Journal of the Korean Physical Society, Vol. 46, No. 3, March 2005, pp. 657∼661

A Noise-Filtering Scheme for a Chaotic Signal

Seung Ki BAEK* and Hie-Tae MOON
Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701

(Received 12 October 2004)

We present a noise-filtering scheme which works on a chaotic signal containing a certain level of noise. Our method exploits the simplicity of Rosa's filtering scheme, as well as the high applicability of the maximum likelihood method. We tested our scheme on signals from various dynamical systems and found that the noise amplitudes could be reduced up to a few percent. In the study, we are learned that our scheme could also be used for inferring the underlying dynamics of a received chaotic signal when no *a-priori* knowledge of the dynamics was given.

PACS numbers: 05.45.-a, 05.40.Ca
Keywords: Nonlinear dynamics, Noise filter

I. INTRODUCTION

How to filter noise from a signal has been of great importance in communication and experimental research. As the broad-band spectrum of signals from nonlinear chaotic systems usually makes traditional linear filters unfeasible, many researchers have studied noise reduction methods applicable to nonlinear systems. After Kostelich and Yorke published a classical work using determinism [1, 2], Hammel, as well as Farmer and Sidorowich, proposed an alternative approach based on *shadowing* [3, 4] (also see Ref. 5 for comparison), Schreiber and Kostelich investigated a simple and common method [6,7], and Davies developed the framework of Bayesian Theory [8]. Recently, Bröcker and Parlitz considered a refinement of the gradient descent method [9]. Most of these methods approach noise reduction by formulating minimization problems.

In the viewpoint of information theory [10, 11], a chaotic system is interpreted as an active process of its dynamical information. As time goes, areas in state space are repeatedly stretched and folded, and this mechanism causes a sensitive dependence on the initial conditions. In this view, we have measuring tools with finitely limited resolution. The stretching process reveals more precisely the initial state that could not be identified with the tools available at that time. If only a stretching process exists, the occupied area in state space, *i.e.*, the energy of the system, diverges to infinity as the precision infinitely increases, as Brillouin claimed in Ref. 11. The folding process prevents this divergence, inevitably removing some stored information of the past, which is why we cannot discriminate every detail of the past only by observing the present state. Thus, sufficient observation should resolve where true data points lie in the absence of noise [12]. The purpose of this article is to present how this property is exploited in noise filtering and how this method becomes possible in high dimensional systems.

II. THEORETICAL BACKGROUND

It is widely known that there are two kinds of noise: *Measurement noise* means corruption of data in the observation process without interfering with the dynamics itself while *dynamical noise* denotes a perturbation of the system coupled to dynamics and occurring at each time step. In this paper, we treat measurement noise, which we can pretend to be dynamical and vice versa [4]. There exists a *true* orbit $\{Y_k\}_{k=1}^N$ satisfying a certain dynamics $Y_{k+1} = M(Y_k)$ for $1 \le k \le N-1$ and we observe only a *noisy* orbit $\{X_k\}_{k=1}^N$ given by $X_k = Y_k + \eta_k$ for small $|\eta_k| < \delta$, where η_k and δ denotes the *noise* and the *noise level*, respectively. We would like to obtain a less noisy orbit $\{X'_k\}_{k=1}^N$, and most approaches to this problem minimize a target function with constraints, such as

$$S = \sum_{k=1}^{N} |X'_k - X_k|^2 + \sum_{k=1}^{N-1} \{M(X'_k) - X'_{k+1}\}\lambda_k, \quad (1)$$

where λ_k is a Lagrangian multiplier [4]. Minimizing S corresponds to maximizing the *likelihood* function $\bar{P}$ within a time interval $[t-\alpha, t+\beta]$:

$$\bar{P}(M^{\alpha}(X_{t-\alpha}), \ldots, M^{-\beta}(X_{t+\beta})) \propto \prod_{j=\alpha}^{j=-\beta} \exp\left(-\frac{1}{2\sigma^2}\left|\frac{M^j(X_{t-j}) - Y_t}{dM^j(Y_{t-j})}\right|^2\right), \quad (2)$$

*E-mail: garuda@kaist.ac.kr; Fax: +82-42-869-2510

where dM is the derivative of M under the assumption that the sequence $\{\eta_k\}$ is independently Gaussian distributed with standard deviation σ. The probability distributions of position at different times are transported to a particular time and distorted by a chaotic dynamics M; the true data point is restricted to their intersection. Thus, the maximum of the joint probability function $\bar{P}$ estimates the position of the true data point at that time. We shall discuss how this calculation is simplified if we consider information aspects as in the area of communication.

