九州大学学術情報リポジトリ
Kyushu University Institutional Repository

Estimating the Lyapunov Exponent from Chaotic Time Series with Dynamic Noise

Yonemoto, Koji
Department of Medicine and Clinical Science,Graduate School of Medical Sciences, Kyushu University

Yanagawa, Takashi
Graduate School of Mathematics, Kyushu University

https://hdl.handle.net/2324/3343

出版情報：Statistical Methodology. 4 (4), pp.461-480, 2007-10. Elsevier
バージョン：
権利関係：© 2007 Elsevier B.V.

MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

Estimating the Lyapunov exponent from chaotic time series with dynamic noise

K. Yonemoto & T. Yanagawa

MHF 2004-1

(Received February 10, 2004)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Estimating the Lyapunov Exponent from Chaotic Time Series with Dynamic Noise

Koji Yonemoto[1] and Takashi Yanagawa[2]

[1]Department of Medicine and Clinical Science,
Graduate School of Medical Sciences, Kyushu University
yonemoto@intmed2.med.kyushu-u.ac.jp

[2]Graduate School of Mathematics, Kyushu University
yanagawa@math.kyushu-u.ac.jp

Abstract

In this paper, we propose an estimator of the Lyapunov exponent of the skeleton for chaotic time series with dynamic noise and prove the consistency of the estimator under some assumptions.

Keywords: embedding dimension, delay time, skeleton, Nadaraya-Watson kernel type estimator, ϕ-mixing, Lyapunov exponent

1 Introduction.

Methods of analyzing experimental and observational data for evidence of chaos have been applied to data in physics, geology, astronomy, neurobiology, ecology and economics. Among them is the Lyapunov exponent which often plays a key role of detecting chaos. Conventional methods for estimating the Lyapunov exponent are reliable if the data are abundant, if measurement error is near 0, and if the data really come from a deterministic system. However, with limited data or a system subject to nonnegligible stochastic perturbations, it is well known that the estimates may be incorrect or ambiguous. We consider the following non-linear autoregressive system with additive noise

$$X_t = F(X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-d\tau}) + \varepsilon_t,$$

where F is unknown non-linear function, d and τ are unknown positive integers called embedding dimension and delay time, respectively, and $\{\varepsilon_t\}$ is a sequence of random noise which is called dynamic noise.

Several methods have been developed to overcome the difficulty. Kostelich and York (1990) approximated F by polynomials and separated the signal from noise, and Pikovsky (1986), Landa and Rosenblum (1989), Cawley and Hsu (1992), and Sauer (1992) filtered out the noise by using linear filters. McCaffrey et al. (1992) employed non-parametric estimation of F, but they assumed identical noises. Yao and Tong (1994) explored alternative measures of detecting chaos in observational data. We take the same approach as Pikovsky (1986), Landa and Rosenblum (1989), Cawley and Hsu (1992), and Sauer (1992), but use kernel type estimators for filtering out the noise. Our goal is to estimating the Lyapunov exponent of F by observing $\{X_t\}$. Our method consists of three steps. First, estimate the embedding dimension and delay time. Second, estimate the skeleton by the Nadaraya-Watson kernel type estimator by using the estimated embedding dimension and delay time in Step 1. Last, generate the data from the estimated skeleton and estimate the Lyapunov exponent by using the generated data and partial derivative of the estimated skeleton. The consistency of the proposed estimator is proved.

The present paper is organized as follows. We propose a method for estimating the Lyapunov exponent in Section 2. In Section 3, the consistency of the estimator is proved. In Section 4, the behavior of the procedure is evaluated numerically.

2 Estimator of the Lyapunov exponent.

2.1 Basic definitions.

We consider $\{X_t\}_{t=1,2,\ldots,N}$ generated from

$$X_t = F(X_{t-\tau_0}, X_{t-2\tau_0}, \ldots, X_{t-d_0\tau_0}) + \varepsilon_t, \tag{1}$$

where d_0 and τ_0 are unknown positive integers and $F : R^{d_0} \to R$ is unknown non-linear function such that $\{Y_t\}$ is ergodic where $Y_t = F(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-d_0})$. We assume that $\{X_t\}$ is a discrete-time strictly stationary time series with $EX_t^2 < \infty$ and that for any positive integer t,

$$E[\varepsilon_t|\mathcal{A}_1^{t-1}(X)] = 0, \text{ almost surely}, \tag{2}$$

and

$$E[\varepsilon_t^2|\mathcal{A}_1^{t-1}(X)] = \sigma^2, (\sigma > 0), \text{ almost surely},$$

where $\mathcal{A}_s^t(X)$ denotes the sigma algebra generated by $(X_s, \ldots, X_t)$. It follows that $F(X_{t-\tau_0}, \ldots, X_{t-d_0\tau_0}) = E[X_t|X_{t-\tau_0}, \ldots, X_{t-d_0\tau_0}]$. The embedding dimension and delay time are defined as follows.

Definition 2.1. *The time series $\{X_t\}$ is said to have the embedding dimension d_0 and the delay time τ_0 if and only if there exist non-negative integers $d_0 < \infty$ and $\tau_0 < \infty$ such that*

$$E[X_t|X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-d\tau}] \neq E[X_t|X_{t-\tau_0}, X_{t-2\tau_0}, \ldots, X_{t-d_0\tau_0}] \text{ a.e.} \tag{3}$$

for any $d < d_0$, and any $\tau > 0$, and

$$E[X_t|X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-d\tau}] = E[X_t|X_{t-\tau_0}, X_{t-2\tau_0}, \ldots, X_{t-d_0\tau_0}] \text{ a.e.} \tag{4}$$

for any $(d, \tau) \in \mathbf{B}(d_0, \tau_0)$, where $\mathbf{B}(d_0, \tau_0) = \{(d, \tau)|\{\tau_0, 2\tau_0, \ldots, d_0\tau_0\} \subset \{\tau, 2\tau, \ldots, d\tau\}\ \}$.

Model (1) may be represented as

$$\mathbf{X}_t^{(d_0,\tau_0)} = \mathbf{F}(\mathbf{X}_{t-\tau_0}^{(d_0,\tau_0)}) + \mathbf{e}_t, \tag{5}$$

where

$$\begin{aligned}
\mathbf{F}(\mathbf{x}) &= {}^t(F(\mathbf{x}), x_1, \ldots, x_{d_0-1}) \text{ for } \mathbf{x} = {}^t(x_1, x_2, \ldots, x_{d_0}) \in R^{d_0}, \\
\mathbf{X}_t^{(d_0,\tau_0)} &= {}^t(X_t, X_{t-\tau_0}, \ldots, X_{t-(d_0-1)\tau_0}), \text{ and} \\
\mathbf{e_t} &= {}^t(\varepsilon_t, 0, \ldots, 0).
\end{aligned}$$

Then skeleton of model (5) is defined as follows.

Definition 2.2 (skeleton). *We refer to the following model as the skeleton of model (5).*

$$\mathbf{Y}_t^{(d_0,1)} = \mathbf{F}(\mathbf{Y}_{t-1}^{(d_0,1)}), \tag{6}$$

where

$$\mathbf{Y}_t^{(d_0,1)} = {}^t(Y_t, Y_{t-1}, \ldots, Y_{t-d_0+1}).$$

Since $\{\mathbf{Y}_t^{(d_0,1)}\}$ is ergodic from the assumption, the Lyapunov exponent of the skeleton (6) is defined as follows.

Definition 2.3 (Lyapunov exponent). *The Lyapunov exponent λ of the skeleton (6) is defined as*

$$\lambda = \lim_{M \to \infty} \frac{1}{M} \log \| D\mathbf{F}(\mathbf{Y}_M^{(d_0,1)}) D\mathbf{F}(\mathbf{Y}_{M-1}^{(d_0,1)}) \cdots D\mathbf{F}(\mathbf{Y}_1^{(d_0,1)}) \|,$$

where $D\mathbf{F}(\mathbf{Y}_t^{(d_0,1)})$ is the matrix of partial derivatives of the map $\mathbf{F} : \mathbf{R}^{d_0} \to \mathbf{R}^{d_0}$ evaluated at $\mathbf{Y}_t^{(d_0,1)}$,

$$\|T\| = \sup_{\|\mathbf{x}\|=1} \|T\mathbf{x}\| \text{ and}$$

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{d_0} x_i^2\right)^{\frac{1}{2}}$$

for $T : d_0 \times d_0$ matrix and $\mathbf{x} = {}^t(x_1, x_2, \ldots, x_{d_0}) \in \mathbf{R}^{d_0}$.

