Generalized Formula of Physical Channel Capacities[1]

Bohdan Hejna
Navigatoru 615/15, 161 00 Praha 6, Prague, Czech Republic
e-mail: bhejna@cpoj.cz bhejna@seznam.cz

Abstract
We consider a process of information transmission in the environment described by the canonical distribution of noise, considered as the transmission channel in the Information Theory. We derive the generalized form of the noisy entropy of this environment. Then, the generalized information capacity formula is derived for the geometric distribution of the output variable under the condition of greater or equal than the minimal value of the mathematical expectation of the input variable. We also state the hypotheses the capacity, when this input parameter is less or equal than that minimal (critical, extreme) value, is just defined by the lower capacity estimation which is the capacity for the limit distribution with this minimal input parameter. This paper generalizes the paper [6] presented on CASYS'98.

Keywords: Information, Capacity, Channel, Noise, Distribution

1 Introduction

In the present paper we study an information transmission in the environment described by the canonical distribution of noise

$$Pr(s) = qp^s \quad 0 \leq s \leq r, \quad q = \frac{1-p}{1-p^{r+1}},$$

which is transformed into the Fermi–Dirac distribution if we put $p = e^{\frac{-\varepsilon}{kT}}$ and $r = 1$, or into the Bose–Einstein distribution if $p = e^{\frac{-\varepsilon}{kT}}$ and $r = \infty$.

We derive the generalized form of the noise entropy of this environment which we consider as the information transmission channel of Information Theory,

$$H_r(p) = \frac{h(p)}{1-p} - \frac{h(p^{r+1})}{1-p^{r+1}},$$

where $h(\cdot)$ is the entropy function $-p\ln p - (1-p)\ln(1-p)$. Information capacity is then evaluated as

$$C(p, r|W) = \frac{h(x)}{1-x} - H_r(p)$$

[1]This paper is an excerpt of Chapter 6 in [7].

for $0 < p < 1,\ r \geq 1$ and geometrical distribution $x(j) = (1-x)x^j,\ j \geq 0$ of the output variable under the condition $W \geq W_{crit}$ for the mathematical expectation W of the input variable. This condition is considered to be the main result of this paper. We also state the limit distribution of the input variable for its extreme expectation value

$$W = W_{crit} = \frac{(r+1)p^{r+1}}{1-p^{r+1}}.$$

The condition binding the output and input variable together is expressed in terms of their expectations as

$$E = \frac{p(1-c)}{1-p} + W, \quad \text{where} \quad c = \frac{(r+1)p^r(1-p)}{1-p^{r+1}}.$$

The lower capacity estimation for $W \leq W_{crit}$ is then stated as

$$C_*(p,r|W) = \frac{h(p^{r+1})}{1-p^{r+1}}, \quad \text{where} \quad \frac{p^{r+1}}{1-p^{r+1}} = \frac{W}{r+1}.$$

We express the hypotheses that $C(p,r|W) = C_*(p,r|W)$ for $W \leq W_{crit}$.

2 Generalized Formula of Physical Channel Capacities

2.1 Definition and Formulation of the Problem

Let us consider the Hermitian operator of energy $\boldsymbol{\varepsilon}$ of quantum particles with spectrum $\mathcal{S}(\boldsymbol{\varepsilon})$ of eigenvalues ε_i (energetic levels of the particle) in pure states $\boldsymbol{\theta}_i$ of the system Ψ under consideration. We assume that a variable $\boldsymbol{\alpha}$ with a spectrum $\mathcal{S}(\boldsymbol{\alpha}) = \{\alpha_0, \alpha_1, \ldots\}$ is measured with the probabilities

$$p(\alpha_j|\boldsymbol{\alpha}|\boldsymbol{\theta}_i) = \begin{cases} p(j-i) & \text{for} \quad j \geq i \\ 0 & \qquad\quad j < i, \end{cases}$$

where $\{p(0), p(1), p(2), \ldots\} = \Pr(\cdot)$ is a probability distribution defined on the set $\{0, 1, \ldots\}$. This situation occurs when, for example, a particle is excited by a random interaction from the energetic level ε_i to ε_{i+s}, and the spectral energetic jump s is random with the distribution $\Pr(s)$, $s = 0, 1, 2, \ldots$. Then, energy of the excited particle is measured. The excitement can occur as a result of interaction with another particle or a wave. If the particle energy is $\varepsilon_i = i\varepsilon$, then, for example, after interaction with a wave with energy $\varepsilon_s = s\varepsilon$, and after absorbing its energy, the *additive* energetic jump to the level $\varepsilon_{i+s} = (i+s)\varepsilon$ occurs. In the example stated, it is the energy distribution $\Pr(\cdot)$ in the environment interacting with the emitted particle that plays the key role. For an observer capable of measuring the particle energy, that particle represents an information signal.

The distribution $Pr(s)$, $s \geq 0$, also has a similar meaning in a general case, when an observer of the variable $\boldsymbol{\alpha}$ can obtain information of the state $\boldsymbol{\theta}$ of the system Ψ. If $Pr(s) = 0$ for all $s \neq 0$ and $Pr(0) = 1$, then, based upon the measured value $\boldsymbol{\alpha} = \alpha_k$, the observer can determine, without errors, that the system is in the pure state $\boldsymbol{\theta}_k$. Otherwise the observer's determination will be less accurate. If we have a possibility to bring gradually the system Ψ into arbitrary pure states $\boldsymbol{\theta}_{i_1}, \boldsymbol{\theta}_{i_2}, \ldots, \boldsymbol{\theta}_{i_n}$ from a state subspace $\Theta_0 \subset \Theta$, then the observer who takes n independent observations of the variable $\boldsymbol{\alpha}$, thus obtaining independent random values $\alpha_{k_1}, \alpha_{k_2}, \ldots, \alpha_{k_n}$ in these states, receives a certain amount of information $I(\boldsymbol{\theta}; \boldsymbol{\alpha})$.

Let us assume we would like to transfer one of the messages from a source $\mathcal{Z}_n = \{1, 2, \ldots, M_n\}$, where the number of messages M_n increases with numbers of observations n as the integer-related portion of the number e^{Rn}, where $R > 0$ is an invariable called *information rate* of the source. It follows from the Shannon coding theorem (see [3]), that a message from the source $\mathcal{Z}_n$ can be transferred to the observer with an arbitrary small error probability when the number of observations n is great enough, provided that $R < C$, where

$$C \triangleq C(\Psi|\{\boldsymbol{\alpha}\}, \Theta_0) = \sup_{\boldsymbol{\theta} \in \Theta_0} I(\boldsymbol{\theta}; \boldsymbol{\alpha}) \tag{1}$$

is called "relative capacity" of the system Ψ.

