2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

Statistical Tests for the SVD-based Analysis of Dynamical Noise on Chaos

Masaru Todoriki† and Shuichi Hasegawa†

†School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan
Email: todoriki@sys.t.u-tokyo.ac.jp, hasegawa@sys.t.u-tokyo.ac.jp

Abstract—
We have already proposed a method to evaluate the influence of dynamical noise on chaotic systems [1]. It was demonstrated that the influence of dynamical noise on a typical chaotic system Chua's electronic circuit can be extracted by the temporal fluctuation of singular values (TFSV) obtained from singular value decomposition (SVD), independently of the presence of measurement noise. However, the obtained results have not yet been sufficiently verified from the statistical perspectives so far.

In this study, some statistical tests are performed regarding the influence of data length. Among them, the results of famous Akaike's Information Criterion (AIC) [4] are mainly illustrated. As a result, it is found out that the validity of the results of our already proposed method is proved and the adequate conditions to obtain statistically correct results can be determined by the analyses.

1. Introduction

Every physical system is subject to noise in the real world. In general, there are two types of noise in any physical system, namely, measurement noise and dynamical noise. Different from the former, the latter type of noise is said to be realistically intrinsic to a physical system and yields an extremely complicated mechanism accompanied by feedback. As a result, it is quite difficult to analyze both on the theoretical and experimental levels. On the other hand, since a chaos system displays particularly strong nonlinearity and sensitivity to its initial condition, dynamical noise may have a remarkable and fatal influence on a chaos system. Numerous studies concerning dynamical noise in chaos have appeared.

We have already proposed a method to evaluate the influence of dynamical noise on chaotic systems [1]. It was demonstrated that the influence of dynamical noise on a typical chaotic system Chua's electronic circuit can be extracted by the temporal fluctuation of singular values (TFSV) obtained from singular value decomposition (SVD), independently of the presence of measurement noise. However, the obtained results have not yet been sufficiently verified from the statistical perspectives so far.

In this study, some statistical tests are performed regarding the influence of data length. Among them, the results of famous Akaike's Information Criterion (AIC) [4] are mainly illustrated.

As a result, it is found out that the validity of the results of our already proposed method is proved and the adequate conditions to obtain statistically correct results can be determined by the analyses.

2. Proposed Method

2.1. Noise

Generally, dynamical noise and measurement noise are defined for a flow system, respectively, as follows:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \xi^{(\mathbf{D})}), \tag{1}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) + \xi^{(\mathbf{M})}, \tag{2}$$

where $\mathbf{x}$ and $\mathbf{y}$ are, respectively, the underlying state vector and the observed one; $\mathbf{f}$ is a governing function of the system; $\mathbf{g}$ is an observation function; and $\xi^{(\mathbf{D})}$ and $\xi^{(\mathbf{M})}$ are dynamical and measurement noise, respectively.

2.2. Temporal Fluctuation of Singular Values (TFSV)

SVD is the operation to diagonalize the singular matrix. Now, if the $N\times n$ rectangular matrix X is diagonalized, the covariance matrix X^tX can be decomposed into $X^tX = V\Sigma^2V^t$, where Σ^2 is the $n\times n$ diagonal matrix and V and V^t are the $n\times n$ orthogonal matrix and the transposed matrix of V, respectively. Here, $VV^t = V^tV = I_n$ is satisfied using the $n\times n$ unit matrix I_n. As $\Sigma^2 = diag(\sigma^2(1),\sigma^2(2),\ldots,\sigma^2(n))$ is obtained, we can extract singular values (SVs) $\{\sigma(i)|i = 1,2,\ldots,n\}$, which are non-zero. The relatively larger SVs correspond to the principal orthogonal basis of the deterministic system. In general, measured data is frequently obtained as a scalar time series. The procedure of SVD for such data is explained. Now, a (n, J)-window:$\{x_i,x_{i+J},\ldots,x_{i+(n-1)J}\}$ is prepared, where n is the number of elements of the window and J is a sample time in applying the method of delays as described in Ref.[2]. Here, a finite measured time series $\{x_i\in R|i = 1,2,\ldots,N+n-1\}$ is transformed into the $N\times n(N\gg n)$ matrix X and the $n\times n$ covariance matrix X^tX can be obtained.

