Reproduction of Dynamical Systems with the Method of Principal Component Analysis

Manabu YUASA
Research Institute for Science & Technology, Kinki University,
Kowakae 3-4-1, Higashi-Osaka, Osaka 577-8502, Japan
E-mail: yuasa@rist.kindai.ac.jp

(Received 22 December, 2003)

Abstract

Differential equations for Hénon-Heiles' non-linear dynamical system are reproduced by using the principal component analysis (PCA) in twice. The adopted method is concretely proved to be capable of the empirical construction of dynamical systems from observed data sets. In this paper data sets are given by numerical integration of Hénon-Heiles' original differential equations. In the case of the non-chaotic motion, the accuracy of three figures is obtained concerning the reproduced coefficients for the resulting differential equations. Then the reproduction is performed using various data sets including the case of chaotic motion. The accuracy of the reproduction seems to be strongly correlative with the energy of chaotic motion. Also, the influence of the overlapped white noise upon the data is investigated. The accuracy of the reproduction is rapidly lost under the condition of the overlapped white noise of the magnitude of 10^{-3}(= 1% relative noise). The observational data should desirably have at least 3 significant figures(= 0.1% relative observational errors) for the rigorous reproduction of the differential equations.

Key words: Dynamical system, Principal component analysis, Hénon-Heiles' system, Chaos, Noise

1 Introduction

The construction of the dynamical system from observed data sets is an important procedure for developing empirical sciences. If the differential equations for describing given data sets are determined, we can find the driving agencies in the dynamical system and our understandings for the system will be much advanced.

The new method for this procedure was proposed by Unno (1995), in which the principal component analysis(PCA) is employed repeatedly. Since then this method has been applied to the data sets in mathematical economics (Unno et al. 1996; Yuasa,Unno 1996; Yuasa et al. 1997), bio-science (Unno et al. 1997) and astronomy(Yuasa et al. 1999). The complete formulation of the method has been described as a general method for the analysis of observations(Unno, Yuasa 2000). In this paper we first intend to reproduce Hénon-Heiles' system (Hénon, Heiles 1964) in the case of non-chaotic motion by using this method, in order to check the feasibility of the method. Then the reproduction is examined for the various values of Hamiltonian to see the influence of chaotic motion on our method. Finally, the effects of noise on the accuracy of our method are studied. Our purpose is to investigate the practicability of the reproduction method and how chaos or noise affects the accuracy of the reproduced differential equations, using rather simple data sets. Further study to treat several data sets altogether of which each set is constructed from different initial conditions, is of interest and a next purpose to us. The procedure in each section can be summarized as follows.

In section 2, we perform the numerical integration of Hénon-Heiles' original differential equations to obtain data sets composed of the variables of Hénon-Heiles' dynamical system, q_1, p_1, q_2, and p_2 .

In section 3, we execute the descriptive(preliminary) PCA. The arranged data sets are embedded in m-dimensional space, and m principal components are extracted from this space.

In section 4, we carry out the dynamical(extended) PCA. In addition to the m principal components which are calculated by descriptive PCA, the difference of one focused variable (say p_1) of which time dependence should be searched for, are embedded in m+1-dimensional space. Then we extract $m+1$ principal components from this extended space.

In section 5, we reproduce the differential equations of Hénon Heiles' dynamical system. If the given data sets can represent the whole dynamical system in sufficient accuracy, the average value of the minimum principal component of the dynamical PCA nearly equals to zero. By putting the value to zero, we should obtain the differential equation concerning the focused variable. With this procedure, we have been able to obtain the differential equation system which coincides with the original Hénon-Heiles' equation up to three figures.

Furthermore, in section 6, the reproduction method is applied to various data sets including the case of the chaotic motion to search for the effect of the chaotic motion on our method.

Also, in section 7, the influence of the white noise overlapped upon the data is investigated to clarify the limitation of applying our method to observed data sets which inevitably include observational errors.