We will look for the probability distribution $Pr(s)$, $s \geq 0$, within the class of the Gibbs canonical distributions in the form

$$Pr(s) = \frac{e^{-\beta E(s)}}{Z(\beta)}, \quad s \geq 0,$$

where the energy $E(s)$ will be different from zero in all of the states $s \geq 0$ of the "system noise" under consideration, or only in the states $0 \leq s \leq r$, where $r > 0$ is finite. Here, $\beta > 0$ is a parameter and $Z(\beta)$ is the statistical sum. For simplicity, we put $\beta E(s) = \theta s$, where the value of the parameter θ is usually considered to be ε / kT, where $\varepsilon > 0$ is energy, $T > 0$ is temperature and k is the Boltzman constant. Consequently, we consider the distribution

$$Pr(s) = \begin{cases} \frac{e^{-\theta s}}{Z(\theta)} & \text{for} \quad 0 \leq s \leq r \\ 0 & \text{for} \quad s > r, \end{cases} \tag{2}$$

where $\theta > 0$ and $Z(\theta) = (1 - e^{-\theta(r+1)})/(1 - e^{-\theta})$, or alternatively,

$$Pr(s) = \begin{cases} qp^s & \text{for} \quad 0 \leq s \leq r, \quad q = \frac{1-p}{1-p^{r+1}} \\ 0 & \text{for} \quad s \geq r, \end{cases} \tag{3}$$

where $0 < p < 1$. In (2) and (3) $r \geq 1$ is either a finite integer or $r = \infty$. In the case of $r = \infty$ the distribution $Pr(s)$ is always positive and $Z(\theta) = 1/(1 - e^{-\theta})$, respectively $q = 1 - p$. Then, in (3) we substitute

$$p = e^{-\theta} \text{ resp. } p = e^{-\frac{\varepsilon}{kT}}. \tag{4}$$

The extreme cases of the distribution given by relations (2) and (3) correspond for the values $r = 1$ and $r = \infty$ with known distributions of statistical physics

For $r = 1$, the only probabilities different from zero are

$$Pr(0) = \frac{1}{1+p} \quad \text{and} \quad Pr(1) = 1 - Pr(0) = \frac{p}{1+p}. \tag{5}$$

With the substitution (4) we obtain the known Fermi-Dirac distribution, see [9]. Therefore, the distributions (2) and (3) with $r = 1$ are called *the F–D (Fermi-Dirac)-type distributions.* If, instead of the substitution (4), we take

$$p = \frac{e^{-\theta}}{1 - e^{-\theta}} \quad \text{resp.} \quad p = \frac{e^{-\frac{\varepsilon}{kT}}}{1 - e^{-\frac{\varepsilon}{kT}}}, \tag{6}$$

we obtain the Maxwell-Boltzman distribution.

For $r = \infty$, all probabilities are different from zero,

$$Pr(s) = (1 - p)p^s, \ s \geq 0, \tag{7}$$

and when substituting (4) we obtain the known Bose-Einstein distribution, see [9]. Therefore, the distributions (2) and (3) with $r = \infty$ are called *the B–E (Bose-Einstein)-type distributions.*

2.2 Additive Physical Channel Capacity

We will consider only such states $\boldsymbol{\theta}$ the distribution of which $q(i) = q(i|\boldsymbol{\theta})$ holds

$$\sum_{i=0}^{\infty} iq(i) = W, \tag{8}$$

where W is *the power parameter* of the "signal variable" $\boldsymbol{\theta}$ on the channel input (if $\varepsilon_i = i\varepsilon_0$ is the energy considered at the beginning of Subsection 2.1., and if we use $\tau > 0$ as a symbol for "time window", in which "modulation of input" and a consequent process of measuring of the variable $\boldsymbol{\alpha}$ is performed, which is an "input-output communication action", then $\varepsilon_0 W/\tau$ is the power on the input of the channel). In this paper we deal with the capacity

$$C(p, r|W) = \sup_{\boldsymbol{\theta} \in \Theta_W} I(\boldsymbol{\alpha}; \boldsymbol{\theta}), \quad W > 0, \tag{9}$$

where Θ_W is a set of the states $\boldsymbol{\theta}$ the distributions of which satisfies the condition (8) and $I(\boldsymbol{\alpha}; \boldsymbol{\theta})$ is the information transferred.

Lemma 2.1 In the additive physical channel it holds, for the output entropy $H(\boldsymbol{\alpha}\|\boldsymbol{\theta}_i)$ in the input state $\boldsymbol{\theta}_i$, for all $i \geq 0$, $0 < p < 1$ and $1 \leq r \leq \infty$ that

$$H(\boldsymbol{\alpha}\|\boldsymbol{\theta}_i) = \frac{h(p)}{1-p} - \frac{h(p^{r+1})}{1-p^{r+1}} \triangleq H_r(p), \tag{10}$$

where $p^\infty = \lim_{r\to\infty} p^r = 0$ and

$$h(p) = -p\ln p - (1-p)\ln(1-p) \tag{11}$$

is the entropy function of the parameter p. Therefore the capacity of this channel is given by the relation

$$C(p,r|W) = \sup_{\boldsymbol{\theta}\in\Theta_W} H(\boldsymbol{\alpha}\|\boldsymbol{\theta}) - H_r(p) \tag{12}$$

for Θ_W defined above and the output entropy $H(\boldsymbol{\alpha}\|\boldsymbol{\theta})$.

Proof. The relation (10) follows from (3) and from that $H(\boldsymbol{\alpha}\|\boldsymbol{\theta}_i)$ occurs with probabilities $q(i|\boldsymbol{\theta})$. It follows from (10) and from the definition of $I(\boldsymbol{\alpha};\boldsymbol{\theta})$ that

$$I(\boldsymbol{\alpha};\boldsymbol{\theta}) = H(\boldsymbol{\alpha}\|\boldsymbol{\theta}) - H_r(p) \tag{13}$$

and the relation (12) follows from here, from (9) and from the definition of Θ_W.

In the lemma below and further on, we consider a distribution of the state $\boldsymbol{\theta}$

$$q(i) = \begin{cases} q(i|\boldsymbol{\theta}) & \text{when} \quad i \geq 0 \\ 0 & \text{when} \quad i < 0 \end{cases} \tag{14}$$

and the respective distribution of the variable $\boldsymbol{\alpha}$

$$x(j) = \begin{cases} p(\alpha_j|\boldsymbol{\alpha}|\boldsymbol{\theta}) & \text{when} \quad j \geq 0 \\ 0 & \text{when} \quad j < 0 \end{cases} \tag{15}$$

in the system Ψ under consideration.

Lemma 2.2. The following system of equations holds between the distributions (14) and (15)

$$x(j) = q[p^r q(j-r) + p^{r-1}q(j-r+1) + \ldots + pq(j-1) + q(j)], \quad j \geq 0, \tag{16}$$

which is equivalent to the system

$$q(i) = p^{r+1}q(i-r-1) + \frac{x(i) - px(i-1)}{q}, \quad i \geq 0, \tag{17}$$

where p, q, r are the noise parameters appearing in (3).

Proof. By the probability definition $p(\alpha_j|\boldsymbol{\alpha}|\boldsymbol{\theta}_i)$ stated at the beginning of Subsection 2.1, for all $j \geq 0$ it holds that

$$x(j) = \sum_{i \geq 0} p(\alpha_j|\boldsymbol{\alpha}|\boldsymbol{\theta}_i) q(i) = \sum_{i=j-r}^{j} Pr(j-i) q(i). \tag{18}$$

By substituting $Pr(j-i)$ from (3) we obtain (16). By (16)

$$x(j-1) = q[p^r q(j-r-1) + p^{r-1} q(j-r) + \ldots + pq(j-2) + q(j-1)],$$

$$px(j-1) = q[p^{r+1} q(j-r-1) + p^r q(j-r) + \ldots + pq(j-1)] = qp^{r+1} q(j-r-1) + x(j) - qq(j),$$

$$x(j) = px(j-1) + qq(j) - qp^{r+1} q(j-r-1). \tag{19}$$

Then (17) follows directly from the equation (19).