Studies on communication using chaos [13–15] have been carried out for an understanding of chaos control [16,17] and chaos synchronization [18,19]. The main issues in this field are how to encode information using a chaotic signal with dynamics already known to both the transmitter and receiver and how to build a system resistant to noise occurring in the communication channel, which corresponds to measurement noise. Rosa *et al.* [20] illustrated a filtering method using a $2x$ mod 1 map. This method, called *Rosa's method*, is described as follows: One picks a point (X_t, X_{t+1}) and executes a backward iteration on X_{t+1}, resulting in two pre-images, $\hat{X}_t^{Left}$ and $\hat{X}_t^{Right}$; the one closest to X_t is selected as a filtered point of time t. This filter shrinks the noise by a factor of two (*i.e.* the Lyapunov exponent of the map) at each iteration. Andreyev *et al.* [21] investigated information aspects and applications of Rosa's method. They, however, only treated basically 1-dimensional maps because they had to do the inverse mapping directly.

III. NOISE FILTER

Following Rosa *et al.*, we start with the case of a $2x$ mod 1 map as the simplest example of chaotic dynamics and of our scheme. If a binary representation is employed in describing states, each iteration simply shifts the decimal point one space to the right. Let us assume that we introduce noise with such a level that we can guarantee only the first effective number. If the initial state is observed to be $0.a_0xx\ldots$ and the first and the second iterations give $0.a_1xx\ldots$ and $0.a_2xx\ldots$, respectively, noting that digits marked by x may be spurious, we can say that the initial state is, in fact, $0.a_0a_1a_2\cdots$, effectively reducing the noise on the initial state.

The above example involves two conditions: the noise level δ is known, and the dynamics is chaotic. In such cases, we ignore the spoiled parts, and that converts an observed point to a set of candidate points, leading to degeneracy (*e.g.*, all the points whose first digit is a_0). Then, we clarify what the point should be by receiving information from other unspoiled parts of data. Roughly speaking, proper temporal extension can compensate for spatial ambiguity [22]. If a data point X_t is observed, the real value Y_t should lie within a finite neighborhood $I(X_t)$, whose size comes from the noise level δ. The next real value Y_{t+1}, evolving from Y_t deterministically, also belongs to $I(X_{t+1})$ while it does not hold for every point $p_t \in I(X_t)$ and its successor p_{t+1}. Noting that the inverse mapping M^{-1} operates on a set of points, not on a single point where the inverse map cannot be defined, we find the n-th order refinement,

$$I(X_t)^{new}_{(n)} = \bigcap_{i=0}^{n} M^{-i}\left\{I(X_{t+i})\right\}. \tag{3}$$

In terms of the previous example, $M^{-i}\{I(X_i)\}$ with $t = 0$ means the set of binary numbers whose i-th digit is a_i. As the n-th order refinement requires $n+1$ successive measurements, it is obvious that the diameter of a remaining set never increases so that this algorithm is non-divergent:

$$0 < \left|I(X_t)^{new}_{(n)}\right| \le \left|I(X_t)^{new}_{(n-1)}\right|. \tag{4}$$

Equation (3) shares similarity with Eq. (2) of the maximum likelihood method, but the Gaussian assumption turns out to be unnecessary in our scheme. Once δ is defined, other details of the noise are irrelevant. It is also worth noting that Eq. (3) formalizes the basic philosophy of Rosa's method. The difficulty in its application is remedied by rewriting Eq. (3) as

$$M^n\left\{I(X_t)^{new}_{(n)}\right\} \subset \bigcap_{i=0}^{n} M^{i}\left\{I(X_{t+n-i})\right\}. \tag{5}$$

