

Article

Some Convex Functions Based Measures of Independence and Their Application to Strange Attractor Reconstruction

Yang Chen ^{1,*} and Kazuyuki Aihara ^{2,3}

¹ School of Information Science and Engineering, Southeast University, Nanjing 210096, China

² Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan;

E-Mail: aihara@sat.t.u-tokyo.ac.jp

³ Aihara Complexity Modelling Project, ERATO, JST, Saitama 332-0012, Japan

* Author to whom correspondence should be addressed; E-Mail: chen@seu.edu.cn.

Received: 8 February 2011; in revised form: 28 March 2011 / Accepted: 28 March 2011 /

Published: 8 April 2011

Abstract: The classical information-theoretic measures such as the entropy and the *mutual information* (MI) are widely applicable to many areas in science and engineering. Csiszar generalized the entropy and the MI by using the convex functions. Recently, we proposed the *grid occupancy* (GO) and the *quasientropy* (QE) as measures of independence. The QE explicitly includes a convex function in its definition, while the expectation of GO is a subclass of QE. In this paper, we study the effect of different convex functions on GO, QE, and Csiszar's *generalized mutual information* (GMI). A *quality factor* (QF) is proposed to quantify the sharpness of their minima. Using the QF, it is shown that these measures can have sharper minima than the classical MI. Besides, a recursive algorithm for computing GMI, which is a generalization of Fraser and Swinney's algorithm for computing MI, is proposed. Moreover, we apply GO, QE, and GMI to chaotic time series analysis. It is shown that these measures are good criteria for determining the optimum delay in strange attractor reconstruction.

Keywords: entropy; mutual information; convex function; quality factor; strange attractor; delay-coordinate

PACS Codes: 89.70.cf, 05.45.Pq

1. Introduction

The advent of information theory was hallmarked by Shannon's seminal paper [1]. In that paper, some fundamental measures of information were established, among which is the entropy. The entropy is a concept that originally resided in thermodynamics and statistical physics as a measure of the degree of disorder [2]. For a different purpose, Shannon used the entropy to measure the amount of information. He also proposed the prototype of *mutual information* (MI) to measure the amount of information transmitted through a communication channel. The Shannon MI can be viewed as the Kullback divergence (also known as the relative entropy) between the joint *probability density function* (PDF) and the product of marginal PDFs. It reaches its minimum, zero, if and only if the variables are independent. Hence MI can be viewed as a measure of independence. Since the 1960s, generalizations of the Shannon entropy and MI have attracted the attention of many researchers, yielding various forms of non-Shannon information-theoretic measures [2–9]. For example, Renyi suggested some properties that the entropy should satisfy, such as the additivity, and proposed a class of entropies with a parameter. The Shannon entropy is the limit of these entropies when the parameter approaches 1 [3]. Harvda and Charvat proposed a generalization of the Shannon entropy that is different from the Renyi's entropy, which is called structural α -entropy [4]. Using convex functions, Csiszar proposed the f -divergence [5], which is a generalization of the relative entropy. A generalized version of Shannon MI, which is referred to as *generalized mutual information* (GMI) in this paper, can be readily derived from f -divergence [5]. Tsallis postulated a generalized form of entropy [6]. Kapur also gave many definitions of entropies [7]. More recently, we proposed the *grid occupancy* (GO) [8] and the *quasientropy* (QE) [9] as measures of independence. The QE, which explicitly includes a convex function, is a general enough definition that includes the α -entropy, Tsallis entropy, and some of the Kapur's entropies as special cases. For the GO, it can be shown that its expectation is a subclass of QE [8].

As we can see, many information-theoretic measures involve convex functions. Thus, it is interesting to pose the question that how different convex functions affect the behavior of the related measures. Previous studies were usually carried out on a restricted type of convex function. For example, Tsallis studied the power function [6]. In this paper, however, we study the effect of *arbitrary* convex functions. We propose a *quality factor* (QF) to quantify the sharpness of the minima of GO, QE, and GMI. We show that the order of QF with respect to l , the number of quantization levels, correctly predicts the sharpness of minima. Especially, some convex functions yield QFs with higher order than that of the minus Shannon entropy and MI, which implies that the related measures have sharper minima than the minus Shannon entropy and MI do.

Nowadays, applications of entropy and information theory have extended, far beyond the original scope of communication theory, to a wide variety of fields such as statistical inference, non-parametric density estimation, time series analysis, pattern recognition, biological, ecological and medical modeling and so on [10–15]. In this paper, besides the above theoretical results, we also deal with a specific application of the independence measures to chaotic time series analysis, the background of which is described as follows. In many cases, the commonly seen one-dimensional scalar signals or time series are observations of a certain variable of a multivariate dynamical system. An interesting property of dynamical systems is that their multivariate behavior can be approximately “reconstructed”

from observations of just one variable by a method called state space (phase portrait) reconstruction. For chaotic systems, this method is also termed strange attractor reconstruction, which has become a fundamental approach in chaotic signal analysis. Basic reconstruction can be done using delay coordinates [16]. Namely, we take the current observed value of the time series and the values after equally spaced time lags τ , 2τ , 3τ ... to obtain “observations” of several reconstructed variables where τ is the delay. If τ is not appropriate, the effect of reconstruction might not be good. So we should consider how to choose a proper delay. Since reconstruction aims at releasing the information of the whole system “condensed” in one variable, generally the reconstructed variables should be as independent as possible. Thus, a measure of independence can be used as a criterion for choosing τ . For example, Fraser and Swinney used the first minimum of the Shannon MI for choosing delay according to Shaw’s suggestion. They proposed a recursive algorithm for computing Shannon MI, and they showed that the MI is actually more advantageous than the correlation function that only takes into account second order dependence [17]. In this paper, we shall develop a recursive algorithm for computing GMI, which is a generalization of Fraser and Swinney’s algorithm for computing Shannon MI. In addition, we are going to show that GO, QE, and GMI are even better criteria for choosing τ than the Shannon MI. It should be noted that this paper is by no means a thorough treatment of all the measures of independence, there are many not covered such as the Hilbert-Schmidt independence criterion proposed by Gretton *et al.* [18] and the distance correlation proposed by Székely *et al.* [19].

The rest of the paper is organized as follows. In Section 2, we first give the definition of *quasientropy* (QE), and then derive the *quality factor* (QF) of QE. In Section 3, we illustrate the principle of *grid occupancy* (GO), introduce the relation between GO and QE, and define the QF of GO based on the QF of QE. Section 4 reveals the relation between QE and the *generalized mutual information* (GMI), and deduces the QF of GMI, also from the QF of QE. The recursive algorithm for computing GMI is mentioned in Section 4, while the details are placed in the Appendix for better organization of the paper. Section 5 is devoted to a study of the order of QF with respect to l , the number of quantization levels, elucidating the cases when this order is higher than, equal to, and lower than that of the MI’s. These theoretical results are verified by the numerical experiments on delay reconstruction of the Rössler and the Lorenz attractors presented in Section 6. Finally, Section 7 concludes the paper. In the following, the word “entropy,” when appears alone, refers to the Shannon entropy and, the phrase “MI,” when appears alone, refers to the Shannon MI, as is the common usage.

2. Quality Factor (QF) of Quasientropy (QE)

2.1. Quasientropy

Let us consider the problem of how to measure the independence of several variables. We denote by $q_r(\cdot)$ the *cumulative distribution function* (CDF) of variable r . $q_r(u) = \text{Prob}(r \leq u) = \int_{-\infty}^u p_r(v) dv$, where $\text{Prob}(A)$ denotes the probability that event A occurs, and $p_r(\cdot)$ is the *probability density function* (PDF) of r . Without loss of generality, consider two continuous variables r_1 and r_2 . In the past one or two decades, the study of copulas has become a blooming field of statistical research [20]. Copula is the joint CDF of the transformed variables by their respective CDFs. The rationale of the study of copulas is that, to study the relation between two or more variables, we should nullify the

effect of their marginal distributions and concentrate on their joint distribution. Based on this principle, we transform r_1 and r_2 by their respective CDFs as follows:

$$z_1 \equiv q_{r_1}(r_1), \quad z_2 \equiv q_{r_2}(r_2) \quad (1)$$

then, $(z_1, z_2) \in [0,1] \times [0,1]$, and we have [8,9]:

Lemma 1. r_1 and r_2 are independent if and only if (z_1, z_2) is uniformly distributed in $[0,1] \times [0,1]$.