Now, we will derive the relation between the power parameter $W = W(\boldsymbol{\theta})$ of the state $\boldsymbol{\theta}$ from the formula (8) and the similar power parameter of the variable $\boldsymbol{\alpha}$

$$E = E(\boldsymbol{\alpha}) = \sum_{j=0}^{\infty} jx(j) > 0. \tag{20}$$

Lemma 2.3. The power parameters $W = W(\boldsymbol{\theta})$ and $E = E(\boldsymbol{\alpha})$ are related by

$$E = \frac{p(1-c)}{1-p} + W \quad \text{where} \quad , \tag{21}$$

$$c = c_r(p) \triangleq \left(\frac{1 + \frac{1}{p} + \ldots + \frac{1}{p^r}}{r+1} \right)^{-1} = \frac{(r+1)p^r(1-p)}{1-p^{r+1}} \in (0,1). \tag{22}$$

Proof. Multiplying both sides of the equation (19) by $j \geq 0$, we obtain that

$$jx(j) = p(j-1)x(j-1) + px(j-1) + qjq(j) -$$

$$-qp^{r+1}(j-r-1)q(j-r-1) - qp^{r+1}(r+1)q(j-r-1).$$

By summing up both sides over all $j \geq 0$, we obtain

$$E = pE + p + (1-p)W - qp^{r+1}(r+1).$$

According to (3) and (22), for $q = (1-p)/(1-p^{r+1})$ and $c = c_r(p)$ it holds

$$qp^{r+1}(r+1) = \frac{(1-p)p^{r+1}(r+1)}{1-p^{r+1}} = \frac{p(r+1)}{1 + \frac{1}{p} + \ldots + \frac{1}{p^r}} = pc.$$

Therefore it holds $E(1-p) = p(1-c) + (1-p)W$, which implies the relation (21).

In the next part we are interested in a solution x of the equation

$$\frac{x}{1-x} = \frac{p(1-c)}{1-p} + W \quad \text{where} \quad c = c_r(p) \text{ from (22)}. \tag{23}$$

Lemma 2.4. The solution x of the equation (23) satisfies the condition

$$x = \frac{p(1-c)+(1-p)W}{1-pc+(1-p)W} \in (0,1) \tag{24}$$

and meets, for c from (22), the relation $x \geq p$ if and only if

$$W \geq \frac{(r+1)p^{r+1}}{1-p^{r+1}} = \frac{pc}{1-p}. \tag{25}$$

Proof. From the equation (23) we obtain

$$x = \frac{p(1-c)+(1-p)W}{p(1-c)+(1-p)(W+1)} = \frac{p(1-c)+(1-p)W}{1-pc+(1-p)W}.$$

Therefore, the equality in (24) is valid. Also it holds that $0 < x < 1$. Further, we obtain

$$x = 1 - \frac{1-p}{1-pc+(1-p)W},$$

$$\frac{x-p}{1-p} = \left(1 - \frac{1}{1-pc+(1-p)W}\right) = \frac{(1-p)W-pc}{1-pc+(1-p)W}.$$

So that $x-p \geq 0$ if and only if $(1-p)W-pc \geq 0$, which is the condition stated in the equivalence (25).

Lemma 2.5 Let $x(j)$, $j \geq 0$, be an arbitrary probability distribution defined on the set $\{0, 1, \ldots\}$. Its entropy $H = -\sum_{j=0}^{\infty} x(j) \ln x(j)$ achieves, under the condition (20), the maximum

$$H_{max} = \frac{h(x)}{1-x} \tag{26}$$

if and only if

$$x(j) = (1-x)x^j, \quad j \geq 0, \tag{27}$$

where x is the solution of the equation

$$\frac{x}{1-x} = E \tag{28}$$

and $h(x)$ is the entropy function defined in (11). (This key lemma follows from Lagrange multipliers method or from more general Theorem 9.37 in [15].)

Lemma 2.6. If x, given by the relation (24), leads to a non-negative solution $q(i),\ i \geq 0$, of the system of equations

$$q(i) = \begin{cases} \frac{1-x}{q} & \text{for } i = 0 \\ p^{r+1}q(i-r-1) + \frac{(1-x)x^{i-1}(x-p)}{q} & \text{for } i \geq 1, \end{cases} \tag{29}$$

then the supremum appearing in the formula (12) satisfies the equality

$$\sup_{\tilde{\theta} \in \Theta_W} H(\boldsymbol{\alpha} \| \tilde{\boldsymbol{\theta}}) = H(\boldsymbol{\alpha} \| \boldsymbol{\theta}) = \frac{h(x)}{1-x}, \tag{30}$$

where $\boldsymbol{\theta} \in \Theta_W$ is the state with distribution $q(i|\boldsymbol{\theta}) = q(i),\ i \geq 0$.

Proof. By Lemma 2.5 and 2.6, $H^*_{max} \triangleq \sup_W(H(\boldsymbol{\alpha} \| \tilde{\boldsymbol{\theta}}))$ is the maximal entropy of the distribution $x(j),\ \ j \geq 0$, which solves the system of equations (16) under the condition the distribution $q(i),\ i \geq 0$, in (16) satisfies (8). By Lemma 2.3, this condition holds if and only if the solution $x(j),\ j \geq 0$ itself satisfies the condition (20) for positive E defined by the relation (21). Hence, following Lemma 2.5, the value $H^*_{\max}$ is less or equal $H_{\max}$ in (26) and the equality $H^*_{max} = H_{max}$ is achieved if and only if it is possible to find out such a solution $x(j),\ j \geq 0$ of the system (16) satisfying the condition (20) for E given by the formula (21), which is the geometrical distribution (27) with the parameter x, solving the equation (28) for E given by the formula (21). This also means that x solving the equation (28) for E given by (21) means nothing else than that x solves also the equation (23), i.e. that it is given by the relation (24). According the equivalence between the systems of equations (16) and (17), the geometrical distribution (27) satisfies the conditions stated above if and only if the solution $q(i),\ i \geq 0$, of the system (29) is, for x given by (24), the probability distribution. As we can see, this solution always satisfies the equality $\sum_{i=0}^{\infty} q(i) = 1$, so it is obvious that the condition of the probability distribution holds when $q(i) \geq 0$ for all $i \geq 0$. This is the necessary and sufficient for the equality $H^*_{\max} = H_{\max}$. The lemma thus holds when $H_{max} = h(x)/(1-x)$ holds, which is guaranteed by Lemma 2.5.

Theorem 2.1. For all parameters W satisfying the condition

$$W \geq W_{crit} \triangleq \frac{(r+1)p^{r+1}}{1-p^{r+1}} = \frac{pc}{1-p} \tag{31}$$

the capacity (9) of the additive physical channel with parameters $0 < p < 1$ and $1 \leq r \leq \infty$ is given by the formula

$$C(p, r|W) = \frac{h(x)}{1-x} - \frac{h(p)}{1-p} + \frac{h(p^{r+1})}{1-p^{r+1}}, \tag{32}$$

where $x = x(p, r|W) \in (0, 1)$ is the variable stated in (24) and $h(\cdot)$ is the entropy function defined by (11) on the interval (0,1). The capacity is achieved, i.e. the equality $C(p, r|W) = I(\boldsymbol{\alpha}; \boldsymbol{\theta})$ is valid, in the state $\boldsymbol{\theta}$ with the distribution $q(i|\boldsymbol{\theta}), i \geq 0$, given by the relations (14) and (29).

Proof. The condition of non-negativity of the variable $q(i)$, $i \geq 0$, in Lemma 2.6 will be satisfied when the solution x of the equation (23) satisfies the condition $x \geq p$. By Lemma 2.4 this happens when the power parameter W satisfies the condition (31). Under this condition Lemma 2.6 guarantees the equality (30). Now, the required equality (32) follows directly from Lemma 2.1 and from the equality (30). The condition for the capacity achieving follows from the formula (13) for information $I(\boldsymbol{\alpha}; \boldsymbol{\theta})$, and from the formulas (12) and (30).

Note that the power restriction (31) holds

$$W_{crit} \to 0, \qquad \text{more accurately} \qquad W_{crit} = o(rp^r),$$

for $rp^r \to 0$, i.e., for example, for any p and $r \to \infty$, or for any r and $p \to 0$. For the critical value of power parameter

$$W = W_{crit} = \frac{(r+1)p^{r+1}}{1 - p^{r+1}} \tag{33}$$

the capacity formula streamlines, following Theorem 2.1.