2.2 The procedure for estimating the Lyapunov exponent.

Suppose that $X_1, X_2, \ldots, X_N$ are observed. Then we propose a procedure for estimating the Lyapunov exponent of the skeleton from $\{X_t\}_{t=1,2,\ldots,N}$. Outline of the procedure is as follows.

(1). Estimate the embedding dimension d_0 and delay time τ_0 from $\{X_t\}_{t=1,2,\ldots,N}$ by the procedure proposed by Yonemoto and Yanagawa (2001). Denote the estimated embedding dimension and delay time by $\hat{d}_0$ and $\hat{\tau}_0$, respectively.

(2). Estimate the skeleton from $\{X_t\}_{t=1,2,\ldots,N}$ by the Nadaraya - Watson kernel type estimator using $\hat{d}_0$ and $\hat{\tau}_0$, and generating $\{\hat{\mathbf{Y}}_{N,t}\}_{t=1,2,\ldots,M}$ from the estimated skeleton by giving an appropriate initial vector.

(3). Estimate the Lyapunov exponent by using the generated data and partial derivatives of the estimated skeleton.

The details of each steps are given in the following subsections.

2.3 Estimating the embedding dimension and delay time.

Let $\{X_1, \ldots, X_N\}$ be a set of observed data. For positive integers d, τ and $L \geq d\tau$, put

$$CV(d,\tau) = \frac{1}{N-L+1}\sum_{t=L}^{N}(X_t - \hat{F}_{\backslash t(d,\tau)}(X_{t-\tau}, \ldots, X_{t-d\tau}))^2,$$

where $\hat{F}_{\backslash t(d,\tau)}$ denotes the Nadaraya - Watson kernel type estimated regression function (Nadaraya (1964), Watson(1964)) with the $t-$th point deleted , that is,

$$\hat{F}_{\backslash t(d,\tau)}(\mathbf{z}) = \frac{1}{N-L}\sum_{s=L, s\neq t}^{N} X_s K_{d,h}(\mathbf{z} - (X_{s-\tau}, \ldots, X_{s-d\tau}))(\hat{f}_{\backslash t(d,\tau)}(\mathbf{z}))^{-1}$$

for $\mathbf{z} = (z_1, z_2, \ldots, z_d)$, where the summation over s omit t in each case, and

$$\hat{f}_{\backslash t(d,\tau)}(\mathbf{z}) = \frac{1}{N-L}\sum_{s=L, s\neq t}^{N} K_{d,h}(\mathbf{z} - (X_{s-\tau}, \ldots, X_{s-d\tau})),$$

and

$$K_{d,h}(\mathbf{z}) = \frac{1}{h_{d,N}^d} \prod_{i=1}^{d} K\left(\frac{z_i}{h_{d,N}}\right), \tag{7}$$

where $K : \mathbf{R} \to \mathbf{R}$ is taken as a density function of a standard normal distribution in this paper, that is, K is Gaussian kernel. Then Fueda and Yanagawa (2001) proposed a following procedure for estimating the embedding dimension and delay time and showed the consistency of their estimator.

(F-Y 1) Give sufficiently large integers $D(\geq d_0)$ and $T(\geq \tau_0)$, and set $L = DT$. Set $h_{d,N} = h_{d,N}(c) = c \times N^{-1/(2d+1)}$. For each $d \in \{1, 2, \ldots, D\}$ and $\tau \in \{1, 2, \ldots, T\}$, compute $CV(d, \tau)$.

(F-Y 2) For each $\tau \in \{1, 2, \ldots, T\}$, minimize $CV(d, \tau)$ with respect to $d \in \{1, 2, \ldots, D\}$, and denote the minimizer as $\hat{d}_0(\tau)$.

(F-Y 3) Find $\hat{d}_0 = \min_\tau \hat{d}_0(\tau)$ and $\hat{\tau}_0 = \mathrm{argmin}_\tau \hat{d}_0(\tau)$.

However the procedure often fails in practice. Yonemoto and Yanagawa (2001) investigated cause of it and modified the procedure to obtain good estimates from finite data. The modified procedure is as follows.

1. Give sufficiently large integers $D(\geq d_0)$ and $T(\geq \tau_0)$, and set $L = DT$. Set $h_{d,N} = h_{d,N}(c) = c \times N^{-1/(2d+1)}$.

2. For each $d \in \{1, 2, \ldots, D\}$ and $\tau \in \{1, 2, \ldots, T\}$ minimize $CV(d, \tau)|_{h_{d,N} = h_{d,N}(c)}$ with respect to $c \in C$ by the method illustrated in below and put

$$\hat{CV}(d, \tau) = \min_{c \in C} CV(d, \tau)|_{h_{d,N} = h_{d,N}(c)}.$$

3. Then select $d \in \{1, 2, \ldots, D\}$ and $\tau \in \{1, 2, \ldots, T\}$ which attain the 'minimum' value of $\{\hat{CV}(d, \tau) : d \in \{1, 2, \ldots, D\}, \tau \in \{1, 2, \ldots, T\}\}$ as estimators of the embedding dimension $\hat{d}_0$ and delay time $\hat{\tau}_0$ based on the procedure given below.

The detail of the minimization of $CV(d, \tau)|_{h_{d,N}=h_{d,N}(c)}$ with respect to $c \in C$ is given as follows.

a. Give a large real number c_{max}, and for each $d \in \{1, 2, \ldots, D\}$ and $\tau \in \{1, 2, \ldots, T\}$, compute $CV(d, \tau)|_{h_{d,N}=h_{d,N}(c)}$ and $CV(d, \tau)|_{h_{d,N}=h_{d,N}(c+0.1)}$ starting from $c = 0.1$, compare these values, and if $c < c_{max}$ and $CV(d, \tau)|_{h_{d,N}=h_{d,N}(c)} \geq CV(d, \tau)|_{h_{d,N}=h_{d,N}(c+0.1)}$ then put $c = c + 0.1$ and repeat the computation; else if, stop and decide $c = c_1$.

b. In the neighborhood of c_1, find $c_2 = \text{argmin} CV(d, \tau)|_{h_{d,N}=h_{d,N}(c)}$ in class $\mathcal{C}_1 = \{c_1 - 0.09, c_1 - 0.08, \ldots, c_1 + 0.09\}$.

c. Furthermore, in the neighborhood of c_2, compute

$$\hat{CV}(d, \tau) = \min_{c} \{CV(d, \tau)_{h_{d,N}=h_{d,N}(c)} : c \in \mathcal{C}_2 = \{c_2 - 0.009, c_2 - 0.008, \ldots, c_2 + 0.009\}\}.$$

The procedure for selecting $\hat{d}_0$ and $\hat{\tau}_0$ is as follows.

d. Put $CV^*(\tau) = \min\{\hat{CV}(d, \tau) : d \in \{1, 2, \ldots, D\}\}$ for each $\tau \in \{1, 2, \ldots, T\}$, and find for given $\varepsilon > 0$,

$$\hat{d}(\tau) = \min\{d : |\hat{CV}(d, \tau) - CV^*(\tau)| < \varepsilon\}.$$

e. Next put $CV^* = \min\{\hat{CV}(\hat{d}(\tau), \tau) | \tau \in \{1, 2, \ldots, T\}\}$, and then find

$$\hat{d}_0 = \min\{\hat{d}(\tau) : |\hat{CV}(\hat{d}(\tau), \tau) - CV^*| < \varepsilon\}$$

and

$$\hat{\tau}_0 = \text{argmin}\{\hat{d}(\tau) : |\hat{CV}(\hat{d}(\tau), \tau) - CV^*| < \varepsilon\}.$$

The smallest $\hat{\tau}_0$ is employed when $\hat{\tau}_0$ is not unique. The selection of ε is important in the above procedure. We have no answer for it's optimal selection. Yonemoto and Yanagawa (2001) examined the performance of the procedure when ε is $CV^*/10$ and $CV^*/20$ by simulation and recommended to use $\varepsilon = CV^*/20$.

