] have tried our heuristic argument in §2.2 to consider a special system modeling the atmosphere–ocean interaction, which is in the form

$$\begin{aligned} dx_1 &= F_1(x_1, x_2)dt \\ dx_2 &= F_2(x_1, x_2)dt + b_{22} dw_2 \end{aligned}$$

Their purpose is to find $T_{2\to 1}$ namely the information transfer from x_2 to x_1. In this case, since the governing equation for x_1 is deterministic, the result is precisely the same as that of LK05, which is

shown in in §2.2. The problem here is that the approach cannot be extended even to finding $T_{1\to 2}$, since the nice law on which the argument is based, *i.e.*, Equation (12), does not hold for stochastic processes.

Liang (2008) [54] adopted a different approach to give this problem a solution. As in the previous section, the general strategy is also to discretize the system in time, modify the discretized system with x_2 frozen as a parameter on an interval $[t, t+\Delta t]$, and then let Δt go to zero and take the limit. But this time no operator analogous to the F-P operator is sought; instead, we discretize the Fokker–Planck equation and expand $x_{1\cancel{2}(t+\Delta t)}$, namely the first component at $t+\Delta t$ with x_2 frozen at t, using the Euler–Bernstein approximation. The complete derivation is beyond the scope of this review; the reader is referred to [54] for details. In the following, the final result is supplied in the form of a proposition.

Proposition 5.1 *For the 2D stochastic system (47), the time change of the marginal entropy of x_1 with the contribution from x_2 excluded is*

$$\frac{dH_{1\cancel{2}}}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) - \frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) - \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 (g_{11}\rho_1)}{\partial x_1^2}\right) \tag{51}$$

In the equation, the second and the third terms on the right hand side are from the stochastic perturbation. The first term, as one may recall, is precisely the result of Theorem 4.2. The heuristic argument for 2D systems in Equation (13) is successfully recovered here. With this the rate of information flow can be easily obtained by subtracting $dH_{1\cancel{2}}/dt$ from dH_1/dt.

Theorem 5.1 *For the 2D stochastic system (47), the rate of information flow from x_2 to x_1 is*

$$T_{2\to 1} = -E\left(\frac{1}{\rho_1}\frac{\partial (F_1\rho_1)}{\partial x_1}\right) + \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 (g_{11}\rho_1)}{\partial x_1^2}\right) \tag{52}$$

where E is the expectation with respect to $\rho(x_1, x_2)$.

It has been a routine to check for the obtained flow the property of causality or asymmetry. Here in Equation (52), the first term on the right hand side is from the deterministic part of the system, which has been checked before. For the second term, if b_{11}, b_{12}, and hence $g_{11} = \sum_k b_{1k}b_{1k}$ have no dependence on x_2, then the integration with respect to x_2 can be taken inside with ρ/ρ_1 or $\rho(x_2|x_1)$, and results in 1. The remaining part is in a divergence form, which, by the assumption of compact support, gives a zero contribution from the stochastic perturbation. We therefore have the following theorem:

Theorem 5.2 *If, in the stochastic system (47), the evolution of x_1 is independent of x_2, then $T_{2\to 1} = 0$.*

The above argument actually has more implications. Suppose $\boldsymbol{B} = (b_{ij})$ are independent of $\boldsymbol{x}$, *i.e.*, the noises are uncorrelated with the state variables. This model is indeed of interest, as in the real world, a large portion of noises are additive; in other words, b_{ij}, and hence g_{ij}, are constant more often than not. In this case, no matter what the vector field $\boldsymbol{F}$ is, by the above argument the resulting information flows within the system will involve no contribution from the stochastic perturbation. That is to say,

Theorem 5.3 *Within a stochastic system, if the noise is additive, then the information flows are the same in form as that of the corresponding deterministic system.*

This theorem shows that, if only information flows are considered, a stochastic system with additive noise functions just like deterministic. Of course, the resemblance is limited to the form of formula; the marginal density ρ_1 in Equation (52) already takes into account the effect of stochasticity, as can be seen from the integrated Fokker–Planck Equation (49). A more appropriate statement might be that, for this case, stochasticity is disguised within the formula of information flow.

6. Applications

Since its establishment, the formalism of information flow has been applied with a variety of dynamical system problems. In the following we give a brief description of these applications.

6.1. Baker Transformation

The baker transformation as a prototype of an area-conserving chaotic map is one of the most studied discrete dynamical systems. Topologically it is conjugate to another well-studied system, the horseshoe map, and has been be used to model the diffusion process in real physical world.

The baker transformation mimicks the kneading of dough: first the dough is compressed, then cut in half; the two halves are stacked on one another, compressed, and so forth. Formally, it is defined as a mapping on the unit square $\Omega = [0,1] \times [0,1]$, $\Phi : \Omega \to \Omega$,

$$\Phi(x_1, x_2) = \begin{cases} (2x_1,\ \frac{x_2}{2}), & 0 \le x_1 \le \frac{1}{2},\ 0 \le x_2 \le 1 \\ (2x_1 - 1,\ \frac{1}{2}x_2 + \frac{1}{2}), & \frac{1}{2} < x_1 \le 1,\ 0 \le x_2 \le 1 \end{cases} \tag{53}$$

with a Jacobian $J = \det\left[\frac{\partial(\Phi_1(\boldsymbol{x}),\Phi_2(\boldsymbol{x}))}{\partial(x_1,x_2)}\right] = 1$. This is the area-conserving property, which, by Equation (28) yields $\Delta H = E\log|J| = 0$; that is to say, the entropy is also conserved. The nonvanishing Jacobian implies that it is invertible; in fact, it has an inverse

$$\Phi^{-1}(x_1, x_2) = \begin{cases} (\frac{x_1}{2},\ 2x_2), & 0 \le x_2 \le \frac{1}{2},\ 0 \le x_1 \le 1 \\ (\frac{x_1+1}{2},\ 2x_2 - 1), & \frac{1}{2} \le x_2 \le 1,\ 0 \le x_1 \le 1 \end{cases} \tag{54}$$

Thus the F-P operator $\mathcal{P}$ can be easily found

$$\mathcal{P}\rho(x_1, x_2) = \rho\left[\Phi^{-1}(x_1, x_2)\right] \cdot \left|J^{-1}\right| = \begin{cases} \rho(\frac{x_1}{2}, 2x_2), & 0 \le x_2 < \frac{1}{2} \\ \rho(\frac{1+x_1}{2}, 2x_2 - 1), & \frac{1}{2} \le x_2 \le 1 \end{cases} \tag{55}$$

First compute $T_{2\to1}$, the information flow from x_2 to x_1. Let ρ_1 be the marginal density of x_1 at time step τ. Taking integration of Equation (55) with respect to x_2, one obtains the marginal density of x_1 at $\tau + 1$

$$\begin{aligned} (\mathcal{P}\rho)_1(x_1) &= \int_0^{1/2} \rho(\frac{x_1}{2}, 2x_2)\, dx_2 + \int_{1/2}^{1} \rho(\frac{x_1+1}{2}, 2x_2 - 1)\, dx_2 \\ &= \frac{1}{2}\int_0^1 \left[\rho\left(\frac{x_1}{2}, x_2\right) + \rho\left(\frac{x_1+1}{2}, x_2\right)\right] dx_2 \end{aligned}$$

$$= \frac{1}{2}\left[\rho_1\left(\frac{x_1}{2}\right) + \rho_1\left(\frac{x_1+1}{2}\right)\right] \tag{56}$$

One may also compute the marginal entropy $H_1(\tau+1)$, which is an entropy functional of $(\mathcal{P}\rho)_1$. However, here it is not necessary, as will soon become clear.