Consequence 2.1. The capacity (9) of the additive physical channel, with the power critical value (33), is

$$C(p, r|W_{crit}) = \frac{h(p^{r+1})}{1 - p^{r+1}}. \tag{34}$$

It follows from this result that $C(p, r|W) \leq h(p^{r+1})/(1 - p^{r+1})$ for $W \leq W_{crit}$, where the upper bound is small except for small values $1 \leq r \leq \infty$ and great values $0 < p < 1$. However, the formula of the capacity $C(p, r|W)$ of the additive physical channel for the under-critical power norm $W < W_{crit}$ remains to be as an open theoretical problem.

2.3 Capacity Estimations in the Area of $W < W_{crit}$

Now we will concentrate on upper and lower estimations

$$C_*(p, r|W) \leq C(p, r|W) \leq C^*(p, r|W) \tag{35}$$

of the capacity $C(p,r|W)$, given by the formula (12) for the area of the low input power parameter $W < W_{crit}$.

Let $x_* = x_*(p,r|W)$ be a solution of the equation

$$\frac{x_*}{1-x_*} = \frac{W}{r+1}. \tag{36}$$

Then $0 < x_* < 1$ for all $0 < p < 1$, $1 \le r \le \infty$ and $W > 0$. For all mentioned p, r and W, we define

$$C_*(p,r|W) = \frac{h(x_*)}{1-x_*} \quad \text{and} \quad C^*(p,r|W) = \frac{h(x)}{1-x} - \frac{h(p)}{1-p} + \frac{h(p^{r+1})}{1-p^{r+1}}, \tag{37}$$

where $x = x(p,r|W)$ is the solution of the equation (23), i.e., is given by (24).

Theorem 2.2. The functions in (37) are estimations of the capacity $C(p,r|W)$ of the additive channel for the whole range of the parameters $0 < p < 1$, $1 \le r \le \infty$ and $W > 0$.

Proof. The function $C^*(p,r|W)$ features an appearance of $h(x)/(1-x) - H_r(p)$, while for $h(x)/(1-x) = H_{max}$ we argued in the proof of Lemma 2.6 that

$$\sup_{\theta \in \Theta_W} H(\alpha \| \theta) \le \frac{h(x)}{1-x}.$$

Therefore, the inequality $C(p,r|W) \le C^*(p,r|W)$ follows from (12). When proving the other inequality for $C_*(p,r|W)$ we will use a state $\theta_* \in \Theta_W$ defined by the condition that for every $i \ge 0$,

$$q(i|\theta_*) = \begin{cases} (1-x_*)x_*^k & \text{if} \quad i = k(r+1),\ k \ge 0, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

For the state θ_* defined in this way it is easy to see, from the definition (36), that $\theta_* \in \Theta_W$, from the definition (37) and from Theorem 4.4 in [7], that

$$C_*(p,r|W) = \mathcal{H}(\theta_*), \tag{39}$$

where $\mathcal{H}(\theta_*)$ is the Shannon entropy of $q(\cdot|\theta_*)$. Now we will prove that it holds

$$H(\alpha \| \theta_*) = \mathcal{H}(\theta_*) + H_r(p), \tag{40}$$

where $H_r(p)$ is that entropy defined in (10). By the definition of information transferred and by the relation (10), it is obvious that $I(\alpha;\theta_*) = \mathcal{H}(\theta_*)$ and, therefore,

$$\mathcal{H}(\theta_*) \le \sup_{\theta \in \Theta_W} I(\alpha;\theta) = C(p,r|W).$$

This means that (39) is the lower estimation of the capacity, which was left to prove. Then, the problem shrank to proof (40). As for this proof, we will use that for $q(i) = q(i|\boldsymbol{\theta}_*)$, the probabilities $x(j) \triangleq p(\alpha_j|\boldsymbol{\alpha}|\boldsymbol{\theta}_*)$, $j \geq 0$, satisfy

$$x(j) = Pr(s)q(k(r+1)) = qp^s(1-x_*)x_*^k,$$

where $k \geq 0$ and $0 \leq s \leq r$ are unambiguously specified by the equation $j = k(r+1)+s$. As we denote $\ell_{sk} \triangleq \ln[qp^s(1-x_*)x_*^k] = \ln(qp^s) + \ln((1-x_*)x_*^k)$, it is easy to find out that

$$\begin{aligned} H(\boldsymbol{\alpha}\|\boldsymbol{\theta}_*) &= -\sum_{j=0}^{\infty} x(j)\ln x(j) = -\sum_{k=0}^{\infty}\sum_{s=0}^{r} qp^s(1-x_*)x_*^k\ell_{sk} \\ &= -\sum_{s=0}^{r} qp^s\ln(qp^s) - \sum_{k=0}^{\infty}(1-x_*)x_*^k\ln((1-x_*)x_*^k) \\ &= H_r(p) + H(q(\cdot|\boldsymbol{\theta}_*)) = H_r(p) + \mathcal{H}(\boldsymbol{\theta}_*). \end{aligned}$$

Theorem 2.3. The probability distribution (38) we used for the lower estimation $C_*(p,r|W)$ of the capacity $C(p,r|W)$ in the area $W \leq W_{crit}$ is the limit of the optimal distribution (29) for $x \downarrow p$, i.e. for $W \downarrow W_{crit}$ and for $x = p$, i.e. for $W = W_{crit}$ it is identical with this distribution. Therefore, it holds for $W = W_{crit}$ that the optimal state $\boldsymbol{\theta}$, in which according to Theorem 2.1 the capacity is achieved, is for this $W = W_{crit}$ identical with the state $\boldsymbol{\theta}_*$ used in the proof of Theorem 2.2. Therefore, the lower as well as the upper estimations satisfy the relation

$$C_*(p,r|W_{crit}) = C(p,r|W_{crit}) = C^*(p,r|W_{crit}). \tag{41}$$

Proof. It is obvious from (29) that for $x = p$ it holds for all $i \geq 0$

$$q(i) = \begin{cases} (1-p^{r+1})(p^{r+1})^k & \text{if} \quad i = (r+1)k,\ k \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

However, for the state $\boldsymbol{\theta}$ with the distribution $q(i) = q(i|\boldsymbol{\theta})$ falls into the set Θ_W (i.e. for the distribution $q(i)$, $i \geq 0$ that satisfies the condition (8)) it has to hold

$$(r+1)\frac{p^{r+1}}{1-p^{r+1}} = (r+1)(1-p^{r+1})\sum_{k=0}^{\infty} k(p^{r+1})^k = W,$$

i.e. p^{r+1} has to be identical with the number x_* defined in (36). In other words, the distribution (29) has to be identical with the distribution (38). As the distribution (29) is a continuous function of the variable $x \in [p,1)$, in all common norms (e.g. in L_1-norm), it has to hold that $\lim_{x\downarrow p}\sum_{i=0}^{\infty}|q(i) - q(i|\boldsymbol{\theta}_*)| = 0$.

The significance of Theorem 2.3 is in the fact that it supports our hypothesis saying that it holds

$$C(p,r|W) = C_*(p,r|W) \quad \text{for all} \quad 0 \le W \le W_{crit}. \tag{42}$$

A rigorous proof of our hypotheses (42) calls for an application of the Kuhn-Tucker conditions in infinite-dimensional spaces, and therefore it is left for a later research.

2.4 F–D Channel Capacity

In this subsection we will concentrate on the special channel case, when $r = 1$ in the general model studied till now.