2.4 Estimating the skeleton.

Suppose that $X_1, X_2, \ldots, X_N$ are observed. Then we estimate $F(\mathbf{x})$ by

$$\hat{F}_N(\mathbf{x}) = \frac{\sum_{i=(\hat{d}_0-1)\hat{\tau}_0+1}^{N-\hat{\tau}_0} K_{\hat{d}_0,h}(\mathbf{x} - \mathbf{X}_i^{(\hat{d}_0,\hat{\tau}_0)}) X_{i+\hat{\tau}_0}}{\sum_{i=(\hat{d}_0-1)\hat{\tau}_0+1}^{N-\hat{\tau}_0} K_{\hat{d}_0,h}(\mathbf{x} - \mathbf{X}_i^{(\hat{d}_0,\hat{\tau}_0)})}.$$

For $\mathbf{x} = {}^t(x_1, x_2, \ldots, x_{\hat{d}_0})$, put

$$\hat{\mathbf{F}}_N(\mathbf{x}) = {}^{\mathbf{t}}(\hat{F}_N(\mathbf{x}), x_1, \ldots, x_{\hat{d}_0-1}).$$

Giving an appropriate initial vector $\hat{\mathbf{Y}}^0_{N,0}$, we generate $\{\hat{\mathbf{Y}}_{N,t}\}_{t=1,2,\ldots,M}$ by

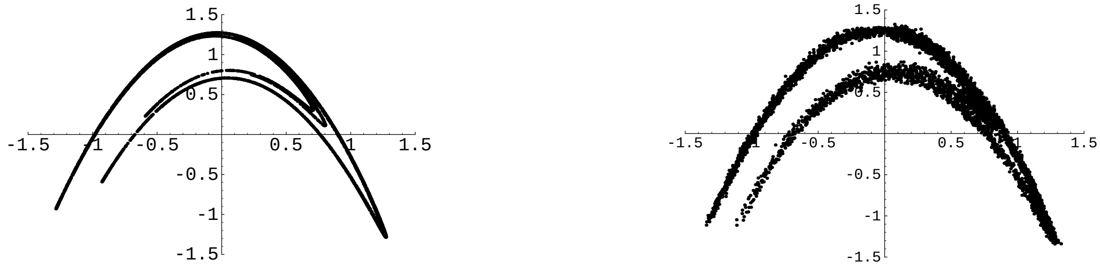
$$\hat{\mathbf{Y}}^{(\hat{d}_0,1)}_{N,t} = \hat{\mathbf{F}}_N(\hat{\mathbf{Y}}^{(\hat{d}_0,1)}_{N,t-1}), \quad t = 1, 2, \ldots, M, \tag{8}$$

where

$$\hat{\mathbf{Y}}^{(\hat{d}_0,1)}_{N,t} = {}^t(\hat{Y}_{N,t}, \hat{Y}_{N,t-1}, \ldots, \hat{Y}_{N,t-\hat{d}_0+1}).$$

Note that $K_{\hat{d}_0,h}(\mathbf{x})$ is defined in Section 2.3 with $d = \hat{d}_0$.

Example 1: Figure 1 (a) shows the plots of (X_t, X_{t+1}) where $\{X_t\}$ is generated from $X_t = 1 - 1.4X_{t-1}^2 + 0.3X_{t-2}$, and Figure 1 (b) shows the plots of (X_t, X_{t+1}) for $\{X_t\}$ generated from $X_t = 1 - 1.4X_{t-1}^2 + 0.3X_{t-2} + \varepsilon_t$ with $\varepsilon_t \sim N(0, 0.04^2)$.



(a) (b)

Figure 1: Plots of (X_t, X_{t+1})

Figure 2 shows the plots of $(\hat{Y}_{N,t}, \hat{Y}_{N,t+1})$ when $N = 5000$. The figure shows that the procedure reproduced the skeleton fairly well.

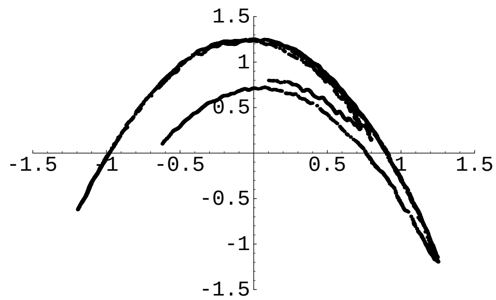


Figure 2: Plots of $(\hat{Y}_{5000,t}, \hat{Y}_{5000,t+1})$

2.5 Estimating the Lyapunov exponent.

We propose an estimator of the Lyapunov exponent as follows.

$$\hat{\lambda}_{N,M} = \frac{1}{M} \log \| D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}^{(\hat{d}_0,1)}_{N,M}) D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}^{(\hat{d}_0,1)}_{N,M-1}) \cdots D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}^{(\hat{d}_0,1)}_{N,1}) \|,$$

where $D\hat{\mathbf{F}}_N$ is the matrix of partial derivatives of $\hat{\mathbf{F}}_N$.

3 Consistency of the proposed estimator.

In this section, we prove the consistency of the estimator. The consistency of the estimators of the embedding dimension and delay time is proved in Fueda and Yanagawa (2001). Hence in this section we suppose that the embedding dimension and delay time are known.

Theorem 3.1. *Under the assumptions that are given in the following subsection, it follows that for any $\varepsilon > 0$,*

$$\lim_{M\to\infty}\lim_{N\to\infty} P(|\hat{\lambda}_{N,M} - \lambda| > \varepsilon) = 0.$$

3.1 Assumptions.

We assume the following assumptions for Theorem 3.1.

Assumption 3.1. *There exists a compact set $G \subset \mathbf{R}^{d_0}$ such that*

for any $\mathbf{x} \in G$,

$$\mathbf{F}(\mathbf{x}) + \mathbf{e} \in G \ \textit{a.s.}.$$

where $\mathbf{e} = {}^t(\varepsilon, 0, \ldots, 0) \in \mathbf{R}^{d_0}$ and ε is identically distributed as $\{\varepsilon_t\}$.

Remark 1. *Many observed chaotic data scattered around in compact spaces and satisfy this assumption. Simulation in Example 1 and Section 4 uses a Gaussian distribution for $\mathbf{e}$, then the generated data often diverge. If this is the case, we modify the generation of data to satisfy Assumption 3.1; see the details in Section 4.*

Assumption 3.2. *F is C^1 class on G.*

Assumption 3.3. *For $\mathbf{X}_t^{(d_0,\tau_0)} = {}^t(X_t, X_{t-\tau_0}, \ldots, X_{t-(d_0-1)\tau_0})$,*

we assume $\mathbf{X}_t^{(d_0,\tau_0)} \in G$ a.s. for any $t \in \{(d_0-1)\tau_0+1, (d_0-1)\tau_0+2, \ldots, d_0\tau_0\}$.

Lemma 3.1. *Under Assumption 3.1 and 3.3, it follows that*

$\mathbf{X}_t^{(d_0,\tau_0)} \in G$ a.s. for any integer $t \in \{(d_0-1)\tau_0+1, (d_0-1)\tau_0+2, \ldots, N\}$.