This allows one to avoid calculating an inverse mapping, which is hardly possible in high-dimensional systems. We deduce that if a point does not belong to the set of the right-hand side of Eq. (5), it cannot lie in the set of the left-hand side. Then, what has to be done is only selecting points within $I(X_t)$ which satisfy the right-hand side after n times of mapping. Henceforth, we iterate all nearby grid points around the observed data which approximate $I(X_t)$ in a discrete manner and reject false ones outside the next expected intervals, $I(X_{t+1})$. We repeat the same procedure until the number of remaining points is less than a certain threshold, *i.e.*, $\left|I(X_t)^{new}_{(m)}\right| < R_{th}$. X_t is then corrected to $X'_k = \left\langle I(X_t)^{new}_{(m)}\right\rangle$, the average of those remaining points. The number of steps, m, required to reach this threshold R_{th}, measures the performance of the noise filter, and we define this quantity as *decay time.* Since each point has its m, we obtain another sequence of decay times $\{m_k\}_{k=1}^{N}$ after refinement. A system with short m is so sensitive that wrong guesses are easily rejected; thus, it is easy to clean the noise. One can expect the average decay time $m_{avg} = N^{-1}\sum_{k=1}^{N} m_k$ to be related to the Lyapunov exponent h, and the relationship is depicted in Fig. 1 for a typical initial condition in the generalized Baker's map. This dependency, $m_{avg} \propto h^{-1}$, comes from Shannon-McMillan-Breiman Theorem [23]: if M is an ergodic transformation on a probability space

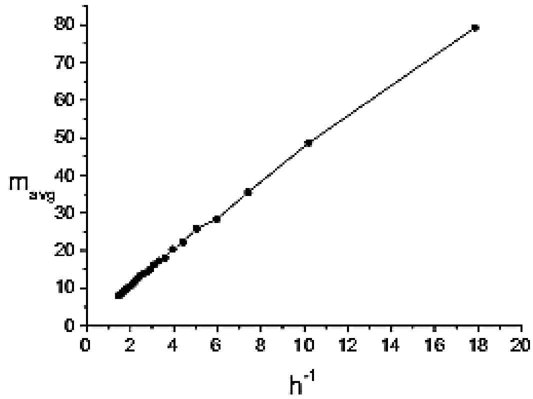

Fig. 1. The average decay time m_{avg} is inversely proportional to the Lyapunov exponent for the generalized baker's map.

(Z, μ) with a finite generating partition P, then for almost every $w = (w_1 w_2 w_3 \ldots) \in Z$,

$$- \lim_{n\to\infty} \frac{\log \mu\left(P_n(w)\right)}{n} = h(M), \tag{6}$$

where $P_n(w) = \{y \in Z : y_1 = w_1, \cdots, y_n = w_n\}$. By choosing $n \simeq m$, we fix μ at a constant value and obtain $m_{avg} \propto h^{-1}$. The proportionality appears to depend on both the system and the noise characteristics and is to be studied more. Later in section IV. , we use this concept of average decay time in a different context; that is, fast decay implies a large deviation from the true dynamics.

Figure 2 demonstrates the result of this scheme for a Lorenz system:

$$\dot{x} = \sigma(y - x), \tag{7}$$

$$\dot{y} = rx - y - xz, \tag{8}$$

$$\dot{z} = xy - bz, \tag{9}$$

where $\sigma = 10$, $r = 28$ and $b = 8/3$. The *noisy* orbit $\{X_k\}$ is generated in Fig. 2(a) by introducing noise of $\delta \approx 5\%$ of whole system size, which is enough to destroy most important characteristics of the attractor [17]. Our scheme corrects each point X_k into X'_k, as depicted in Fig. 2(b), where $20 \times 20 \times 20$ neighboring grid points are constructed for each data point and R_{th} is set to be 10 throughout the calculation. We define the *relative variance* as

$$e = \frac{\sum_{k=1}^{N}(Y_k - X'_k)^2}{\sum_{k=1}^{N}(Y_k - X_k)^2} \tag{10}$$

to quantify the performance of the scheme, where $e < 1$ means that noise is reduced ($e = 0$ for total noise removal). This demonstration yields $e \approx 0.05$, which implies a high point-to-point correspondence so that following Ref. 4, this scheme can be categorized as *detailed noise reduction.* The value of e can be controlled by