Theorem 2.4 For all parameters of the power W and parameters $0 < p < 1$ satisfying the condition[2]

$$W \ge W_{crit} \triangleq \frac{2p^2}{1-p^2}, \tag{43}$$

the capacity (9) of the respective F–D channel is given by the formula

$$C_{FD} = C_{FD}(p|W) = \frac{h(x)}{1-x} - h\left(\frac{p}{1+p}\right), \tag{44}$$

where

$$x = \frac{W + p(W+1)}{W+1+p(W+2)} \in (p,1) \tag{45}$$

and $h(\cdot)$ is the entropy function defined by the relation (11) on the interval $(0,1)$. The capacity is achieved, i.e. the equality $C_{FD} = I(\boldsymbol{\alpha};\boldsymbol{\theta})$ holds, in the state $\boldsymbol{\theta}$ with the distribution

$$q(i|\boldsymbol{\theta}) = \frac{(1+p)(1-x)}{x+p}(x^{i+1} - (-p)^{i+1}), \quad i \ge 0. \tag{46}$$

Proof. When we substitute $c = 2p/(1+p)$, we obtain the condition (43) from (31) and the condition (45) from (24). By (32), we obtain the formula

$$C_{FD} = \frac{h(x)}{1-x} - \frac{h(p)}{1-p} + \frac{h(p^2)}{1-p^2}.$$

Hence, after a verification of the equality

$$\frac{h(p)}{1-p} - \frac{h(p^2)}{1-p^2} = h\left(\frac{p}{1+p}\right)$$

valid for all $0 < p < 1$, we obtain (44).
The distribution (46), for which information in the variable $\boldsymbol{\alpha}$ is maximal, is not a

[2]Formula (43), and (31), corrects an inaccurate statement for W in Theorem 10 in [6].

canonical one in the sense of the exponentiality. On the pure states $\boldsymbol{\theta}_i$ of an even order we speak about the addition of two canonical distributions which, after the normalization to only "even states", takes on the form

$$q_{even}(i) = \frac{(1-x^2)(1-p^2)}{(x+p)(1-xp)}(x^{i+1}+p^{i+1}), \quad i \in \{0, 2, 4, \ldots\}. \tag{47}$$

On the other hand, on the states $\boldsymbol{\theta}_i$ of an odd order we speak about the subtraction of two canonical distributions which, after an application of the respective normalization process on the set of "odd states", takes on the form

$$q_{odd}(i) = \frac{(1-x^2)(1-p^2)}{(x+p)(x-p)}(x^{i+1}-p^{i+1}), \quad i \in \{1, 3, 5, \ldots\}. \tag{48}$$

Now, let us have a close look at these upper and lower estimations C_{*FD} and C^*_{FD} of the capacity C_{FD} in the area $0 \leq W \leq 2p^2/(1-p^2)$. The upper bound is identical to the right-hand side of (44) and we include it for the sake of completeness only, while, with regard of the hypothesis at the end of Subsection 2.2, the lower bound is more interesting and of a far greater importance.

Theorem 2.5. The upper and the lower estimations of the capacity C_{FD} are

$$C^*_{FD} = \frac{h(x)}{1-x} - h\left(\frac{p}{1+p}\right) \quad \text{and} \quad C_{*FD} = \frac{h(x_*)}{1-x_*}) \quad \text{for} \quad x_* = \frac{W}{2+W}$$

in the whole range of the parameters $0 < p < 1$ and $W > 0$, x is the variable given by the relation (45) and $h(\cdot)$ is the same entropy function as in the previous theorem.

Proof. This result follows from Theorem 2.2. It is sufficient to verify that x is a solution of the equation (23) for $r = 1$ and x_* satisfies (36) for $r = 1$.

The formula for the lower estimation which follows from Theorem 2.5 is

$$C_{*FD} = \frac{W}{2}\ln\left(1+\frac{2}{W}\right) + \ln\left(1+\frac{W}{2}\right) \quad \text{for all} \quad W > 0.$$

Our hypothesis says that, for $0 < W < 2p^2/(1-p^2)$, the capacity is exactly on this bound. At the same time, for very low powers, the following formula is applicable,

$$C_{*FD} = W\frac{1}{2}\left(\ln\frac{2}{W}+1\right) + o(W^2) \quad \text{for} \quad W \downarrow 0. \tag{49}$$

By substituting $p = e^{\frac{-\epsilon}{kT}}$ from the relation (4), and by introducing *"the effective temperature"* $T_W > 0$ from the condition

$$e^{\frac{-\epsilon}{kT_W}} = \frac{W+p(W+1)}{W+1+p(W+2)} = x, \quad \text{where} \quad x < 1,$$

we obtain from (44), (45), the capacity expressed in the form

$$C_{FD} = \alpha\left(\frac{\varepsilon}{kT_W}\right) - \beta\left(\frac{\varepsilon}{kT}\right), \tag{50}$$

where[3] for $t > 0$

$$\alpha(t) = \frac{t}{e^t - 1} - \ln(1 - e^{-t}) \tag{51}$$

and

$$\beta(t) = \frac{1}{e^t + 1}\ln(1 + e^t) + \frac{1}{e^{-t} + 1}\ln(1 + e^{-t}). \tag{52}$$

It holds for temperature T_W that $T_W > T$ when $x > p$, i.e. $W > W_{crit}$, $T_W = T$ when $W = W_{crit}$ and $T_W < T$ when $W < W_{crit}$.

2.5 B-E Channel Capacity

In this subsection we will study the other extreme case for which the value $r = \infty$ in the general model.

Theorem 2.6. For all the power parameters $W > 0$ and parameters $0 < p < 1$, the capacity of the respective B–E channel is given by the formula

$$C_{BE} = C_{BE}(p|W) = \frac{h(x)}{1-x} - \frac{h(p)}{1-p}, \tag{53}$$

where

$$x = \frac{p + (1-p)W}{1 + (1-p)W} \in (0,1) \tag{54}$$

and $h(\cdot)$ is the entropy function from (11). The capacity is achieved, i.e. the equality $C_{BE} = I(\boldsymbol{\alpha}; \boldsymbol{\theta})$ holds, in the state $\boldsymbol{\theta}$ with the distribution

$$q(i|\boldsymbol{\theta}) = \begin{cases} \frac{1-x}{1-p} & \text{if} \quad i = 0 \\ \frac{(1-x)(x-p)x^{i-1}}{1-p} & \text{if} \quad i \geq 1. \end{cases} \tag{55}$$

[3]Formula (51) corrects a small misprint in formula (26) in [6].

Proof. It is sufficient to apply Theorem 2.1 and notice that in this case W_{crit} is reduced to 0, and that c from the relations (24) and (22) also shrinks to 0. With regards of these features the condition (31) is reduced to $W \geq 0$, the variable x stated in (24) shrinks to the value (54), and $h(p^{r+1})/(1-p^{r+1})$ from (10) is reduced to 0. Therefore the formula (53) follows from (32). The formula (55) follows from (14) and from the equations (29) after setting $r = \infty$ and $q = 1-p$, as it follows from (3) for q.

The effective temperature $T_W > 0$ can be established from the condition

$$e^{\frac{-\varepsilon}{kT_W}} = \frac{p + (1-p)W}{1 + (1-p)W} = x, \quad \text{where} \quad x < 1.$$

As x is now a solution of the equation $x/(1-x) = p/(1-p) + W$, see (23) with $c = 0$, where $W > 0$, it holds

$$\frac{e^{-\frac{\varepsilon}{kT_W}}}{1 - e^{-\frac{\varepsilon}{kT_W}}} > \frac{e^{-\frac{\varepsilon}{kT}}}{1 - e^{-\frac{\varepsilon}{kT}}},$$

i.e. in the B–E channels the effective temperature T_W of the variable α in the state of the maximum information rate is always higher than the temperature T of the whole system.

3 Conclusion

We presented the generalized formula (32) for the capacity of the physical information transmission channel with the additive system noise with an arbitrary, finite or infinite, number of discrete levels, which is considered to be a novelty. This noise follows the Gibbs canonical distribution (3). All our results we demonstrated for the special cases of the B–E and F–D noise.

