Available online at www.sciencedirect.com

ScienceDirect

Physica D 227 (2007) 173–182

PHYSICA D

www.elsevier.com/locate/physd

A rigorous formalism of information transfer between dynamical system components. II. Continuous flow

X. San Liang*, Richard Kleeman

Courant Institute of Mathematical Sciences, New York, NY 10012, USA

Received 24 January 2006; received in revised form 17 December 2006; accepted 22 December 2006
Available online 29 January 2007
Communicated by C.K.R.T. Jones

Abstract

The transfer of information between dynamical system components is formalized with causality faithfully represented. In a continuous system with many components, information transfer is a mechanism controlling the marginal entropy evolution of the target component. It is measured by the rate of entropy thus transferred, which is obtained through freezing the source component instantaneously, and comparing the entropy increases between the original system and the so modified system. The resulting transfer measure is consistent with our earlier 2D formalism derived in Liang and Kleeman [X.S. Liang, R. Kleeman, Information transfer between dynamical system components, Phys. Rev. Lett. 95 (24) (2005) 244101] using different methods; it also possesses a property of unidirectionalism which has been emphasized by Schreiber [T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2) (2000) 461–464]. We apply our formalism to a two-mode (four-dimensional) truncated Burgers–Hopf system. No significant information exchange is identified between the four components, save for a transfer from the cosine direction of mode 2 to the sine direction of mode 1. This transfer occurs continuously and at a nearly constant rate. The present work should serve as a starting point for the development of a rigorous dynamics-free formalism for the information transfer of multivariate time series.

Keywords: Information transfer; Causality; Entropy evolution; Continuous dynamical system; Truncated Burgers–Hopf system

1. Introduction

Information transfer is an important concept in communication, neuroscience, atmosphere–ocean science, among other fields, and has been a subject of active research since two decades ago. A brief review was given in the first part of this study, Liang and Kleeman [7] or Part I for short. As mentioned in there, many previous studies are based on mutual information, which lacks a representation of transfer directionality or causality between events, as emphasized by Schreiber [11]. To solve the problem, we developed in Part I a rigorous formalism with causality naturally embedded. It has been tested with the benchmark baker transformation and Hénon map, and the resulting information transfers are just as one may expect through intuitive physical reasoning.

In reality, many dynamical systems are continuous. One may find all kinds of examples in different disciplines (e.g., atmosphere–ocean science). This paper extends the discrete formalism of Part I to continuous dynamical systems. It should be pointed out that we already have such a formalism for two-dimensional (2D) flows [6]. However, as we will see, the 2D formalism, and moreover, the strategy for the 2D formalism, does not work for flows of dimensionality greater than 2. In this sense, this study is a generalization, but not a simple generalization, of what we have done in [6].

We first summarize the 2D formalism established in [6], then propose the strategy for this study (Section 2). A map Φ is built through discretizing the flow under consideration to make contact with the discrete formalism in Part I (Section 3). Properties of Φ are explored and the evolution of marginal entropy derived (Section 4). The central part of this study is Section 5, where the entropy increase of one component under Φ with another component fixed as a parameter is formulated. The information transfer, which is measured by the rate of entropy transfer, is obtained by taking limits (Section 6). In Section 7, we show that this formalism satisfies the

* Corresponding author.
E-mail address: sanliang@cims.nyu.edu (X.S. Liang).

0167-2789/$ - see front matter
doi:10.1016/j.physd.2006.12.012

unidirectionalism requirement, and verifies the 2D formalism obtained through a completely different avenue. Following this (Section 8), we investigate how information flows among the components of an interesting application, namely the truncated Burger–Hopf system. This study is concluded in Section 9.

2. The problem

Consider a generic n-dimensional autonomous system[1]:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = F_1(x_1, x_2, \ldots, x_n), \tag{1}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = F_2(x_1, x_2, \ldots, x_n), \tag{2}$$

$$\vdots \qquad \vdots$$

$$\frac{\mathrm{d}x_n}{\mathrm{d}t} = F_n(x_1, x_2, \ldots, x_n), \tag{3}$$

or in a vectorial form,

$$\frac{\mathrm{d}\underline{\mathbf{x}}}{\mathrm{d}t} = \underline{\mathbf{F}}(\underline{\mathbf{x}}). \tag{4}$$

Let $\{\underline{\mathbf{X}}; t\}$ be a stochastic process, where $\underline{\mathbf{X}} = (X_1, X_2, \ldots, X_n) \in \Omega$ are the random variables corresponding to $(x_1, x_2, \ldots, x_n)$, and Ω is the sample space; let $\rho = \rho(x_1, x_2, \ldots, x_n)$ be the joint probability density of $\underline{\mathbf{X}}$. As in Part I, we assume that Ω is equal to a Cartesian product of $\Omega_1, \Omega_2, \ldots,$ and Ω_n, the sample spaces of $X_1, X_2, \ldots,$ and X_n, respectively, and write

$$\Omega_{jn} \equiv \Omega_j \times \Omega_{j+1} \times \cdots \times \Omega_n, \quad j = 1, 2, \ldots, n-1 \tag{5}$$

throughout for notational convenience. Further assume that ρ vanishes at the boundaries of Ω, i.e., the extreme events have a measure of zero in the probability space. In [6], these assumptions have been justified for most real problems. More generic formalisms can be established, but the result is fundamentally the same except for some extra terms related to boundary fluxes. We will leave that to future work.

Without loss of generality, consider the information transfer from component X_2 to component X_1. As we did in [6], we first find the evolution law for the marginal entropy H_1 of X_1:

$$H_1 = H_1(X_1) = -\int_{\Omega_1} \rho_1 \log \rho_1 \, \mathrm{d}x_1, \tag{6}$$

where $\rho_1 = \rho_1(x_1) = \int_{\Omega_{2n}} \rho(x_1, x_2, \ldots, x_n) \mathrm{d}x_2 \ldots \mathrm{d}x_n$ is the marginal probability density of X_1, then select out the transfer among many mechanisms that govern the evolution. Usually these mechanisms are intertwined. When the system is of dimension 2, however, they can be naturally decomposed into a transfer from X_2, $T_{2\to 1}$, plus the evolution of X_1 on its own, $\frac{\mathrm{d}H_1^*}{\mathrm{d}t}$. Here the key is to find $\frac{\mathrm{d}H_1}{\mathrm{d}t}^*$. In [6], we have obtained

$$\frac{\mathrm{d}H_1^*}{\mathrm{d}t} = E\left(\frac{\partial F_1}{\partial x_1}\right). \tag{7}$$

[1] Autonomy is not essential, as we will see in the derivations below. We deal with autonomous systems in this paper in order to make the notation more concise and readable.

It thus follows that

$$T_{2\to 1} = \frac{\mathrm{d}H_1}{\mathrm{d}t} - E\left(\frac{\partial F_1}{\partial x_1}\right). \tag{8}$$

As in the discrete case of Part I, this strategy does not apply when $n \geq 3$; otherwise the transfer thus obtained would be the transfer to X_1 from all other components. But on the other hand, the above $\frac{\mathrm{d}H_1^*}{\mathrm{d}t}$ may be equally understood as the evolution of H_1 with the influence from X_2 excluded, namely, the time rate of change of H_1 with x_2 frozen at time t. In this spirit, we may then partition the mechanisms governing the evolution of H_1 into two disjoint subsets: one is the transfer from X_2, another one the evolution without influence from X_2, which we will hereafter denote as $\frac{\mathrm{d}H_{1\not{2}}}{\mathrm{d}t}$. Such a partitioning of course causes no restraints on the dimensionality of the system. The transfer is then the difference between $\frac{\mathrm{d}H_1}{\mathrm{d}t}$ and $\frac{\mathrm{d}H_{1\not{2}}}{\mathrm{d}t}$, i.e.,

$$T_{2\to 1} = \frac{\mathrm{d}H_1}{\mathrm{d}t} - \frac{\mathrm{d}H_{1\not{2}}}{\mathrm{d}t}. \tag{9}$$

Since $\frac{\mathrm{d}H_1}{\mathrm{d}t}$ is easy to obtain given the dynamics, the key to the formalism is finding $\frac{\mathrm{d}H_{1\not{2}}}{\mathrm{d}t}$. Based on the foregoing argument, entropy transfer is an instantaneous event, and $\frac{\mathrm{d}H_{1\not{2}}}{\mathrm{d}t}$ measures the rate of change of H_1 under an dynamical system modified instantaneously at time t. Mathematically this may be written as

$$\frac{\mathrm{d}H_{1\not{2}}}{\mathrm{d}t} = \lim_{\Delta t\to 0} \frac{H_{1\not{2}}(t+\Delta t) - H_1(t)}{\Delta t},$$

where $H_{1\not{2}}(t + \Delta t)$ obeys an evolution law invoked at time t, which is different from that $H_1(t)$ obeys. This makes it impossible to derive $\frac{\mathrm{d}H_{1\not{2}}}{\mathrm{d}t}$ from some known equations, such as the Liouville equation, where the dynamics is consistent throughout. We have to go back to the definition of derivatives and do the derivation.