If, on the other hand, x_2 is frozen as a parameter, the transformation (53) then reduces to a dyadic mapping in the stretching direction, $\Phi_1 : [0,1] \to [0,1]$, $\Phi_1(x_1) = 2x_1 \pmod 1$. For any $0 < x_1 < 1$, The counterimage of $[0, x_1]$ is

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

So

$$\begin{aligned}
(\mathcal{P}_{\text{ɧ}}\rho)_1(x_1) &= \frac{\partial}{\partial x_1}\int_{\Phi^{-1}([0,x_1])} \rho(s)\, ds \\
&= \frac{\partial}{\partial x_1}\int_0^{x_1/2} \rho(s)\, ds + \frac{\partial}{\partial x_1}\int_{1/2}^{(1+x_1)/2} \rho(s)\, ds \\
&= \frac{1}{2}\left[\rho\left(\frac{x_1}{2}\right) + \rho\left(\frac{1+x_1}{2}\right)\right]
\end{aligned}$$

Two observations: (1) This result is exactly the same as Equation (56), *i.e.*, $(\mathcal{P}_{\text{ɧ}}\rho)_1$ is equal to $(\mathcal{P}\rho)_1$. (2) The resulting $(\mathcal{P}_{\text{ɧ}}\rho)_1$ has no dependence on the parameter x_2. The latter helps to simplify the computation of $H_{1\text{ɧ}}(\tau+1)$ in Equation (22): Now the integration with respect to x_2 can be taken inside, giving $\int \rho(x_2|x_1)dx_2 = 1$. So $H_{1\text{ɧ}}(\tau+1)$ is precisely the entropy functional of $(\mathcal{P}_{\text{ɧ}}\rho)_1$. But $(\mathcal{P}_{\text{ɧ}}\rho)_1 = (\mathcal{P}\rho)_1$ by observation (1). Thus $H_1(\tau+1) = H_{1\text{ɧ}}(\tau+1)$, leading to a flow/transfer

$$T_{2\to 1} = 0 \tag{57}$$

The information flow in the opposite direction is different. As above, first compute the marginal density

$$(\mathcal{P}\rho)_2(x_2) = \int_0^1 \mathcal{P}\rho(x_1,x_2)\, dx_1 = \begin{cases} \int_0^1 \rho\left(\frac{x_1}{2}, 2x_2\right)\, dx_1, & 0 \le x_2 < \frac{1}{2} \\ \int_0^1 \rho\left(\frac{x_1+1}{2}, 2x_2 - 1\right)\, dx_1, & \frac{1}{2} \le x_2 \le 1 \end{cases} \tag{58}$$

The marginal entropy increase of x_2 is then

$$\begin{aligned}
\Delta H_2 &= -\int_0^1\int_0^1 \mathcal{P}\rho(x_1,x_2)\cdot\left[\log\left(\int_0^1 \mathcal{P}\rho(\lambda,x_2)d\lambda\right)\right] dx_1dx_2 \\
&\quad + \int_0^1\int_0^1 \rho(x_1,x_2)\cdot\left[\log\left(\int_0^1 \rho(\lambda,x_2)d\lambda\right)\right] dx_1dx_2,
\end{aligned} \tag{59}$$

which is reduced to, after some algebraic manipulation,

$$\Delta H_2 = -\log 2 + (I + \mathrm{II}) \tag{60}$$

where

$$I = \int_0^1\int_0^{1/2} \rho(x_1,x_2)\cdot\left[\log\frac{\int_0^1 \rho(\lambda,x_2)d\lambda}{\int_0^{1/2} \rho(\lambda,x_2)d\lambda}\right] dx_1dx_2 \tag{61}$$

$$\mathrm{II} = \int_0^1\int_{1/2}^1 \rho(x_1,x_2)\cdot\left[\log\frac{\int_0^1 \rho(\lambda,x_2)d\lambda}{\int_{1/2}^1 \rho(\lambda,x_2)d\lambda}\right] dx_1dx_2 \tag{62}$$

To compute $H_{2\not{1}}$, freeze x_1. The transformation is invertible and the Jacobian J_2 is equal to a constant $\frac{1}{2}$. By Theorem 3.2,

$$\Delta H_{2\not{1}} = E \log \frac{1}{2} = -\log 2 \tag{63}$$

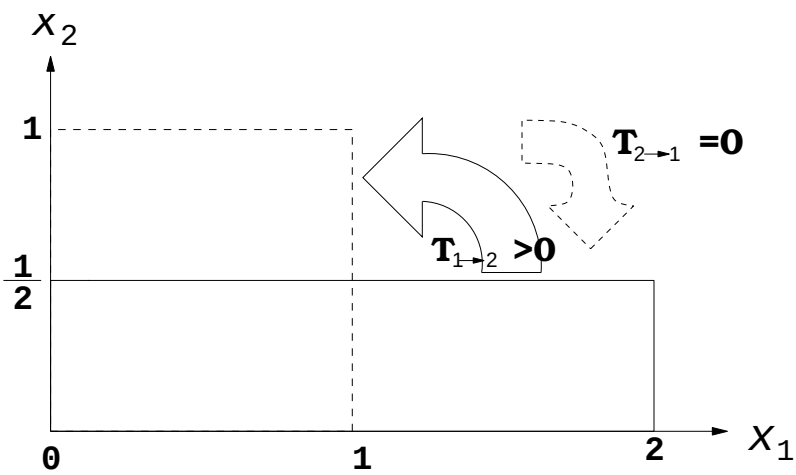
So,

$$T_{1\to 2} = \Delta H_2 - \Delta H_{2\not{1}} = I + II \tag{64}$$

In the expressions for I and II, since both ρ and the terms within the brackets are nonnegative, $I + II \geq 0$. Furthermore, the two brackets cannot vanish simultaneously, hence $I + II > 0$. By Equation (64) $T_{1\to 2}$ is strictly positive; in other words, there is always information flowing from x_1 to x_2.

To summarize, the baker transformation transfers information asymmetrically between the two directions x_1 and x_2. As the baker stretches the dough, and folds back on top the other, information flows continuously from the stretching direction x_1 to the folding direction x_2 ($T_{1\to 2} > 0$), while no transfer occurs in the opposite direction ($T_{2\to 1} = 0$). These results are schematically illustrated in Figure 2; they are in agreement with what one would observe in daily life, as described in the beginning of this review.

Figure 2. Illustration of the unidirectional information flow within the baker transformation.



6.2. Hénon Map

The Hénon map is another most studied discrete dynamical systems that exhibit chaotic behavior. Introduced by Michel Hénon as a simplified Poincaré section of the Lorenz system, it is a mapping $\Phi = (\Phi_1, \Phi_2) : \mathbb{R}^2 \mapsto \mathbb{R}^2$ defined such that

$$\begin{cases} \Phi_1(x_1, x_2) = 1 + x_2 - ax_1^2 \\ \Phi_2(x_1, x_2) = bx_1 \end{cases} \tag{65}$$

with $a > 0$, $b > 0$. When $a = 1.4$, $b = 0.3$, the map is termed "canonical," for which initially a point will either diverge to infinity, or approach an invariant set known as the Hénon strange attractor. Shown in Figure 3 is the attractor.