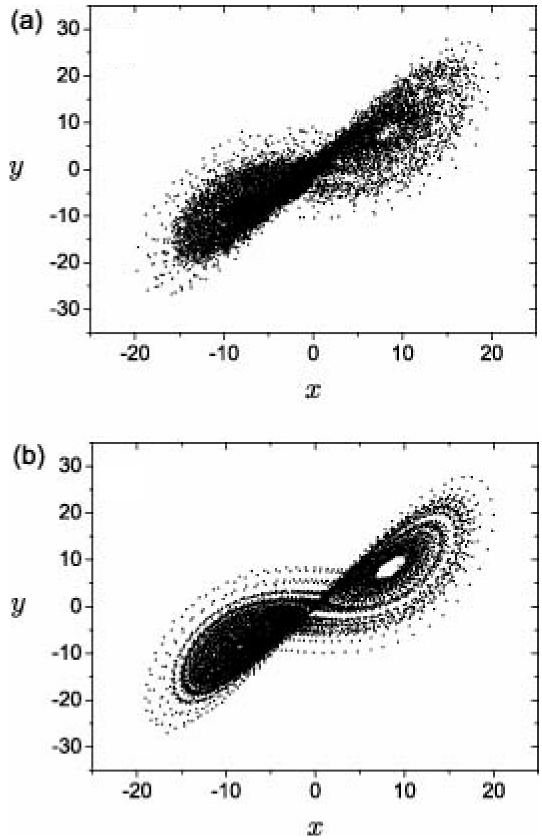

Fig. 2. (a) Lorenz attractor with 5 % noise added and (b) refined data (100,000 iterations for each). The relative variation becomes reduced to about 0.05.

the number of neighboring grid points and the threshold value R_{th}. Similar results are obtained for the logistic map, Rössler's system, and the 4-dimensional model introduced in Ref. 24.

The condition of chaoticity often involves directionality in some maps where invariant manifolds are aligned with specific directions. For example, we apply the above procedure to the generalized Baker's map (Fig. 3). In that histogram, hollow and filled bars represent the error distributions of the original data set and the refined one, respectively. It is evident that noise gets very close to zero only along the unstable direction. The other direction can be properly treated by a backward iteration. We also find that this directionality highlights the existence of homoclinic tangencies, which are suggested as a solution to the generating partition problem in nonhyperbolic systems [25,26].

IV. INFERRING DYNAMICS

So far, a full knowledge of dynamics has been assumed. Although this assumption may be valid in some areas, in general, we need to infer dynamics from given raw data. Farmer and Sidorowich pointed out that how much noise

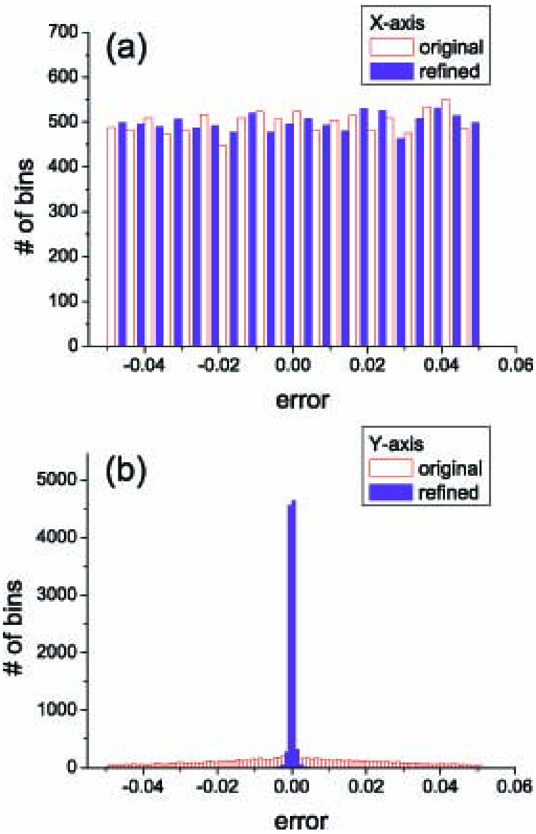

Fig. 3. Directionality of noise reduction in the generalized baker's map. The noise remains at the same level along the x-axis (stable direction), while it shrinks to zero, as expected, along the y-axis (unstable direction).