The main result and novelty of this paper, however, is the formula (31) for the expectation value (8) of the input variable (the average energy level of the input signal, input parameter W) for an arbitrary number of levels of the system noise.

The further result, also of importance and novelty, is in the hypotheses (42) that the capacity C exists even when the input parameter W does not follow the condition (31), and that the lower capacity estimation C_* from (37) is then defined by (34) for the critical value W_{crit} from (33) and, that is the capacity C for this case. For the case the input parameter W is above the critical value W_{crit}, the capacity C is equal to the upper capacity estimation C^* in (37). Both these estimations streamline together when the equation (33) holds as stated in (41). The probability distributions (47), (48), of not canonical type, also are to be considered as a novelty and of interest similarly as the distribution (29) and its limit (38).

As mentioned at the end of Section 2 the hypotheses about the existence and properties of the non zero and positive capacity under that condition of the critical and lower value of the input parameter W is to be studied more deeply in a mathematical way. The same is for a physical description of this phenomenon.

References

[1] Ballescu R. (1975): *Equilibrium and Nonequilibrium Statistical Mechanics.* Wiley, New York.

[2] Brillouin L. (1963): *Science and Information Theory.* Academia Press, New York.

[3] Cover T. M., Thomas J. B. (1991): *Elements of Information Theory.* Wiley, New York.

[4] Cholevo A. S. (1979): On the capacity of quantum communication channel. *Problems of Information Transmission.* 15, N. 4, 3–11.

[5] Fuchs C. A., van de Graaf J. (1999): Cryptographic Distinguishability Measures for Quantum-Mechanical States. *IEEE Transactions on Information Theory*, vol. 45, No. 4.

[6] Hejna B., Vajda I. (1999): Information Transmission in Stationary Stochastic Systems. *Proceedings of CASYS'98*, 405–417. American Institute of Physics, Woodbury N. Y.

[7] Hejna B. (2000): *Informační kapacita stacionárních fyzikálních systémů*, Doctoral disertation, UTIA AV ČR, Prague and FJFI ČVUT, Prague (2000).

[8] Helström C. W. (1976): *Quantum Detection and Estimation Theory.* Pergamon Press, London.

[9] Landau L. D., Lifschitz E. M. (1969): *Statistical Physics.* Second Edition. Pergamon Press, Oxford.

[10] Lavenda B. H. (1991): *Statistical Physics.* Wiley, New York.

[11] Lebedev D. L., Levitin L. B. (1966): Information transmission by electromagnetic field. *Infor-mation and Control* 9, 1–20.

[12] Peres A. (1993): Quantum Theory: Concepts and Methods. Kluwer, Dordrecht.

[13] Rényi A. (1984): *A Diary on Information Theory.* Academiai Kiadó, Budapest.

[14] Shannon C. E. (1948): A mathematical theory of communication. *The Bell Systems Technical Journal* 27, 379–423, 623–656.

[15] Vajda I. (1989): *Theory of Statistical Inference and Information.* Kluwer Academic Publishers, Dordrecht.

[16] Watanabe S. (1969): *Knowing and Guessing.* Wiley, New York.

Further references

[17] Chintschin A. J., Faddejew D. K., Kolmogoroff A. N., Rényi A. und Balatoni J. (1957): *Arbeiten zur Informationstheorie.* VEB Verlag, Berlin.

[18] Khinchin A. I. (1957): *Mathematical Foundations of Information theory.* Dover, New York.

Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction

Ma Hongguang, Han Chongzhao, Li Xihai; Wang Guohua, Xu Jianfeng, Zhu Xiaofei
Xi'an Jiaotong University, School of Electronic & Information Engineering, Xi'an 710049, P.R. China; Research Institute of High Technology, Xi'an 710025, P.R. China
FAX. 0086- 29-82668775, E-mail: mhg_xian@163.com, http://www.xjtu.edu.cn

Abstract
Embedding techniques provide a powerful advance in the development of experiment chaos. However there seems no universal method to find the best set of parameters to use. In this paper, we analyze the drawback of an algorithm of automatic embedding dimension and time delay presented in Reference[1](Massayuki Otani and Antonia Jones, Oct. 2000), and propose a new approach for computing the embedding dimension and delay time based on the multiple autocorrelation and Γ-test. This approach is provided with a sound theoretic basis, and its computing complexity is relatively lower and not strongly depended on the data length. The experimental results indicate that a near optimum embedding dimension and delay time can be estimated by using this approach, and the accuracy of invariants in phase space reconstruction is efficiently improved.
Keywords: Phase Space Reconstruction, Embedding Dimension, Delay Time, Multiple Autocorrelation, Γ-test.

1 Introduction

The characteristics of strange attractors of a chaotic system can be analyzed by sampling a part of the output chaotic time series of system. The method in common use is the state space reconstruction in delay coordinate proposed by Packad[2]. It can be proved by Takens' theorem[3] that the unstable periodic obits (strange attractor) could be recovered properly in an embedding space whenever a suitable embedding dimension $m \geqslant 2d+1$, (d is the dimension of chaotic system) were found out, i.e. the obits in the reconstructed space R^m keeps a differential homeomorphism with the original system.

It is very important to select a suitable pair of embedding dimension m and time delay τ when performing the phase space reconstruction. For doing this there are two different points of view: one is that m and τ are not correlated with each other, i.e. m and τ can be selected independently (Takens has proved that m and τ are independent in a chaotic time series with infinite length and noiseless). Under this golden rule, a commonly used approach, G-P algorithm for calculating the embedding dimension m was proposed by Grassberger and Procacia[4]. For the time delay τ, there are three criterions to select it: 1. Series correlation approaches, such as Autocorrelation [4], Mutual Information [5], and High-order Correlations [6], etc. 2. Approaches of phase space extension, e.g. Fill Factor [7], Wavering Product [8], Average Displacement [9] and SVF [10], etc. 3. Multiple Autocorrelation and Non-bias Multiple Autocorrelation [11].

The second viewpoint is that m and τ are closely related, because the time series in the real world could not be the infinite long, and hardly avoid being noised. A great deal of experiments indicate that m and τ tie tightly up with the time window $t_w= (m-1)\tau$ for the reconstruction of phase space. For a given chaotic time series, t_w is relatively steadfast. An irrelevant partnership of m and τ will directly impact the equivalence between the original system and the reconstructed phase space. Therefore, the combination approaches for computing m and τ are accordingly come into being, e.g. small-window solution [12], C-C method [13] and automated embedding [1]. We consider that the second viewpoint is more practical and reasonable than the first one in the engineering practice. The research on the combination algorithm of embedding dimension and delay time will become a hotspot in the category of the chaotic time series analysis.

2 Automated embedding algorithm

This algorithm was proposed by Masayuki Otani and Antonia Jones in Oct. 2000, which is based on the Average Displacement Method (AD) and Γ-test [14]. By means of this algorithm, a near optimum embedding dimension and delay time can be estimated. A brief description about this algorithm is given as follows.