2.3. In the presence of Measurement Noise

In the presence of measurement noise, each SV uniformly increases, since the underlying state vectors and the noise are uncorrelated, as explained in Ref.[2]. If the system remains steady, uniformly increased SVs are expected to be nearly constant independent of the passage of time.

2.4. In the presence of Dynamical Noise

On the other hand, in the presence of dynamical noise, the result is utterly different from the case of measurement noise [1, 5]. As a result, SVs temporally fluctuate for consecutive time series. Thus, the influence of dynamical noise on chaos can be extracted with a different form from that of measurement noise. This result means that the influences of dynamical noise and measurement noise can be distinguished even in the case of the noise-mixed data composed of both noises.

2.5. Performance Index S

In practice, TFSV can be estimated as follows (see Fig.1). First, temporally consecutive time series data sets are prepared. Each of sets is called an "*interval*" $\{I_k|k = 1,2,\ldots,N_{int}\}$ in this study. In each I_k, N elements are included such as $\{x_{N(k-1)},x_{N(k-1)+1},\ldots,x_{N(k-1)+N}\}$. Second, SVs $\{\sigma_k(i)|i = 1,2,\ldots,n\}$ are calculated in each I_k, where i is an "*Index*" of SVs lined in descending order. Third, each standard deviation $S(i)$ of SVs over all *intervals* is calculated every ith Index. Finally, S_{av}, the average of $S(i)$ over all "*Indices*", is obtained as concrete expression of TFSV.

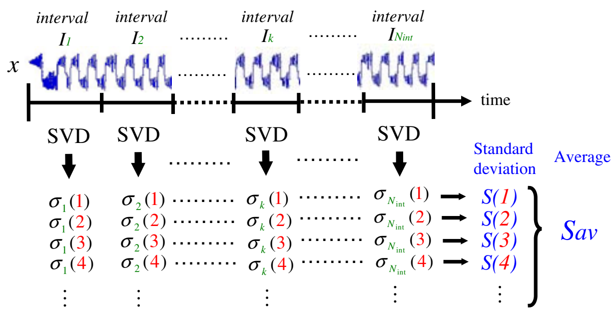

Figure 1: The concept of the extraction of TFSV. Temporally consecutive time series data sets are prepared. Each of sets is called an "*interval*" $\{I_k|k = 1,2,\ldots,N_{int}\}$. N elements; $\{x_{N(k-1)},x_{N(k-1)+1},\ldots,x_{N(k-1)+N}\}$ are included in each I_k. Meanwhile, each performance index $S(i)$ of SVs over all *intervals* is calculated for every ith Index. TFSV can be estimated by the average S_{av} as a proxy of $S(i)$.

$$S(i) = \sqrt{\frac{\sum_{k=1}^{N_{int}}(\sigma_k(i)-\overline{\sigma(i)})^2}{N_{int}}},\quad S(i)\in[0,\infty], \tag{3}$$

$$S_{av} = \frac{\sum_{i=1}^{n}S(i)}{n},\quad S_{av}\in[0,\infty], \tag{4}$$

where $\overline{\sigma(i)}$ is the average of $\sigma_k(i)$ over all "*intervals*" for ith Index.

3. Numerical Analysis

3.1. Preparation

Chua's electronic circuit is used as a typical chaos system, which is described by the 3-dimensional ordinary differential equations that follow [3],

$$C_1\frac{dV_{C_1}}{dt} = \frac{1}{R}(V_{C_2}-V_{C_1}) - f_{N_R}(V_{C_1}), \tag{5}$$