We therefore approach the problem by studying how entropy is increased from time t to time $t + \Delta t$, and then we take the limit $\Delta t \to 0$. That is to say, we are going to utilize what we have established with discrete processes in Part I to establish the formalism. This is fulfilled in Sections 3–6.

3. Mapping Φ

To allow application of the discrete formalism in Part I, we need to construct out of Eqs. (1)–(3) an n-dimensional discrete system, which steers $\underline{\mathbf{x}}(t) = (x_1, x_2, \ldots, x_n)$ to $\underline{\mathbf{x}}(t + \Delta t)$. In order to avoid confusion in later derivations, we will hereafter write $\underline{\mathbf{x}}(t + \Delta t)$ as $\underline{\mathbf{y}} = (y_1, y_2, \ldots, y_n)$. Discretization of (1)–(3) gives a map

$$\Phi : \begin{cases} y_1 = x_1 + \Delta t \cdot F_1(\underline{\mathbf{x}}), \\ y_2 = x_2 + \Delta t \cdot F_2(\underline{\mathbf{x}}), \\ \vdots \qquad \vdots \\ y_n = x_n + \Delta t \cdot F_n(\underline{\mathbf{x}}), \end{cases} \tag{10}$$

which is an approximation of (1)–(3) up to the first order of the time step Δt. For later use, we denote the components of Φ as

$$\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n). \tag{11}$$

Because of its special form, Φ has some interesting properties. First, it is always invertible as Δt goes to zero; so are its components. In fact, the inverse is

$$\Phi^{-1}: \begin{cases} x_1 = y_1 - \Delta t \cdot F_1(\underline{\mathbf{y}}) + O(\Delta t^2), \\ x_2 = y_2 - \Delta t \cdot F_2(\underline{\mathbf{y}}) + O(\Delta t^2), \\ \vdots \quad \vdots \\ x_n = y_n - \Delta t \cdot F_n(\underline{\mathbf{y}}) + O(\Delta t^2). \end{cases} \tag{12}$$

As we are primarily interested in the limiting case when $\Delta t \to 0$, the symbol $O(\Delta t^2)$ may be omitted in future derivations.

The second interesting property we can learn from Φ is about its Jacobian J. By (10),

$$\begin{aligned} J &= \det\left[\frac{\partial(y_1, y_2, \ldots, y_n)}{\partial(x_1, x_2, \ldots, x_n)}\right] \\ &= \det\begin{bmatrix} 1+\Delta t\frac{\partial F_1}{\partial x_1} & \Delta t\frac{\partial F_1}{\partial x_2} & \cdots & \Delta t\frac{\partial F_1}{\partial x_n} \\ \Delta t\frac{\partial F_2}{\partial x_1} & 1+\Delta t\frac{\partial F_2}{\partial x_2} & \cdots & \Delta t\frac{\partial F_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta t\frac{\partial F_n}{\partial x_1} & \Delta t\frac{\partial F_n}{\partial x_2} & \cdots & 1+\Delta t\frac{\partial F_n}{\partial x_n} \end{bmatrix} \\ &= \prod_i \left(1+\Delta t\frac{\partial F_i}{\partial x_i}\right) + O(\Delta t^2) \\ &= 1 + \Delta t \sum_{i=1}^n \frac{\partial F_i}{\partial x_i} + O(\Delta t^2). \end{aligned} \tag{13}$$

Following the same procedure, the inverse of J can also be obtained:

$$J^{-1} = 1 - \Delta t \sum_{i=1}^n \frac{\partial F_i}{\partial x_i} + O(\Delta t^2). \tag{14}$$

Note that in the derivation, we have interchanged the arguments of $\underline{\mathbf{F}}$ freely between $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$. The resulting error is in the higher order terms.

The third useful property of Φ is that its associated Frobenius–Perron operator $\mathcal{P}$ (F–P operator hereafter) can be explicitly evaluated. By definition, $\mathcal{P}$ is an operator driving the probability density function ρ from t to $t + \Delta t$, such that

$$\int_\omega \mathcal{P}\rho(\underline{\mathbf{x}})\,\mathrm{d}\underline{\mathbf{x}} = \int_{\Phi^{-1}(\omega)} \rho(\underline{\mathbf{x}})\,\mathrm{d}\underline{\mathbf{x}}, \tag{15}$$

for any subset of Ω, ω. As established above, Φ is invertible. This yields

$$\begin{aligned} \mathcal{P}\rho(y_1, \ldots, y_n) &= \rho\left(\Phi^{-1}(y_1, \ldots, y_n)\right)|J^{-1}| \\ &= \rho(x_1, x_2, \ldots, x_n)|J^{-1}|, \end{aligned} \tag{16}$$

where J^{-1} is the inverse of the Jacobian given by (14) [5].

We have shown in [6] that, in the 2D case, invertibility of Φ implies that the increase in entropy for the system is given by a simple form $\Delta H = E \log |J|$. This is also true for a system with many dimensions. In this context, the entropy at time $t + \Delta t$ is

$$\begin{aligned} H(t+\Delta t) &= -\int_\Omega \mathcal{P}\rho(\underline{\mathbf{y}}) \log \mathcal{P}\rho(\underline{\mathbf{y}})\,\mathrm{d}\underline{\mathbf{y}} \\ &= -\int_\Omega \rho(\underline{\mathbf{y}} - \Delta t \cdot \underline{\mathbf{F}}(\underline{\mathbf{y}})) \cdot |J^{-1}| \\ &\quad \cdot \log\left[\rho(\underline{\mathbf{y}} - \Delta t \cdot \underline{\mathbf{F}}(\underline{\mathbf{y}})) \cdot |J^{-1}|\right] \mathrm{d}\underline{\mathbf{y}}. \end{aligned}$$

Make a change of variable:

$$\underline{\mathbf{x}} = \underline{\mathbf{y}} - \Delta t \cdot \underline{\mathbf{F}}(\underline{\mathbf{y}}) = \Phi^{-1}(\underline{\mathbf{y}}),$$

and notice the invariance of Ω upon transformation. The above equation becomes

$$\begin{aligned} H(t+\Delta t) &= -\int_\Omega \rho(\underline{\mathbf{x}})|J^{-1}| \cdot \log\left(\rho(\underline{\mathbf{x}})|J^{-1}|\right) \cdot |J|\,\mathrm{d}\underline{\mathbf{x}} \\ &= H(t) + E \log |J|. \end{aligned} \tag{17}$$

So

$$\Delta H = H(t+\Delta t) - H(t) = E \log |J|. \tag{18}$$

Let $\Delta t \to 0$,

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

or

$$\frac{\mathrm{d}H}{\mathrm{d}t} = E(\nabla \cdot \underline{\mathbf{F}}). \tag{19}$$

This is the result we have obtained in [6], via a different approach.

4. Time change of H_1

We now compute the Shannon entropy for X_1 at time $t+\Delta t$. Given ρ at time t, it is

$$H_1(t+\Delta t) = -\int_{\Omega_1} (\mathcal{P}\rho)_1(y_1) \cdot \log(\mathcal{P}\rho)_1(y_1)\,\mathrm{d}y_1, \tag{20}$$

where

$$\begin{aligned} (\mathcal{P}\rho)_1(y_1) &= \int_{\Omega_{2n}} \mathcal{P}\rho(y_1, y_2, \ldots, y_n)\,\mathrm{d}y_2 \ldots \mathrm{d}y_n \\ &= \int_{\Omega_{2n}} \rho(y_1 - \Delta t\ F_1(\underline{\mathbf{y}}), y_2 - \Delta t\ F_2(\underline{\mathbf{y}}), \ldots, y_n \\ &\quad - \Delta t\ F_n(\underline{\mathbf{y}}))|J^{-1}|\mathrm{d}y_2 \ldots \mathrm{d}y_n \end{aligned}$$

is the marginal distribution at time $t + \Delta t$. Make the variable transformation $x_i = y_i - \Delta t\ F_i(\underline{\mathbf{y}})$, for $i = 2, \ldots, n$. The marginal distribution then becomes

$$\begin{aligned} (\mathcal{P}\rho)_1(y_1) &= \int_{\Omega_{2n}} \rho(y_1 - \Delta t\ F_1(\underline{\mathbf{y}}), x_2, \ldots, x_n) \\ &\quad \cdot |J^{-1}| \cdot |J_{2n}|\mathrm{d}x_2 \ldots \mathrm{d}x_n. \end{aligned} \tag{21}$$