Proof. We prove the lemma by mathematical induction. From Assumption 3.3, $\mathbf{X}_t^{(d_0,\tau_0)} \in G$ a.s. for any $t \in \{(d_0-1)\tau_0+1, (d_0-1)\tau_0+2, \ldots, d_0\tau_0\}$. We assume that $\mathbf{X}_{n-\tau_0}^{(d_0,\tau_0)} \in G$ a.s.. From the assumption and Assumption 3.1, $\mathbf{X}_n^{(d_0,\tau_0)} = \mathbf{F}(\mathbf{X}_{n-\tau_0}^{(d_0,\tau_0)}) + \mathbf{e}_n \in G$ a.s.. Hence the assertion follows. $\square$

Assumption 3.4. *There exists* $\gamma > 0$ *such that*

$$P(\mathbf{X}_t^{(d_0,\tau_0)} \in B) \geq \gamma\mu(B)$$

for any $B \in \mathcal{B}_G$ *and* $t \in \{(d_0-1)\tau_0+1, (d_0-1)\tau_0+2, \ldots, N\}$*, where* μ *is the Lebesgue measure and* $\mathcal{B}_G$ *is the Borel sigma algebra on* G*.*

Assumption 3.5. *There exists* $\Gamma < \infty$ *such that*

$$P(\mathbf{X}_t^{(d_0,\tau_0)} \in B) \leq \Gamma\mu(B)$$

for any $B \in \mathcal{B}_G$ *and* $t \in \{(d_0-1)\tau_0+1, (d_0-1)\tau_0+2, \ldots, N\}$*.*

Assumption 3.6. $\{X_t\}$ *is* ϕ*-mixing, that is, for any positive integers* k*,* n *and* N *such that* $k \geq (d_0-1)\tau_0+1, n \geq 1$*,* $k+n \leq N-\tau_0$ *and* $N > d_0\tau_0+1$*, there exists* $\phi_n \downarrow 0$ *such that*

$$|P(A \cap B) - P(A)P(B)| \leq \phi_n P(A)$$

for any $A \in \mathcal{F}_{(d_0-1)\tau_0+1}^{k}$ *and* $B \in \mathcal{F}_{k+n}^{N-\tau_0}$*, where* $\mathcal{F}_s^t$ *is* σ*-algebra generated by* $(\mathbf{X}_s^{(d_0,\tau_0)}, X_{s+\tau_0}), (\mathbf{X}_{s+1}^{(d_0,\tau_0)}, X_{s+1+\tau_0}), \ldots, (\mathbf{X}_t^{(d_0,\tau_0)}, X_{t+\tau_0})$*.*

Assumption 3.7. *For some monotone increasing series* $\{m_N \in \mathbf{Z}\}$ *such that*

$$1 \leq m_N \leq N - d_0\tau_0 - 1 \text{ for any integer } N > d_0\tau_0+1,$$

$h_{d_0,N}$ *given in (7) and* ϕ_n *satisfies the following conditions :*

$$\text{there exists } A > 0 \text{ such that } \frac{N\phi_{m_N}}{m_N} < A \text{ for any integer} N > d_0\tau_0+1,$$

and furthermore it follows that

$$\frac{N h_{d_0,N}^{d_0}}{m_N \log N} \to \infty \text{ as } N \to \infty.$$

Remark 2. *Assumption 3.4 - 3.7 are mathematically involved, but needed for consistency of the kernel estimator that will be shown numerically work well below.*

For $N \in \mathbf{N}$ such that $N \geq d_0\tau_0 + 1$, $t \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d_0\}$, let

$$\mathbf{Z}_{N,t,i} = (X_1, X_2, \ldots, X_N, \hat{Y}_{N,t}, \hat{Y}_{N,t-1}, \ldots, \hat{Y}_{N,t-i+2}, \hat{Y}_{N,t-i}, \ldots, \hat{Y}_{N,t-d_0+1}),$$

$f(y|\mathbf{Z}_{N,t,i})$ be a conditional probability density function of $\hat{Y}_{N,t-i+1}$ given $\mathbf{Z}_{N,t,i}$, and $G_{N,t,i}$ a probability measure of $\mathbf{Z}_{N,t,i}$. For integers $N \geq d_0\tau_0+1$, $t \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d_0\}$, put

$$\mathbf{W}_{N,t,i}(x) = (\hat{Y}_{N,t}, \hat{Y}_{N,t-1}, \ldots, \hat{Y}_{N,t-i+2}, x, \hat{Y}_{N,t-i}, \ldots, \hat{Y}_{N,t-d_0+1}).$$

Let $I_G(\mathbf{x})$ be a defining function of G, that is,

$$I_G(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in G \\ 0 & \text{if } \mathbf{x} \notin G \end{cases}.$$

We use the following lemmas for the proof of Theorem 3.5.

Lemma 3.2. *For given $\mathbf{Z}_{N,t,i}$, if*

$$\left\{ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) > 0 \right\} \neq \emptyset,$$

then under Assumption 3.2, there exist integer $m^{(+)}(\mathbf{Z}_{N,t,i})$ and finite collection $\{g_k^{(+)}(\mathbf{Z}_{N,t,i})\}_{k=1,2,\ldots,m^{(+)}(\mathbf{Z}_{N,t,i})}$ such that each $g_k^{(+)}(\mathbf{Z}_{N,t,i})$ is a non-empty, connected and relative open set in $\{x|I_G(\mathbf{W}_{N,t,i}(x)) = 1\}$, and satisfying

$$\left\{ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) > 0 \right\} = \bigcup_{k=1}^{m^{(+)}(\mathbf{Z}_{N,t,i})} g_k^{(+)}(\mathbf{Z}_{N,t,i}),$$

and $g_j^{(+)}(\mathbf{Z}_{N,t,i}) \cap g_k^{(+)}(\mathbf{Z}_{N,t,i}) = \emptyset$ for $j \neq k$.

Proof. By Assumption 3.2 and the definition of $\hat{F}_N$, $\frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))$ is continuous on G. Hence $\left\{ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) > 0 \right\}$ is a relative open set in

$\{\ x\ | I_G(\mathbf{W}_{N,t,i}(x)) = 1\ \}$. Since G is a compact set, $\{\ x\ | I_G(\mathbf{W}_{N,t,i}(x)) = 1\ \}$ is a compact set, too. Let $g_k^+(\mathbf{Z}_{N,t,i})$ $(k = 1, 2, \ldots, m^{(+)}(\mathbf{Z}_{N,t,i}))$ be non-empty, connected and relative open sets in $\{x | I_G(\mathbf{W}_{N,t,i}(x)) = 1\}$, and satisfying

$$\left\{\ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) > 0\ \right\} = \bigcup_{k=1}^{m^{(+)}(\mathbf{Z}_{N,t,i})} g_k^{(+)}(\mathbf{Z}_{N,t,i}),$$

and $g_j^{(+)}(\mathbf{Z}_{N,t,i}) \cap g_k^{(+)}(\mathbf{Z}_{N,t,i}) = \emptyset$ for $j \neq k$. If $m^{(+)}(\mathbf{Z}_{N,t,i}) = \infty$, then it contradicts compactness of $\{\ x\ | I_G(\mathbf{W}_{N,t,i}(x)) = 1\ \}$. Hence the assertion follows. □

Lemma 3.3. *For given* $\mathbf{Z}_{N,t,i}$*, if*

$$\left\{\ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) < 0\ \right\} \neq \emptyset,$$

then under Assumption 3.2, there exist integer $m^{(-)}(\mathbf{Z}_{N,t,i})$ *and finite collection* $\{g_k^{(-)}(\mathbf{Z}_{N,t,i})\}_{k=1,2,\ldots,m^{(-)}(\mathbf{Z}_{N,t,i})}$ *such that each* $g_k^{(-)}(\mathbf{Z}_{N,t,i})$ *is a non-empty, connected and relative open set in* $\{x | I_G(\mathbf{W}_{N,t,i}(x)) = 1\}$*, and satisfying*

$$\left\{\ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) < 0\ \right\} = \bigcup_{k=1}^{m^{(-)}(\mathbf{Z}_{N,t,i})} g_k^{(-)}(\mathbf{Z}_{N,t,i})$$

and $g_j^{(-)}(\mathbf{Z}_{N,t,i}) \cap g_k^{(-)}(\mathbf{Z}_{N,t,i}) = \emptyset$ *for* $j \neq k$.

Proof. Proof is given similary as that of Lemma 3.2. □

If the condition of the lemmas is violated, we put $m^{(+)}(\mathbf{Z}_{N,t,i}) = 0$ and $m^{(-)}(\mathbf{Z}_{N,t,i}) = 0$, more precisely

$$m^{(+)}(\mathbf{Z}_{N,t,i}) = 0 \text{ if } \left\{\ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) > 0\ \right\} = \emptyset,$$

and

$$m^{(-)}(\mathbf{Z}_{N,t,i}) = 0 \text{ if } \left\{\ x \mid \frac{\partial}{\partial x_i}(F(\mathbf{x}) - \hat{F}_N(\mathbf{x}))|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x)) < 0\ \right\} = \emptyset.$$

We assume the following assumption.

Assumption 3.8. *For any $t \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d_0\}$, there exists $J_{t,i} < \infty$ such that*

for any integer $N \geq d_0\tau_0 + 1$,

$$f(y|\mathbf{Z}_{N,t,i}) < J_{t,i} \text{ a.e. } G_{N,t,i} \text{ except for at most countable number of } y \in \mathbf{R}.$$

For the notation of Lemma 3.2 and Lemma 3.3, we assume following assumption.