$$C_2\frac{dV_{C_2}}{dt} = \frac{1}{R}(V_{C_1}-V_{C_2}) + i_L, \tag{6}$$

$$L\frac{di_L}{dt} = -V_{C_2}, \tag{7}$$

where $f_{N_R}(V_{C_1}) = G_bV_{C_1} + \frac{1}{2}(G_a - G_b)|V_{C_1} + B_p| - |V_{C_1} - B_p|$. V_{C_1}, V_{C_2} and i_L indicate voltage of two capacitors C_1, C_2 and the current of coil L, respectively. $f_{N_R}(V_{C_1})$ denotes the 3-segment odd-symmetric voltage-current characteristic of the nonlinear resistor N_R, by which the system exhibits a large variety of typical chaotic behaviors. The *i.i.d.* dynamical noise ξ with a 0 mean is added to V_{C_1} such as $V_{C_1}\rightarrow V_{C_1} + \xi$ as additive noise. In this work values of parameters giving rise to double-scroll chaos are selected as follows, $C_1 = 10\ nF$, $C_2 = 100\ nF$, $L = 18\ mH$, $1/R = 0.55\ 1/\Omega$, $G_a = -0.758\ mA/V$, $G_b = -0.409\ mA/V$, and $B_p = 1.17\ V$. Here, the analyses are performed for the scalar time series of V_{C_1}. In this study, 4 kinds of time series are prepared, these being noise-free data ($NF-data$), measurement noise data ($M-data$), dynamical noise data ($D-data$), and noise-mixed data composed of both dynamical noise and measurement noise ($DM-data$). Each noise level is given as a ratio of a standard deviation of noise data to that of the time series V_{C_1} in NF_data. The range of the noise amplitude is 0.01%-20.0% for $M-data$ and 0.01%-3.7% for $D-data$, where 3.7% is the maximum, below which a chaotic state can be retained. For $DM-data$, measurement noise with a 20.0% noise level is added to all $D-data$. *The* 4*th-order Runge-Kutta method* is used with a constant time step $\tau_s = 0.000005$. The number of data length N in each *interval* and the number of *intervals* N_{int} are 100,000 and 10,000, respectively. SVD is performed for all intervals to extract TFSV. However, ahead of SVD, an adequate (n,J) window should be determined, satisfying the window length $\tau_w = n\tau_L = nJ\tau_s$, where the lag time τ_L is expressed as $\tau_L = J\tau_s$. In this study, the adequate window length should satisfy $nJ = 60$ and SVD is undertaken for $(n,J) = (4,15)$ [1]. Although $n = 4$ is not satisfied with Takens' embedding theorem, the embedding dimension 4 was selected as the condition, on which the change of TFSV can be more clearly captured, from the preliminary verification with various combinations of (n,J).

3.2. Results of TFSV

In figure 2(a), the distribution of TFSV of the representatives for 4 types of time series is illustrated. From standard deviations of these distributions, each S_{av} is obtained as shown in figure 2(b).

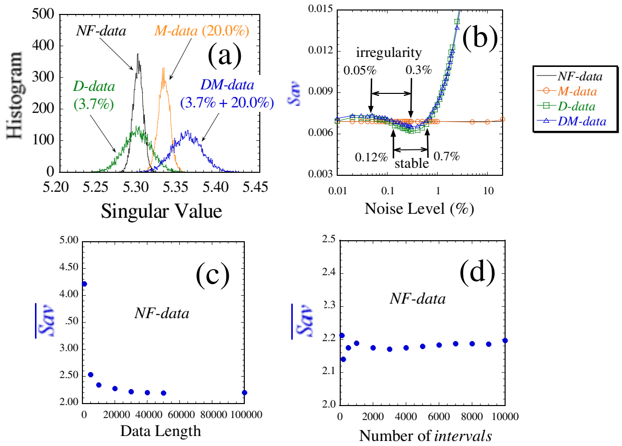

Figure 2: (a) The distribution of TFSV of the representatives for 4 types of time series. (b) Results of TFSV for the 4 representatives. (c) Influence of the data length in each *interval*. (d) Influence of the number of *intervals*.

3.3. Estimation of Data Length N in each Interval

The influence of the data length in each *interval* on S_{av} is estimated with the normalized index $\overline{S_{av}}$, as follows [5],

$$\overline{S_{av}} = \sqrt{N} \times S_{av}, \tag{8}$$

where N is the number of data points in each *interval*. The result can be shown in Figure 2(c) for $NF-data$. The data length from 1,000 to 100,000 are used for this estimation. As the data length increases, $\overline{S_{av}}$ seems to converge to a constant value in the range of more than nearly 30,000 data length. It can be said that the data length 100,000, which has been used in this study, seems to be statistically sufficient to obtain the adequate results. In Figure 2(d), the influence of the number of *intervals* is shawn. Few change of $\overline{S_{av}}$ is recognized as the number increases. However, N and N_{int}should be carefully determined. Next, as one of the famous analytical methods, AIC is introduced for the selection of adequate N and N_{int}.

4. A Statistical Test (AIC)

Some statistical approaches to select an adequate model to a given system or phenomenon have been proposed. Akaike's Information Criterion (AIC) is one of the most famous method among them [4]. In general, AIC can be expressed as following equation,

$$AIC = -2 \times MLL + 2 \times k, \tag{9}$$

where MLL means maximum logarithmic likelihood and k the number of degree of freedom.