Here we use J_{2n} to signify the determinant of the Jacobian $\left[\frac{\partial(y_2,y_3,\ldots,y_n)}{\partial(x_2,x_3,\ldots,x_n)}\right]$. It is evaluated as

$$J_{2n} = \det \begin{bmatrix} 1+\Delta t\dfrac{\partial F_2}{\partial x_2} & \Delta t\dfrac{\partial F_2}{\partial x_3} & \cdots & \Delta t\dfrac{\partial F_2}{\partial x_n} \\ \Delta t\dfrac{\partial F_3}{\partial x_2} & 1+\Delta t\dfrac{\partial F_3}{\partial x_3} & \cdots & \Delta t\dfrac{\partial F_3}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta t\dfrac{\partial F_n}{\partial x_2} & \Delta t\dfrac{\partial F_n}{\partial x_3} & \cdots & 1+\Delta t\dfrac{\partial F_n}{\partial x_n} \end{bmatrix}$$
$$= 1 + \Delta t \sum_{i=2}^{n} \frac{\partial F_i}{\partial x_i} + O(\Delta t^2), \tag{22}$$

which gives

$$|J^{-1}| \cdot |J_{2n}| = \left(1 + \Delta t \sum_{2}^{n} \frac{\partial F_i}{\partial x_i}\right) \cdot \left(1 - \Delta t\ \nabla \cdot \underline{\mathbf{F}}\right) + O(\Delta t^2)$$
$$= 1 - \Delta t \frac{\partial F_1}{\partial x_1} + O(\Delta t^2). \tag{23}$$

Recall that $J_1^{-1} = 1 - \Delta t \frac{\partial F_1}{\partial x_1} + O(\Delta t^2)$. So

$$|J^{-1}| \cdot |J_{2n}| = |J_1^{-1}| + O(\Delta t^2),$$

and hence

$$(\mathcal{P}\rho)_1(y_1) = \int_{\Omega_{2n}} \rho(y_1 - \Delta t\ F_1(\underline{\mathbf{y}}), x_2, \ldots, x_n) \times |J_1^{-1}| \mathrm{d}x_2 \ldots \mathrm{d}x_n. \tag{24}$$

This integral is still difficult to evaluate, as $F_1(\underline{\mathbf{y}})$ is in general a function of all the components of $\underline{\mathbf{y}}$ (and hence $x_2, \ldots, x_n$), which makes the integration with respect to the second through the nth variables intertwined with the first argument of ρ. We solve this problem by expanding $\rho(y_1 - \Delta t\ F_1(\underline{\mathbf{y}}), x_2, \ldots, x_n)$ around $(y_1, x_2, \ldots, x_n)$:

$$\rho(y_1 - \Delta t\ F_1(\underline{\mathbf{y}}), x_2, \ldots, x_n)$$
$$= \rho(y_1, x_2, \ldots, x_n) - \Delta t \frac{\partial \rho(y_1, x_2, \ldots, x_n)}{\partial y_1} F_1(\underline{\mathbf{y}}) + O(\Delta t^2)$$
$$= \rho(y_1, x_2, \ldots, x_n) - \Delta t \frac{\partial \rho}{\partial y_1} F_1(y_1, x_2, \ldots, x_n) + O(\Delta t^2). \tag{25}$$

Here we have used the fact that $F_1(y_1, y_2, \ldots y_n) = F_1(y_1, x_2, \ldots x_n) + O(\Delta t)$. In the last equation, both ρ and F_1 are functions of $(y_1, x_2, \ldots, x_n)$. So Eq. (24) becomes

$$(\mathcal{P}\rho)_1(y_1)$$
$$= \int_{\Omega_{2n}} \left[\rho(y_1, x_2, \ldots, x_n) - \Delta t \frac{\partial \rho}{\partial y_1} F_1(y_1, x_2, \ldots, x_n)\right] \times \left(1 - \Delta t \frac{\partial F_1}{\partial y_1}\right) \mathrm{d}x_2 \mathrm{d}x_3 \ldots \mathrm{d}x_n$$
$$= \rho_1(y_1) - \Delta t \int_{\Omega_{2n}} \left(\rho \frac{\partial F_1}{\partial y_1} + \frac{\partial \rho}{\partial y_1} F_1\right) \mathrm{d}x_2 \ldots \mathrm{d}x_n + O(\Delta t^2)$$
$$= \rho_1(y_1) - \Delta t \int_{\Omega_{2n}} \frac{\partial \rho F_1}{\partial y_1} \mathrm{d}x_2 \ldots \mathrm{d}x_n + O(\Delta t^2), \tag{26}$$

and hence

$$\log(\mathcal{P}\rho)_1(y_1) = \log \rho_1(y_1) - \Delta t \int_{\Omega_{2n}} \frac{1}{\rho_1(y_1)} \frac{\partial \rho F_1}{\partial y_1} \mathrm{d}x_2 \ldots \mathrm{d}x_n + O(\Delta t^2). \tag{27}$$

These can be substituted into Eq. (20) to yield the entropy for component X_1:

$$H_1(t+\Delta t) = -\int_{\Omega_1} \left(\rho_1(y_1) - \Delta t \int_{\Omega_{2n}} \frac{\partial \rho F_1}{\partial y_1} \mathrm{d}x_2 \ldots \mathrm{d}x_n\right) \times \left(\log \rho_1(y_1) - \Delta t \int_{\Omega_{2n}} \frac{1}{\rho_1(y_1)} \frac{\partial \rho F_1}{\partial y_1} \mathrm{d}x_2 \ldots \mathrm{d}x_n\right) \mathrm{d}y_1 + O(\Delta t^2)$$
$$= H_1(t) + \Delta t \int_{\Omega} \log \rho_1(y_1) \frac{\partial \rho F_1}{\partial y_1} \mathrm{d}y_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n + \Delta t \int_{\Omega} \frac{\partial \rho F_1}{\partial y_1} \mathrm{d}y_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n + O(\Delta t^2). \tag{28}$$

Now y_1 is a dummy variable in the integrand. We may write it as x_1 to simplify the notation. So

$$\frac{\mathrm{d}H_1}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{H_1(t+\Delta t) - H_1(t)}{\Delta t}$$
$$= \int_{\Omega} (1 + \log \rho_1) \frac{\partial \rho F_1}{\partial x_1} \mathrm{d}\underline{\mathbf{x}}. \tag{29}$$

Using the zero-probability assumption at the boundaries, this is also

$$\frac{\mathrm{d}H_1}{\mathrm{d}t} = \int_{\Omega} \log \rho_1 \frac{\partial \rho F_1}{\partial x_1} \mathrm{d}\underline{\mathbf{x}}. \tag{30}$$

5. Time change of H_1 with x_2 as a parameter

The key to the formalism of information transfer is to find the entropy increase in direction 1 from t to $t + \Delta t$ under Φ with x_2 frozen instantaneously at time t, given $X_1(t)$. Denote the entropy thus obtained at $t + \Delta t$ as $H_{1\not 2}$. As in Part I, here we use the symbol $\not j$ to signify that component j is frozen, or that component j is excluded from a set of n independent variables. For example,

$$\rho_{\not 2} = \rho_{\not 2}(x_1, x_3, \ldots, x_n) = \int_{\Omega_2} \rho(x_1, x_2, x_3, \ldots, x_n) \mathrm{d}x_2, \tag{31}$$

$$\rho_{\not 1 \not 2} = \rho_{\not 1 \not 2}(x_3, \ldots, x_n) = \int_{\Omega_1 \times \Omega_2} \rho(x_1, x_2, x_3, \ldots, x_n) \mathrm{d}x_1 \mathrm{d}x_2. \tag{32}$$

This convention will be used throughout the paper without further clarification.

Based on the foregoing argument, the marginal entropy for the first component evolved from H_1 with contribution from X_2 excluded since time t is

$$H_{1\not 2}(t+\Delta t) = -\int_\Omega (\mathcal{P}\rho)_{1\not 2}(y_1)\cdot \log(\mathcal{P}\rho)_{1\not 2}(y_1) \cdot\rho(x_2|x_1,x_3,\ldots,x_n)\cdot \rho_{3\ldots n}(x_3,\ldots,x_n) \times \mathrm{d}y_1\mathrm{d}x_2\mathrm{d}x_3\ldots\mathrm{d}x_n, \tag{33}$$

where $y_1 = x_1 + \Delta t\ F_1(\underline{\mathbf{x}})$. In this equation, $(\mathcal{P}\rho)_{1\not 2}(y_1)$ represents the marginal density of X_1 at time $t+\Delta t$, as the density evolves from $\rho_{\not 2}$ at time t to time $t+\Delta t$ under the transformation

$$\Phi_{\not 2} : \begin{cases} y_1 = x_1 + \Delta t\cdot F_1(\underline{\mathbf{x}}) \\ y_3 = x_3 + \Delta t\cdot F_3(\underline{\mathbf{x}}) \\ \vdots \quad \vdots \\ y_n = x_n + \Delta t\cdot F_n(\underline{\mathbf{x}}) \end{cases}, \tag{34}$$

i.e., the map Φ in (10) with x_2 frozen instantaneously at t as a parameter. As Φ and its components are invertible, $\Phi_{\not 2}$ is also invertible, and its inverse is the same as Φ^{-1} in (12) except that now x_2 is a parameter. By (16),

$$\begin{aligned}(\mathcal{P}\rho)_{1\not 2}(y_1) &= \int_{\Omega_{3n}} \rho_{\not 2}(y_1-\Delta t\ F_1, y_3-\Delta t\ F_3,\ldots,y_n-\Delta t\ F_n) \times |J_{\not 2}^{-1}|\mathrm{d}y_3\ldots\mathrm{d}y_n \\ &= \int_{\Omega_{3n}} \rho_{\not 2}(\underline{\mathbf{y}}_{\not 2} - \Delta t\ \underline{\mathbf{F}}_{\not 2})|J_{\not 2}^{-1}|\mathrm{d}y_3\ldots\mathrm{d}y_n,\end{aligned} \tag{35}$$

where $\underline{\mathbf{F}}_{\not 2} = (F_1, F_3, \ldots, F_n)$ is understood as functions of $(y_1, x_2, y_3, \ldots, y_n)$, and

$$\begin{aligned} J_{\not 2}^{-1} &= \det\left[\frac{\partial(x_1,x_3,\ldots,x_n)}{\partial(y_1,y_3,\ldots,y_n)}\right] \\ &= 1-\Delta t\sum_{i\neq 2}\frac{\partial F_i}{\partial x_i} + O(\Delta t^2).\end{aligned} \tag{36}$$