Assumption 3.9. *For any $t \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d_0\}$, there exist $m_{t,i}^{(+)} < \infty$ and $m_{t,i}^{(-)} < \infty$ such that*

$$m^{(+)}(\mathbf{Z}_{N,t,i}) \leq m_{t,i}^{(+)} \text{ a.e. } G_{N,t,i} \text{ and } m^{(-)}(\mathbf{Z}_{N,t,i}) \leq m_{t,i}^{(-)} \text{ a.e. } G_{N,t,i}$$

for any $N \in \mathbf{N}$ such that $N \geq d_0\tau_0 + 1$.

Remark 3. *Assumption 3.8 - 3.9 are used for the consistency of the Lyapunov exponent.*

3.2 Theorems.

Theorem 3.2 (Collomb (1984)). *Under Assumption 3.1 - 3.7, it follows that for any $\epsilon > 0$,*

$$P\left(\sup_{\mathbf{x}\in G} |F(\mathbf{x}) - \hat{F}_N(\mathbf{x})| > \epsilon\right) \to 0 \text{ as } N \to \infty.$$

Select an initial vector $\mathbf{Y}_0^{(d_0,1)}$ randomly from G with uniform probability, and set $\hat{\mathbf{Y}}_{N,0}^{(d_0,1)} = \mathbf{Y}_0^{(d_0,1)}$. Then the following theorems hold.

Theorem 3.3. *Under Assumption 3.1 - 3.7, it follows that for any $t \in \{1, 2, \ldots, M\}$,*

$$\hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \to \mathbf{Y}_t^{(d_0,1)} \text{ in probability as } N \to \infty.$$

Proof. We prove the theorem by mathematical induction. From Theorem 3.2, for any $\epsilon > 0$,

$$
\begin{aligned}
P(|Y_1 - \hat{Y}_{N,1}| > \epsilon) &= P(|F(\mathbf{Y}_0^{(d_0,1)}) - \hat{F}_N(\hat{\mathbf{Y}}_{N,0}^{(d_0,1)})| > \epsilon) \\
&= P(|F(\mathbf{Y}_0^{(d_0,1)}) - \hat{F}_N(\mathbf{Y}_0^{(d_0,1)})| > \epsilon) \\
&\leq P\left(\sup_{\mathbf{x}\in G} |F(\mathbf{x}) - \hat{F}_N(\mathbf{x})| > \epsilon\right) \\
&\rightarrow 0 \text{ as } N \rightarrow \infty.
\end{aligned}
$$

Hence $P(\|\mathbf{Y}_1^{(d_0,1)} - \hat{\mathbf{Y}}_{N,1}^{(d_0,1)}\| > \epsilon) \leq P\left(\sum_{i=0}^{d_0-1} |Y_{1-i} - \hat{Y}_{N,1-i}| > \epsilon\right) \rightarrow 0$ as $N \rightarrow \infty$, that is, $\hat{\mathbf{Y}}_{N,1}^{(d_0,1)} \rightarrow \mathbf{Y}_1^{(d_0,1)}$ in probability as $N \rightarrow \infty$. We assume that $\mathbf{Y}_{N,n-1}^{(d_0,1)} \rightarrow \mathbf{Y}_{n-1}^{(d_0,1)}$ in probability as $N \rightarrow \infty$. From Assumption 3.1 and $\mathbf{Y}_0^{(d_0,1)} \in G$, $\mathbf{Y}_{n-1}^{(d_0,1)} \in G^\circ$. Hence $P(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)} \notin G) \rightarrow 0$ as $N \rightarrow \infty$. Since F is continuous on the compact set G, $F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)}) \rightarrow F(\mathbf{Y}_{n-1}^{(d_0,1)})$ in probability as $N \rightarrow \infty$. Therefore for any $\epsilon > 0$,

$$
\begin{aligned}
& P(|Y_n - \hat{Y}_{N,n}| > \epsilon) \\
= \;& P(|F(\mathbf{Y}_{n-1}^{(d_0,1)}) - \hat{F}_N(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \epsilon) \\
\leq \;& P(|F(\mathbf{Y}_{n-1}^{(d_0,1)}) - F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| + |F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)}) - \hat{F}_N(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \epsilon) \\
\leq \;& P(|F(\mathbf{Y}_{n-1}^{(d_0,1)}) - F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \tfrac{\epsilon}{2}) + P(|F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)}) - \hat{F}_N(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \tfrac{\epsilon}{2}) \\
\leq \;& P(|F(\mathbf{Y}_{n-1}^{(d_0,1)}) - F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \tfrac{\epsilon}{2}) \\
& + P(|F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)}) - \hat{F}_N(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \tfrac{\epsilon}{2}, \hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)} \in G) \\
& + P(|F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)}) - \hat{F}_N(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \tfrac{\epsilon}{2}, \hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)} \notin G)
\end{aligned}
$$

$$\begin{aligned} &\leq P(|F(\mathbf{Y}_{n-1}^{(d_0,1)}) - F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \tfrac{\epsilon}{2}) \\ &\quad + P(|F(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)}) - \hat{F}_N(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)})| > \tfrac{\epsilon}{2}, \hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)} \in G) \\ &\quad + P(\hat{\mathbf{Y}}_{N,n-1}^{(d_0,1)} \notin G) \\ &\to 0 \text{ as } N \to \infty. \end{aligned}$$

Hence $P(\|\mathbf{Y}_n^{(d_0,1)} - \hat{\mathbf{Y}}_{N,n}^{(d_0,1)}\| > \epsilon) \leq P\left(\sum_{i=0}^{d_0-1} |Y_{2-i} - \hat{Y}_{N,2-i}| > \epsilon\right) \to 0$ as $N \to \infty$.

Hence $\mathbf{Y}_n^{(d_0,1)} \to \hat{\mathbf{Y}}_{N,n}^{(d_0,1)}$ in probability as $N \to \infty$. □

Theorem 3.4. *Under Assumption 3.1-3.7, it follows that*

$$E\left[\sup_{\mathbf{x}\in G} |F(\mathbf{x}) - \hat{F}_N(\mathbf{x})|\right] \to 0 \text{ as } N \to \infty.$$

Proof. By Lemma 3.1, there exists $K_1 < \infty$ such that $|X_t| < K_1$ a.s. for any $t \in \{1, 2, \ldots, N\}$.

Since the Gaussian kernel is used, it follows that for any integer $N \geq d_0\tau_0 + 1$,

$$\sup_{\mathbf{x}\in G} |\hat{F}_N(\mathbf{x})| < K_1 \text{ a.s..}$$

Since F is continuous on the compact set G by Assumption 3.2, there exists $K_2 < \infty$ such that

$$\sup_{\mathbf{x}\in G} |F(\mathbf{x})| < K_2.$$

Hence for any integer $N \geq d_0\tau_0 + 1$,

$$E\left[\sup_{\mathbf{x}\in G} |F(\mathbf{x}) - \hat{F}_N(\mathbf{x})|\right] \leq K_1 + K_2 < \infty.$$

From this inequality, for any integer $N \geq d_0\tau_0 + 1$ and any $\epsilon > 0$, there exists $c(\epsilon)$ such that $\epsilon < c(\epsilon)$ and

$$E\left[\sup_{\mathbf{x}\in G} \left|F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right| \cdot I_{\left\{\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|>c(\epsilon)\right\}}\right] < \epsilon.$$

Therefore

$$
\begin{aligned}
& E\left[\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|\right] \\
= \quad & E\left[\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|\cdot I_{\left\{\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|>c(\epsilon)\right\}}\right] \\
& +E\left[\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|\cdot I_{\left\{\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|\leq\epsilon\right\}}\right] \\
& +E\left[\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|\cdot I_{\left\{\epsilon<\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|\leq c(\epsilon)\right\}}\right] \\
\leq \quad & \epsilon+\epsilon+c(\epsilon)\cdot P\left(\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|>\epsilon\right) \\
\to \quad & 2\epsilon \text{ as } N\to\infty.
\end{aligned}
$$

Consequently $E\left[\sup_{\mathbf{x}\in G}\left|F(\mathbf{x})-\hat{F}_N(\mathbf{x})\right|\right]\to 0$ as $N\to\infty$. □