If a stochastic density distribution of the target system $f(x_k,\theta)$ { x_k $|k = 0,1, \cdots, n$ } , θ:*parameters* is given, a logarithmic likelihood function LL can be defined as $LL = logL = \Sigma_{k=1}^{n} log f(x_k,\theta)$. Accordingly, MLL is decided by θ_{opt} satisfying $\frac{\partial}{\partial\theta} LL\,|_{\theta_{opt}} = \Sigma_{k=1}^{n}\, log f(x_k,\theta)\,|_{\theta_{opt}} = 0$.

In this work, AIC is used to statistically verify the validity of the results in Sections 3.2 and 3.3 and find the minimum data length in each *interval* and the minimum number of *intervals*, where statistically stable results can be obtained. Two types of tests are performed corresponding to the purposes. The first is to determine the minimum data length, where $\overline{S_{av}}$ converges to a constant value corresponding to the results of Section 3.3. The second is to determine the minimum number of *intervals*, where the results can be regarded as statistically stable.

As each set $\{S(i)\}$ indicates a normal distribution as shown in Figure. 2(a), the four types of models to compare the characteristics of two distributions are prepared for AIC test, as follows.

$$\begin{aligned} Model1\ (M1) &: \mu_1 = \mu_2, \nu_1 = \nu_2,\\ Model2\ (M2) &: \mu_1 \neq \mu_2, \nu_1 = \nu_2,\\ Model3\ (M3) &: \mu_1 = \mu_2, \nu_1 \neq \nu_2,\\ Model4\ (M4) &: \mu_1 \neq \mu_2, \nu_1 \neq \nu_2, \end{aligned}$$

where μ_1 and μ_2 are averages, and ν_1 and ν_2 are variances for the two distributions, respectively.

Here, a general procedure is explained for this case. Two sample sets such as $\{x_1,x_2,\cdots,x_n\}$ and $\{y_1,y_2,\cdots,y_n\}$ are prepared. Each set has a normal distribution $N(\mu_1,\nu_1)$ and $N(\mu_2,\nu_2)$, where μ_1 and μ_2 are averages, and ν_1 and ν_2 are variances for x_k and y_k ($k = 0,1,\cdots,n$), respectively.

The stochastic density distributions can be expressed,

$$f(x_k,\mu_1,\nu_1) = \frac{1}{\sqrt{2\pi\nu_1}} e^{-\frac{(x_k-\mu_1)^2}{2\nu_1}} \tag{10}$$

$$f(y_k,\mu_2,\nu_2) = \frac{1}{\sqrt{2\pi\nu_2}} e^{-\frac{(x_k-\mu_2)^2}{2\nu_2}}. \tag{11}$$

Accordingly, logarithmic likelihood function can be given,

$$LL(\mu_1,\mu_2,\nu_1,\nu_2) =$$

$$\Sigma_{k=1}^{n} \log f(x_k,\mu_1,\nu_1) + \Sigma_{k=1}^{n} \log f(y_k,\mu_2,\nu_2). \tag{12}$$

For the two given distributions, the four AIC values are calculated corresponding to the four above-mentioned models. The adequate model to the given distributions can be determined by selecting the model with the minimum AIC value among the four calculated AIC values.

For the first test, two distributions of SVs corresponding to the two different time series of NF-*data* with a deferent N are compared. One is the distribution at the data length $N = 100,000$ and the other is at another N value. If $\overline{S_{av}}$ converges to a constant value as data length increases, the selected model should be M1 or M2, which has a relation $\nu_1 = \nu_2$. Accordingly, from the model selection, it becomes

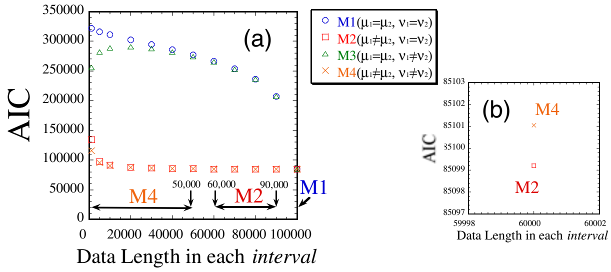

Figure 3: A selection of a model indicating the smallest AIC value in determination of the minimum data length in each *interval* in *NF-data*. (a) Selected models are different dependent on data length in each *interval*. (b) A model M2 is selected at the data length 60,000 for example.