Make the change of variables, $x_i = y_i - \Delta t\cdot F_i(y_1, x_2, y_3,\ldots,y_n)$, for $i=3,4,\ldots,n$. The Jacobian associated with this transformation is

$$\begin{aligned} J_{3n} &= \det\left[\frac{\partial(y_3,y_4,\ldots,y_n)}{\partial(x_3,x_4,\ldots,x_n)}\right] \\ &= 1+\Delta t\sum_{i=3}^{n}\frac{\partial F_i}{\partial x_i} + O(\Delta t^2),\end{aligned} \tag{37}$$

which gives

$$|J_{\not 2}^{-1}|\cdot|J_{3n}| = 1-\Delta t\frac{\partial F_1}{\partial x_1} + O(\Delta t^2). \tag{38}$$

With these substituted (35) becomes

$$\begin{aligned}(\mathcal{P}\rho)_{1\not 2}(y_1) &= \int_{\Omega_{3n}} \rho_{\not 2}(y_1-\Delta t\ F_1, x_3,\ldots,x_n) \cdot|J_{\not 2}^{-1}|\cdot|J_{3n}|\mathrm{d}x_3\ldots\mathrm{d}x_n \\ &= \int_{\Omega_{3n}} \rho_{\not 2}(y_1-\Delta t\ F_1, x_3,\ldots,x_n) \cdot\left(1-\Delta t\frac{\partial F_1}{\partial x_1}\right)\mathrm{d}x_3\ldots\mathrm{d}x_n + O(\Delta t^2) \\ &= \left[\rho_{\not 2}(y_1,x_3,\ldots,x_n) - \Delta t\frac{\partial\rho_{\not 2}}{\partial y_1}F_1\right]\left(1-\Delta t\frac{\partial F_1}{\partial x_1}\right) \times\mathrm{d}x_3\ldots\mathrm{d}x_n + O(\Delta t^2) \\ &= \rho_1(y_1) - \Delta t\cdot\int_{\Omega_{3n}}\left[\frac{\partial F_1}{\partial x_1}\rho_{\not 2}(y_1,x_3,\ldots,x_n) + F_1\frac{\partial\rho_{\not 2}(y_1,x_3,\ldots,x_n)}{\partial y_1}\right]\mathrm{d}x_3\ldots\mathrm{d}x_n + O(\Delta t^2).\end{aligned} \tag{39}$$

Since x_1 and y_1 are interchangeable up to an order of Δt, the two terms in the bracket can be combined, with the residual going to the higher order terms. That is to say,

$$(\mathcal{P}\rho)_{1\not 2}(y_1) = \rho_1(y_1) - \Delta t\cdot\int_{\Omega_{3n}}\frac{\partial F_1\rho_{\not 2}}{\partial y_1}\mathrm{d}x_3\ldots\mathrm{d}x_n + O(\Delta t^2). \tag{40}$$

Hence

$$\log(\mathcal{P}\rho)_{1\not 2}(y_1) = \log\rho_1(y_1) - \Delta t\cdot\int_{\Omega_{3n}}\frac{1}{\rho_1(y_1)}\frac{\partial F_1\rho_{\not 2}}{\partial y_1}\mathrm{d}x_3\ldots\mathrm{d}x_n + O(\Delta t^2). \tag{41}$$

So far we have evaluated the two parts of $H_{1\not 2}(t+\Delta t)$ in (33), $(\mathcal{P}\rho)_{1\not 2}(y_1)$ and $\log(\mathcal{P}\rho)_{1\not 2}(y_1)$. The evaluation is around $(y_1, x_3, x_4, \ldots x_n)$. In (33), there is another term $\rho(x_2 \mid x_1, x_3, \ldots, x_n)$ which is the probability of x_2 at time t conditioned on $(x_1, x_3, \ldots, x_n)$. Notice that the first component on which it is conditioned is x_1, i.e., a state of X_1 at time t. We need to make it around y_1, a state of X_1 at time $t+\Delta t$ to facilitate the integration;

$$\begin{aligned}\rho(x_2\mid x_1,x_3,\ldots,x_n) &= \frac{\rho(\underline{\mathbf{x}})}{\rho_{\not 2}(x_1,x_3,\ldots,x_n)} \\ &= \frac{\rho(y_1,x_2,x_3,\ldots,x_n) - \Delta t\frac{\partial\rho}{\partial y_1}F_1}{\rho_{\not 2}(y_1,x_3,\ldots,x_n) - \frac{\partial\rho_{\not 2}}{\partial y_1}F_1\Delta t} + O(\Delta t^2) \\ &= \frac{1}{\rho_{\not 2}(y_1,x_3,\ldots,x_n)}\left[\rho(y_1,x_2,x_3,\ldots,x_n) - \Delta t\frac{\partial\rho}{\partial y_1}F_1\right] \times\left[1+\frac{1}{\rho_{\not 2}(y_1,x_3,\ldots,x_n)}\frac{\partial\rho_{\not 2}}{\partial y_1}F_1\Delta t\right] + O(\Delta t^2) \\ &= \rho(x_2\mid y_1,x_3,\ldots,x_n) + \rho(x_2\mid y_1,x_3,\ldots,x_n) \cdot\frac{\partial\log\rho_{\not 2}}{\partial y_1}\cdot F_1\Delta t - \frac{1}{\rho_{\not 2}}\frac{\partial\rho}{\partial y_1}F_1\Delta t + O(\Delta t^2).\end{aligned} \tag{42}$$

Substituting (40)–(42) into (33), we get

$$\begin{aligned} H_{1\not 2}(t+\Delta t) = -\int_\Omega &\left\{\rho_1(y_1) - \Delta t \times\int_{\Omega_{3n}}\frac{\partial F_1\rho_{\not 2}}{\partial y_1}(y_1,x_3,\ldots,x_n)\,\mathrm{d}x_3\ldots\mathrm{d}x_n\right\} \\ &\cdot\left\{\log\rho_1(y_1) - \Delta t\cdot\int_{\Omega_{3n}}\frac{1}{\rho_1(y_1)}\frac{\partial F_1\rho_{\not 2}}{\partial y_1}\mathrm{d}x_3\ldots\mathrm{d}x_n\right\}\end{aligned}$$

$$
\begin{aligned}
&\cdot \Big\{ \rho(x_2 \mid y_1, x_3, \ldots, x_n) + \rho(x_2 \mid y_1, x_3, \ldots, x_n) \\
&\cdot \frac{\partial \log \rho_{\not 2}}{\partial y_1} \cdot F_1 \Delta t - \frac{1}{\rho_{\not 2}} \frac{\partial \rho}{\partial y_1} F_1 \Delta t \Big\} \\
&\cdot \rho(x_3, x_4, \ldots, x_n)\, \mathrm{d}y_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n + O(\Delta t^2) \\
= &- \int_\Omega \rho_1(x_1) \log \rho_1(x_1) \\
&\times \rho(x_2 | x_1, x_3, \ldots, x_n) \rho_{3\ldots n}(x_3, \ldots, x_n)\, \mathrm{d}\underline{\mathbf{x}} \\
&+ \Delta t \cdot \int_\Omega \log \rho_1(x_1) \left(\int_{\Omega_{3n}} \frac{\partial F_1 \rho_{\not 2}}{\partial x_1} \mathrm{d}x_3 \ldots \mathrm{d}x_n \right) \\
&\cdot \rho(x_2 \mid x_1, x_3, \ldots, x_n) \rho_{3\ldots n}(x_3, \ldots, x_n)\, \mathrm{d}\underline{\mathbf{x}} \\
&+ \Delta t \cdot \int_\Omega \rho_1(x_1) \left(\int_{\Omega_{3n}} \frac{\partial F_1 \rho_{\not 2}}{\partial x_1} \frac{1}{\rho_1(x_1)} \mathrm{d}x_3 \ldots \mathrm{d}x_n \right) \\
&\cdot \rho(x_2 \mid x_1, x_3, \ldots, x_n) \rho_{3\ldots n}(x_3, \ldots, x_n)\, \mathrm{d}\underline{\mathbf{x}} \\
&- \Delta t \cdot \int_\Omega \rho_1(x_1) \log \rho_1(x_1) \cdot F_1 \cdot \rho(x_2 \mid x_1, x_3, \ldots, x_n) \\
&\cdot \frac{\partial \log \rho_{\not 2}}{\partial x_1} \rho_{3\ldots n}(x_3, \ldots, x_n)\, \mathrm{d}\underline{\mathbf{x}} \\
&+ \Delta t \cdot \int_\Omega \rho_1(x_1) \log \rho_1(x_1) \cdot F_1 \cdot \frac{1}{\rho_{\not 2}} \frac{\partial \rho}{\partial x_1} \\
&\times \rho_{3\ldots n}(x_3, \ldots, x_n)\, \mathrm{d}\underline{\mathbf{x}} + O(\Delta t^2) \\
\equiv &\ \mathrm{(I)} + \mathrm{(II)} + \mathrm{(III)} + \mathrm{(IV)} + \mathrm{(V)} + O(\Delta t^2).
\end{aligned}
$$