Like the baker transformation, the Hénon map is invertible, with an inverse

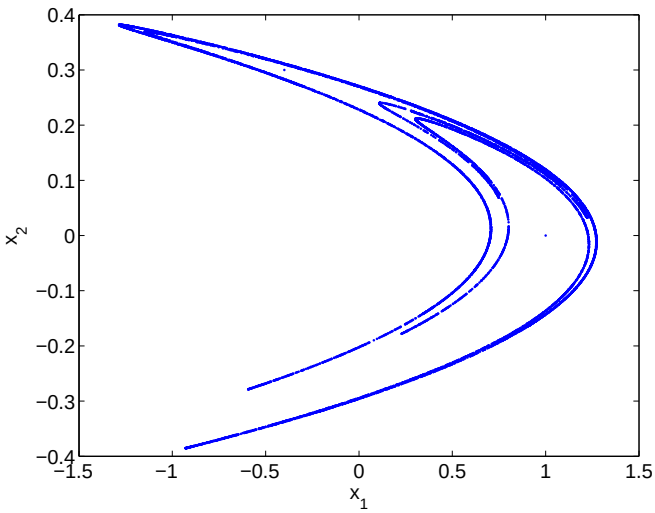
$$\Phi^{-1}(x_1, x_2) = \left(\frac{x_2}{b},\ x_1 - 1 + \frac{a}{b^2}x_2^2\right) \tag{66}$$

The F-P operator thus can be easily found from Equation (18):

$$\begin{aligned} \mathcal{P}\rho(x_1, x_2) &= \rho(\Phi^{-1}(x_1, x_2))|J^{-1}| \\ &= \frac{1}{b} \cdot \rho\left(\frac{x_2}{b},\ x_1 - 1 + \frac{a}{b^2}x_2^2\right) \end{aligned} \tag{67}$$

In the following, we compute the flows/transfers between x_1 and x_2.

Figure 3. A trajectory of the canonical Hénon map ($a = 1.4$, $b = 0.3$) starting at $(x_1, x_2) = (1, 0)$.



First, consider $T_{2\to 1}$, *i.e.*, the flow from the linear component x_2 to the quadratic component x_1. By Equation (23), we need to find the marginal density of x_1 at step $\tau + 1$ with and without the effect of x_2, *i.e.*, $(\mathcal{P}\rho)_1$ and $(\mathcal{P}\rho)_{1\not{2}}$. With the F-P operator obtained above, $(\mathcal{P}\rho)_1$ is

$$\begin{aligned} (\mathcal{P}\rho)_1(x_1) &= \int_{\mathbb{R}} \mathcal{P}\rho(x_1, x_2)\, dx_2 \\ &= \int_{\mathbb{R}} \frac{1}{b} \cdot \rho\left(\frac{x_2}{b}, x_1 - 1 + \frac{a}{b}x_2^2\right)\, dx_2 \\ &= \int_{\mathbb{R}} \rho(\eta, x_1 - 1 + a\eta^2)\, d\eta \qquad\qquad (x_2/b \equiv \eta) \end{aligned}$$

If $a = 0$, this integral would be equal to $\rho_2(x_1 - 1)$. Note it is the marginal density of x_2, but the argument is $x_1 - 1$. But here $a > 0$, the integration is taken along a parabolic curve rather than a straight line. Still the final result will be related to the marginal density of x_2; we may as well write it $\tilde{\rho}_2(x_1)$, that is

$$(\mathcal{P}\rho)_1(x_1) = \tilde{\rho}_2(x_1) \tag{68}$$

Again, notice that the argument is x_1.

To compute $(\mathcal{P}_{\not{2}}\rho)_1$, let

$$y_1 \equiv \Phi_1(x_1) = 1 + x_2 - ax_1^2$$

following our convention to distinguish variables at different steps. Modify the system so that x_2 is now a parameter. As before, we need to find the counterimage of $(-\infty, y_1]$ under the transformation with x_2 frozen:

$$\Phi_1^{-1}((-\infty, y_1]) = \left(-\infty,\ -\sqrt{(1+x_2-y_1)/a}\right] \cup \left[\sqrt{(1+x_2-y_1)/a},\ \infty\right)$$

Therefore,

$$\begin{aligned}
(\mathcal{P}_{\not{2}}\rho)_1(y_1) &= \frac{d}{dy_1}\int_{\Phi_1^{-1}((-\infty,y_1])} \rho_1(s)\, ds \\
&= \frac{d}{dy_1}\int_{-\infty}^{-\sqrt{(1+x_2-y_1)/a}} \rho_1(s)\, ds + \frac{d}{dy_1}\int_{\sqrt{(1+x_2-y_1)/a}}^{\infty} \rho_1(s)\, ds \\
&= \frac{1}{2\sqrt{a(1+x_2-y_1)}}\left[\rho_1\left(-\sqrt{(1+x_2-y_1)/a}\right) + \rho_1\left(\sqrt{(1+x_2-y_1)/a}\right)\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (y_1 < 1+x_2) \\
&= \frac{1}{2a|x_1|}\left[\rho_1(-x_1) + \rho_1(x_1)\right]. \qquad\qquad (\text{recall } y_1 = 1+x_2-ax_1^2)
\end{aligned}$$

Denote the average of $\rho_1(-x_1)$ and $\rho_1(x_1)$ as $\bar{\rho}_1(x_1)$ to make an even function of x_1. Then $(\mathcal{P}_{\not{2}}\rho)_1$ is simply

$$(\mathcal{P}_{\not{2}}\rho)_1(y_1) = \frac{\bar{\rho}_1(x_1)}{a|x_1|} \tag{69}$$

Note that the parameter x_2 does not appear in the arguments. Furthermore, $J_1 = \det\left(\frac{\partial \Phi_1}{\partial x_1}\right) = -2ax_1$. Substitute all the above into Equation (23) to get

$$\begin{aligned}
T_{2\to 1} &= -\int_{\mathbb{R}} (\mathcal{P}\rho)_1(x_1)\cdot\log(\mathcal{P}\rho)_1(x_1)\, dx_1 \\
&\quad + \int_{\mathbb{R}^2} (\mathcal{P}_{\not{2}}\rho)_1(y_1)\log(\mathcal{P}\rho)_{1\not{2}}(y_1)\cdot\rho(x_2|x_1)\cdot|J_1|\, dx_1 dx_2 \\
&= -\int_{\mathbb{R}} \tilde{\rho}_2(x_1)\log\tilde{\rho}_2(x_1)\, dx_1 \\
&\quad + \int_{\mathbb{R}} \frac{\bar{\rho}_1(x_1)}{a|x_1|}\log\frac{\bar{\rho}_1(x_1)}{a|x_1|}\cdot|-2ax_1|\cdot\left[\int_{\mathbb{R}}\rho(x_2|x_1)dx_2\right] dx_1
\end{aligned}$$