Theorem 3.5. *Under Assumption 3.1-3.9, it follows that*

$$\lambda_M-\hat{\lambda}_{N,M}=o_p(1) \text{ as } N\to\infty,$$

where

$$\lambda_M=\frac{1}{M}\log\|D\mathbf{F}(\mathbf{Y}_M^{(d_0,1)})D\mathbf{F}(\mathbf{Y}_{M-1}^{(d_0,1)})\cdots D\mathbf{F}(\mathbf{Y}_1^{(d_0,1)})\|.$$

Proof.

$$
\begin{aligned}
& \left|\lambda_M-\hat{\lambda}_{N,M}\right| \\
= \quad & \Big|\tfrac{1}{M}\log\|D\mathbf{F}(\mathbf{Y}_M^{(d_0,1)})D\mathbf{F}(\mathbf{Y}_{M-1}^{(d_0,1)})\cdots D\mathbf{F}(\mathbf{Y}_1^{(d_0,1)})\| \\
& -\tfrac{1}{M}\log\|D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,M}^{(d_0,1)})D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,M-1}^{(d_0,1)})\cdots D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,1}^{(d_0,1)})\|\Big| \\
\leq \quad & \tfrac{1}{M}\Big(\Big|\log\|D\mathbf{F}(\mathbf{Y}_M^{(d_0,1)})D\mathbf{F}(\mathbf{Y}_{M-1}^{(d_0,1)})\cdots D\mathbf{F}(\mathbf{Y}_1^{(d_0,1)})\| \\
& -\log\|D\mathbf{F}(\hat{\mathbf{Y}}_{N,M}^{(d_0,1)})D\mathbf{F}(\hat{\mathbf{Y}}_{N,M-1}^{(d_0,1)})\cdots D\mathbf{F}(\hat{\mathbf{Y}}_{N,1}^{(d_0,1)})\|\Big|\Big) \\
+ \quad & \tfrac{1}{M}\Big(\Big|\log\|D\mathbf{F}(\hat{\mathbf{Y}}_{N,M}^{(d_0,1)})D\mathbf{F}(\hat{\mathbf{Y}}_{N,M-1}^{(d_0,1)})\cdots D\mathbf{F}(\hat{\mathbf{Y}}_{N,1}^{(d_0,1)})\| \\
& -\log\|D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,M}^{(d_0,1)})D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,M-1}^{(d_0,1)})\cdots D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,1}^{(d_0,1)})\|\Big|\Big).
\end{aligned}
$$

First term in the right hand side of the inequality converges to 0 in probability as $N \to \infty$, because $D\mathbf{F}$ is continuous on G by Assumption 3.2 and $\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}$ converges to $\mathbf{Y}_t^{(d_0,1)}$ in probability as $N \to \infty$ for any $t \in \{1, 2, \ldots, M\}$ by Theorem 3.3. We show that the second term converges to 0 in probability as $N \to \infty$. Note that it is equivalent to show that for any $t \in \{1, 2, \ldots, M\}$,

$$\|D\mathbf{F}(\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}) - D\hat{\mathbf{F}}_{\mathbf{N}}(\hat{\mathbf{Y}}_{N,t}^{(d_0,1)})\| \to 0$$

in probability as $N \to \infty$. Since for $\mathbf{x} = {}^t(x_1, x_2, \ldots, x_{d_0})$, $(1, i)$ elements of $D\mathbf{F}(\hat{\mathbf{Y}}_{N,t}^{(d_0,1)})$ and $D\hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,t}^{(d_0,1)})$ are given respectively by

$$\frac{\partial F(\mathbf{x})}{\partial x_i}\Big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} \text{ and } \frac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}\Big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}},$$

it may be proved if we show that for any $t \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d_0\}$,

$$\left|\frac{\partial F(\mathbf{x})}{\partial x_i}\Big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} - \frac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}\Big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| \to 0 \text{ in probability as } N \to \infty.$$

For any $t \in \{1, 2, \ldots, M\}$, $i \in \{1, 2, \ldots, d_0\}$ and $\varepsilon > 0$,

$$\begin{aligned}
& P\left(\left|\tfrac{\partial F(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} - \tfrac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| > \varepsilon\right) \\
= \; & P\left(\left|\tfrac{\partial F(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} - \tfrac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| > \varepsilon, \hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \in G\right) \\
& + P\left(\left|\tfrac{\partial F(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} - \tfrac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| > \varepsilon, \hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \notin G\right) \\
\leq \; & P\left(\left|\tfrac{\partial F(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} - \tfrac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}\big|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| > \varepsilon, \hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \in G\right) \\
& + P\left(\hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \notin G\right).
\end{aligned}$$

By Theorem 3.3 and Assumption 3.3, it follows that

$$P\left(\hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \notin G\right) \to 0 \text{ as } N \to \infty.$$

We next consider the first term of the right hand side of the inequality.

By Theorem 3.4, Lemma 3.2, Lemmma 3.3, Assumption 3.8 and Assumption 3.9, for any $N \in \mathbf{N}$ such that $N \geq d_0\tau_0 + 1$, $t \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d_0\}$, we have

$$
\begin{aligned}
& E\left[\left|\frac{\partial}{\partial x_i}\left(F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right)|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| \cdot I_{\{\hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \in G\}}\right] \\
= \; & E\left[E\left[\left|\frac{\partial}{\partial x_i}\left(F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right)|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| \cdot I_{\{\hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \in G\}} | \mathbf{Z}_{N,t,i}\right]\right] \\
= \; & E\left[\int \left|\frac{\partial}{\partial x_i}\left(F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right)|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)}\right| \cdot I_G(\mathbf{W}_{N,t,i}(x)) f(x|\mathbf{Z}_{N,t,i}) dx\right] \\
\leq \; & J_{t,i} \cdot E\left[\int \left|\frac{\partial}{\partial x_i}\left(F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right)|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} \cdot I_G(\mathbf{W}_{N,t,i}(x))\right| dx\right] \\
= \; & J_{t,i} \cdot E\left[\sum_{k=1}^{m^{(+)}(\mathbf{Z}_{N,t,i})} \int_{g_k^{(+)}(\mathbf{Z}_{N,t,i})} \frac{\partial}{\partial x_i}\left(F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right)|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} dx\right. \\
& \qquad \left. - \sum_{k=1}^{m^{(-)}(\mathbf{Z}_{N,t,i})} \int_{g_k^{(-)}(\mathbf{Z}_{N,t,i})} \frac{\partial}{\partial x_i}\left(F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right)|_{\mathbf{x}=\mathbf{W}_{N,t,i}(x)} dx\right] \\
= \; & J_{t,i} \cdot E\left[\sum_{k=1}^{m^{(+)}(\mathbf{Z}_{N,t,i})}\left(\sup_{x \in g_k^{(+)}(\mathbf{Z}_{N,t,i})}\left\{F(\mathbf{W}_{N,t,i}(x)) - \hat{F}_N(\mathbf{W}_{N,t,i}(x))\right\}\right.\right. \\
& \qquad \left. - \inf_{x \in g_k^{(+)}(\mathbf{Z}_{N,t,i})}\left\{F(\mathbf{W}_{N,t,i}(x)) - \hat{F}_N(\mathbf{W}_{N,t,i}(x))\right\}\right) \\
& \qquad - \sum_{k=1}^{m^{(-)}(\mathbf{Z}_{N,t,i})}\left(\inf_{x \in g_k^{(-)}(\mathbf{Z}_{N,t,i})}\left\{F(\mathbf{W}_{N,t,i}(x)) - \hat{F}_N(\mathbf{W}_{N,t,i}(x))\right\}\right. \\
& \qquad \left.\left. - \sup_{x \in g_k^{(-)}(\mathbf{Z}_{N,t,i})}\left\{F(\mathbf{W}_{N,t,i}(x)) - \hat{F}_N(\mathbf{W}_{N,t,i}(x))\right\}\right)\right] \\
\leq \; & 2J_{t,i} \cdot E\left[(m^{(+)}(\mathbf{Z}_{N,t,i}) + m^{(-)}(\mathbf{Z}_{N,t,i})) \cdot \sup_{\mathbf{x} \in G}\left|F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right|\right] \\
\leq \; & 2J_{t,i} \cdot (m_{t,i}^{(+)} + m_{t,i}^{(-)}) \cdot E\left[\sup_{\mathbf{x} \in G}\left|F(\mathbf{x}) - \hat{F}_N(\mathbf{x})\right|\right] \\
\to \; & 0 \text{ as } N \to \infty.
\end{aligned}
$$