From step 1 to step 2, we have replaced y_1 by x_1 (and the dependency of F_1 and ρ on y_1 becomes on x_1). This is legal as in this definite integration, y_1 is now a dummy variable. Notice that

$$
\rho(x_2 \mid x_1, x_3, \ldots, x_n) = \frac{\rho(x_1, x_2, \ldots, x_n)}{\rho(x_1, x_3, \ldots, x_n)} = \frac{\rho}{\rho_{\not 2}},
$$

$$
\rho_{3\ldots n}(x_3, \ldots, x_n) = \rho_{\not 1 \not 2} = \int_{\Omega_1 \times \Omega_2} \rho(\underline{\mathbf{x}})\, \mathrm{d}x_1 \mathrm{d}x_2. \tag{43}
$$

Also notice that $\int_{\Omega_{3n}} \frac{\rho}{\rho_{\not 2}} \rho_{\not 1 \not 2} \mathrm{d}x_3 \ldots \mathrm{d}x_n$ represents a kind of conditional density: When (X_1, X_2) are independent of $(X_3, \ldots, X_n)$, it is just the density of X_2 conditioned on X_1. For this reason we write

$$
\int_{\Omega_{3n}} \frac{\rho}{\rho_{\not 2}} \rho_{\not 1 \not 2}\, \mathrm{d}x_3 \ldots \mathrm{d}x_n \equiv \Theta_{2|1}(x_1, x_2), \tag{44}
$$

signifying a generalized conditional density of X_2 on X_1. Eqs. (43) and (44) allow for a simplification of the above five terms (I)–(V):

$$
\begin{aligned}
\mathrm{(I)} &= - \int_{\Omega_1} \rho_1(x_1) \log \rho_1(x_1) \\
&\quad \times \left(\int_{\Omega_{2n}} \frac{\rho}{\rho_{\not 2}} \rho_{\not 1 \not 2}\, \mathrm{d}x_2 \mathrm{d}x_3 \ldots \mathrm{d}x_n \right) \mathrm{d}x_1 \\
&= - \int_{\Omega_1} \rho_1(x_1) \log \rho_1(x_1) \cdot 1\, \mathrm{d}x_1 \\
&= H_1(t);
\end{aligned} \tag{45}
$$

$$
\begin{aligned}
\mathrm{(II)} &= \Delta t \int_{\Omega_1 \times \Omega_2} \log \rho_1(x_1) \left(\int_{\Omega_{3n}} \frac{\partial F_1 \rho_{\not 2}}{\partial x_1} \mathrm{d}x_3 \ldots \mathrm{d}x_n \right) \\
&\quad \times \left(\int_{\Omega_{3n}} \frac{\rho}{\rho_{\not 2}} \rho_{\not 1 \not 2}\, \mathrm{d}x_3 \ldots \mathrm{d}x_n \right) \mathrm{d}x_1 \mathrm{d}x_2 \\
&= \Delta t \int_\Omega \log \rho_1(x_1) \frac{\partial F_1 \rho_{\not 2}}{\partial x_1}\ \Theta_{2|1}(x_1, x_2)\, \mathrm{d}\underline{\mathbf{x}};
\end{aligned} \tag{46}
$$

$$
\begin{aligned}
\mathrm{(III)} &= \Delta t \int_{\Omega_1 \times \Omega_2} \left(\int_{\Omega_{3n}} \frac{\partial F_1 \rho_{\not 2}}{\partial x_1} \mathrm{d}x_3 \ldots \mathrm{d}x_n \right) \\
&\quad \times \left(\int_{\Omega_{3n}} \frac{\rho}{\rho_{\not 2}} \rho_{\not 1 \not 2}\, \mathrm{d}x_3 \ldots \mathrm{d}x_n \right) \mathrm{d}x_1 \mathrm{d}x_2 \\
&= \Delta t \int_\Omega \frac{\partial F_1 \rho_{\not 2}}{\partial x_1}\ \Theta_{2|1}(x_1, x_2)\, \mathrm{d}\underline{\mathbf{x}};
\end{aligned} \tag{47}
$$

$$
\mathrm{(IV)} = -\Delta t \int_\Omega \rho_1 \log \rho_1 \frac{\partial \log \rho_{\not 2}}{\partial y_1} \cdot F_1 \cdot \frac{\rho}{\rho_{\not 2}} \cdot \rho_{\not 1 \not 2}\, \mathrm{d}\underline{\mathbf{x}} \tag{48}
$$

$$
\begin{aligned}
\mathrm{(V)} &= \Delta t \int_\Omega \rho_1 \log \rho_1 \cdot F_1 \frac{1}{\rho_{\not 2}} \frac{\partial \rho}{\partial x_1} \rho_{\not 1 \not 2}\, \mathrm{d}\underline{\mathbf{x}} \\
&= \Delta t \int_\Omega \rho_1 \log \rho_1 \cdot F_1 \cdot \frac{1}{\rho} \frac{\partial \rho}{\partial x_1} \frac{\rho}{\rho_{\not 2}} \cdot \rho_{\not 1 \not 2}\, \mathrm{d}\underline{\mathbf{x}} \\
&= \Delta t \int_\Omega \rho_1 \log \rho_1 \cdot F_1 \frac{\partial \log \rho}{\partial x_1} \frac{\rho}{\rho_{\not 2}} \rho_{\not 1 \not 2}\, \mathrm{d}\underline{\mathbf{x}}.
\end{aligned} \tag{49}
$$

Combination of (II) and (III), and (IV) and (V), gives

$$
\mathrm{(II)} + \mathrm{(III)} = \Delta t \int_\Omega (1 + \log \rho_1) \cdot \frac{\partial F_1 \rho_{\not 2}}{\partial x_1} \cdot \Theta_{2|1}\, \mathrm{d}\underline{\mathbf{x}},
$$

$$
\begin{aligned}
\mathrm{(IV)} + \mathrm{(V)} &= \Delta t \int_\Omega \rho_1 \log \rho_1 \cdot F_1 \cdot \frac{\partial}{\partial x_1} \log \left(\frac{\rho}{\rho_{\not 2}} \right) \\
&\quad \cdot \left(\frac{\rho}{\rho_{\not 2}} \rho_{\not 1 \not 2} \right) \mathrm{d}\underline{\mathbf{x}} \\
&= \Delta t \int_\Omega \rho_1 \log \rho_1 \cdot F_1 \cdot \frac{\partial(\rho/\rho_{\not 2})}{\partial x_1} \cdot \rho_{\not 1 \not 2}\, \mathrm{d}\underline{\mathbf{x}}.
\end{aligned}
$$

So

$$
\begin{aligned}
\frac{\mathrm{d}H_{1\not 2}}{\mathrm{d}t} &= \lim_{\Delta t \to 0} \frac{H_{1\not 2}(t + \Delta t) - H_1(t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\mathrm{(II)} + \mathrm{(III)} + \mathrm{(IV)} + \mathrm{(V)}}{\Delta t} \\
&= \int_\Omega (1 + \log \rho_1) \cdot \frac{\partial F_1 \rho_{\not 2}}{\partial x_1} \cdot \Theta_{2|1}\, \mathrm{d}\underline{\mathbf{x}} \\
&\quad + \int_\Omega \rho_1 \log \rho_1 \cdot F_1 \cdot \frac{\partial(\rho/\rho_{\not 2})}{\partial x_1} \cdot \rho_{\not 1 \not 2}\, \mathrm{d}\underline{\mathbf{x}}.
\end{aligned} \tag{50}
$$