Lemma 1 shows that we can measure the independence of r_1 and r_2 by measuring the uniformity of the distribution of (z_1, z_2) in $[0,1] \times [0,1]$. To this end, let us partition the region $[0,1] \times [0,1]$ into an $l \times l$ uniform grid, and denote by $\xi(i, j)$ the (i, j) th square in the grid:

$$\xi(i, j) = \left[\frac{i-1}{l}, \frac{i}{l} \right] \times \left[\frac{j-1}{l}, \frac{j}{l} \right], \quad (i, j) \in \{1, \dots, l\} \times \{1, \dots, l\} \quad (2)$$

Denote by $p(i, j)$ the probability that (z_1, z_2) belongs to $\xi(i, j)$, i.e.,

$$p(i, j) = \text{Prob} \{ (z_1, z_2) \in \xi(i, j) \} \quad (3)$$

then, the QE for measuring the independence of r_1 and r_2 is defined as [9]:

$$\beta(r_1, r_2) = \sum_{i=1}^l \sum_{j=1}^l f(p(i, j)) \quad (4)$$

where $f(\cdot)$ is a differentiable strictly convex function on $[0, 1/l]$. Clearly, when $f(u) = u \log u$, QE is nothing but the entropy of $p(i, j)$ up to a minus sign. When $f(u) = (1/l^2 - u^q)/(1 - q)$, up to a minus sign, QE becomes the Tsallis entropy [6] of $p(i, j)$.

Due to the Jensen's inequality, we have [9]:

$$\beta(r_1, r_2) \geq l^2 f\left(\frac{1}{l^2}\right) \quad (5)$$

with equality if and only if $p(i, j)$ is uniform in $\{1, \dots, l\} \times \{1, \dots, l\}$.

If r_1 and r_2 are independent, then (z_1, z_2) is uniform in $[0,1] \times [0,1]$ and, thus, $p(i, j)$ is uniform in $\{1, \dots, l\} \times \{1, \dots, l\}$. Then the equality in (5) holds and $\beta(r_1, r_2)$ reaches its minimum $l^2 f(1/l^2)$. Conversely, if r_1 and r_2 are not independent, then there exists l_0 such that for any $l > l_0$, $\beta(r_1, r_2)$ cannot reach its minimum [9]. Thus, for a large enough l , the minimal $\beta(r_1, r_2)$ implies independent r_1 and r_2 .

2.2. Quality Factor (QF) of QE

There are infinitely many strictly convex functions. Let us consider the effect of different convex functions on the performance of QE. In [9], we proved that:

$$\beta(r_1, r_2) \leq (l^2 - l) f(0) + l f\left(\frac{1}{l}\right) \quad (6)$$

In order to better differentiate between independence and “near” independence, it is preferable that QE have high sensitivity around its minimum to the variance of the uniformity of probability density. Thus, we write $p(i, j)$ in the following form:

$$p(i, j) = \frac{1}{l^2} + \Delta p(i, j), \quad (i, j) \in \{1, \dots, l\} \times \{1, \dots, l\} \quad (7)$$

Then, the Taylor expansion of (4) gives [9]:

$$\beta(r_1, r_2) \approx l^2 f\left(\frac{1}{l^2}\right) + \frac{1}{2} f''\left(\frac{1}{l^2}\right) \sum_{i=1}^l \sum_{j=1}^l [\Delta p(i, j)]^2 \quad (8)$$

Based on (5), (6), and (8), we can define the quality factor (QF) of QE as:

$$Q_\beta(f(u)) = \frac{f''\left(\frac{1}{l^2}\right)}{\left(l^2 - l\right) f(0) + l f\left(\frac{1}{l}\right) - l^2 f\left(\frac{1}{l^2}\right)} \quad (9)$$

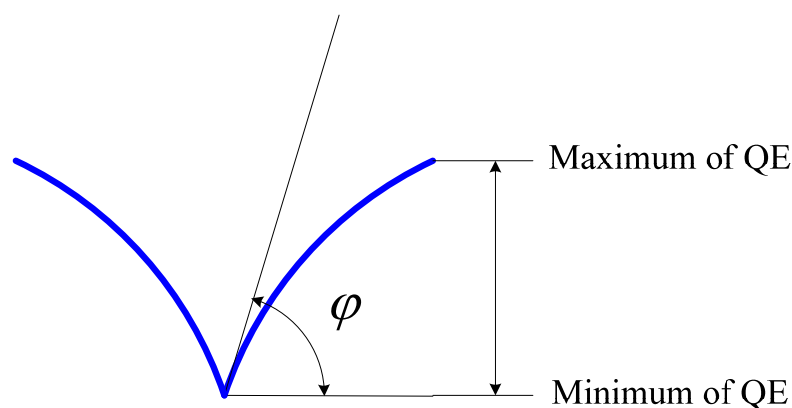
where, the numerator is twice the amplification rate of the variance between $p(i, j)$ and the uniform probability distribution in (8), and the denominator is the maximum dynamic range of $\beta(r_1, r_2)$ derived from (6) and (5). The greater the QF, the more sensitive is QE to the change in the uniformity of probability distribution around its minimum and, thus, the sharper is the shape of the minimum of QE.

An illustration of the QF of QE is shown in Figure 1. Suppose that QE is plotted *versus* $\sum_{i=1}^l \sum_{j=1}^l [\Delta p(i, j)]^2$, and φ is the angle that is tangent to the curve of QE at the minimum of QE, then:

$$Q_\beta = \frac{2 \cdot \tan \varphi}{\max \beta - \min \beta} \quad (10)$$

Clearly, when Q_β is large (small), QE is sharp (blunt) around its minimum.

Figure 1. Illustration of quality factor (QF) of QE.



3. QF of Grid Occupancy (GO)

In [8], we described an interesting phenomenon that independence can be measured by simply counting the number of occupied squares: Given N realizations (observed points) of (r_1, r_2) denoted by $(r_1(t), r_2(t))$, $t = 1, \dots, N$, where $N > 1$, let:

$$z_1(t) = q_{r_1}(r_1(t)), \quad z_2(t) = q_{r_2}(r_2(t)), \quad t = 1, \dots, N \quad (11)$$

and let:

$$m(i, j) = \begin{cases} 1 & \text{if } \exists t_0, 1 \leq t_0 \leq N \text{ s.t. } (z_1(t_0), z_2(t_0)) \in \xi(i, j), \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

where $\xi(i, j)$ is as defined in (2), and $(i, j) \in \{1, \dots, l\} \times \{1, \dots, l\}$. Then, the independence measure GO is defined as:

$$\alpha(r_1, r_2) = -\frac{\sum_{i=1}^l \sum_{j=1}^l m(i, j)}{l^2} \quad (13)$$

A visual illustration of GO is depicted in Figure 2. Figures 2a and 2b plot 500 observed points of two independent variables (r_1, r_2) and two dependent variables $(r_1^\#, r_2^\#)$, respectively. Figures 2c and 2d plot 500 observed points of (z_1, z_2) and $(z_1^\#, z_2^\#)$, respectively, where $z_1 \equiv q_{r_1}(r_1)$, $z_2 \equiv q_{r_2}(r_2)$, $z_1^\# \equiv q_{r_1^\#}(r_1^\#)$, and $z_2^\# \equiv q_{r_2^\#}(r_2^\#)$. It is easy to see that the points in Figure 2c are uniformly distributed whereas those in Figure 2d are not. Partition the region $[0, 1] \times [0, 1]$ in Figures 2c and 2d into a grid of $l \times l$, say, 10×10 , same-sized squares. A square is said to be occupied if there is at least one point in it. Then, being uniform is the most efficient way to occupy maximum number of squares, and this is confirmed in Figures 2e and 2f where the situations of Figures 2c and 2d are shown, respectively. (The occupied squares are shaded.) The grid occupancy (GO) defined in (13) is exactly minus the ratio of occupied squares. As Figures 2e and 2f clearly show, $\alpha(r_1, r_2) = -0.99 < \alpha(r_1^\#, r_2^\#) = -0.89$. Therefore, r_1 and r_2 are more independent than $r_1^\#$ and $r_2^\#$.