Hence for any $t \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d_0\}$,

$$
E\left[\left|\frac{\partial F(\mathbf{x})}{\partial x_i}|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} - \frac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| \cdot I_{\{\hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \in G\}}\right] \to 0 \text{ as } N \to \infty.
$$

Thus for any $t \in \{1, 2, \ldots, M\}$, $i \in \{1, 2, \ldots, d_0\}$ and $\varepsilon > 0$, it follows that

$$P\left(\left|\frac{\partial F(\mathbf{x})}{\partial x_i}|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}} - \frac{\partial \hat{F}_N(\mathbf{x})}{\partial x_i}|_{\mathbf{x}=\hat{\mathbf{Y}}_{N,t}^{(d_0,1)}}\right| > \varepsilon, \hat{\mathbf{Y}}_{N,t}^{(d_0,1)} \in G\right) \to 0 \text{ as } N \to \infty.$$

Therefore $|\lambda_M - \hat{\lambda}_{N,M}| = o_p(1)$ as $N \to \infty$. □

Proof. Proof of Theorem 3.1

For any $\varepsilon > 0$,

$$\begin{aligned} & P(|\lambda - \hat{\lambda}_{N,M}| > \varepsilon) \\ \leq \; & P(|\lambda - \lambda_M| + |\lambda_M - \hat{\lambda}_{N,M}| > \varepsilon) \\ \leq \; & P\left(|\lambda - \lambda_M| > \tfrac{\varepsilon}{2}\right) + P\left(|\lambda_M - \hat{\lambda}_{N,M}| > \tfrac{\varepsilon}{2}\right). \end{aligned}$$

Hence from the definition of the Lyapunov exponent and Theorem 3.5, for any $\varepsilon > 0$,

$$\lim_{M\to\infty} \lim_{N\to\infty} P(|\lambda - \hat{\lambda}_{N,M}| > \varepsilon) = 0.$$

□

4 Numerical evaluation.

In this section, we evaluate numerically the behavior of the proposed procedure using stochastic Henon map

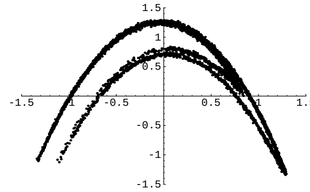
$$X_t = 1 - 1.4X_{t-1}^2 + 0.3X_{t-2} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \sigma^2)$.

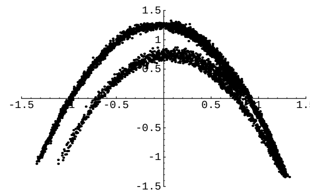
Because dynamic noise follows normal distribution, data may diverge. To stay in the neighborhood of attractor of the skeleton, we generate data as follows. For each $\sigma \in \{0.02, 0.04, 0.06, 0.08, 0.1, 0.3\}$, the data of size $N \in \{1000, 3000, 5000\}$ is generated by giving an initial value (X_2, X_1) which is selected randomly from quadrilateral $ABCD$, where $A = (-1.33, 1.4)$,

$B = (1.32, 0.443)$, $C = (1.245, -0.466)$, $D = (-1.06, -1.666)$. This quadrilateral is introduced in Henon (1976) as a region where X_t does not diverge as $t \to \infty$ when $\sigma^2 = 0$. When $\sigma^2 \neq 0$, divergence could occur even if the initial values are selected from the quadrilateral. To avoid it we discarded $X_3, X_4, \ldots, X_i$ if $|X_i - X_{i-1}| > 3$ for some $i \leq 12$, and generated new $X_3, X_4, \ldots$ using the same initial values. Also we discarded $X_{i-9}, X_{i-8}, \ldots, X_i$ if $|X_i - X_{i-1}| > 3$ for some $i \geq 13$ and generated new $X_{i-9}, X_{i-8}, \ldots$ using (X_{i-10}, X_{i-11}), but keeping $X_1, X_2, \ldots, X_{i-10}$. $N+1000$ points are generated by the above procedure, first 1000 points are discarded, and then remaining points are used as the data of size N. At first, we estimate the embedding dimension and delay time from $\{X_t\}$ by the method described in Section 2.3. For all combination of σ and N, the true values of d_0 and τ_0 are obtained, that is, $\hat{d}_0 = 2$ and $\hat{\tau}_0 = 1$. In estimating the skeleton for each combination of σ and N, the same value of $h_{\hat{d}_0,N}$ is employed as that used for estimating the embedding dimension and delay time. For the initial vector $(Y^0_{N,0}, Y^0_{N,-1})$, we used the one that is randomly selected from data $\{\mathbf{X}_2^{(2,1)}, \mathbf{X}_3^{(2,1)}, \ldots, \mathbf{X}_N^{(2,1)}\}$.

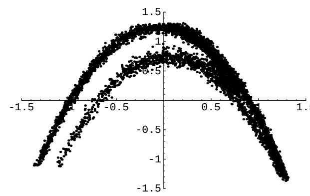
Figure 3 shows plots of (X_t, X_{t+1}) $(t = 1, 2, \ldots, 5000)$, whereas Figure 4 demonstrates plots of $(\hat{Y}_{N,t}, \hat{Y}_{N,t+1})$ $(t = 1, 2, \ldots, 30000)$. Figure 4 shows that when $N = 3000$, $\sigma = 0.06$ or $\sigma = 0.3$, the attractors are limit cycle and the corresponding figures in Figure 3 are not reproduced. Since our method of estimating the Lyapunov exponent is based on the samples from the attractors, it is impossible to estimate the Lyapunov exponent in such cases and we exclude these cases from further consideration.



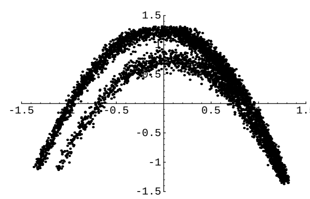
$\sigma = 0.02$



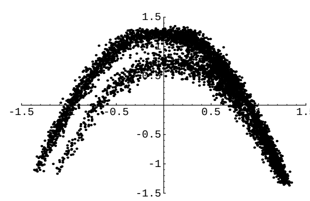
$\sigma = 0.04$



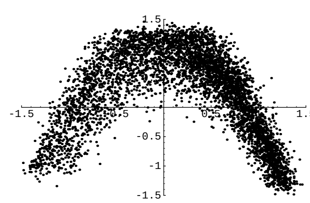
$\sigma = 0.06$



$\sigma = 0.08$



$\sigma = 0.1$



$\sigma = 0.3$

Figure 3: Plots of (X_t, X_{t+1}).