6. Entropy transfer

According to what we established in Section 2, the rate of entropy transfer from X_2 to X_1 is obtained by subtracting (50) from (29). That is to say,

$$
\begin{aligned}
T_{2\to 1} &= \int_\Omega (1 + \log \rho_1) \cdot \left(\frac{\partial(F_1 \rho)}{\partial x_1} - \frac{\partial(F_1 \rho_{\not 2})}{\partial x_1}\ \Theta_{2|1} \right) \mathrm{d}\underline{\mathbf{x}} \\
&\quad - \int_\Omega \rho_1 \log \rho_1 \cdot F_1 \cdot \frac{\partial \left(\frac{\rho}{\rho_{\not 2}} \right)}{\partial x_1} \cdot \rho_{\not 1 \not 2}\, \mathrm{d}\underline{\mathbf{x}}.
\end{aligned} \tag{51}
$$

Using the zero-boundary-probability assumption, this may be further simplified. Integrating by parts, and noticing that $\rho_{\not 1 \not 2}$ does not depend on x_1, we can change the second integral to

$$+\int_\Omega \left(\frac{\rho}{\rho_{\not 2}}\rho_{\not 1 \not 2}\right) \cdot \frac{\partial(F_1\rho_1 \log\rho_1)}{\partial x_1}\,\mathrm{d}\underline{\mathbf{x}}.$$

The transfer therefore becomes

$$T_{2\to 1} = \int_\Omega (1+\log\rho_1)\left(\frac{\partial(F_1\rho)}{\partial x_1} - \frac{\partial(F_1\rho_{\not 2})}{\partial x_1}\cdot \Theta_{2|1}\right)\mathrm{d}\underline{\mathbf{x}} + \int_\Omega \frac{\partial(F_1\rho_1\log\rho_1)}{\partial x_1}\cdot\theta_{2|1}\,\mathrm{d}\underline{\mathbf{x}}, \tag{52}$$

where

$$\theta_{2|1} = \theta_{2|1}(x_1, x_2, x_3, \ldots, x_n) = \frac{\rho}{\rho_{\not 2}}\rho_{\not 1 \not 2}, \tag{53}$$

and $\Theta_{2|1} = \int_{\Omega_{3n}} \theta_{2|1}(\underline{\mathbf{x}})\,\mathrm{d}x_3 \ldots \mathrm{d}x_n$ reminds one of the conditional density of X_2 on X_1.

The above transfer measure, between X_1 and X_2, can be easily extended to that between any two components. Replacing 1 by i and 2 by j in (51), it follows the entropy transfer from X_j to X_i

$$T_{j\to i} = \int_\Omega (1+\log\rho_i)\left(\frac{\partial(F_i\rho)}{\partial x_i} - \frac{\partial(F_i\rho_{\not j})}{\partial x_i}\cdot \Theta_{j|i}\right)\mathrm{d}\underline{\mathbf{x}} + \int_\Omega \frac{\partial(F_i\rho_i\log\rho_i)}{\partial x_i}\cdot\theta_{j|i}\,\mathrm{d}\underline{\mathbf{x}}. \tag{54}$$

Here

$$\theta_{j|i} = \theta_{j|i}(\underline{\mathbf{x}}) = \frac{\rho}{\rho_{\not j}}\rho_{\not i \not j}, \tag{55}$$

and

$$\Theta_{j|i} = \Theta_{j|i}(x_i, x_j) = \int_{\prod_{\nu\neq i,j}\Omega_\nu} \theta_{j|i}(\underline{\mathbf{x}}) \prod_{\nu\neq i,j}\mathrm{d}x_\nu \tag{56}$$

is the "generalized conditional density" of X_j on X_i.

7. Properties for the entropy transfer rate

A 2D system is very special in that $\frac{\mathrm{d}H_{\not 1 \not 2}}{\mathrm{d}t}$ is just the change of H_1 when X_1 evolves on its own. As detailed in [6], the latter is simply $E\left(\frac{\partial F_1}{\partial x_1}\right)$, a result which can be directly obtained by integrating the Liouville equation. Since we are dealing with a system with arbitrarily many dimensions, we expect our formalism (51) or (54) to verify this result. This forms the following theorem, which corresponds to the Theorem 1 in Part I.

Theorem 1. *When $n = 2$, $\frac{\mathrm{d}H_{\not 1 \not 2}}{\mathrm{d}t} = E\left(\frac{\partial F_1}{\partial x_1}\right)$.*

Proof. Mathematically what makes a 2D system special is that, when $n = 2$, $\rho_{\not 2} = \rho_1$, and $\Theta_{2|1}$ is just the conditional distribution of X_2 given X_1, $\frac{\rho}{\rho_1} = \rho(x_2|x_1)$. Eq. (50) thereby can be greatly simplified:

$$\begin{aligned}
\frac{\mathrm{d}H_{\not 1 \not 2}}{\mathrm{d}t} &= \int_\Omega (1+\log\rho_1)\frac{\partial F_1\rho_1}{\partial x_1}\cdot\frac{\rho}{\rho_1}\,\mathrm{d}\underline{\mathbf{x}} + \int_\Omega \rho_1\log\rho_1\cdot F_1\cdot\frac{\partial(\frac{\rho}{\rho_1})}{\partial x_1}\,\mathrm{d}\underline{\mathbf{x}} \\
&= \int_\Omega \frac{\partial(F_1\rho_1)}{\partial x_1}\frac{\rho}{\rho_1}\,\mathrm{d}\underline{\mathbf{x}} + \int_\Omega \log\rho_1 \cdot\left[\frac{\partial(F_1\rho_1)}{\partial x_1}\left(\frac{\rho}{\rho_1}\right) + (F_1\rho_1)\frac{\partial\rho/\rho_1}{\partial x_1}\right]\mathrm{d}\underline{\mathbf{x}} \\
&= \int_\Omega \frac{\partial(F_1\rho_1)}{\partial x_1}\frac{\rho}{\rho_1}\,\mathrm{d}\underline{\mathbf{x}} + \int_\Omega \log\rho_1\cdot\frac{\partial(F_1\rho)}{\partial x_1}\,\mathrm{d}\underline{\mathbf{x}} \\
&= \int_\Omega \left(\rho\frac{\partial F_1}{\partial x_1} + F_1\frac{\partial\rho_1}{\partial x_1}\cdot\frac{\rho}{\rho_1}\right)\mathrm{d}\underline{\mathbf{x}} + 0 - \int_\Omega (F_1\rho)\frac{\partial(\log\rho_1)}{\partial x_1}\,\mathrm{d}\underline{\mathbf{x}} \\
&= \int_\Omega \rho\left(\frac{\partial F_1}{\partial x_1}\right)\mathrm{d}\underline{\mathbf{x}} \\
&= E\left(\frac{\partial F_1}{\partial x_1}\right).
\end{aligned} \tag{57}$$

One important property that $T_{j\to i}$ should have is asymmetry, as emphasized by Schreiber [11] and later by us in the context of a 2D dynamical system [6]. For a system of arbitrary dimensionality, this is also true, as we hereafter establish.

Theorem 2. *If F_i is independent of x_j, then $T_{j\to i} = 0$.*

Proof. Without loss of generality, we need only consider the case when $i = 1$ and $j = 2$.

If F_1 is independent of x_2,

$$\begin{aligned}
&\int_\Omega (1+\log\rho_1)\frac{\partial(F_1\rho)}{\partial x_1}\,\mathrm{d}\underline{\mathbf{x}} \\
&= \int_{\Omega_1\times\Omega_{3n}} (1+\log\rho_1)\cdot\frac{\partial(F_1\int\rho\,\mathrm{d}x_2)}{\partial x_1}\,\mathrm{d}x_1\mathrm{d}x_3\ldots\mathrm{d}x_n \\
&= \int_{\Omega_1\times\Omega_{3n}} (1+\log\rho_1)\cdot\frac{\partial(F_1\rho_{\not 2})}{\partial x_1}\,\mathrm{d}x_1\mathrm{d}x_3\ldots\mathrm{d}x_n,
\end{aligned} \tag{58}$$

and

$$\begin{aligned}
&\int_\Omega (1+\log\rho_1)\frac{\partial(F_1\rho_{\not 2})}{\partial x_1}\,\Theta_{2|1}\,\mathrm{d}\underline{\mathbf{x}} \\
&= \int_{\Omega_1\times\Omega_{3n}} (1+\log\rho_1)\cdot\frac{\partial(F_1\rho_{\not 2})}{\partial x_1}\left(\int_{\Omega_2}\Theta_{2|1}\,\mathrm{d}x_2\right) \times\mathrm{d}x_1\mathrm{d}x_3\ldots\mathrm{d}x_n \\
&= \int_{\Omega_1\times\Omega_{3n}} (1+\log\rho_1)\cdot\frac{\partial(F_1\rho_{\not 2})}{\partial x_1}\,\mathrm{d}x_1\mathrm{d}x_3\ldots\mathrm{d}x_n,
\end{aligned} \tag{59}$$

since $\int\Theta_{2|1}\,\mathrm{d}x_2 = 1$. These two are equal, so the first integral on the right hand side of (51) vanishes. For the second integral in (51), the independence of F_1 of x_2 implies that ρ in the derivative is the only variable depending on x_2, and hence can be integrated out with respect to x_2. But $\int_{\Omega_2}\rho\,\mathrm{d}x_2 = \rho_{\not 2}$, which makes the derivative function a constant 1. So the second integral in (51) also vanishes. Therefore, when F_1 is independent of x_2, $T_{2\to 1} = 0$. □