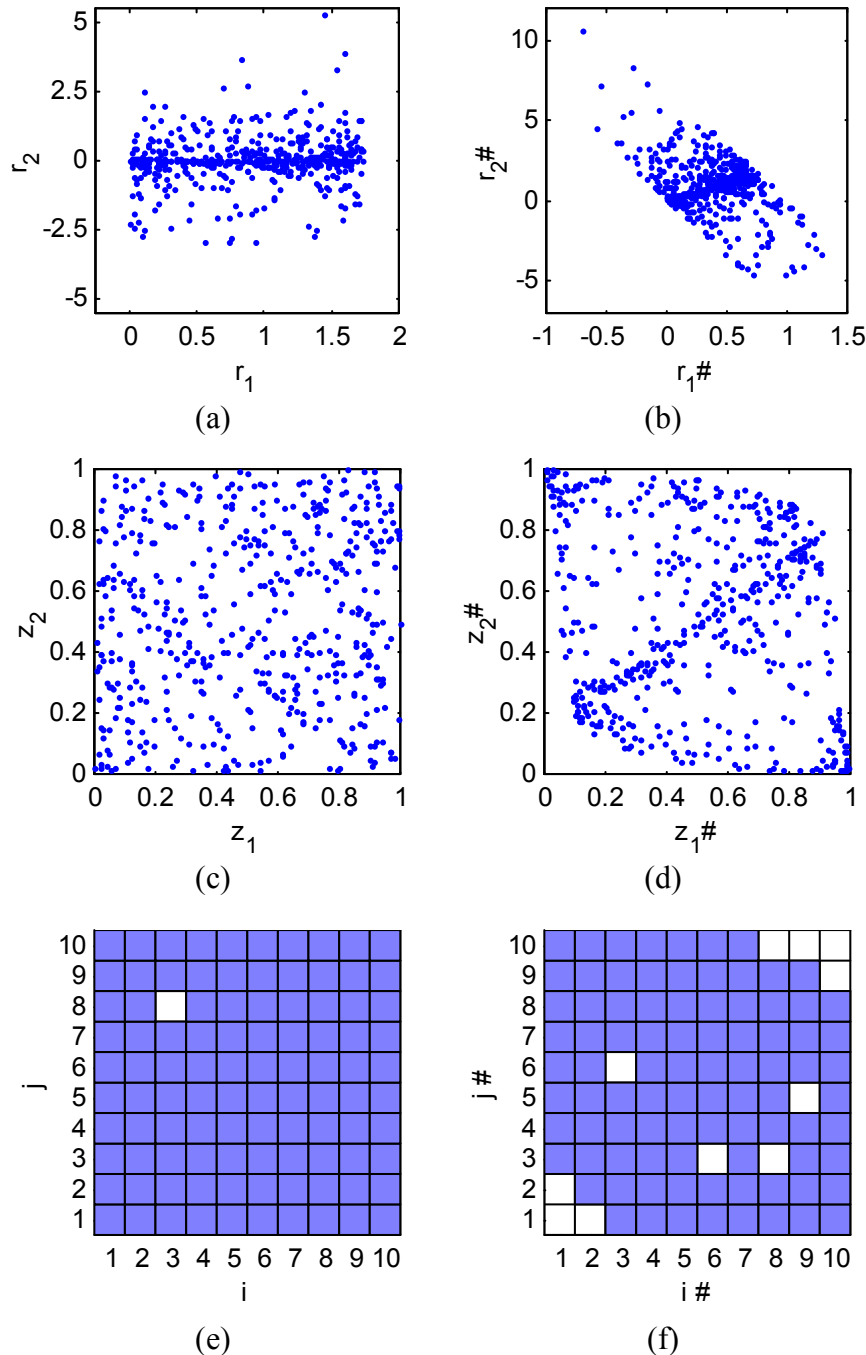
Indeed, we can prove that the expectation of GO is a subclass of QE [8]:

$$E[\alpha(r_1, r_2)] = \beta(r_1, r_2) \Big|_{f(u) = -\frac{1-(1-u)^N}{l^2}} \quad (14)$$

where $E[\cdot]$ denotes the mathematical expectation. Therefore, we can define the QF of GO as:

$$Q_\alpha = Q_\beta \left(-\frac{1-(1-u)^N}{l^2} \right) = \frac{N(N-1) \left(1 - \frac{1}{l^2} \right)^{N-2}}{l \left\{ \left(1 - \frac{1}{l} \right)^N - 1 + l \left[1 - \left(1 - \frac{1}{l^2} \right)^N \right] \right\}} \quad (15)$$

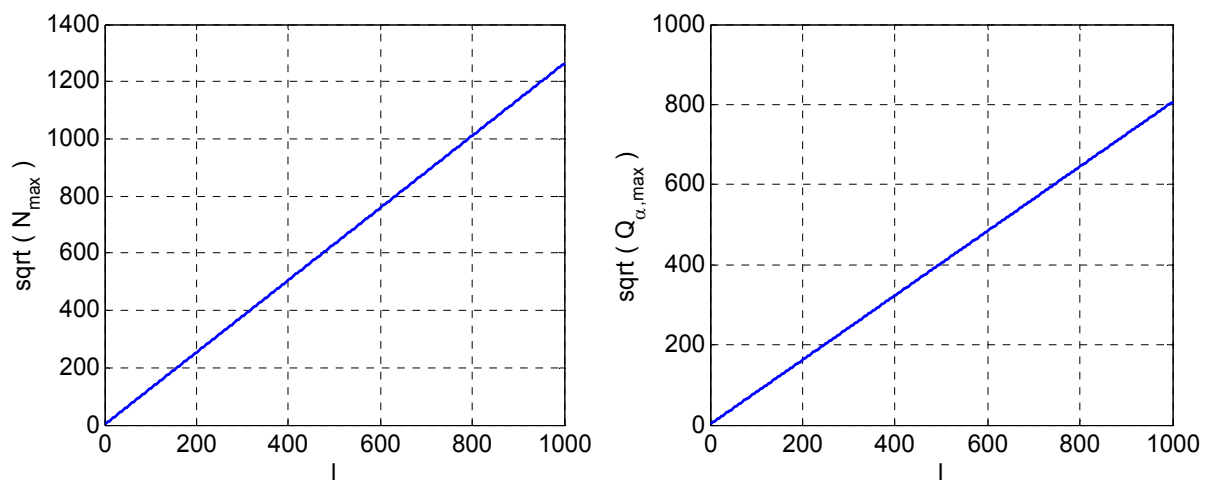
Figure 2. Illustration of grid occupancy (GO). (a) r_1 and r_2 are independent. (b) $r_1^\#$ and $r_2^\#$ are dependent. (c) (z_1, z_2) transformed from (r_1, r_2) is uniform. (d) $(z_1^\#, z_2^\#)$ transformed from $(r_1^\#, r_2^\#)$ is not uniform. (e) $\alpha(r_1, r_2) = -0.99$. (f) $\alpha(r_1^\#, r_2^\#) = -0.89$.



Since Q_α is not easy to be analyzed, we study it using numerical methods as shown in Figure 3. We vary l from 2 to 1,000. For each l , find the N , denoted $N_{\max}$, that maximizes Q_α to $Q_{\alpha, \max}$. In Figure 3, we can find:

$$N_{\max} \approx \left(\frac{5}{4}l\right)^2, \quad Q_{\alpha, \max} \approx \left(\frac{4}{5}l\right)^2 = \mathcal{O}(l^2) \quad (16)$$

Figure 3. Numerical study of QF of GO. Left: $\sqrt{N_{\max}}$ versus l . Right: $\sqrt{Q_{\alpha,\max}}$ versus l .



4. QF of Generalized Mutual Information (GMI)

4.1. QF of GMI

Let us now consider the *generalized mutual information* (GMI) proposed by Csiszar [5]:

$$\gamma(r_1, r_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{r_1}(u) p_{r_2}(v) f\left(\frac{p_{r_1 r_2}(u, v)}{p_{r_1}(u) p_{r_2}(v)}\right) dudv \quad (17)$$

where $f(\cdot)$ is strictly convex on $[0, \infty)$. Clearly, when $f(u) = u \log u$, GMI reduces to the classical MI:

$$I(r_1, r_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{r_1 r_2}(u, v) \log \frac{p_{r_1 r_2}(u, v)}{p_{r_1}(u) p_{r_2}(v)} dudv \quad (18)$$

Like MI, GMI is invariant under componentwise invertible transformations [21]. Therefore:

$$\gamma(r_1, r_2) = \gamma(z_1, z_2) = \int_0^1 \int_0^1 f(p_{z_1 z_2}(u, v)) dudv \quad (19)$$

where z_1 and z_2 are as defined in (1), and we have utilized the following facts [9]:

$$p_{z_1}(u) = 1, \quad \forall u \in [0, 1], \quad p_{z_2}(v) = 1, \quad \forall v \in [0, 1] \quad (20)$$

It is easy to verify that:

$$\lim_{l \rightarrow \infty} \sum_{i=1}^l \sum_{j=1}^l f\left(\frac{p(i, j)}{\frac{1}{l} \cdot \frac{1}{l}}\right) \cdot \frac{1}{l} \cdot \frac{1}{l} = \int_0^1 \int_0^1 f(p_{z_1 z_2}(u, v)) dudv = \gamma(r_1, r_2) \quad (21)$$

where $p(i, j)$ is as defined in (3). As we are mainly concerned about the situation when l is fairly large, according to (21), we can define the QF of GMI as:

$$Q_\gamma(f(u)) = Q_\beta\left(\frac{f(ul^2)}{l^2}\right) = \frac{l^3 f''(1)}{(l-1)f(0) + f(l) - lf(1)} \quad (22)$$

We have developed a recursive algorithm for computing GMI. The algorithm is described in the Appendix, where some related issues are also addressed.