The taking of the integration with respect to x_2 inside the integral is legal since all the terms except the conditional density are independent of x_2. With the fact $\int_{\mathbb{R}}\rho(x_2|x_1)dx_2 = 1$, and the introduction of notations $\tilde{H}$ and $\bar{H}$ for the entropy functionals of $\tilde{\rho}$ and $\bar{\rho}$, respectively, we have

$$T_{2\to 1} = \tilde{H}_2 - 2\bar{H}_1 - \log|ax_1| \tag{70}$$

Next, consider $T_{1\to 2}$, the flow from the quadratic component to the linear component. As a common practice, one may start off by computing $(\mathcal{P}\rho)_2$ and $(\mathcal{P}_{\not{1}}\rho)_2$. However, in this case, things can be much

simplified. Observe that, for the modified system with x_1 frozen as a parameter, the Jacobian of the transformation $J_2 = \det\left[\frac{\partial \Phi_2}{\partial x_2}\right] = 0$. So, by Equation (24),

$$\begin{aligned} T_{1\to 2} &= -\int_{\mathbb{R}} (\mathcal{P}\rho)_2(x_2)\cdot\log(\mathcal{P}\rho)_2(x_2)\,dx_2 \\ &\quad + \int_{\mathbb{R}} (\mathcal{P}_{\backslash 1}\rho)_2(y_2)\cdot\log(\mathcal{P}_{\backslash 1}\rho)_2(y_2)\cdot\rho(x_1|x_2)\cdot|J_2|\;dx_1dx_2, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (y_2 \equiv \Phi_2(x_1,x_2)) \\ &= -\int_{\mathbb{R}} (\mathcal{P}\rho)_2(x_2)\cdot\log(\mathcal{P}\rho)_2(x_2)\,dx_2 \end{aligned}$$

with Equation (67), the marginal density

$$\begin{aligned} (\mathcal{P}\rho)_2(x_2) &= \int_{\mathbb{R}} \mathcal{P}\rho(x_1,x_2)\,dx_1 \\ &= \int_{\mathbb{R}} \frac{1}{b}\rho\left(\frac{x_2}{b},\; x_1 - 1 + a\frac{x_2^2}{b^2}\right)\,dx_1 \\ &= \frac{1}{b}\int_{\mathbb{R}} \rho(y,\xi)\,d\xi = \frac{1}{b}\rho_1\left(\frac{x_2}{b}\right) \end{aligned}$$

allowing us to arrive at an information flow from x_1 to x_2 in the amount of:

$$\begin{aligned} T_{1\to 2} &= -\int_{\mathbb{R}} \frac{1}{b}\rho_1\left(\frac{x_2}{b}\right)\cdot\log\left[\frac{1}{b}\rho_1\left(\frac{x_2}{b}\right)\right]\,dx_2 \\ &= H_1 + \log b \end{aligned} \tag{71}$$

That is to say, the flow from x_1 to x_2 has nothing to do with x_2; it is equal to the marginal entropy of x_1, plus a correction term due to the factor b.

The simple result of Equation (71) is remarkable; particularly, if $b = 1$, the information flow from x_1 to x_2 is just the entropy of x_1. This is precisely what what one would expect of the mapping component $\Phi_2(x_1,x_2) = bx_1$ in Equation (65). While the information flow is interesting *per se*, it also serves as an excellent example for the verification of our formalism.

6.3. Truncated Burgers–Hopf System

In this section, we examine a more complicated system, the Truncated Burgers–Hopf system (TBS hereafter). Originally introduced by Majda and Timofeyev [55] as a prototype of climate modeling, the TBS results from a Galerkin truncation of the Fourier expansion of the inviscid Burgers' equation, *i.e.*,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0 \tag{72}$$

to the nth order. Liang and Kleeman [51] examined such a system with two Fourier modes retained, which is governed by 4 ordinary differential equations:

$$\frac{dx_1}{dt} = F_1(\boldsymbol{x}) = x_1x_4 - x_3x_2 \tag{73}$$

$$\frac{dx_2}{dt} = F_2(\boldsymbol{x}) = -x_1x_3 - x_2x_4 \tag{74}$$

$$\frac{dx_3}{dt} = F_3(\boldsymbol{x}) = 2x_1x_2 \tag{75}$$

$$\frac{dx_4}{dt} = F_4(\boldsymbol{x}) = -x_1^2 + x_2^2 \tag{76}$$

Despite its simplicity, the system is intrinsically chaotic, with a strange attractor lying within

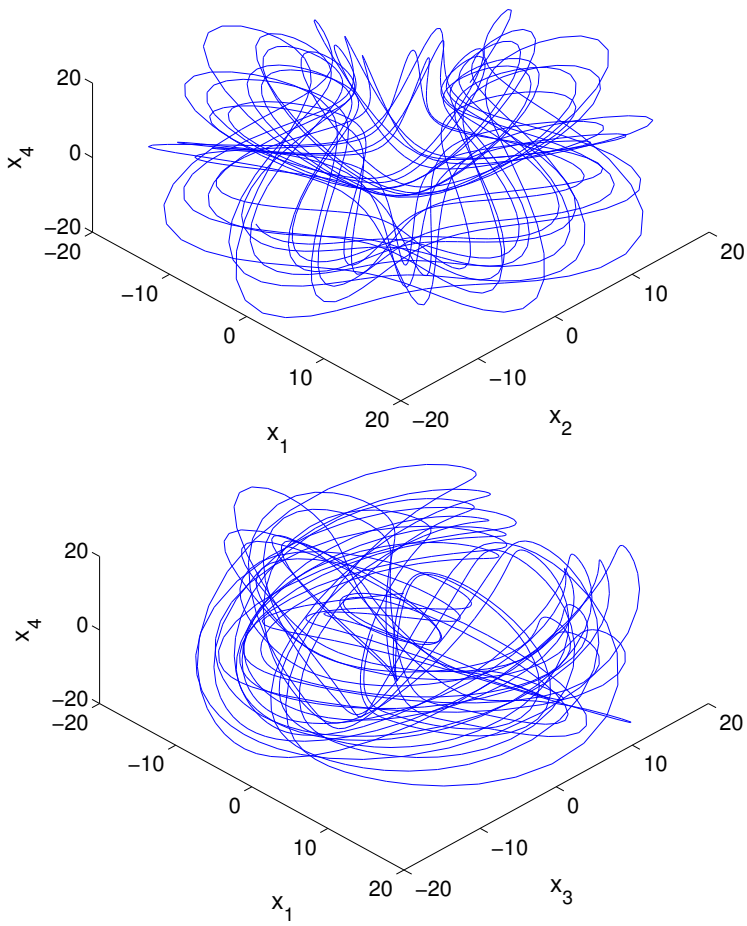
$$[-24.8, 24.6] \times [-25.0, 24.5] \times [-22.3, 21.9] \times [-23.7, 23.7]$$

Shown in Figure 4 are its projections onto the x_1-x_2-x_4 and x_1-x_3-x_4 subspaces, respectively.

Finding the information flows within the TBS system turns out to be a challenge in computation, since the Liouville equation corresponding to Equations (73)–(76) is a four-dimensional partial differential equation. In [51], Liang and Kleeman adopt a strategy of ensemble prediction to reduce the computation to an acceptable level. This is summarized in the following steps:

1. Initialize the joint density of (x_1, x_2, x_3, x_4) with some distribution ρ_0; make random draws according to ρ_0 to form an ensemble. The ensemble should be large enough to resolve adequately the sample space.
2. Discretize the sample space into "bins."
3. Do ensemble prediction for the system (73)–(74).
4. At each step, estimate the probability density function ρ by counting the bins.
5. Plug the estimated ρ back to Equation (39) to compute the rates of information flow at that step.

Figure 4. The invariant attractor of the truncated Burgers–Hopf system (73)–(76). Shown here is the trajectory segment for $2 \leq t \leq 20$ starting at $(40, 40, 40, 40)$. (a) and (b) are the 3-dimensional projections onto the subspaces x_1-x_2-x_3 and x_2-x_3-x_4, respectively.