8. Application

We now apply the above formalism to study how information is exchanged within a dynamical system, namely the truncated

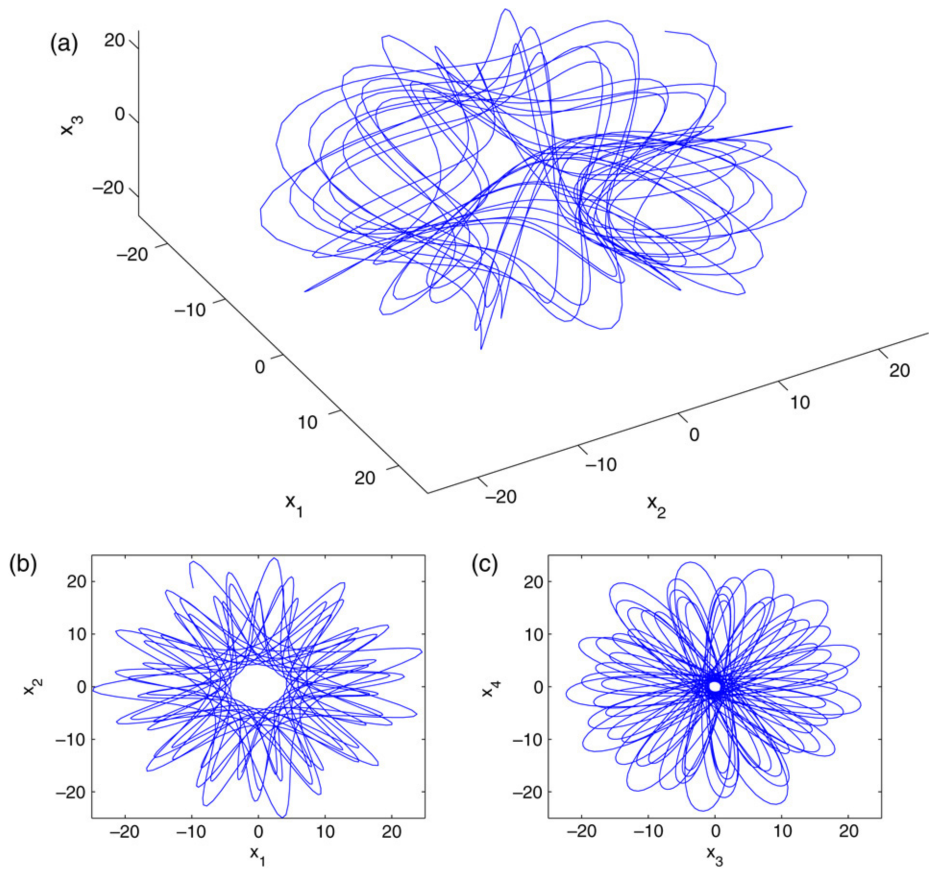

Fig. 1. The trajectory of (64)–(67) from $t = 2$ to $t = 20$ with step size $\Delta t = 0.01$ and initial condition $x_1 = x_2 = x_3 = x_4 = 40$. Shown in (a) is a 3D plot for x_1, x_2, and x_3, and in (b) and (c) are projections on the planes of x_1–x_2 and x_3–x_4, respectively. The trajectory before $t = 2$ lies outside the domain and is not displayed.

Burgers–Hopf system (TBS). This is a Galerkin truncation of the inviscid Burgers or Hopf equation:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0. \tag{60}$$

It was first introduced by Majda and Timofeyev [8] to test their systematic stochastic mode reduction strategies for the large eddy simulation of climate processes (see [10]). In contrast to its shock and soliton counterparts, which are completely integrable, the TBS is intrinsically chaotic. Remarkable statistical behaviors have been discovered by Majda and Timofeyev [8,9] and Abramov et al. [1].

To form the TBS, let P_N (N some positive integer) be an operator such that, for $u = u(x,t)$,

$$P_N u = \sum_{|k|\le N} \hat{u}_k(t)\mathrm{e}^{\mathrm{i}kx}, \tag{61}$$

where $\hat{u}_k$ is the Fourier transform of u, and $\hat{u}_{-k} = \hat{u}_k^*$, the conjugate of $\hat{u}_k$. The Galerkin truncation of (60) is

$$\frac{\partial P_N u}{\partial t} + P_N\left(P_N u\frac{\partial P_N u}{\partial x}\right) = 0. \tag{62}$$

This yields, thanks to the orthogonality of the Fourier basis,

$$\frac{\mathrm{d}\hat{u}_k}{\mathrm{d}t} = -\frac{\mathrm{i}k}{2}\sum_{\substack{k+p+q=0\\ |p|,|q|\le 2}} \hat{u}_p^*\hat{u}_q^*. \tag{63}$$

Obviously $\hat{u}_0$ is a constant. This trivial mode is generally neglected by setting $\hat{u}_0 = 0$. In this study, we only consider two modes, $|k| = 1$ and $|k| = 2$. Let $\hat{u}_1 = x_1 + \mathrm{i}x_2$, $\hat{u}_2 = x_3 + \mathrm{i}x_4$. Eq. (63) gives an autonomous system of dimensionality 4:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = x_1x_4 - x_3x_2 \tag{64}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -x_1x_3 - x_2x_4, \tag{65}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = 2x_1x_2, \tag{66}$$

$$\frac{\mathrm{d}x_4}{\mathrm{d}t} = -x_1^2 + x_2^2. \tag{67}$$

Eqs. (64)–(67) are solved using the second-order Runge–Kutta method, with a step size $\Delta t = 0.01$. We find that the precision is enough for our purpose while the computation is not too expensive. Fig. 1 shows the trajectory of an example thus computed. This scheme is applied to perform an ensemble prediction for the joint density of $\underline{\mathbf{X}}$, $\rho(\underline{\mathbf{x}})$. ρ is defined on $\mathbb{R}^4$, but for practicality the computational domain is limited to

$$\Omega_d = [-d,d]\times[-d,d]\times[-d,d]\times[-d,d],$$

with $d = 30$. As pointed out by Majda and Timofeyev [8], there exists an invariant low dimensional attractor for the TBS system. Our computation with (64)–(67) shows that the attractor lies within Ω_{25} as set above. As an example, Fig. 1 is

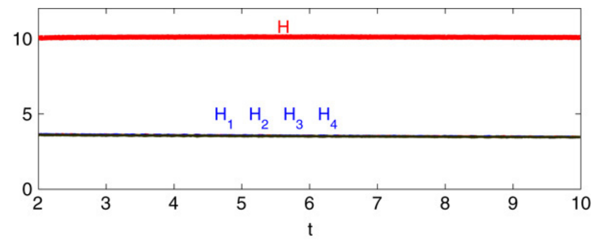

Fig. 2. Time evolution of the joint entropy (H) and the marginal entropies (H_1, H_2, H_3, H_4). The initial distribution has a parametric combination of $\mu_k = 9, \sigma_k^2 = 9, k = 1, 2, 3, 4$.

generated with an initial condition outside Ω_d, but the trajectory enters entirely into Ω_d in less than 200 time steps ($t < 2$).

We partition the domain Ω_d uniformly into $30 \times 30 \times 30 \times 30$ bins, with a spacing of 2 in each dimension. An ensemble of initial condition of size 40^4 is first generated, through drawing randomly according to a preset distribution $\rho_0(\underline{\mathbf{x}})$. We adopt an ensemble size of 40^4 instead of 30^4 to ensure more than one draw per bin on average. Suppose $\underline{\mathbf{X}}$ is initially distributed as a Gaussian $N(\underline{\mu}, \underline{\underline{\mathbf{\Sigma}}})$, with a mean $\underline{\mu}$ and a covariance matrix $\underline{\underline{\mathbf{\Sigma}}}$:

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

The parameters $\underline{\mu}$ and σ_k^2 ($k = 1, 2, 3, 4$) may be tuned to cater for different experiments. Using these initial conditions, Eqs. (64)–(67) are integrated forward. At every time step we obtain an ensemble of $\underline{\mathbf{X}}$, and therefrom a four-variable joint distribution $\rho(\underline{\mathbf{x}})$ through bin counting. This ρ describes the statistical behavior of the system and allows for the computation of all the statistical properties that we are interested in.