4.2. Existence of GMI

From (19), $f(p_{z_1 z_2}(u, v))$ being continuous on $[0, 1] \times [0, 1]$ is a sufficient condition to ensure the existence of GMI. Sometimes, however, $f(p_{z_1 z_2}(u, v))$ might not be continuous, so let us investigate

the existence of GMI in more depth. As already done in (22), the $\sum_{i=1}^l \sum_{j=1}^l f\left(\frac{p(i, j)}{\frac{1}{l} \cdot \frac{1}{l}}\right) \cdot \frac{1}{l} \cdot \frac{1}{l}$ in (21) can be treated as the QE with convex function $\frac{1}{l^2} f(ul^2)$. Thus, (5) and (6) can be applied to this QE, which yield:

$$l^2 \frac{1}{l^2} f\left(\frac{1}{l^2} \cdot l^2\right) \leq \sum_{i=1}^l \sum_{j=1}^l f\left(\frac{p(i, j)}{\frac{1}{l} \cdot \frac{1}{l}}\right) \cdot \frac{1}{l} \cdot \frac{1}{l} \leq (l^2 - l) \frac{1}{l^2} f(0 \cdot l^2) + l \cdot \frac{1}{l^2} f\left(\frac{1}{l} \cdot l^2\right) \quad (23)$$

Namely,

$$f(1) \leq \sum_{i=1}^l \sum_{j=1}^l f\left(\frac{p(i, j)}{\frac{1}{l} \cdot \frac{1}{l}}\right) \cdot \frac{1}{l} \cdot \frac{1}{l} \leq f(0) \left(1 - \frac{1}{l}\right) + \frac{f(l)}{l} \quad (24)$$

Taking the limit as $l \rightarrow \infty$ and applying (21), we get:

$$f(1) \leq \gamma(r_1, r_2) \leq f(0) + \lim_{l \rightarrow \infty} \frac{f(l)}{l} \quad (25)$$

Equation (25) gives the lower and upper bounds of GMI. The lower bound is reached when r_1 and r_2 are independent [5]. The upper bound is reached when, for example, $r_1 \equiv r_2$, which is one of the most dependent cases. When $f(u) = u \log u$, (24) gives the lower and upper bounds of Shannon MI of l -level uniformly quantized versions of z_1 and z_2 , denoted $I_l(z_1, z_2)$, as follows:

$$0 \leq I_l(z_1, z_2) \leq \log l \quad (26)$$

When $z_1 \equiv z_2 \equiv z$, $I_l(z, z) = \log l$, which reaches the upper bound shown in (26). The MI of continuous variables is the limit of the MI of their quantized versions as the number of quantization levels goes to infinity [9]. Therefore, when $r_1 \equiv r_2 \equiv r$ and thus $z_1 \equiv z_2 \equiv z$,

$$I(r, r) = I(z, z) = \lim_{l \rightarrow \infty} \log l = \infty \quad (27)$$

We see that the MI between a continuous variable and itself is infinity. This is because that $p_{zz}(u, v)$ is infinite along the diagonal $u = v$ and causes the integration in (19) to diverge. Compared with the Shannon MI, there are more stable forms of GMI. For example, when $f(u) = \exp(-u)$, (25) becomes:

$$\exp(-1) \leq \gamma(r_1, r_2) \leq 1 \quad (28)$$

This shows that the GMI with $f(u) = \exp(-u)$ is lower bounded by $\exp(-1)$ (about 0.3679) and upper bounded by 1 and never diverge. It takes the finite value, 1, even for the most dependent case.

5. Orders of QFs with Respect to l

Table 1 shows the Q_β 's and Q_γ 's of some convex functions and their orders with respect to l as $l \rightarrow \infty$. The function $-\sin \pi u$, which appears in the last line, is not convex on $[0, \infty)$, so it cannot be used in GMI and Q_γ does not exist. However, it is convex on $[0, 1]$, so it can be used in QE and Q_β can be calculated. Note the Q_β and Q_γ of $-u^a$ ($0 < a < 1$), and the Q_γ of a^u ($0 < a < 1$). Their orders are all $\mathcal{O}(l^2)$, which are higher than $\mathcal{O}(l^2/\ln l)$, the order of the minus entropy and the MI. For the case of GO, (16) already shows that it can have the QF of $\mathcal{O}(l^2)$. This means that, when l is large enough, these measures will have sharper minima than the minus entropy and the MI.

Table 1. QFs of QE and GMI.

$f(u)$	$Q_\beta(f(u))$	$Q_\gamma(f(u))$
$-u^a$ ($0 < a < 1$)	$\frac{a(1-a)l^2}{1-l^{a-1}} = \mathcal{O}(l^2)$	$\frac{a(1-a)l^2}{1-l^{a-1}} = \mathcal{O}(l^2)$
$u \log u$	$\frac{l^2}{\ln l} = \mathcal{O}\left(\frac{l^2}{\ln l}\right)$	$\frac{l^2}{\ln l} = \mathcal{O}\left(\frac{l^2}{\ln l}\right)$
u^a ($a > 1$)	$\frac{a(a-1)l^2}{l^{a-1}-1} = \mathcal{O}(l^{3-a})$	$\frac{a(a-1)l^2}{l^{a-1}-1} = \mathcal{O}(l^{3-a})$
a^u ($a > 0, a \neq 1$)	$\frac{a^{\frac{1}{l^2}} \ln^2 a}{l^2 - l + la^{\frac{1}{l}} - l^2 a^{\frac{1}{l^2}}} = \mathcal{O}(l)$	$\frac{l^3 a \ln^2 a}{l-1+a^l-al} = \begin{cases} \mathcal{O}(l^2) & 0 < a < 1 \\ \mathcal{O}\left(\frac{l^3}{a^l}\right) & a > 1 \end{cases}$
$-\sin \pi u$	$\frac{\pi^2 \sin\left(\frac{\pi}{l^2}\right)}{-l \sin\left(\frac{\pi}{l}\right) + l^2 \sin\left(\frac{\pi}{l^2}\right)} = \mathcal{O}(1)$	

6. Numerical Experiments

In this section, we show some numerical results of applying GO, QE, and GMI to delay reconstruction of strange attractors. Figure 4 shows the time evolution $(x(t), y(t))$ of Rössler chaotic system [20], where t is the sampling number. The sampling time interval is $\pi/100$. The observed sequence $e(t)$ is obtained by adding a white noise uniformly distributed in $[-0.1, 0.1]$ to $x(t)$. We use $(e(t), e(t+\tau))$ to reconstruct the original chaotic attractor, where τ is the delay to be determined. We examine the curves of $\alpha(e(t), e(t+\tau))$, $\beta(e(t), e(t+\tau))$, and $\gamma(e(t), e(t+\tau))$ versus τ . τ can be chosen at the minima of these curves. So we hope these curves provide apparent minima.

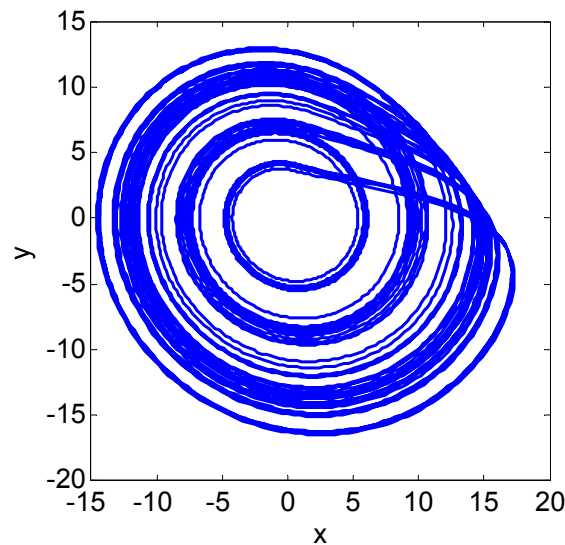
Figure 4. The Rössler attractor.