Notice that the invariant attractor in Figure 4 allows us to perform the computation on a compact subspace of $\mathbb{R}^4$. Denote by $[-d,d]^4$ the Cartesian product $[-d,d]\times[-d,d]\times[-d,d]\times[-d,d]$. Obviously, $[-30,30]^4$ is large enough to cover the whole attractor, and hence can be taken as the sample space. Liang and Kleeman [51] discretize this space into 30^4 bins. With a Gaussian initial distribution $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu}=\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{bmatrix},\qquad \boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1^2 & 0 & 0 & 0\\ 0 & \sigma_2^2 & 0 & 0\\ 0 & 0 & \sigma_3^2 & 0\\ 0 & 0 & 0 & \sigma_4^2\end{bmatrix}$$

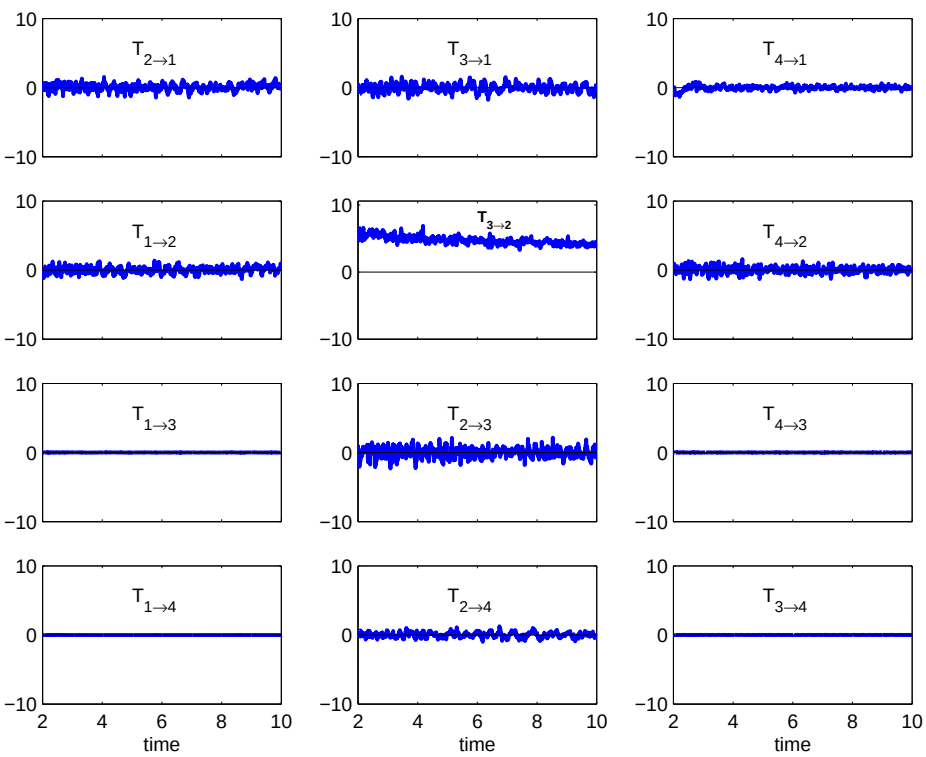
they generate an ensemble of 2,560,000 members, each steered independently under the system (73)–(76). The details about the sample space discretization, probability estimation, *etc*., are referred to [51]. Shown in the following are only the major results.

Between the four components of the TBS system, pairwise there are 12 information flows, namely,

$$\begin{array}{lll} T_{2\to1}, & T_{3\to1}, & T_{4\to1}\\ T_{1\to2}, & T_{3\to2}, & T_{4\to2}\\ T_{1\to3}, & T_{2\to3}, & T_{4\to3}\\ T_{1\to4}, & T_{2\to4}, & T_{3\to4}\end{array}$$

To compute these flows, Liang and Kleeman [51] have tried different parameters $\boldsymbol{\mu}$ and σ_k^2 $(k=1,2,3,4)$, but found the final results are the same after $t=2$ when the trajectories are attracted into the invariant set. It therefore suffices to show the result of just one experiment: $\mu_k=9$ and $\sigma_k^2=9$, $k=1,2,3,4$.

Figure 5. Information flows within the 4D truncated Burgers-Hopf system. The series prior to $t=2$ are not shown because some trajectories have not entered the attractor by that time.



Plotted in Figure 5 are the 12 flow rates. First observe that $T_{3\to4} = T_{4\to3} = 0$. This is easy to understand, as both F_3 and F_4 in Equations (75) and (76) have no dependence on x_3 nor on x_4, implying a zero flow in either direction between the pair (x_3, x_4) by the property of causality. What makes the result remarkable is, besides $T_{3\to4}$ and $T_{4\to3}$, essentially all the flows, except $T_{3\to2}$, are negligible, although obvious oscillations are found for $T_{2\to1}$, $T_{3\to1}$, $T_{1\to2}$, $T_{4\to1}$, $T_{2\to3}$, and $T_{2\to4}$. The only significant flow, *i.e.*, $T_{3\to2}$, means that, within the TBS system, it is the fine component that causes an increase in uncertainty in a coarse component but not conversely. Originally the TBS was introduced by Majda and Timofeyev [55] to test their stochastic closure scheme that models the unresolved high Fourier modes. Since additive noises are independent of the state variables, information can only be transferred from the former to the latter. The transfer asymmetry observed here is thus reflected in the scheme.

6.4. Langevin Equation

Most of the applications of information flow/transfer are expected with stochastic systems. Here we illustrate this with a simple 2D system, which has been studied in reference [54] for the validation of Equation (52):

$$d\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}dt + \boldsymbol{B}d\boldsymbol{w} \tag{77}$$

where $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$ are 2×2 constant matrices. This is the linear version of Equation (47). Linear systems are particular in that, if initialized with a normally distributed ensemble, then the distribution of the variables will be a Gaussian subsequently (e.g., [56]). This greatly simplifies the computation which, as we have seen in the previous subsection, is often a formidable task. Let $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ is the mean vector, and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$ the covariance matrix; they evolve as

$$d\boldsymbol{\mu}/dt = \boldsymbol{A}\,\boldsymbol{\mu} \tag{78a}$$

$$d\boldsymbol{\Sigma}/dt = \boldsymbol{A}\,\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\,\boldsymbol{A}^T + \boldsymbol{B}\,\boldsymbol{B}^T \tag{78b}$$

($\boldsymbol{B}\boldsymbol{B}^T$ is the matrix (g_{ij}) we have seen in Section 5), which determine the joint density of $\boldsymbol{x}$:

$$\rho(\boldsymbol{x}) = \frac{1}{2\pi\left(\det \boldsymbol{\Sigma}\right)^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \tag{79}$$

By Theorem 5.1, the rates of information flow thus can be accurately computed.

Several sets of parameters have been chosen in [54] to study the model behavior. Here we just look at one such choice: $\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\boldsymbol{A} = \begin{pmatrix} -0.5 & 0.1 \\ 0 & -0.5 \end{pmatrix}$. Its corresponding mean and covariance approach to an equilibrium: $\boldsymbol{\mu}(\infty) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\boldsymbol{\Sigma}(\infty) = \begin{pmatrix} 2.44 & 2.2 \\ 2.2 & 2 \end{pmatrix}$. Shown in Figure 6 are the time evolutions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ initialized with $\boldsymbol{\mu}(0) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\boldsymbol{\Sigma}(0) = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$, and a sample

path of $\boldsymbol{x}$ starting from $(1, 2)$. The computed rates of information flow, $T_{2\to1}$ and $T_{1\to2}$, are plotted in Figure 7a and b. As time moves on, $T_{2\to1}$ increases monotonically and eventually approaches a constant; on the other hand, $T_{1\to2}$ vanishes throughout. While this is within one's expectations, since $dx_2 = -0.5x_2dt + dw_1 + dw2$ has no dependence on x_1 and hence there should be no transfer of information from x_1 to x_2, it is interesting to observe that, in contrast, the typical paths of x_1 and x_2 could be highly correlated, as shown in Figure 6c. In other words, for two highly correlated time series, say $x_1(t)$ and $x_2(t)$, one series may have nothing to do with the other. This is a good example illustrating how information flow extends the classical notion of correlation analysis, and how it may be potentially utilized to identify the causal relation between complex dynamical events.