1. Let $\boldsymbol{X}=\{x_i(t)\}$, i=1,2...N, be a part of chaotic time series whose evolution through time is described by a d-dimension dynamical system. Set an initial value for the embedding dimension, i.e. let $m = m_0$. Take the time delay τ as a variable and let it increase by one for each iteration. At each determinate value of τ, reconstruct X into $M=N-(m-1)\tau$ dimensions of vectors $\{\boldsymbol{x}_i\}$, $i=1,2,\ldots M$, $\boldsymbol{x}_i= (x_i, x_{i+1}, \ldots, x_{i+(m-1)\tau})$ $\boldsymbol{x}_i \in R^m$. Then calculate the average displacement of entire vector space by using formula (1).

   $$S(\tau)=\frac{1}{M}\sum_{i=1}^{M}\sqrt{\sum_{j=1}^{m-1}[x_{i+j\tau}-x_i]^2} \tag{1}$$

   Where M is the number of data points used for the estimation. As the delay time increases from zero, the reconstructed trajectory expands from the diagonal and $S(\tau)$ increases accordingly until it reaches a plateau. With large values of m, reconstruction expansion reaches a plateau at smaller value of the delay time, which maintains the time span approximately constant. The corresponding value of delay time when $S(\tau)$ gets in saturation is the near optimum τ under the certain value of m.
2. Take the result of step 1 as a constant and let embedding dimension m is a variable. Estimate the near optimum m by means of Γ-test, which can estimate the best mean squared output error of a continuous or smooth underlying input/output model without overfitting, i.e. suppose the samples of chaotic time series are generated by a continuous function f: $R^m \rightarrow R$, and let y be defined as $y=f(x_1,\cdots x_m)+\gamma$. Where γ represents an indeterminable part, which may be due to noise or lack of functional determination in the input/output relationship. At each given value of m, reconstruct X into $M=N-(m-1)\tau$ dimensions of vectors $\{x_i\}$, and construct the

input/output pairs $\{\xi_i, y_i\}$ as follows:

$$\xi_i = \{x(i), x((i+1)\tau), \cdots, x((i+m-1)\tau\} \tag{2}$$
$$y_i = x((i+m)\tau, \quad i = 1,2,\cdots M$$

Then find out the p[th] nearest neighbour $\xi_i(N(i,p))$ to ξ_i (p_{max}=20～50) and compute the distances by means of the formula 3.

$$dx(h) = \frac{1}{p}\sum_{h=1}^{p}\frac{1}{M}\sum_{i=1}^{M}\left|\xi(N(i,p)) - \xi(i)\right|^2$$
$$dy(h) = \frac{1}{p}\sum_{h=1}^{p}\frac{1}{2M}\sum_{i=1}^{M}(y(N(i,p)) - y(i))^2 \tag{3}$$

Perform a least squares fit on the coordinates (*dx*, *dy*) to obtain a regression line in the form of $(dy = Adx + \overline{\Gamma})$, where $\overline{\Gamma}$ is the estimated value of γ.

Increase the value m by one gradually and repeat steps 1 and 2. The estimated value of γ will decrease accordingly until it is much closed to zero. At this moment, the values of m and τ are the near optimum embedding dimension and time delay for the given chaotic time series. By chance if the estimated value of γ is not close to zero, the data set is non-deterministic; therefore we cannot hope to reconstruct the attractor accurately. This may happen if the SNR is lower, or the choice of time delay is poor.

The experimental results indicate that this algorithm is very efficient for the continuous chaotic time series. But the computing accuracy of this algorithm is tightly depended on that of AD algorithm. The average displacements of Lorenz and Rossler flows are depicted in fig.1 and fig.2. It can be seen clearly that the time delay is decreasing with the increasing of embedding dimension, and also there are some waviness when the waveshapes get into saturation.

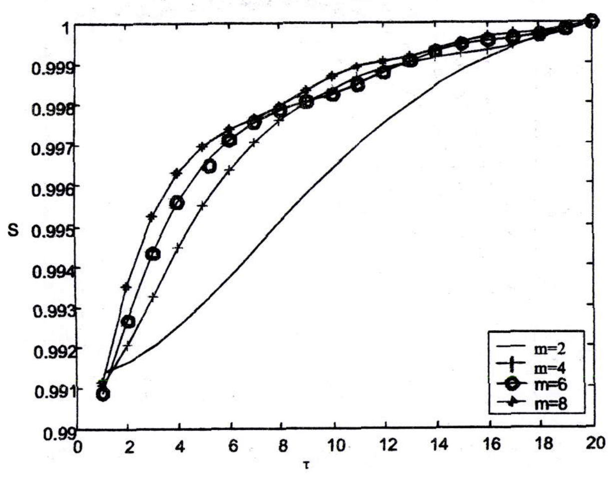

Figure 1: Average Displacement of Lorenz Flow.

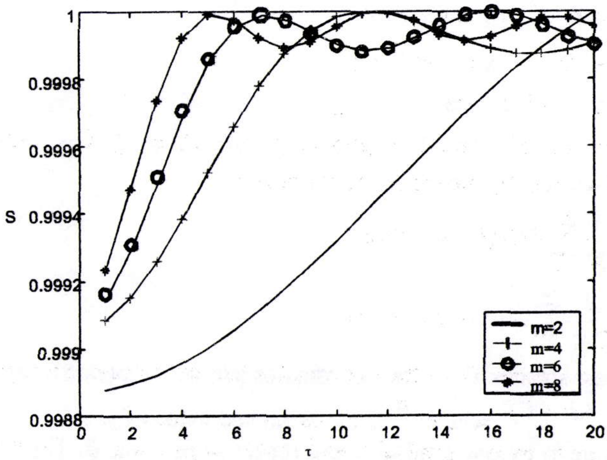

Figure 2: Average Displacement of Rossler Flow.

However, this algorithm cannot directly process the discrete chaotic time series, such as Henon, Logistic and Quadratic, etc. The major causation is that the sampling spacing of the discrete chaotic time series is "too large" that make the relativity between the data change so swiftly, and it seems that those maps behave like the random series. Hence, the discrete chaotic time series must be interpolated before processing. Fig.3 and 4 depict the average displacements of Henon and Quadratic maps after the interpolation with spline function. The data are the 10 times more than that of the originals.

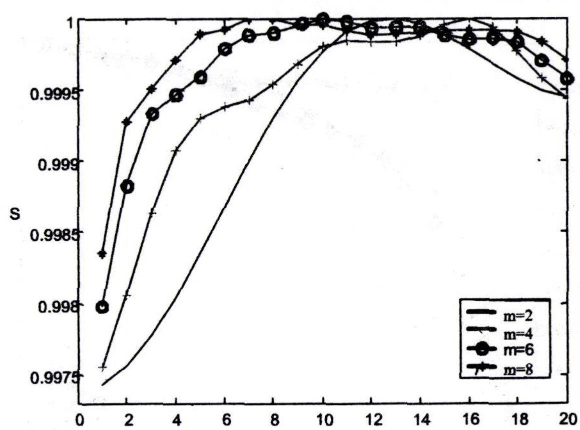

Figure 3: Average Displacement of Henon Map.

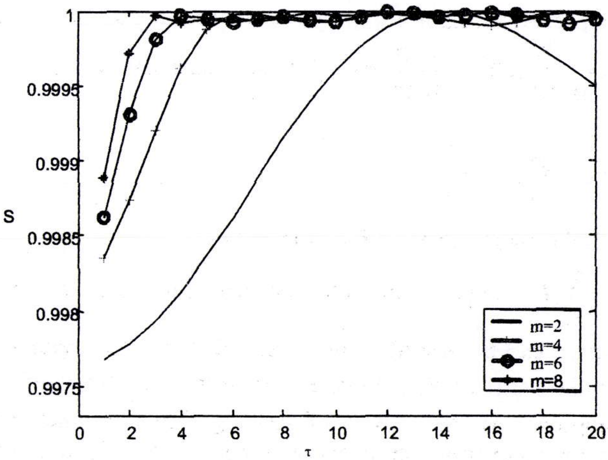

Figure 4: Average Displacement of Quadratic Map.