Figure 5 shows these curves where several different convex functions are used for QE and GMI. We can see that the order of QF correctly predicts the shape of minima: The higher (lower) the order, the sharper (blunter) the minima. Figures 5b–5g are plots of QE, where we have arranged them so that their orders of QFs decrease from high to low. Observing Figures 5b–5g, we can see the process that the minima of these curves of QE vary from sharp to blunt, and until very vague. In [9], it is shown that the order of QF should be at least $\mathcal{O}(l)$ to ensure that β keeps a well-defined minimum as l increases. The orders of QFs of Figures 5b–5d are all higher than $\mathcal{O}(l)$. The minima in these plots are all very prominent. Figure 5c is minus the entropy of $p(i, j)$ (labeled $-H$ in the plot). Adding $2\log_2 l = 2\log_2 100 = 13.29$ bits to its ordinate approximately leads to the plot of MI shown in Figure 5i. The QFs of Figures 5e and 5f are both $\mathcal{O}(l)$. Hence the shapes of their minima are basically identical. The order $\mathcal{O}(l)$ reaches the theoretical lower bound proposed in [9]. However, we can see that the neighboring areas of the minima are too flat and the positions of minima are not easy to locate. This is due to that $e(t)$ and $e(t+\tau)$ cannot be completely independent. Therefore, generally we should choose convex functions whose orders of QFs are higher than $\mathcal{O}(l)$. Figure 5e is indeed the variance between $p(i, j)$ and the uniform probability distribution. Figure 5e shows that such form of variance, though easily conceived of, is obviously not a good measure of independence. The shape of the QE plot of $f(u) = u^2$ is identical with Figure 5e. According to Table 1, under the precondition $a > 1$, the smaller a , the higher the order of QF of u^a . Thus the nearer a approaches 1, the better the effect of QE. For instance, Figures 5d and 5e show that the effect of $u^{1.001}$ is better than that of u^2 . Tsallis proved that the limit of his entropy is the Shannon entropy when the parameter q in his entropy, which corresponds to the a in $f(u) = u^a$ here, approaches 1 [6]. Figures 5c and 5d are good manifestations of that statement. We can see that the shapes of curves in Figures 5c and 5d look very close. What is more, that statement can be testified by the order of QFs. The order of QF of u^a where $a > 1$, $\mathcal{O}(l^{3-a})$, is lower than but approaches the order of QF of $u \log u$, $\mathcal{O}(l^2/\ln l)$, as a approaches 1. We have used in Figure 5g the sine function $-\sin \pi u$. This can be traced back to Kapur who used

trigonometric functions to create entropies [7]. The QF of $-\sin \pi u$ is of constant order $\mathcal{O}(1)$. Therefore, the neighborhoods of the minima in Figure 5g are even more flat than those in Figures 5e and 5f, and the positions of minima are hardly located. Finally, we can see that the minima in Figures 5a, 5b, and 5h, whose order is $\mathcal{O}(l^2)$, are sharper, and their positions are more definite than those in Figures 5c and 5i whose order is $\mathcal{O}(l^2/\ln l)$. Namely, these measures outperform the minus entropy and the MI in the prominence of minima. Of course, the minima in Figures 5a–5d, 5h, and 5i are all distinct enough, and their positions coincide. The first (leftmost) minimum, $\tau = 47$, is a good choice of the delay for reconstruction. To verify this point, Figures 6a, 6b, and 6c show the delay portraits using the first (leftmost) minimum, a non-minimum, and the seventh minimum, respectively. Compared with Figure 4, it is clear that Figure 6a reproduces well the folding structure of the original Rössler attractor.

Figure 5. GO, QE, and GMI of delay-coordinate variables of Rössler chaotic system *versus* delay τ . All plots are calculated using 65,536 sample points. $l = 100$ is used for GO and QE. Delay is measured in sampling numbers. The convex functions $f(u)$'s used in (b)–(i) and the orders of QFs of (a)–(i) are labeled in their captions.

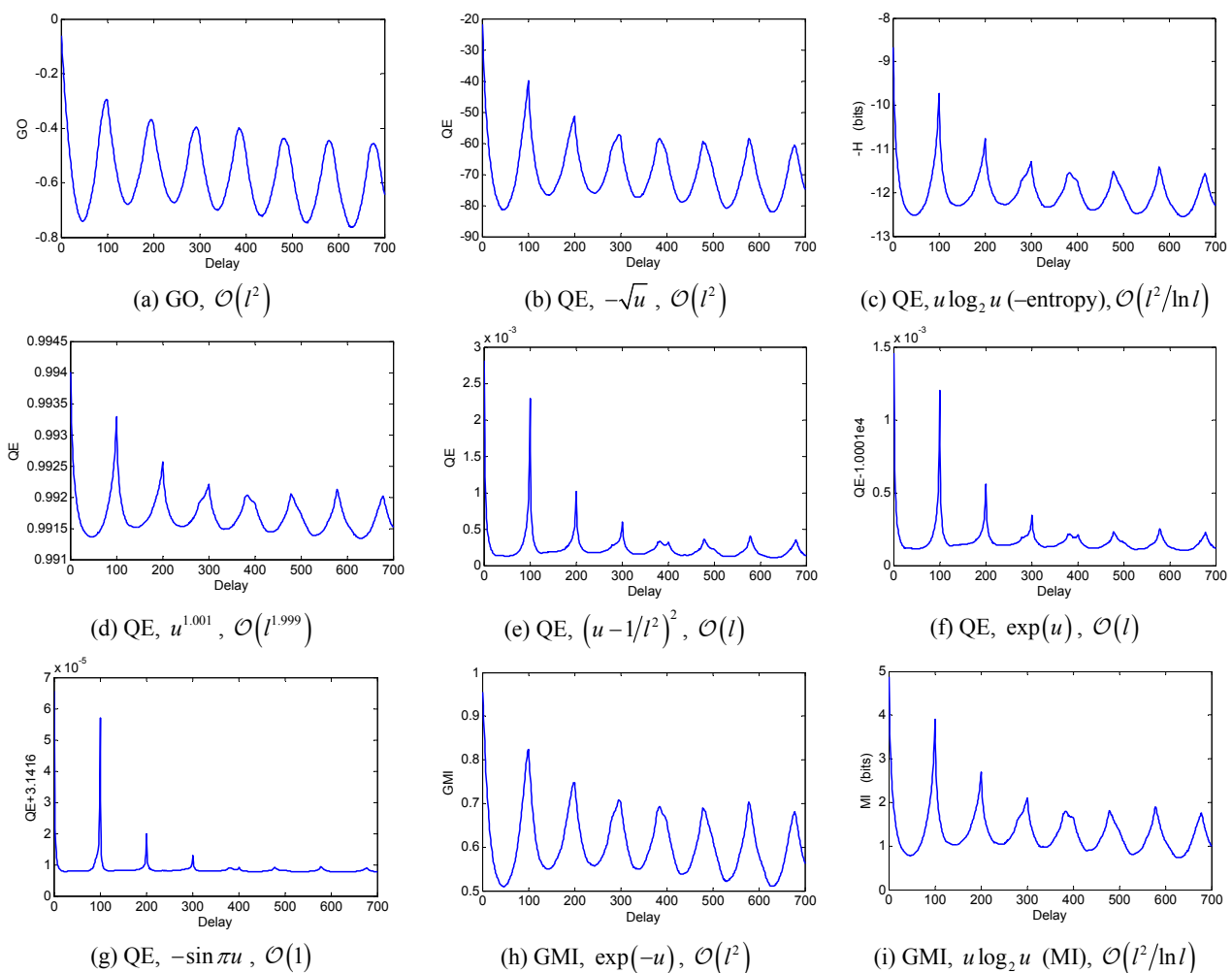
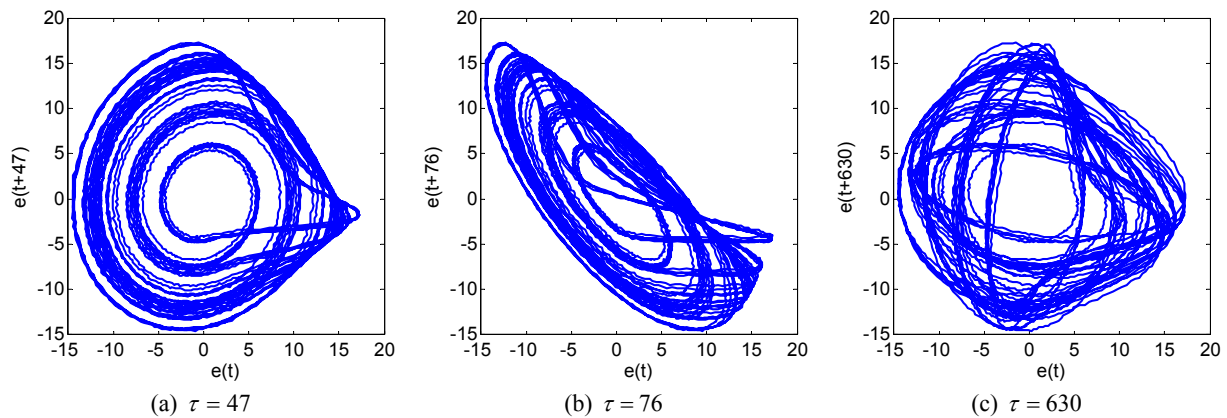
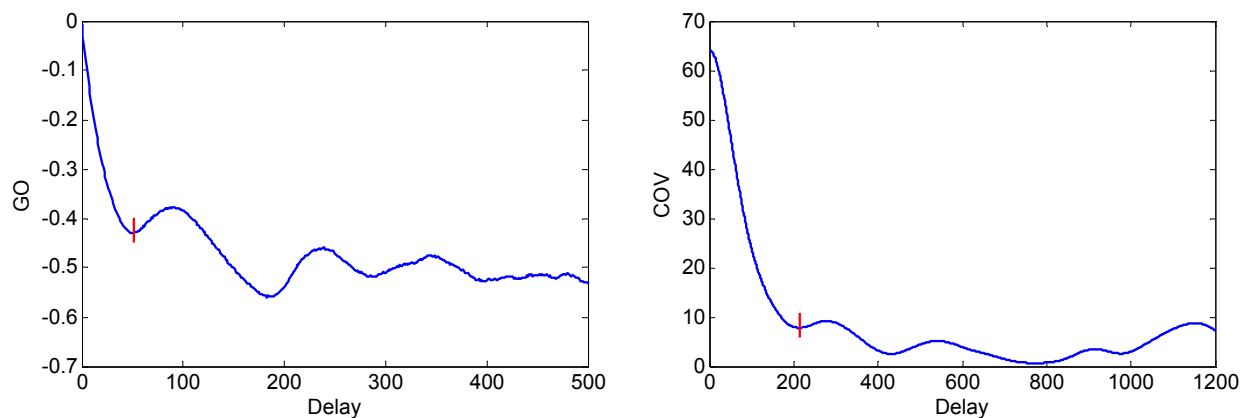