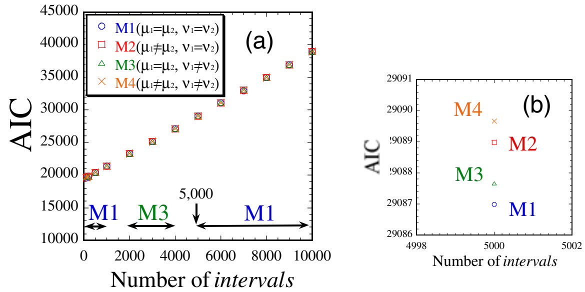

Figure 4: A selection of a model indicating the smallest AIC value in determination of the minimum number of *intervals* in *NF-data*. (a) Selected models are different dependent on the number of *intervals*. (b) A model M1 is selected at the number of *intervals* 5,000 for example.

possible to determine the minimum data length, where the statistically stable results can be obtained. For the second one, two distributions of SVs of *NF-data* are also compared. One is the distribution at N_{int} = 10,000 and the other is at another N_{int} value. If $\overline{S_{av}}$ becomes stable as the number of *intervals* increases, the selected model should be M1 or M2 analogously to the first case. Accordingly, from the model selection, it becomes possible to determine the minimum number of *intervals*, where statistically stable results can be obtained.

The results of the analyses are illustrated in Figures 3 and 4 only for a set $\{S(1)\}$. The selected model is indicated for each data length and each number of *intervals*, respectively. As in other sets $\{S(i)\}$ similar tendencies can be seen and the first SV can be regarded as dominant in the system, it is enough to estimate only the results of a set $\{S(1)\}$. In Figures 3, a model M1 or M2, which has a relation $\nu_1 = \nu_2$, is selected in the range of more than the data length 60,000 in each *interval* under the fixed number of *intervals* 10,000. Furthermore, in 4, a model M1, which has a relation $\nu_1 = \nu_2$, is selected in the range of more than 5,000 *intervals* under the fixed data length in each *interval* 10,000. Accordingly, it is found out that the previous calculation condition $N = 100,000$ and $N_{int} = 10,000$ are reliable enough and the minimum numbers can be determined as $N = 60,000$ under the fixed data length 10,000 and $N_{int} = 5,000$ under the fixed number of *intervals* 10,000, respectively. Similar results are obtained by other statistical tests; *F-test* and *Kolmogorov−Smirnov test* as a conventional hypothesis test. These tests, however, need a significance level and some preprocesses. While, AIC is free from those operations. Accordingly AIC is more convenient than other methods and hence the results of statistical analyses are illustrated by only AIC.

5. Conclusions

Some statistical methods are introduced for the verification of the results by the proposed method, particularly, the analyses of AIC is mainly explained here. The adequate models are selected from the smallest AIC value for both the minimum data length in each *interval* and the number of *intervals*. As a result, it is found out that the validity of the results of our already proposed method is proved and the adequate conditions to obtain statistically stable results can be determined by the analyses.

Acknowledgments

We would like to thank Professor K. Aihara and his colleagues, especially Dr. Y. Hirata at The University of Tokyo for fruitful comments and stimulating discussion. This study was supported by a Grant-in-Aid for Scientific Research of JSPS.

References

[1] M. Todoriki, H. Nagayoshi, and A. Suzuki, Temporal Fluctuation of Singular Values caused by Dynamical Noise in Chaos, *Phys. Rev. E*, **72**, pp. 036207, 2005.

[2] D. S. Broomhead, and G. P. King, Extracting Qualitative Dynamics from Experimental Data, *Physica D*, **20**, pp. 217, 1986.

[3] R. N. Madan, Chua's Circuit: A Paradigm for Chaos, *World Scientific Nonlinear Science Ser. B*, World Scientific vol.1, 1993.

[4] H. Akaike, A New Look at the Statistical Model Identification, *IEEE Trans. Autom. Control*, **19**, pp. 716, 1974.

[5] M. Todoriki, and S. Hasegawa, Further Investigation for the SVD-based Analysis of Dynamical Noise on Chaos, *Proc. Int. Symo. NOLTA2007, Vancouver*, pp. 473, 2007.