Figure 6. A solution of Equation (78), the model examined in [54], with $a_{21} = 0$ and initial conditions as shown in the text: (**a**) $\boldsymbol{\mu}$; (**b**) $\boldsymbol{\Sigma}$; and (**c**) a sample path starting from (1,2).

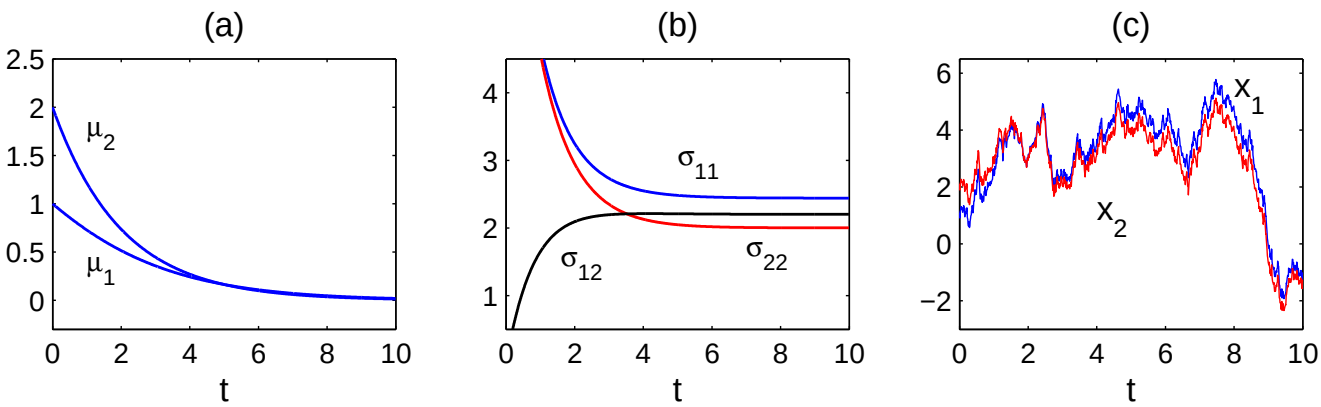
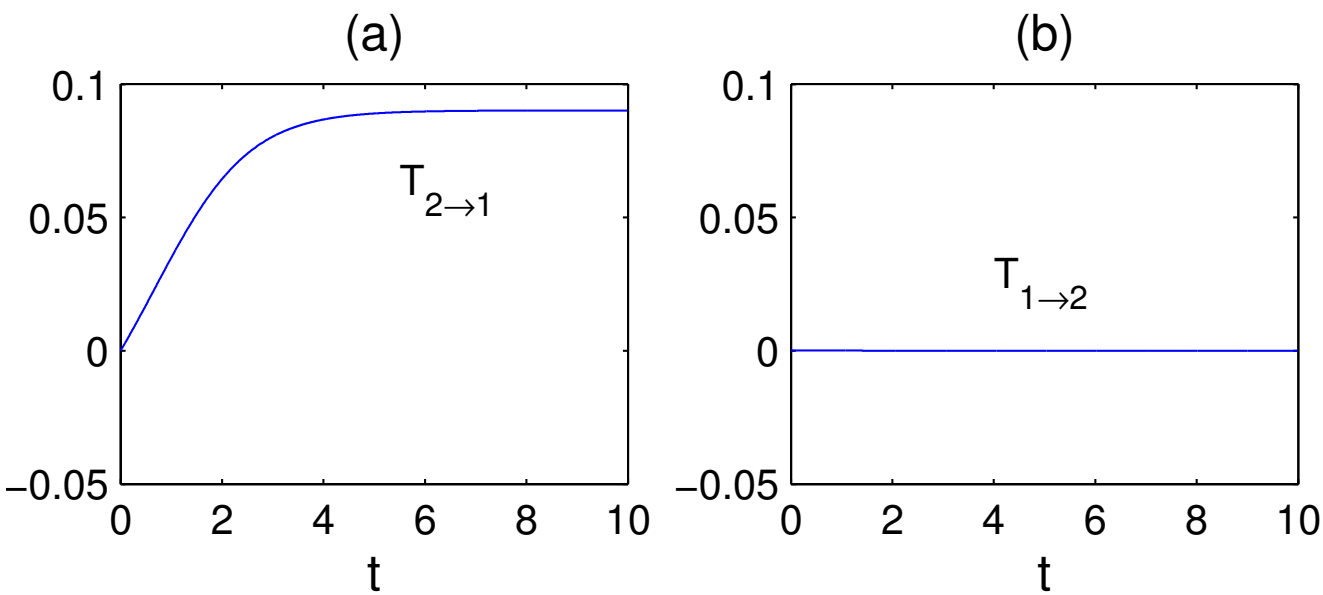


Figure 7. The computed rates of information flow for the system (77): (a) $T_{2\to1}$, (b) $T_{1\to2}$.



7. Summary

The past decades have seen a surge of interest in information flow (or information transfer, as it is sometimes called) in different fields of scientific research, mostly in the appearance of some empirical/half-empirical form. We have shown that, given a dynamical system, deterministic or stochastic, this important notion can actually be formulated on a rigorous footing, with flow measures explicitly derived. The general results are summarized in the theorems in Sections 3, 4 and 5. For two-dimensional systems, the result is fairly tight. In fact, if writing such a system as

$$\begin{cases} dx_1 = F_1(\boldsymbol{x},t)dt + b_{11}(\boldsymbol{x},t)dw_1 + b_{12}(\boldsymbol{x},t)dw_2 \\ dx_2 = F_2(\boldsymbol{x},t)dt + b_{21}(\boldsymbol{x},t)dw_1 + b_{22}(\boldsymbol{x},t)dw_2 \end{cases}$$

where (w_1, w_2) are standard Wiener processes, we have a rate of information flowing from x_2 to x_1,

$$T_{2\to 1} = -E\left(F_1 \frac{\partial \log \rho_1}{\partial x_1}\right) - E\left(\frac{\partial F_1}{\partial x_1}\right) + \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 (g_{11}\rho_1)}{\partial x_1^2}\right)$$

This is an alternative expression of that in Theorem 5.1; $T_{1\to 2}$ can be obtained by switching the subscripts 1 and 2. In the formula, $g_{11} = \sum_k b_{1k}^2$, ρ_1 is the marginal density of x_1, and E stands for mathematical expectation with respect to ρ, *i.e.*, the joint probability density. On the right-hand side, the third term is contributed by the Brownian notion; if the system is deterministic, this term vanishes. In the remaining two terms, the first is the tendency of H_1, namely the marginal entropy of x_1; the second can be interpreted as the rate of H_1 increase on x_1 its own, thanks to the law of entropy production (12) [49], which we restate here:

For an n-dimensional system $\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x},t)$, its joint entropy H evolves as $\frac{dH}{dt} = E(\nabla \cdot \mathbf{F})$

This interpretation lies at the core of all the theories along this line. It illustrates that the marginal entropy increase of a component, say, x_1, is due to two different mechanisms: the information transferred from some component, say, x_2, and the marginal entropy increase associated with a system without taking x_2 into account. On this ground, the formalism is henceforth established, with respect to discrete mappings, continuous flows, and stochastic systems, respectively. Correspondingly, the resulting measures are summarized in Equations (24), (39) and (52).