The average displacement algorithm is a geometry-based approach that can overcome the drawbacks of the autocorrelation-based methods, since the autocorrelation can ensure x_i and $x_{i+\tau}$, $x_{i+\tau}$ and $x_{i+2\tau}$ are not correlated respectively, but it cannot guarantee that x_i and $x_{i+2\tau}$are not correlated, too. Therefore, the autocorrelation-based method cannot be generalized in the high-order dimensions. So the AD algorithm looks like a suitable approach for the high-order system. In practice, the sloping variation of statistic S(τ) should be measured to figure out the corresponding delay time, usually we take the time point at the slope decreases to 40% of its initial value as the near optimum time delay. But from Figures 1 and 2 we can see that there intermix some wobbles in the entire variation of S(τ). Thereby, using the changing slope to determine the time delay sometimes will introduce a non-ignored error, and this error will influence the computing accuracy of embedding dimension in Γ-test. Hence, a modification should be done for the algorithm of time delay.

3 Multiple Autocorrelation Approach[11]

The multiple autocorrelation approach is derived from autocorrelation and average displacement. From formula 1 we can rewrite the statistic S(τ) of the chaotic time series $\{x_i\}$ in m dimension as follows:

$$S_m^2(\tau) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{m-1}(x(i+j\tau)-x(i))^2 \tag{4}$$

Extend the left part of formula 4 and ignore the errors caused by the border data. Consider that $E=\frac{1}{M}\sum_{i=1}^{M}x(i)^2=\frac{1}{M}\sum_{i=1}^{M}x(i+j\tau)^2$ is a constant within $1\leq j\leq m-1$, we can get:

$$S_m^2(\tau)=2(m-1)E-2\sum_{j=1}^{m-1}R_{xx}(j\tau) \quad (5)$$

Where, $R_{xx}(j\tau)$ is the autocorrelation function of $\{x_i\}$.

Define $R_{xx}^m(\tau)=\sum_{j=1}^{m-1}R_{xx}(j\tau)$, the multiple autocorrelation approach for the series $\{x_i\}$ in m dimension space can be described like that: select the corresponding time as the time delay τ when the value of $R_{xx}^m(\tau)$ decreasing to the $1\text{-}e^{-1}$ times of its initial value. Obviously, this approach is the ecdysis of AD algorithm. It inherits the geometric property of AD in the reconstruction of phase space. Meanwhile, it can be regarded as the extension of autocorrelation approach in the high-order dimensions. It overcomes the drawback of the autocorrelation, i.e. the multiple autocorrelation not only guarantees that x_i and $x_{i+\tau}$, $x_{i+\tau}$ and $x_{i+2\tau}$ are not correlated with each other respectively, but also ensures that x_i and $x_{i+2\tau}$ are not correlated. Therefore, the multiple autocorrelation has a sound theoretic basis.

Finally, the algorithm we adopt to replace the AD algorithm is the "Non-bias Multiple Autocorrelation":

$$\begin{aligned}C_{xx}^m(\tau)&=\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{m-1}(x(i)-\bar{x})(x(i+j\tau)-\bar{x})\\&=R_{xx}^m(\tau)-(m-1)(\bar{x})^2\end{aligned} \quad (6)$$

Where, $\bar{x}$ is the mean value of $\{x_i\}$. So employing the non-bias multiple autocorrelation for $\{x_i\}$ to select a near optimum time delay τ in m dimension of phase space is to choose the corresponding time when $C_{xx}^m(\tau)$ goes to zero at first time. The strongpoint of this approach is that it is endow with the merit of AD algorithm but gets rid of its drawback. The mathematic expression is sententious and easy to computation.

In order to validate the accuracy of the improved approach, we took Henon map and Lorenz flow as examples to reconstruct them with AD and non-bias multiple autocorrelation plus Γ-test respectively. Thereinto, the data of Henon map has been interpolated 10 times with spline function and then took out 500 data to be in for experiment, for Lorenz flow, we firstly generated 10,000 data and then chosen 1,000 points between 5,000 and 6,000 for experiment. Then calculate the correlation dimensions of them and made a comparison with their nominal values [15] to figure out the errors. The experimental results are shown in table 1.

Table 1. Experimental Results

Modle	Sample Period	AD+Γ-test			C_{xx}+Γ-test			Nominal Value
		Embedding Dimension Time Delay	Correlation Dimension	Error	Embedding Dimension Time Delay	Correlation Dimension	Error	
Henon (a=1.4,b=0.3)	0.1	m=3 τ=0.8	1.3158	0.0558	m=3 τ=0.7	1.2734	0.0134	1.26
Lorenz (a=10,b=8/3, γ=28)	0.01	m=5 τ=0.35	2.0772	0.0172	m=5 τ=0.25	2.0539	0.0061	2.06

4 Conclusion

We have described an efficient method for choosing a pair of delay time and embedding dimension which facilitates an accurate reconstruction of the high dimensional dynamics. This technique is based on the non-bias multiple autocorrelation and Γ-test methods, the combination of which is computationally inexpensive. The choices of delay time and embedding dimension are important, as a good choice can reduce both the amount of data required and the effect of noise. Throughout our experiments, we have consistently found that the delay time and embedding dimension are tightly correlated. Choosing a near optimum pair of them can effectively describe the strange attractors in a nonlinear chaotic system. Since the embedding techniques are widely employed to model a physical system in cases where the mathematical description is unknown, such an automated reconstruction has a wide applicability.

References

[1]Otani Masayuki, Jones Antonia (2000). Automated Embedding and Creep Phenomenon in Chaotic Time Series, http://www.cs.cf.ac.uk.

[2]Packard N.H., Crutchfield,J.P. Farmer J.D. and Shaw R.S. (1980). Geometry from a time series, Phys. Rev. Lett. , 45, pp.712-716.

[3]Takens F. (1980). Detecting strange attractor in turbulence. Lecture notes in Mathematics, 898, pp.361-381.

[4]Albano A.M., et al.,(1988) SVD and Grassberger-Procaccia algorithm. Phy. Rev. A, 38, pp.3017-3026.

[5]Fraser A.M.(1989). Information and entropy in strange attractors. IEEE Tr. On IT, Mar. 35(2), pp.245-262.

[6] Albano A.M., et al.(1991). Using high-order correlations to define an embedding window. Physica D,54, pp.85-97

[7] Pfister Buzug T.(1992). Optimal delay time and embedding dimension for delay-

time coordinates by analysis of the global static and local dynamical behavior of strange attractors. Phy. Rev. A, 45, pp.7073-7084.
[8] Pfister Buzug T.(1992). Comparison of algorithms calculating optimal embedding parameters for delay time coordinates. Physica D, 58, pp.127-137.
[9] Rossenstein M. T., Colins J. J., Luca C.J.De (1994). Reconstruction Expansion as a Geometry-based Framework for Choosing proper Delay Times, Physica D 73, pp. 82-98.
[10] Kember G., Fowler A.C.(1993). A correlation function for choosing time delays in phase portrait reconstructions. Phy. Lett. A, 179, pp.72-80.
[11]林嘉宇，王跃科，黄芝平，沈振康。(1999) 语音信号相空间重构中时间延迟的选择—复自相关法，信号处理，Vol. 15, pp.220-225.
(Lin Jiayu, Wang Yueke, Huang Zhiping and Shen Zhenkang.(1999) Selection of Proper Time –Delay in Phase Space Reconstruction of Speech Signals, Signal Processing, Vol.15,pp.220-225)
[12]Kugiumtzis D. (1996). State Space Reconstruction Parameters in the Analysis of Chaotic Time Series—the Role of the Time Window Length. Physica D, 95, pp.13-28.
[13] Kim H. S., Eykholt R., Salas J. D.(1999). Nonlinear dynamics, delay times, and embedding windows, Physica D 127, pp. 48-60.
[14]Stefansson A., Koncar N., Jones A. J. (1997). A Note on the Gamma test, Neural Computing & Applications, 5, pp.131-133.
[15]Kugiumtzis D.,Lillekjendlie B.,Christophersen N.(1994).Chaotic time series Part I:Estimation of some invariant properties in state space, Modelling Identification and Control, 15(4), pp. 205-224.