Figure 6. Delay reconstruction portraits of the Rössler attractor using time delays corresponding to (a) the first minimum, (b) a non-minimum, and (c) the seventh minimum, respectively, of the curves in Figures 5a–5d, 5h, and 5i.



The covariance function reflects the correlation between signals and sometimes can also be used for choosing τ [23]. However, generally speaking, the independence measures GO, QE, and GMI, like MI, are better criteria than the covariance. To illustrate this point, let us see the second example, which is on the Lorenz system [22,24]. The sampling time interval is $\pi/1000$. The observed sequence $e(t)$ is also obtained by adding noise to $x(t)$. Figures 7a and 7b plot the curves of $\alpha(e(t), e(t+\tau))$ and the auto-covariance function $\text{cov}(e(t), e(t+\tau))$ versus τ . Figure 8 shows that the reconstruction portrait with the first minimum of GO reproduces the double wing structure of the Lorenz attractor fairly well. The results are similar when using the QE or GMI. By contrast, none of the minima of the covariance function in Figure 7b are good choices. For instance, Figure 8c is the delay portrait using the first minimum of covariance function. It looks much more complex than Figures 8a and 8b. Actually, the curve of covariance function misses the best choice of τ , which appears earlier than its first minimum.

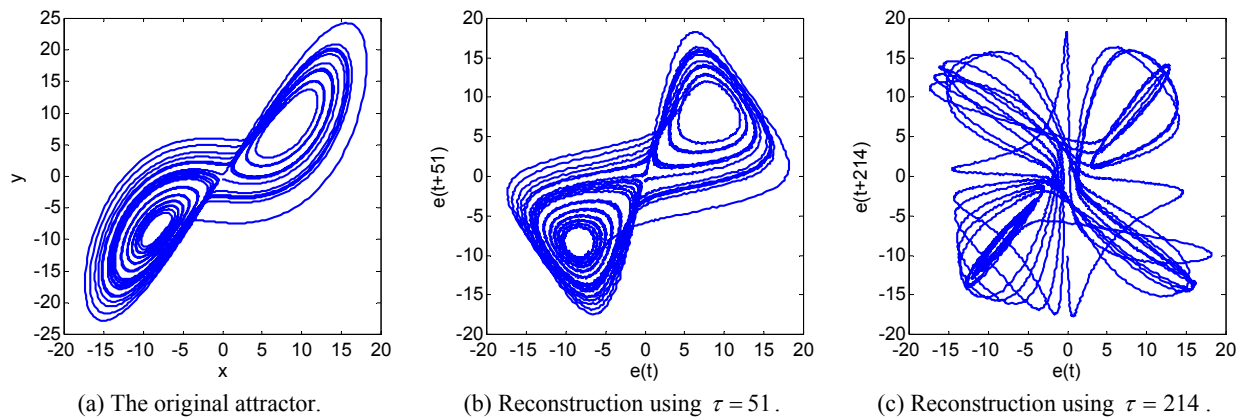
Figure 7. GO and auto-covariance function of reconstructed variables of the Lorenz attractor versus τ . $N = 50000$ sample points are used in both. $l = 200$ in GO.



(a) GO, the short red stem indicates the first minimum $\tau = 51$.

(b) Auto-covariance function, the short red stem indicates the first minimum $\tau = 214$.

Figure 8. Comparison of reconstruction effects of the Lorenz attractor. The reconstruction delays of (b) and (c) are the first minima of GO and auto-covariance function in Figure 7, respectively.



Note that GO and QE are much more efficient to compute than MI and GMI so that they work well with real-time applications such as independent component analysis [8] and blind source separation [9]. MI and GMI, however, can give more accurate estimate of the degree of independence.

7. Conclusions

We have studied some convex functions based measures of independence. Essentially, the independence of variables is equivalent to the uniformity of the joint PDF of the variables obtained by transforming the original variables by their respective CDFs. The MI just uses the convex function $u \log u$ to penalize the nonuniformity of this joint PDF, and thus measures the independence. Similarly, the QE uses an arbitrary convex function to penalize the nonuniformity of the discretized version of this PDF. Since any convex functions can appear in it, the QE is a broad enough definition that actually includes quite a few existing definitions of entropies. In this paper, we have focused on how different selections of convex functions affect the performance of the corresponding independence measure. When the QE is plotted *versus* the variance from the uniform distribution, the QF of QE is twice the slope of QE at its minimum over the dynamic range of QE and, therefore, indicates the sharpness of QE around its minimum. The expectation of GO is a subclass of QE, and the Csiszar's GMI is the limit of a special form of QE as the number of quantization levels approaches infinity. Based on these two facts, the definition of QF can be generalized to be made also applicable for GO and GMI. The order of QF with respect to the number of quantization levels well predicts the sharpness of the minima of these measures of independence. Furthermore, it indicates that GO, QE, and GMI can have more prominent minima than the minus entropy and the MI.

Besides theoretical study, we have also applied the convex functions based measures of independence to chaotic signal processing. Initiated by Fraser and Swinney, the MI has become the most widely used criterion for choosing delay in strange attractor reconstruction. However, there is no evidence to show that it is the optimum criterion. Indeed, GO, QE, and GMI are good alternative and sometimes even better criteria for choosing delay. We have proposed a recursive algorithm for computing GMI that is a generalization of Fraser and Swinney's algorithm for computing MI. In

addition, we have conducted numerical experiments on chaotic systems that well exemplify the theory of QF that we have developed.

Finally we note that whether a measure of information is good or not is application dependent. In some applications, a sharp minimum might not be crucial. We shall describe such application in a forthcoming paper. Future work also includes applying the convex function based measures to real world problems such as EEG analysis. To conclude, the results presented in this paper is instructive in understanding the behavior of convex functions based information measures and helping selecting the most suitable measures for various applications.

Acknowledgements

The authors thank H. Shimokawa for his valuable comments. This work was partly supported by the National Natural Science Foundation of China under Grant 60872074 and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry. The authors also thank the anonymous reviewers for their valuable comments and suggestions. This work is an invited contribution to the special issue of *Entropy* on advances in information theory.

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423; 623–656.
2. Zemansky, M.W. *Heat and Thermodynamics*; McGraw-Hill: New York, NY, USA, 1968.
3. Renyi, A. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
4. Havrda, J.; Charvat, F. Quantification method of classification processes. *Kybernetika* **1967**, *1*, 30–34.
5. Csiszar, I. Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.* **1967**, *2*, 299–318.
6. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
7. Kapur, J.N. *Measures of Information and Their Applications*; John Wiley & Sons: New York, NY, USA, 1994.
8. Chen, Y. A novel grid occupancy criterion for independent component analysis. *IEICE Trans. Fund. Electron. Comm. Comput. Sci.* **2009**, *E92-A*, 1874–1882.
9. Chen, Y. Blind separation using convex functions. *IEEE Trans. Signal Process.* **2005**, *53*, 2027–2035.
10. Kapur, J.N. *Maximum-Entropy Models in Science and Engineering*; John Wiley & Sons: New York, NY, USA, 1989.
11. Xu, J.; Liu, Z.; Liu, R.; Yang, Q. Information transmission in human cerebral cortex. *Physica D* **1997**, *106*, 363–374.
12. Kolarczyk, B. Representing entropy with dispersion sets. *Entropy* **2010**, *12*, 420–433.
13. Takata, Y.; Tagashira, H.; Hyono, A.; Ohshima, H. Effect of counterion and configurational entropy on the surface tension of aqueous solutions of ionic surfactant and electrolyte mixtures. *Entropy* **2010**, *12*, 983–995.