The above-obtained measures possess several interesting properties, some of which one may expect based on daily life experiences. The first one is a property of flow/transfer asymmetry, which has been set as the basic requirement for the identification of causal relations between dynamical events. The information flowing from one event to another event, denoted respectively as x_2 and x_1, may yield no clue about its counterpart in the opposite direction, *i.e.*, the flow/transfer from x_1 to x_2. The second says that, if the evolution of x_1 is independent of x_2, then the flow from x_2 to x_1 is zero. The third one is about the role of stochasticity, which asserts that, if the stochastic perturbation to the receiving component does not rely on the given component, the flow measure then has a form same as that for the corresponding deterministic system. As a direct corollary, when the noise is additive, then in terms of information flow, the stochastic system functions in a deterministic manner.

The formalism has been put to application with benchmark dynamical systems. In the context of the baker transformation, it is found that there is always information flowing from the stretching

direction to the folding direction, while no flow exists conversely. This is in agreement with what one would observe in kneading dough. Application to the Hénon map also yields a result just as expected on physical grounds. In a more complex case, the formalism has been applied to the study of the scale–scale interaction and information flow between the first two modes of the chaotic truncated Burgers equation. Surprisingly, all the twelve flows are essentially zero, save for one strong flow from the high-frequency mode to the low-frequency mode. This demonstrates that the route of information flow within a dynamical system, albeit seemingly complex, could be simple. In another application, we test how one may control the information flow by tuning the coefficients in a two-dimensional Langevin system. A remarkable observation is that, for two highly correlated time series, there could be no transfer from one certain series, say x_2, to the other (x_1). That is to say, the evolution of x_1 may have nothing to do with x_2, even though x_1 and x_2 are highly correlated. Information flow/transfer analysis thus extends the traditional notion of correlation analysis and/or mutual information analysis by providing a quantitative measure of causality between dynamical events, and this quantification is based firmly on a rigorous mathematical and physical footing.

The above applications are mostly with idealized systems; this is, to a large extent, intended for the validation of the obtained flow measures. Next, we would extend the results to more complex systems, and develop important applications to realistic problems in different disciplines, as envisioned in the beginning of this paper. The scale–scale information flow within the Burgers–Hopf system in § 6.3, for example, may be extended to the flow between *scale windows*. By a scale window we mean, loosely, a subspace with a range of scales included (cf. [57]). In atmosphere–ocean science, important phenomena are usually defined on scale windows, rather than on individual scales (e.g., [58]). As discussed in [53], the dynamical core of the atmosphere and ocean general circulation models is essentially a quadratically nonlinear system, with the linear and nonlinear operators possessing certain symmetry resulting from some conservation properties (such as energy conservation). Majda and Harlim [53] argue that the state space may be decomposed into a direct sum of scale windows which inherit evolution properties from the quadratic system, and then information flow/transfer may be investigated between these windows. Intriguing as this conceptual model might be, there still exist some theoretical difficulties. For example, the governing equation for a window may be problem-specific; there may not be such governing equations as simply written as those like Equation (3) for individual components. Hence one may need to seek new ways to the derivation of the information flow formula. Nonetheless, central at the problem is still the aforementioned classification of mechanisms that govern the marginal entropy evolution; we are expecting new breakthroughs along this line of development.

The formalism we have presented thus far is with respect to Shannon entropy, or absolute entropy as one may choose to refer to it. In many cases, such as in the El Niño case where predictability is concerned, this may need to be modified, since the predictability of a dynamical system is measured by relative entropy. Relative entropy is also called Kullback–Leibler divergence; it is defined as

$$D(\rho\|q) = E_\rho\left[\log\left(\frac{\rho}{q}\right)\right]$$

i.e., the expectation of the logarithmic difference between a probability ρ and another reference probability q, where the expectation is with respect to ρ. Roughly it may be interpreted as the "distance" between ρ and q, though it does not satisfy all the axioms for a distance functional.

Therefore, for a system, if letting the reference density be the initial distribution, its relative entropy at a time t informs how much additional information is added (rather than how much information it has). This provides a natural choice for the measure of the utility of a prediction, as pointed out by Kleeman (2002) [59]. Kleeman also argues in favor of relative entropy because of its appealing properties, such as nonnegativity and invariance under nonlinear transformations [60]. Besides, in the context of a Markov chain, it has been proved that it always decreases monotonically with time, a property usually referred to as the generalized second law of thermodynamics (e.g., [60,61]). The concept of relative entropy is now a well-accepted measure of predictability (e.g., [59,62]). When predictability problems (such as those problems in atmosphere-ocean science and financial economics as mentioned in the introduction) are dealt with, it is necessary to extend the current formalism to one with respect to the relative entropy functional. For all the dynamical system settings in this review, the extension should be straightforward.

Acknowledgments

This study was supported by the National Science Foundation of China (NSFC) under Grant No. 41276032 to NUIST, by Jiangsu Provincial Government through the "Jiangsu Specially-Appointed Professor Program" (Jiangsu Chair Professorship), and by the Ministry of Finance of China through the Basic Research Funding to China Institute for Advanced Study.

References

1. Baptista, M.S.; Garcia, S.P.; Dana S.K.; Kurths, J. Transmission of information and synchronization in a pair of coupled chaotic circuits: An experimental overview. *Eur. Phys. J.-Spec. Top.* **2008**, *165*, 119–128.
2. Baptista, M.D.S.; Kakmeni, F.M.; Grebogi, C. Combined effect of chemical and electrical synapses in Hindmarsh-Rose neural networks on synchronization and the rate of information. *Phys. Rev. E* **2010**, *82*, 036203.
3. Bear, M.F.; Connors, B.W.; Paradiso, M.A. *Neuroscience: Exploring the Brain*; 3rd ed.; Lippincott Williams & Wilkins: Baltimore, MD, USA, 2007; p. 857.
4. Vakorin, V.A.; MiAiA, B.; Krakovska, O.; McIntosh, A.R. Empirical and theoretical aspects of generation and transfer of information in a neuromagnetic source network. *Front. Syst. Neurosci.* **2011**, *5*, 96.
5. Ay, N.; Polani, D. Information flows in causal networks. *Advs. Complex Syst.* **2008**, *11*, doi: 10.1142/S0219525908001465.
6. Peruani, F.; Tabourier, L. Directedness of information flow in mobile phone communication networks. *PLoS One* **2011**, *6*, e28860.
7. Sommerlade, L.; Amtage, F.; Lapp, O.; Hellwig, B.; Licking, C.H.; Timmer, J.; Schelter, B. On the estimation of the direction of information flow in networks of dynamical systems. *J. Neurosci. Methods* **2011**, *196*, 182–189.
8. Donner, R.; Barbosa, S.; Kurths, J.; Marwan, N. Understanding the earth as a complex system-recent advances in data analysis and modelling in earth sciences. *Eur. Phys. J.* **2009**, *174*, 1–9.