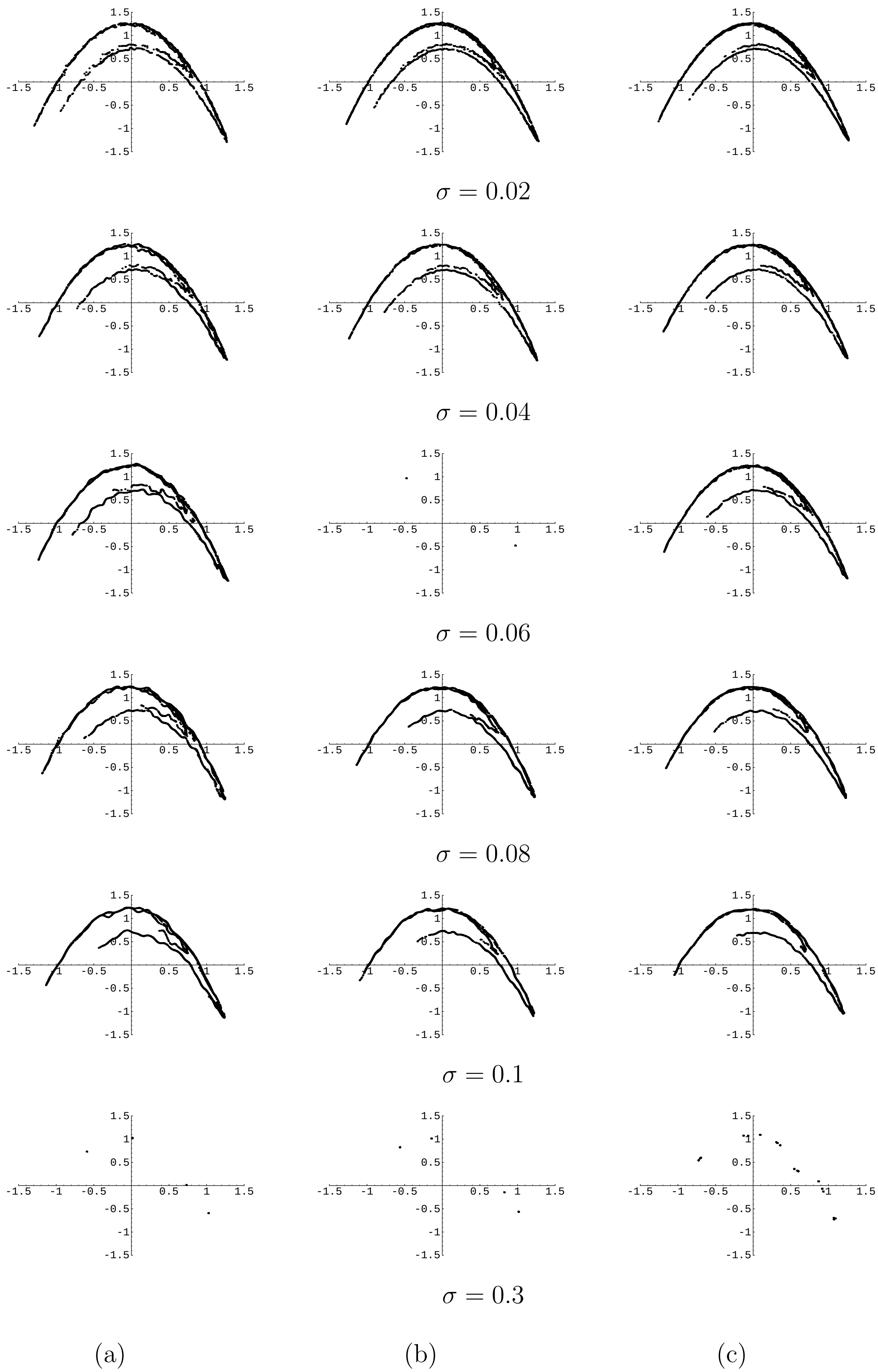


Figure 4: Column (a), (b) and (c) show plots of $(\hat{Y}_{N,t}, \hat{Y}_{N,t+1})$ with N=1000, 3000 and 5000, respectively.

Now the Lyapunov exponents is estimated using the data of size $M = 1000$ sampled from each attractor. Table 1 (a), (b) and (c) summarize the results. The tables show that the estimated Lyapunov exponents are in the range of $0.301 \sim 0.389$, relatively stable, except for $\sigma = 0.02$, 0.1 when $N = 1000$, indicating that $N = 1000$ is not enough for estimating the Lyapunov exponents. When $N \geq 3000$ and $\sigma \leq 0.08$, the range of the estimated Lyapunov exponents is $0.339 \sim 0.365$ when $N = 3000$, and $0.316 \sim 0.372$ when $N = 5000$, showing that the method works well. Note that the Lyapunov exponent of the Henon map without noise is known as about 0.418. In stochastic Henon map where the additive noise is involved, the time series often diverge and we must stop or contract the series when it is diverged. This changes the shape of attractors as is seen in Figure 3; for example, when $\sigma = 0.1$ the shape is flattened around the vertical axis. Thus the Lyapunov exponent of the stochastic Henon map is reasonably anticipated to be smaller than the deterministic case. Our estimates support this speculation.

σ	$\hat{\lambda}$
0.02	0.264
0.04	0.359
0.06	0.389
0.08	0.342
0.1	0.408

(a) $N = 1000$

σ	$\hat{\lambda}$
0.02	0.339
0.04	0.351
0.08	0.365
0.1	0.334

(b) $N = 3000$

σ	$\hat{\lambda}$
0.02	0.345
0.04	0.372
0.06	0.350
0.08	0.316
0.1	0.301

(c) $N = 5000$

Table 1: Estimates of the Lyapunov exponent.

Remark 4. *The procedure for reproducing the skeleton did not work well when* $\sigma = 0.06$ *and* $N = 3000$*. We consider that the initial vector was not appropriate for the estimated skeleton when* $\sigma = 0.06$ *and* $N = 3000$*. Hence if the procedure for reproducing the skeleton does not work well, then we recommend to regenerate data from the estimated skeleton by giving an another initial vector and estimate the Lyapunov exponent from the regenerated data.*

References

Cawley, R. and Hsu, G. (1992) Local-geometric-projection method for noise reduction in chaotic maps and flows. *Physical Review A,* **46,** 3057-3082.

Collomb,G. (1984) Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahrscheinlichkeitstheorie verw. Geviete*,**66**,441–460.

Fueda, K. and Yanagawa, T. (2001) Estimating the embedding dimension and delay time from chaotic time series with dynamic noise, *Journal of the Japan Statistical Society*, **31**, No.1, 27-38.

Kostelich, K. and Yorke, J. A. (1990) Noise reduction-finding the simplest dynamical system consistent with the data. *Physica D,* **41** 183-196.

Landa,P. S. and Rozenblum, M. G. (1989) A comparison of methods for constructing a phase space and determining the dimension of an attractor from experimental data. *Sov. Phys.-Tech. Phys.,* **34** 1229-1232.

McCaffrey, D. F., Ellner, S., Gallant, D., and Nychka, D. (1992) Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. *J. of Americ. Statist. Assoc.,* **87** 682-695.

Nadayara, E. A. (1964) On estimating regression. *Theory of probability and its applications,* **9,** 141-142.

Pikovsky, A. S. (1986). Discrete-time dynamic noise filtering. *Radio. Eng. Electron. Phys.,* **9,** 81.

Sauer, T. (1992) A noise reduction method for signals from nonlinear systems. *Physica, D,* **58** 193-201.

Watson, G. S. (1964) Smooth regression analysis. *Sankhyā,* Ser. A, **26,** 359-372.

Wolff, R. C. L. (1992) Local Lyapunov exponents: looking closely at chaos. *J. R. Statist. Soc.* **B54** 353-371.

Yao, Q. and Tong, H. (1994) Quantifying the influence of initial value on nonlinear prediction. *J. R. Statist. Soc.* **B 56,** 701-725

Yonemoto, K. and Yanagawa, T. (2001) Estimating the embedding dimension and delay time of chaotic time series by an autoregressive model *Bulletin of Informatics and Cybernetics* **33** No.1-2, 53-62.

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF

2003-1 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Yoshitaka WATANABE
A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems

2003-2 Masahisa TABATA & Daisuke TAGAMI
Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients

2003-3 Tomohiro ANDO, Sadanori KONISHI & Seiya IMOTO
Adaptive learning machines for nonlinear classification and Bayesian information criteria

2003-4 Kazuhiro YOKOYAMA
On systems of algebraic equations with parametric exponents

2003-5 Masao ISHIKAWA & Masato WAKAYAMA
Applications of Minor Summation Formulas III, Plücker relations, Lattice paths and Pfaffian identities

2003-6 Atsushi SUZUKI & Masahisa TABATA
Finite element matrices in congruent subdomains and their effective use for large-scale computations

2003-7 Setsuo TANIGUCHI
Stochastic oscillatory integrals - asymptotic and exact expressions for quadratic phase functions -

2003-8 Shoki MIYAMOTO & Atsushi YOSHIKAWA
Computable sequences in the Sobolev spaces

2003-9 Toru FUJII & Takashi YANAGAWA
Wavelet based estimate for non-linear and non-stationary auto-regressive model

2003-10 Atsushi YOSHIKAWA
Maple and wave-front tracking — an experiment

2003-11 Masanobu KANEKO
On the local factor of the zeta function of quadratic orders

2003-12 Hidefumi KAWASAKI
Conjugate-set game for a nonlinear programming problem

2004-1 Koji YONEMOTO & Takashi YANAGAWA
Estimating the Lyapunov exponent from chaotic time series with dynamic noise