First take a look at the time evolution of the joint entropy H, and the entropies of the individual components, H_k, $k = 1, 2, 3, 4$. Plotted in Fig. 2 are the series with $\mu_k = 9$ and $\sigma_k^2 = 9, k = 1, 2, 3, 4$. Clearly, the four marginal entropies are almost the same, and the joint entropy is invariant with time. The latter is consistent with the Liouville property as was pointed out by Majda and Timofeyev [8]. In fact, the flow of (64)–(67) is incompressible, yielding a zero joint entropy increase by (19). Our numerical computation is accurate enough to reproduce this result.

Next look at the transfers. There are twelve series to compute:

$$T_{j\to1}, \quad j = 2, 3, 4;$$
$$T_{j\to2}, \quad j = 1, 3, 4;$$
$$T_{j\to3}, \quad j = 1, 2, 4;$$
$$T_{j\to4}, \quad j = 1, 2, 3,$$

among which it is easily shown that

$$T_{3\to4} = T_{4\to3} = 0.$$

In fact, the third and fourth components of the system (64)–(67), $F_3 = 2x_1x_2$ and $F_4 = -x_1^2 + x_2^2$, are both independent of x_3 and x_4, implying a zero transfer in either direction between X_3 and X_4 by Theorem 2. Our computation produces precisely this analytical result (within the machine accuracy), in every test we have tried.

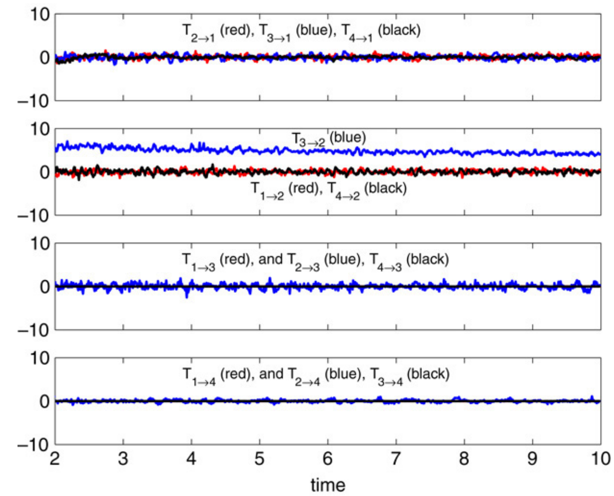

Fig. 3. Information transfer between different pairs of components. The series prior to $t = 2$ are not shown because some trajectories are still outside the computational domain by that time.

The other transfers have only numerical solutions. Remarkably, all of them except $T_{3\to2}$ are virtually nil, whatever μ_k and σ_k^2 are used. Fig. 3 plots the result of one experiment (parameters the same as those for Fig. 2). No significant transfers can be identified except for the one already identified, although some small scale oscillations do still exist.

The only nonzero transfer is the transfer from X_3 to X_2, $T_{3\to2}$. It generally depends on the shape of $\rho_0(\underline{\mathbf{x}})$ or how the $\underline{\mathbf{X}}$ are initially distributed. When $\rho_0(\underline{\mathbf{x}})$ is symmetric about the origin, it also goes to zero; but when $\underline{\mu}$ deviates even by a quite small amount from the origin, it appears distinctly different from zero. Moreover, it evolves linearly with time (actually almost steady). Although the magnitude may vary with different $\underline{\mu}$, the trend stays invariant.

A further examination of the equations of the system (64)–(67) may help better understand the constant transfer out of the highly variable trajectories. In this case, all the marginal entropies are invariant (cf. Fig. 2). So, according to (9), the information transfer from X_3 to X_2 is

$$T_{3\to2} = -\frac{\mathrm{d}H_2}{\mathrm{d}t}.$$

For illustrative purposes, suppose the system contains only two components X_3 and X_2; by Theorem 1, the above transfer is then simply $-E\left(\frac{\partial F_2}{\partial x_2}\right) = E(x_4)$. Of course, the system is actually four dimensional, and there is no such simple analytical expression for $T_{3\to2}$, but this offers us a clue that the transfer is somehow related to the mean of the trajectories, which could be invariant even though the individuals in the ensemble are highly variable. Indeed, the ensemble mean of the trajectories almost converges to a point in the phase space soon after entering the invariant attractor (figure not shown).

Physically what is happening is that a fine component is causing an increase in uncertainty in a coarse component but not conversely. This is a good illustration of transfer asymmetry or unidirectionalism. Such a situation is conceptually what occurs in a dynamical system when one replaces fine grained components by stochastic forcing since there the coarse grain elements of the system have their uncertainty increased by the stochastic forcing which is intended to represent the neglected fine scales. Note that in an additive stochastic forcing model the forcing is unaffected by the retained coarse grain elements just as is occurring here. These considerations would suggest that perhaps the TBM might be well approximated by a stochastic model, something which certainly appears to be the case (I. Timofeyev, private communication).

9. Discussion and conclusions

A rigorous formalism of information transfer was established for continuous dynamical systems of arbitrary dimensionality. The resulting measure (54) possesses the property of asymmetry or causality, and is consistent with our previous 2D formalism which was obtained by different methods.

We have applied the formalism to the study of the information exchange among the components of a two-mode truncated Burgers–Hopf system. We found there that information transfer only exists from the second mode to the first mode, and specifically, from the cosine part of the second mode to the sine part of the first mode. An observation is: the transfer rate stays (almost) invariant with time.

The above remarkable result motivates one to think more about the significance of the notion of information transfer. We know by Theorem 2 that if the evolution of one variable X_1 is independent of another one, say X_2, then there is no information transfer from X_2 to X_1, i.e., $T_{2\to1} = 0$. One may naturally ask whether the converse is also true. By the above result it is obviously not: All the transfers except one are virtually zero but the variables in the system (64)–(67) are mostly interdependent. But what if in the above question we change "dependence" to "decouplability", and ask whether the variables can be decoupled if the transfer at least in one direction is zero? In other words, we want to know whether the following is true: If $T_{1\to2} = 0$ or $T_{2\to1} = 0$, X_1 and X_2 are decouplable; if neither transfer is zero then the two cannot be decoupled. So far we have no idea about the correctness of the hypothesis, but the result of the example system (64)–(67) seems to support it: The truncated system (64)–(67) is indeed integrable [8]. We will leave the resolution of this subtle issue to future work.

Our formalism has been constructed in a general dynamical setting, and real applications can be performed in a straightforward way. However, it should be pointed out that the transfer measure (54) requires knowledge of the flow (the vector $\underline{\mathbf{F}}$ in the governing Eqs. (1)–(3)) and the density distribution simultaneously. While this is the case in many physical problems (for example weather prediction), a large number of problems are characterized by observational data, without *a priori* knowledge of dynamics. In order for this to be studied we require a dynamics-free formulation which is currently in development and will be applied to the problem of multivariate time series analysis.

Another issue is that, in deriving (54), we have adopted Shannon's definition of entropy as the starting point. Shannon entropy has been widely used but in many contexts its utility is limited because it does not possess some properties desirable for a good physical measure such as invariance under nonlinear transformation. On the other hand relative entropy has this property (and others of interest) and is consequently often more useful (see, e.g., [2–4]). We are therefore currently extending our formalism to this functional.

Acknowledgments

XSL first learned about the remarkable stochastic dynamics of the truncated Burgers–Hopf system from Andrew Majda, whom he sincerely thanks. The intuitive thoughts from a referee on the possible relation between information exchange and the decouplability of dynamical system variables are greatly appreciated. This work was supported by NSF under CMG Grant 0417728 to Courant Institute of Mathematical Sciences.

References

[1] R.V. Abramov, G. Kovacic, A.J. Majda, Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers–Hopf equation, Comm. Pure Appl. Math. LVI (2003) 1–46.
[2] R. Kleeman, Measuring dynamical prediction utility using relative entropy, J. Atmospheric Sci. 59 (2002) 2057–2072.
[3] R. Kleeman, A.J. Majda, Predictability in a model of geostrophic turbulence, J. Atmospheric Sci. 62 (2005) 2864–2879.
[4] R. Kleeman, Information flow in ensemble weather predictions, J. Atmospheric Sci. (in press).
[5] A. Lasota, M.C. Mackey, Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics, Springer, New York, 1994.
[6] X.S. Liang, R. Kleeman, Information transfer between dynamical system components, Phys. Rev. Lett. 95 (24) (2005) 244101.
[7] X.S. Liang, R. Kleeman, A rigorous formalism of information transfer between dynamical system components. I. Discrete mapping, Physica D (in press).
[8] A.J. Majda, I. Timofeyev, Remarkable statistical behavior for truncated Burgers–Hopf dynamics, Proc. Natl. Acad. Sci. USA 97 (23) (2000) 12413–12417.
[9] A.J. Majda, I. Timofeyev, Statistical mechanics for truncations of the Burgers–Hopf equation. A model for intrinsic stochastic behavior with scaling, Milan J. Math. 70 (1) (2002) 39–96.
[10] A.J. Majda, I. Timofeyev, E. Vanden-Eijnden, Systematic strategies for stochastic mode reduction in climate, J. Atmospheric Sci. 60 (2003) 1705–1722.
[11] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2) (2000) 461–464.