9. Kleeman, R. Information flow in ensemble weather prediction. *J. Atmos. Sci.* **2007**, *64*, 1005–1016.
10. Materassi, M.; Ciraolo, L.; Consolini, G.; Smith, N. Predictive space weather: An information theory approach. *Adv. Space Res.* **2011**, *47*, 877–885.
11. Tribbia, J.J. Waves, Information and Local Predictability. In Proceedings of the Workshop on Mathematical Issues and Challenges in Data Assimilation for Geophysical Systems: Interdisciplinary Perspectives, IPAM, UCLA, 22–25 February 2005.
12. Chen, C.R.; Lung, P.P.; Tay, N.S.P. Information flow between the stock and option markets: Where do informed traders trade? *Rev. Financ. Econ.* **2005**, *14*, 1–23.
13. Lee, S.S. Jumps and information flow in financial markets. *Rev. Financ. Stud.* **2012**, *25*, 439–479.
14. Sommerlade, L.; Eichler, M.; Jachan, M.; Henschel, K.; Timmer, J.; Schelter, B. Estimating causal dependencies in networks of nonlinear stochastic dynamical systems. *Phys. Rev. E* **2009**, *80*, 051128.
15. Zhao, K.; Karsai, M.; Bianconi, G. Entropy of dynamical social networks. *PLoS One* **2011**, doi:10.1371/journal.pone.0028116.
16. Cane, M.A. The evolution of El Niño, past and future. *Earth Planet. Sci. Lett.* **2004**, *164*, 1–10.
17. Jin, F.-F. An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. *J. Atmos. Sci.* **1997**, *54*, 811–829.
18. Philander, S.G. *El Niño, La Niña, and the Southern Oscillation*; Academic Press: San Diego, CA, USA, 1990.
19. Ghil, M.; Chekroun, M.D.; Simonnet, E. Climate dynamics and fluid mechanics: Natural variability and related uncertainties. *Physica D* **2008**, *237*, 2111–2126.
20. Mu, M.; Xu, H.; Duan, W. A kind of initial errors related to "spring predictability barrier" for El Niño events in Zebiak-Cane model. *Geophys. Res. Lett.* **2007**, *34*, L03709, doi:10.1029/2006GL027412.
21. Zebiak, S.E.; Cane, M.A. A model El Niño-Southern Oscillation. *Mon. Wea. Rev.* **1987**, *115*, 2262–2278.
22. Chen, D.; Cane, M.A. El Niño prediction and predictability. *J. Comput. Phys.* **2008**, *227*, 3625–3640.
23. Mayhew, S.; Sarin, A.; Shastri, K. The allocation of informed trading across related markets: An analysis of the impact of changes in equity-option margin requirements. *J. Financ.* **1995**, *50*, 1635–1654.
24. Goldenfield, N.; Woese, C. Life is physics: Evolution as a collective phenomenon far from equilibrium. *Ann. Rev. Condens. Matt. Phys.* **2011**, *2*, 375–399.
25. Küppers, B. *Information and the Origin of Life*; MIT Press: Cambridge, UK, 1990.
26. Murray, J.D. *Mathematical Biology*; Springer-Verlag: Berlin, Germany, 2000.
27. Allahverdyan, A.E.; Janzing, D.; Mahler, G. Thermodynamic efficiency of information and heat flow. *J. Stat. Mech.* **2009**, *PO9011*, doi:10.1088/1742-5468/2009/09/P09011.
28. Crutchfield, J.P., Shalizi, C.R. Thermodynamic depth of causal states: Objective complexity via minimal representation. *Phys. Rev. E* **1999**, *59*, 275–283.
29. Davies, P.C.W. The Physics of Downward Causation. In *The Re-emergence of Emergence*; Clayton, P., Davies, P.C.W., Eds.; Oxford University Press: Oxford, UK, 2006; pp. 35–52.

30. Ellis, G.F.R. Top-down causation and emergence: Some comments on mechanisms. *J. R. Soc. Interface* **2012**, *2*, 126–140.
31. Okasha, S. Emergence, hierarchy and top-down causation in evolutionary biology. *J. R. Soc. Interface* **2012**, *2*, 49–54.
32. Walker, S.I.; Cisneros, L.; Davies, P.C.W. Evolutionary transitions and top-down causation. arXiv:1207.4808v1 [nlin.AO], **2012**.
33. Wu, B.; Zhou, D.; Fu, F.; Luo, Q.; Wang, L.; Traulsen, A. Evolution of cooperation on stochastic dynamical networks. *PLoS One* **2010**, *5*, e11187.
34. Pope, S. *Turbulent Flows*; 8th ed.; Cambridge University Press: Cambridge, UK, 2011.
35. Faes, L.; Nollo, G.; Erla, S.; Papadelis, C.; Braun, C.; Porta, A. Detecting Nonlinear Causal Interactions between Dynamical Systems by Non-uniform Embedding of Multiple Time Series. In Proceedings of the Engineering in Medicine and Biology Society, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 102–105.
36. Kantz, H.; Shreiber, T. *Nonlinear Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2004.
37. Schindler-Hlavackova, K.; Palus, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approach in time series analysis. *Phys. Rep.* **2007**, *441*, 1–46.
38. Granger, C. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438.
39. McWilliams, J.C. The emergence of isolated, coherent vortices in turbulence flows. *J. Fluid Mech.* **1984**, *146*, 21–43.
40. Salmon, R. *Lectures on Geophysical Fluid Dynamics*; Oxford University Press: Oxford, UK, 1998; p. 378.
41. Bar-Yam, Y. *Dynamics of Complex Systems*; Addison-Welsley Press: Reading, MA, USA, 1997; p. 864.
42. Crutchfield, J.P. The calculi of emergence: computation, dynamics, and induction induction. "Special issue on the Proceedings of the Oji International Seminar: Complex Systems-From Complex Dynamics to Artifical Reality". *Physica D* **1994**, *75*, 11–54.
43. Goldstein, J. Emergence as a construct: History and issues. *Emerg. Complex. Org.* **1999**, *1*, 49–72.
44. Corning, P.A. The re-emergence of emergence: A venerable concept in search of a theory. *Complexity* **2002**, *7*, 18–30.
45. Vastano, J.A.; Swinney, H.L. Information transport in sptiotemporal systems. *Phys. Rev. Lett.* **1988**, *60*, 1773–1776.
46. Kaiser, A.; Schreiber, T. Information transfer in continuous processes. *Physica D* **2002** *166*, 43–62.
47. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461.
48. Liang, X.S.; Kleeman, R. A rigorous formalism of information transfer between dynamical system components. I. Discrete mapping. *Physica D* **2007**, *231*, 1–9.
49. Liang, X.S.; Kleeman, R. Information transfer between dynamical system components. *Phys. Rev. Lett.* **2005**, *95*, 244101.
50. Lasota, A.; Mackey, M.C. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*; Springer: New York, NY, USA, 1994.

51. Liang, X.S.; Kleeman, R. A rigorous formalism of information transfer between dynamical system components. II. Continuous flow. *Physica D* **2007**, *227*, 173–182.
52. Liang, X.S. Uncertainty generation in deterministic fluid flows: Theory and applications with an atmospheric stability model. *Dyn. Atmos. Oceans* **2011**, *52*, 51–79.
53. Majda, A.J.; Harlim, J. Information flow between subspaces of complex dynamical systems. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9558–9563.
54. Liang, X.S. Information flow within stochastic dynamical systems. *Phys. Rev. E* **2008**, *78*, 031113.
55. Majda, A.J.; Timofeyev, I. Remarkable statistical behavior for truncated Burgers-Hopf dynamics. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12413–12417.
56. Gardiner, C.W. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*; Springer-Verlag: Berlin/Heidelberg, Germany, 1985.
57. Liang, X.S.; Anderson, D.G.M.A. Multiscale window transform. *SIAM J. Multiscale Model. Simul.* **2007**, *6*, 437–467.
58. Liang, X.S.; Robinson, A.R. Multiscale processes and nonlinear dynamics of the circulation and upwelling events off Monterey Bay. *J. Phys. Oceanogr.* **2009**, *39*, 290–313.
59. Kleeman, R. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.* **2002**, *59*, 2057–2072.
60. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.
61. Ao, P. Emerging of stochastic dynamical equalities and steady state thermodynamics from Darwinian dynamics. *Commun. Theor. Phys.* **2008**, *49*, 1073–1090.
62. Tang, Y.; Deng, Z.; Zhou, X.; Cheng, Y.; Chen, D. Interdecadal variation of ENSO predictability in multiple models. *J. Clim.* **2008**, *21*, 4811–4832.