14. Zupanovic, P.; Kuic, D.; Losic, Z.B.; Petrov, D.; Juretic, D.; Brumen, M. The maximum entropy production principle and linear irreversible processes. *Entropy* **2010**, *12*, 996–1005.
15. Van Dijck, G.; Van Hulle, M.M. Increasing and decreasing returns and losses in mutual information feature subset selection, *Entropy* **2010**, *12*, 2144–2170.
16. Takens, F. Detecting strange attractors in turbulence. In *Warwick 1980 Lecture Notes in Mathematics*; Springer-Verlag: Berlin, Germany, 1981; Volume 898, pp. 366–381.
17. Fraser, A.M.; Swinney, H.L. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **1986**, *33*, 1134–1140.
18. Gretton, A.; Bousquet, O.; Smola, A.; Scholkopf B. Measuring statistical dependence with Hilbert-Schmidt norms. In Proceedings of the 16th International Conference on Algorithmic Learning Theory, Singapore, October 2005; pp. 63–77.
19. Szekely, G.J.; Rizzo, M.L.; Bakirov, N.K. Measuring and testing dependence by correlation of distances. *Ann. Stat.* **2007**, *35*, 2769–2794.
20. Nelsen, R.B. *An Introduction to Copulas, Lecture Notes in Statistics 139*; Springer-Verlag: New York, NY, USA, 1999.
21. Chen, Y. On Theory and Methods for Blind Information Extraction. Ph.D. dissertation, Southeast University, Nanjing, China, 2001.
22. Rössler, O.E. An equation for continuous chaos. *Phys. Lett. A* **1976**, *57*, 397–398.
23. Longstaff, M.G.; Heath, R.A. A nonlinear analysis of the temporal characteristics of handwriting. *Hum. Movement Sci.* **1999**, *18*, 485–524.
24. Lorenz, E.N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **1963**, *20*, 130–141.

Appendix: Recursive Algorithm for Computing GMI and Related Issues

A. Recursive Algorithm for Computing GMI

Recall (21) and rewrite it as follows:

$$\lim_{l \rightarrow \infty} \sum_{i=1}^l \sum_{j=1}^l f \left(\frac{p(i,j)}{\frac{1}{l} \cdot \frac{1}{l}} \right) \cdot \frac{1}{l} \cdot \frac{1}{l} = \int_0^1 \int_0^1 f(p_{z_1 z_2}(u,v)) du dv = \gamma(r_1, r_2) \quad (29)$$

where $p(i, j)$ is as defined in (3). Thus, when l is fairly large, $\gamma(r_1, r_2) \approx \sum_{i=1}^l \sum_{j=1}^l f(p(i, j) l^2) / l^2$.

Alternatively, a more delicate approach can be taken. Namely, different numbers of quantization levels are used according to the bumpiness of $p_{z_1 z_2}$ in different local regions. Larger l should be taken in more fluctuant regions to avoid the estimated γ from being too small, whereas smaller l should be taken in rather flat regions to avoid the estimated γ from being too large due to limited sample size.

Specifically, let $l = 2^m$, $m = 0, 1, 2, \dots$, then:

$$\gamma_m = \sum_{i=1}^{2^m} \sum_{j=1}^{2^m} \frac{f(p(i, j) 4^m)}{4^m} \rightarrow \gamma(r_1, r_2) \text{ as } m \rightarrow \infty \quad (30)$$

As shown in Figure 9, each square in the $2^m \times 2^m$ grid consists of four squares in the $2^{m+1} \times 2^{m+1}$ grid. Let us denote a certain square in the $2^m \times 2^m$ grid by S and denote the four squares in the $2^{m+1} \times 2^{m+1}$ grid that constitute S by S_1 , S_2 , S_3 , and S_4 . Also, let:

$$\begin{aligned} \text{Prob}\{ (z_1, z_2) \in S \} &= P \\ \text{Prob}\{ (z_1, z_2) \in S_i \} &= P_i, \quad i = 1, 2, 3, 4 \end{aligned} \quad (31)$$

If (z_1, z_2) is uniformly distributed over S , then:

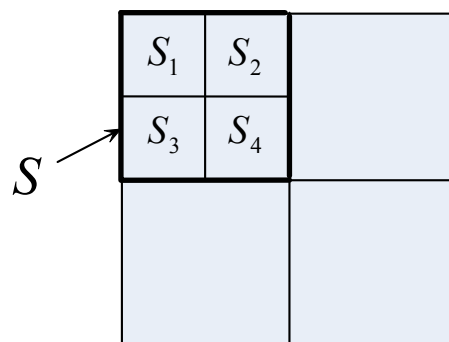
$$\begin{aligned} P_1 = P_2 = P_3 = P_4 &= \frac{P}{4} \\ \frac{f(P \times 4^m)}{4^m} &= \sum_{i=1}^4 \frac{f(P_i \times 4^{m+1})}{4^{m+1}} \end{aligned} \quad (32)$$

Namely,

$$\text{Component of } \gamma_m \text{ on } S = \text{Component of } \gamma_{m+1} \text{ on } S \quad (33)$$

Hence there is no need to further divide S into S_1 , S_2 , S_3 , and S_4 . Thus, the following recursive algorithm for computing $\gamma(r_1, r_2)$ is obtained.

Figure 9. Each square in the $2^m \times 2^m$ grid is composed of four squares in the $2^{m+1} \times 2^{m+1}$ grid.



Let $\xi_m(i, j)$ denote the square in the i th column and the j th row of the $2^m \times 2^m$ uniform grid over $[0, 1] \times [0, 1]$, i.e.,

$$\xi_m(i, j) = \left[\frac{i-1}{2^m}, \frac{i}{2^m} \right] \times \left[\frac{j-1}{2^m}, \frac{j}{2^m} \right] \quad (34)$$

and let:

$$p_m(i, j) = \text{Prob}\{ (z_1, z_2) \in \xi_m(i, j) \} \quad (35)$$

Then,

$$\gamma(r_1, r_2) = F(\xi_0(1, 1)) \quad (36)$$

where, if $p_{z_1 z_2}$ is uniform in $\xi_m(i, j)$, then

$$F(\xi_m(i, j)) = f(4^m p_m(i, j)) \quad (37)$$

If $p_{z_1 z_2}$ is not uniform in $\xi_m(i, j)$, then we further divide $\xi_m(i, j)$ into four squares in the $2^{m+1} \times 2^{m+1}$ grid, noting that:

$$\xi_m(i, j) = \bigcup_{\substack{k=0, 1 \\ d=0, 1}} \xi_{m+1}(2i-k, 2j-d) \quad (38)$$

yielding:

$$F(\xi_m(i, j)) = \frac{1}{4} \sum_{k=0}^1 \sum_{d=0}^1 F(\xi_{m+1}(2i-k, 2j-d)) \quad (39)$$

It is easy to verify that, when $f(u) = u \log u$, the above algorithm reduces to Fraser and Swinney's recursive algorithm for computing MI [17].

B. Uniformity Test

In the above recursive algorithm for computing GMI, we need to judge whether $p_{z_1 z_2}$ is uniform in $\xi_m(i, j)$. This can be done using the χ^2 test proposed in [17]. However, the 20% confidence level in [17] was chosen arbitrarily. Experiments show that the uniformity test with the 20% confidence level may be too stringent and cause the GMI estimate to be too large. To solve this problem, we here propose a simple method for choosing the confidence level. We mix two independent variables, s_1 and s_2 , with identical distributions by a 2×2 rotation matrix with rotation angle θ to get two variables, r_1 and r_2 , *i.e.*,

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (40)$$

The MI of r_1 and r_2 is then a function of θ , *i.e.*,

$$I(\theta) = I(r_1, r_2) \quad (41)$$

Referring to the χ^2 distribution table, we choose five confidence levels, 20, 10, 5, 2, and 1%, to estimate $I(\theta)$. If s_1 and s_2 have some standard distributions, then standard values of $I(\theta)$ are easy to obtain. Examining in turn the two cases where s_1 and s_2 are both uniform variables and both Laplacian variables, we can spot that the 5% confidence level produces the minimum estimation error in both cases. Comparing the estimation results using 20% and 5% confidence levels with the standard curves, we can see that the results with the 20% confidence level are obviously too large, whereas those with the 5% confidence level are quite accurate, as shown in Figures 10 and 11. Therefore, 5% is a better choice for the confidence level.

