Front. Electr. Electron. Eng. China (2006) 1: 111–114
DOI 10.1007/s11460-005-0023-7

RESEARCH ARTICLE

MA Hong-guang, HAN Chong-zhao

Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction

Abstract A new algorithm is proposed for computing the embedding dimension and delay time in phase space reconstruction. It makes use of the zero of the nonbias multiple autocorrelation function of the chaotic time series to determine the time delay, which efficiently depresses the computing error caused by tracing arbitrarily the slop variation of average displacement (AD) in AD algorithm. Thereafter, by means of the iterative algorithm of multiple autocorrelation and Γ test, the near-optimum parameters of embedding dimension and delay time are estimated. This algorithm is provided with a sound theoretic basis, and its computing complexity is relatively lower and not strongly dependent on the data length. The simulated experimental results indicate that the relative error of the correlation dimension of standard chaotic time series is decreased from 4.4% when using conventional algorithm to 1.06% when using this algorithm. The accuracy of invariants in phase space reconstruction is greatly improved.

Keywords phase space reconstruction, embedding dimension, delay time, multiple autocorrelation, Γ test

1 Introduction

The characteristics of the strange attractors of a chaotic system can be analyzed by sampling a part of the output chaotic time series of a system. The method that is commonly used is the state space reconstruction in delay coordinate proposed by Packard *et al*. [1]. It can be proved through Takens' theorem [2] that the unstable periodic obits (strange attractor) could be recovered properly in an embedding space whenever a suitable embedding dimension $m \geq 2d+1$ (d is the dimension of chaotic system) is detected; that is, 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. The precision of τ and m is directly related with the accuracy of the invariables of the described characteristics of the strange attractors in 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; that is, 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 no noise). Under this golden rule, a commonly used approach, G–P algorithm, for calculating the embedding dimension m was proposed by Albano et al. [3]. For the time delay τ, there are three criterions for its selection: (1) series correlation approaches, such as autocorrelation, mutual information [4], high-order correlations [5], *etc*., (2) approaches of phase space extension, *e.g*., fill factor [6], wavering product [7], average displacement (AD; [8]), SVF [9], *etc*., and (3) multiple autocorrelation and nonbias multiple autocorrelation [10].

The second viewpoint is that m and τ are closely related because the time series in the real world could not be infinitely long and could 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 the phase space. For a given chaotic time series, t_w is relatively steadfast. An irrelevant partnership between 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 accordingly come into being, *e.g*., small-window solution [11], C–C method [12], and automated embedding [13]. Most researchers consider that the second viewpoint is more practical and reasonable than the first one in the engineering practice. 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 October 2000, which is based on the AD method and Γ test [14]. By means of this algorithm, a near-optimum embedding dimension and delay time can be

Translated from *Journal of Xi'an Jiaotong University*, 2004, 38(4): 335–338 (in Chinese)

MA Hong-guang (✉), HAN Chong-zhao
Research Institute of High Technology,
Xi'an 710025, China
E-mail: mhg_xian@163.com

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; that is, let $m=m_0$. Take the time delay τ as a variable and let it increase by one for each iteration. At each determinated value of τ, reconstruct $\boldsymbol{X}$ into $M=N-(m-1)\tau$ dimensions of vectors $\{\boldsymbol{x}_i\}$, i=1, 2,..., M, $\boldsymbol{x}_i=(x_i, x_{i+1},..., x_{i+(m-1)\tau})$ $\boldsymbol{x}_i \in R^m$. Then, calculate the AD of the entire vector space using:

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

   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 a smaller value of the delay time, which maintains the time span as approximately constant. The corresponding value of delay time when $S(\tau)$ gets into saturation is the near-optimum τ under a certain value of m.
2. Take the result of step 1 as a constant and let embedding dimension m become a variable. Estimate the near-optimum m by means of the Γ test, which can estimate the best mean-squared output error of a continuous or smooth underlying input/output model without overfitting. That is, suppose that the samples of chaotic time series are generated by a continuous function $f:R^m \rightarrow R$, and let y be defined as $y=f(x_1,..., x_m)+\gamma$, 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 $\boldsymbol{X}$ into $M=N-(m-1)\tau$

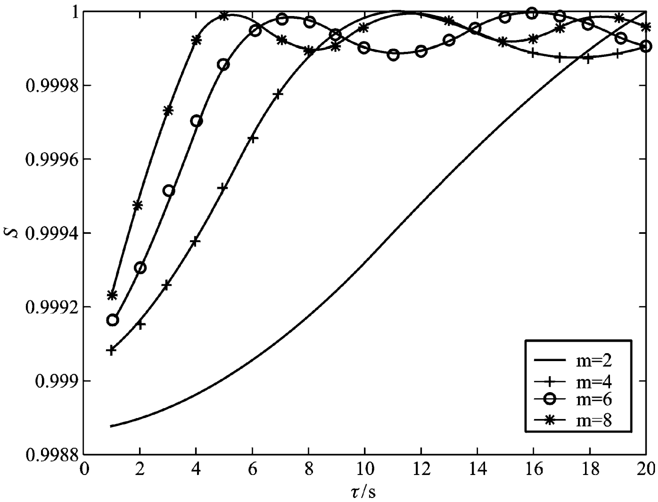

Fig. 2 Average displacement of Rossler flow

dimensions of vectors $\{\boldsymbol{x}_i\}$ and construct the input/output pairs $\{\xi_i, y_i\}$ as follows:

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

Then, find out the pth nearest neighbor $\xi_i(N(i, p))$ to $\xi_i(p_{\max}$=20~50) and compute the distances using:

$$\begin{aligned}dx(h) &= \frac{1}{p}\sum_{h=1}^{p}\frac{1}{M}\sum_{i=1}^{M}|\xi(N(i,p))-\xi(i)|^2 \\ dy(h) &= \frac{1}{p}\sum_{h=1}^{p}\frac{1}{2M}\sum_{i=1}^{M}(y(N(i,p))-y(i))^2\end{aligned} \quad (3)$$

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

Increase gradually the value of m by one and repeat steps 1 and 2. The estimated value of γ will decrease accordingly

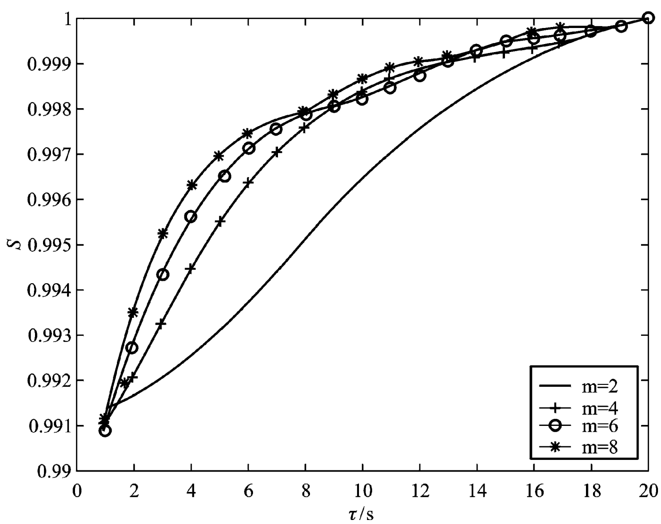

Fig. 1 Average displacement of Lorenz flow

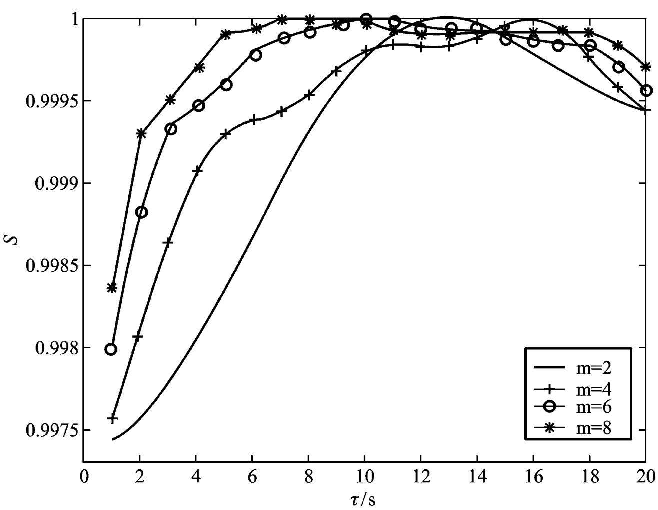

Fig. 3 Average displacement of Henon map

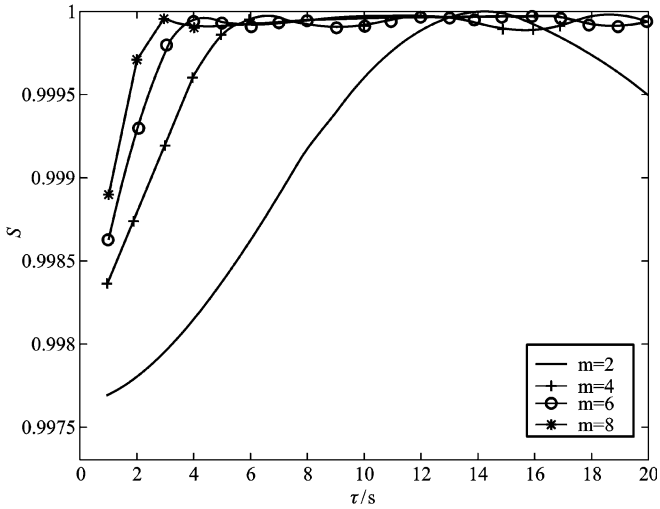

Fig. 4 Average displacement of quadratic map

until it is very close to zero. At this moment, the values of m and τ are the near-optimum embedding dimension and time delay, respectively, for the given chaotic time series. By chance, if the estimated value of γ is not close to zero, the data set is nondeterministic; therefore, we cannot hope to reconstruct the attractor accurately. This may happen if the signal-to-noise ratio (SNR) is lower or if the choice of time delay is poor.