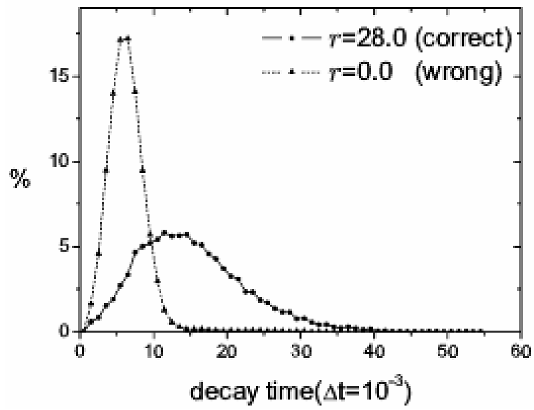

Fig. 4. Distribution shapes of the decay time in a Lorenz system. On average, the decay time decreases with the deviation from correct r.

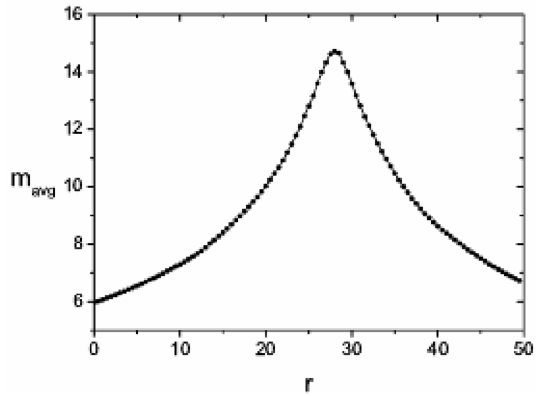

Fig. 5. m_{avg} at some r values in a Lorenz system. A peak appears when the correct value is chosen.

one can reduce is limited by the accuracy of the approximation to the true dynamics [4]. At first, we tried to find local linear dynamics as Kostelich and Yorke did [1], but it was not quite satisfactory because determining the size of the neighborhood was troublesome; that is, a size that was too small often decreased the statistical confidence and one that was too large could not capture the fine structure of the attractor. Looking for alternatives consistent with the above scheme, we noted that the true dynamics would be the most accurate approximation among other candidate models and that our getting closer to the true dynamics could be expressed by a longer m on average.

Let us suppose that the parameter r in Eq. (8), representing the Rayleigh number in the convection problem [27], is unknown. Even though we are given the same data as Fig. 2(a), we should test many Lorenz systems with different r values until finding $r = 28$. We already know that application of the scheme yields an m_k for each data point X_k, resulting in a series of $\{m_k\}$. It is natural that the statistical properties of $\{m_k\}$ change with the different r value. Figure 4 shows how the choice of r changes the distribution of $\{m_k\}$. We depicted only two cases, $r = 28$ (correct) and $r = 0$ (wrong), though we observed the same tendency for intermediate values of r. The average decay time, m_{avg}, rises to 14.71 for $r = 28$ while it rises to only 5.96 for $r = 0$. The distribution looks Maxwellian in the vicinity of the true dynamics, and this Maxwellian region can be reached by processing raw data. This distribution form interests us, but the reason remains to be explained later. The values of m_{avg} are plotted in Fig. 5 for various parameters. It is obvious in a Lorenz system that the correct parameter values can be chosen from the peak. Rössler's system also gives the desired results, though it requires quite many data points to find a peak.

Let us consider two extreme cases to elucidate the basic nature of the distribution. If the underlying dynamics is so trivial (*e.g.*, stable periodic motion) that one can easily discover it, the future orbit is highly predictable, and the distribution will be drawn to infinity. As a nonchaotic system contains little information, our scheme becomes ineffective with diverging m_{avg}. Conversely, if the dynamics looks totally unpredictable based on our knowledge, the distribution will collapse to a zero point. We again see that noise is not reduced at all because the accuracy of the approximation sets an upper bound of reducing performance, as stated above.