Figure 10. MI of rotational mixtures of two independent identical uniform variables *versus* rotation angle.

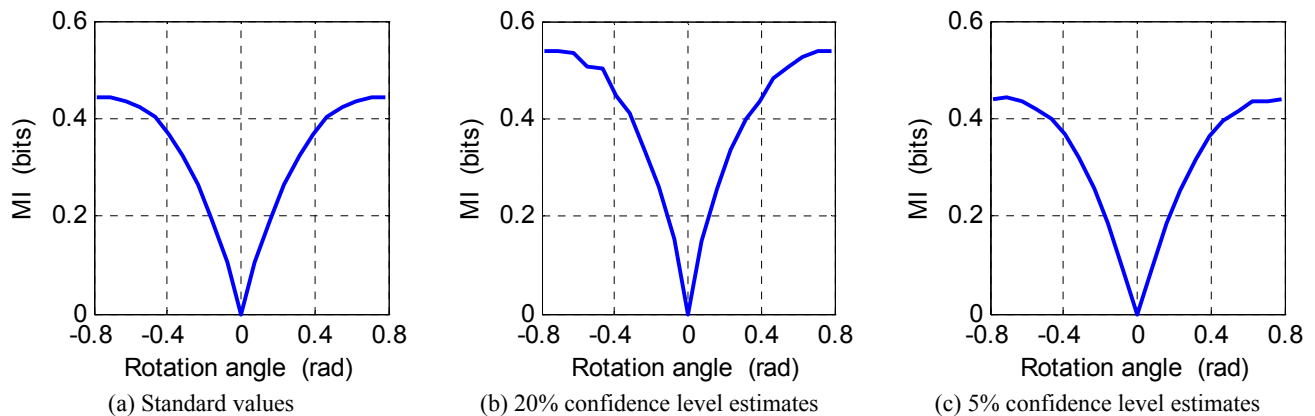
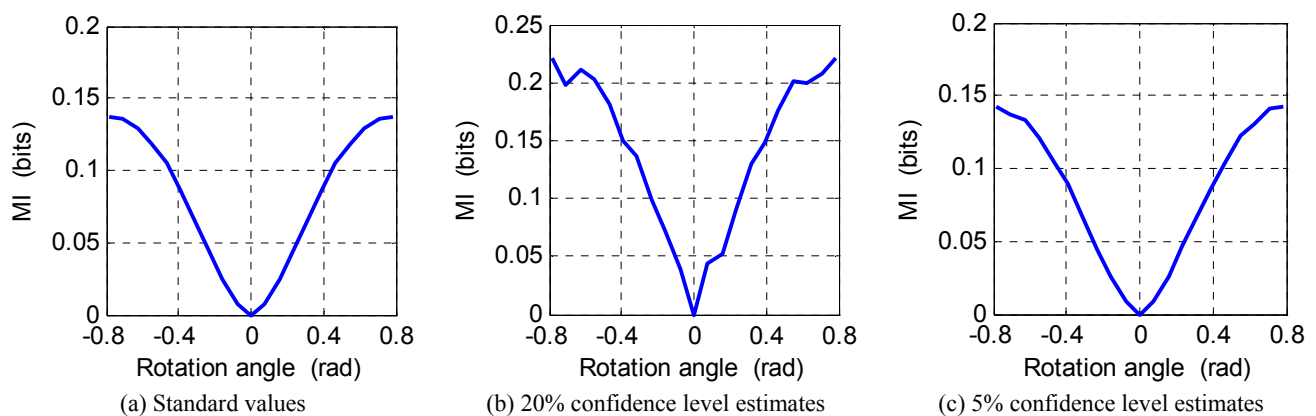


Figure 11. MI of rotational mixtures of two independent identical Laplacian variables *versus* rotation angle.



C. Practical Implementation

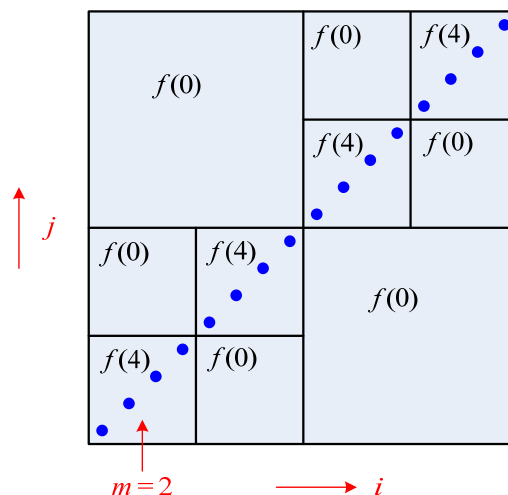
In computing GO, QE, and GMI, the original variables should be transformed by their CDFs. In practice, the analytical forms of CDFs are usually unknown. However, the lack of analytical CDFs does not impede the computation of these measures. Using the definition of CDF, if we sort the N observations of variable r denoted by $r(1), \dots, r(N)$ in ascending order and among them $r(t)$ is in the i th place, then i/N is an unbiased estimate of $q_r(r(t))$. Indeed, when computing GO and QE, we may directly map the sample points of (r_1, r_2) to the indices of the squares that the corresponding sample points of (z_1, z_2) belong to [8,9]. When computing GMI, a sample point of (r_1, r_2) is first mapped to an order pair (i, j) . Then the square that (i, j) belongs to is determined according to the present partitioned grid. These operations (mainly sorting) do not involve floating point operations and can achieve high efficiency. The $p(i, j)$ ($p_m(i, j)$) in QE (GMI) is estimated by the ratio of the sample points in the square $\xi(i, j)$ ($\xi_m(i, j)$) to the total number of sample points.

D. Output of the Recursive Algorithm of GMI

If r_1 and r_2 are independent, (z_1, z_2) will be uniform in $[0,1] \times [0,1]$. The recursive algorithm for computing GMI will terminate at the $m=0$ hierarchy and produce the minimal output $f(1)$. The maximal output is generated for the most dependent case, e.g., $r_1 \equiv r_2$. Assume that we use $N = 2^k$ sample points of two same variables $r_1 \equiv r_2$ to compute GMI. The corresponding sample points of (z_1, z_2) will line up along the diagonal. The χ^2 uniformity tests, no matter using 20, 10, or 5% confidence levels, all lead to the same final partition that divides the sample points of (z_1, z_2) into every four points in a square. For example, Figure 12 shows the case of $N = 16$ sample points. In this case, the computation of GMI proceeds as follows:

$$\begin{aligned}
 \gamma &= F(\xi_0(1,1)) = \frac{1}{4} \cdot [F(\xi_1(1,1)) + F(\xi_1(1,2)) + F(\xi_1(2,1)) + F(\xi_1(2,2))] \\
 &= \frac{1}{4} \cdot \left\{ \frac{1}{4} \cdot [F(\xi_2(1,1)) + F(\xi_2(1,2)) + F(\xi_2(2,1)) + F(\xi_2(2,2))] + f(0) + f(0) \right. \\
 &\quad \left. + \frac{1}{4} \cdot [F(\xi_2(3,3)) + F(\xi_2(3,4)) + F(\xi_2(4,3)) + F(\xi_2(4,4))] \right\} \\
 &= \frac{1}{4} \times \left\{ \frac{1}{4} \times [f(4) + f(0) + f(0) + f(4)] + f(0) + f(0) + \frac{1}{4} \times [f(4) + f(0) + f(0) + f(4)] \right\} \\
 &= \frac{3}{4} f(0) + \frac{1}{4} f(4)
 \end{aligned} \tag{42}$$

Figure 12. Partition of $N = 16$ sample points of (z_1, z_2) of two same variables.



The result of (42) can be generalized to the case of $N = 2^k$ sample points:

$$\gamma = \left(1 - \frac{1}{2^m}\right) f(0) + \frac{1}{2^m} f(2^m) \tag{43}$$

where:

$$m = k - 2 = \log_2 N - 2 \tag{44}$$

Thus, we obtain the lower and upper bounds of GMI computed using the recursive algorithm as follows.

$$f(1) \leq \gamma(r_1, r_2) \leq \left(1 - \frac{1}{2^m}\right) f(0) + \frac{1}{2^m} f(2^m) \quad (45)$$

We can check whether overflow might occur with (45) when a convex function such as $f(u) = a^u$ ($a > 1$) is used.

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).