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

However, this algorithm cannot directly process the discrete chaotic time series, such as Henon, logistic, and quadratic, etc. The major reason is that the sampling spacing of the discrete chaotic time series is "too large", which makes 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 Fig. 4 depict the ADs of Henon and quadratic maps, respectively, after the interpolation with spline function. The data are 10 times more than that of the originals.

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

3 Multiple autocorrelation approach [13]

The multiple autocorrelation approach is derived from autocorrelation and AD. From Eq. (1), we can rewrite the statistic $S(\tau)$ of the chaotic time series $\{\boldsymbol{x}_i\}$ 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 Eq. (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 $1\le j\le m-1$, we can get

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

where $R_{xx}(j\tau)$ is the autocorrelation function of $\{\boldsymbol{x}_i\}$.

Define $R_{xx}^m(\tau)=\sum_{j=1}^{m-1}R_{xx}(j\tau)$. The multiple autocorrelation approach for the series $\{\boldsymbol{x}_i\}$ in m dimension space can be described as such: select the corresponding time as the time delay τ when the value of $R_{xx}^m(\tau)$ decreases to the $1-e^{-1}$ 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.

Table 1 Experimental results

Model	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

Meanwhile, it can be regarded as the extension of autocorrelation approach in the high-order dimensions. It overcomes the drawback of the autocorrelation; that is, the multiple autocorrelation not only guarantees that x_i and $x_{i+\tau}$ and that $x_{i+\tau}$ and $x_{i+2\tau}$ are not correlated with each other, 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 nonbias 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 $\{\boldsymbol{x}_i\}$. Thus, employing the nonbias multiple autocorrelation for $\{\boldsymbol{x}_i\}$ 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 the first time. The strongpoint of this approach is that it inherits the merit of AD algorithm but gets rid of its drawback. The mathematic expression is sententious and easy to compute.

To validate the accuracy of the improved approach, we took the Henon map and Lorenz flow as examples to reconstruct them with AD and nonbias multiple autocorrelation plus Γ test, respectively. Thereinto, the data of the Henon map have been interpolated 10 times with spline function, and then 500 data were taken out for the experiment. For Lorenz flow, we firstly generated 10,000 data and then chose 1,000 points between 5,000 and 6,000 for experiment. We then calculated their correlation dimensions and made a comparison with their nominal values [15] to figure out the errors. The experimental results are shown in Table 1.

4 Conclusions

We have described an efficient method for choosing a pair of delay time and embedding dimension that facilitates an accurate reconstruction of the high-dimensional dynamics. This technique is based on the nonbias 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 from 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. Packard N. H., Crutchfield J. P. Farmer J. D. and Shaw R. S., Geometry from a time series, Phys. Rev. Lett., 1980, 45: 712–716
2. Takens F., Detecting strange attractor in turbulence, Lect. Notes Math., 1981, 898: 361–381
3. Albano, A. M. et al, SVD and Grassberger–Procaccia algorithm, Phys. Rev. A, 1988, 38(6): 3017–3026
4. Fraser A. M., Information and entropy in strange attractors[J]. IEEE Trans. Inf. Theory, Mar. 1989, 35(2): 245–262
5. Albano A. M. et al, Using high-order correlations to define an embedding window, Physica D, 1991, 54(1–2): 85–97
6. Pfister Buzug T., Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behaviour of strange attractors, Phys. Rev. A, 1992, 45(10): 7073–7084
7. Pfister Buzug T., Comparison of algorithms calculating optimal embedding parameters for delay time coordinates, Physica D, 1992, 58(1–4): 127–137
8. Rossenstein M. T., Colins J. J. and de Luca C. J., Reconstruction expansion as a geometry-based framework for choosing proper delay times, Physica D, 1994, 73(1): 82–98
9. Kembe G. and Fowler A. C., A correlation function for choosing time delays in phase portrait reconstructions, Phys. Lett. A, 1993, 179(2): 72–80
10. Lin J., Wang Y., Huang Z. and Shen Z., Selection of proper time-delay in phase space reconstruction of speech signals, Signal Process., 1999, 15(2): 220–225
11. Kugiumtzis D., State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length, Physica D, 1996, 95(1): 13–28
12. Kim H. S., Eykholt R. and Salas J. D., Nonlinear dynamics, delay times, and embedding windows, Physica D, 1999, 127 (1): 48–60
13. Otani M. and Jones A., Automated Embedding and Creep Phenomenon in Chaotic Time Series [EB/OL], http://users.cs.cf.ac.uk/Antonia.J.Jones/UnpublishedPapers/Creep.pdf, 2000
14. Stefansson A., Koncar N. and Jones A. J., A note on the gamma test, Neural Comput. Appl., 1997, 5(1): 131–133
15. Kugiumtzis D., Lillekjendlie B. and Christophersen N., Chaotic time series part I: estimation of some invariant properties in state space, Model. Identif. Control, 1994, 15(4): 205–224