V. CONCLUSION

In summary, we suggest a nonlinear noise filtering scheme, which requires the two conditions of chaoticity and a well-defined noise level. Topological considerations and an information-theoretic approach are combined in our scheme and provide a concise and easily applied way to reduce noise. In the example of a Lorenz model, this scheme yields a good result in which the variance from the original orbit becomes about a twentieth. The value of relative variance, a performance measure of the filter, can be controlled by using the parameters of the scheme. The example also clearly demonstrates that Rosa's method can be extended to a high dimensional system if properly formulated.

We introduce the concept of the decay time and propose its average m_{avg}, an extension of the Lyapunov exponent, as a quantifier for inferring dynamics. We note that inference is but the other side of noise reduction. Since the quantifier is expected to be highest when the parameter under consideration hits its true value, one can guess the unknown value of the parameter from its peaks. These results imply that an information theoretic approach can help one to get a fruitful perspective on dynamical systems.

ACKNOWLEDGMENTS

We would like to thank T-W. Ko for helpful comments. This work is supported by grant No. R01-1999-000-00019-0 from the Basic Research Program of the Korea Science and Engineering Foundation.

REFERENCES

[1] E. J. Kostelich and J. A. Yorke, Phys. Rev. A **38**, 1649 (1988).
[2] E. J. Kostelich and J. A. Yorke, Physica **41D**, 183 (1989).
[3] S. M. Hammel, Phys. Lett. **148A**, 421 (1990).
[4] J. D. Farmer and J. J. Sidorowich, Physica **47D**, 373 (1991).
[5] M. Davies, Physica **79D**, 174 (1994).
[6] T. Schreiber, Phy. Rev. E **47**, 2401 (1993).
[7] E. J. Kostelich and T. Schreiber, Phy. Rev. E **48**, 1752 (1993).
[8] M. Davies, Chaos **8**, 775 (1998).
[9] J. Bröcker and U. Parlitz, Chaos **11**, 319 (2001).
[10] C. E. Shannon, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949) p. 18.
[11] L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1962) p. 235.
[12] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 1993) p. 142.
[13] S. Hayes, C. Grebogi and E. Ott, Phys. Rev. Lett. **70**, 3031 (1993).
[14] K. M. Cuomo and A. V. Oppenheim, Phys. Rev. Lett. **71**, 65 (1993).
[15] E. Bollt, Y-C. Lai and C. Grebogi, Phys. Rev. Lett. **79**, 3787 (1997).
[16] E. Ott, C. Grebogi and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990).
[17] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, (Cambridge University Press, Cambridge, 1997) p. 152.
[18] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).
[19] J-H. Cho, D-U. Hwang, C-M. Kim and Y-J. Park, J. Korean Phys. Soc. **39**, 378 (2001); W. Lim and S.-Y. Kim, J. Korean Phys. Soc. **44**, 510 (2004); J. Korean Phys. Soc. **45**, 287 (2004).
[20] E. Rosa, Jr., S. Hayes and C. Grebogi, Phys. Rev. Lett. **78**, 1247 (1997).
[21] Y. V. Andreyev, A. S. Dmitriev, E. V. Efremova and A. N. Anagnostopoulos, Chaos Solitons & Fractals **17**, 531 (2003).
[22] S. D. Pethel, N. J. Corron, Q. R. Underwood and K. Myneni, Phys. Rev. Lett. **90**, 254101 (2003).
[23] G. H. Choe, *Computational Ergodic Theory* (to be published by Springer-Verlag).
[24] H-T. Moon, Phys. Lett. **325A**, 324 (2004); H-T. Moon, Phys. Rev. Lett. **79**, 403 (1997).
[25] P. Grassberger and H. Kantz, Phys. Lett. **113A**, 235 (1985); M. Eisele, J. Phys. A: Math. Gen. **32**, 1533 (1999).
[26] K. T. Park and S. H. Yi, J. Korean Phys. Soc. **44**, 569 (2004).
[27] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Addison-Wesley, Reading, Massachusetts, 1994) p. 310.