2 Arrangements of data sets

The differential equations of Hénon-Heiles' system are written as follows;

$$\frac{dq_1}{dt} = \frac{\partial H}{\partial p_1} = p_1, \quad (1)$$

$$\frac{dp_1}{dt} = -\frac{\partial H}{\partial q_1} = -q_1 - 2q_1q_2, \quad (2)$$

$$\frac{dq_2}{dt} = \frac{\partial H}{\partial p_2} = p_2, \quad (3)$$

$$\frac{dp_2}{dt} = -\frac{\partial H}{\partial q_2} = -q_2 - q_1^2 + q_2^2, \quad (4)$$

where Hamiltonian H is given by

$$H = \frac{1}{2}(p_1^2 + p_2^2) + \frac{1}{2}(q_1^2 + q_2^2 + 2q_1^2q_2 - \frac{2}{3}q_2^3). \quad (5)$$

We have performed the numerical integration of this system by Runge-Kutta method, adopting the initial condition $q_1(0) = 0.1$, $p_1(0) = -0.3$, $q_2(0) = 0.3$, $p_2(0) = 0.1$ and the step width 0.01 .

Fig.1 shows the orbit between t=0 and t=40 in the $q_1 - q_2$ plane. The horizontal axis is q_1 and the vertial axis is q_2. We have selected the initial condition to give the non-chaotic obedient trajectory. The orbit is limited within the triangular region $-\frac{1}{2} \le q_2 \le \pm\sqrt{3}q_1 + 1$, because our initial condition gives $H = 0.0940 < \frac{1}{6}$. If the value of H exceeds $\frac{1}{6}$, the orbit can not be limited within the finite space.

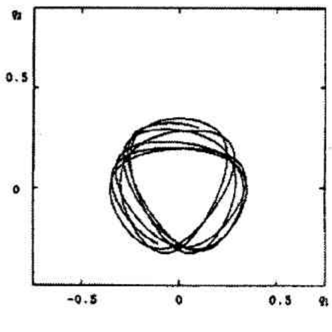

Fig.1. The orbit of Hénon-Heiles' system between $t = 0$ and $t = 40$ in $q_1 - q_2$ plane. The initial conditions are $q_1(0) = 0.1$, $p_1(0) = -0.3$, $q_2(0) = 0.3$ and $p_2(0) = 0.1$. The step width of the numerical integration is 0.01.

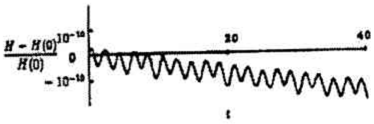

Fig.2. The error of the numerical integration. The initial conditions are the same as the case of Figure 1. The horizontal axis is time and the vertical axis is the relative error of Hamiltonian in the unit of 10^{-10}.

Fig.2 shows the error of the numerical integration. The horizontal axis is time and the vertical axis is the relative error of Hamiltonian in the unit of 10^{-10}. If the dimension of the dynamical system is d and the each data set has address number n, then the dimension m of the embedding space should satisfy the following inequality:

$$2d + 1 \leq m \leq \log_2 n. \qquad (6)$$

Hénon-Heiles' system has two degrees of freedom consisting of second order differential equations, then the system can be written by four simultaneous first order differential equations. On the other side the system has an integral $H = const.$ Therefore, the dynamical system dimension of Henon-Heiles' equations must be $d = 3$. By the substitution of $d = 3$ to the condition (6), we have adopted $m = 7$ and $n = 1332$. In the 7-dimensional space we have embedded the value of the seven quantities $q_1, p_1, q_2, p_2, q_1^2, q_2^2, q_1q_2$ at every $t = 0.03$ between t=0 and t=40.

3 Descriptive(Preliminary) PCA

Before proceeding to the PCA, we have standardized the each data set so as to have mean value 0 and standard deviation 1.

Namely, the standardized variables X_i are given as follows:

$$X_i = (x_i - \overline{x_i})/\delta_{x_i} \quad (i = 1, 2, \cdots, 7), \qquad (7)$$

$$\text{where} \quad x_1 = q_1, x_2 = p_1, x_3 = q_2, x_4 = p_2, \quad x_5 = q_1^2, x_6 = q_2^2, x_7 = q_1q_2 \qquad (8)$$

and $\overline{x_i}, \delta_{x_i}$ represent the mean values and standard deviations of x_i's, respectively concerning 1332 addresses.

The results of descriptive PCA in the embedded space $(q_1, p_1, q_2, p_2, q_1^2, q_2^2, q_1q_2)$ are as follows:

eigen value	*eigen vector*
1.980	$\vec{a_1} = (\ 0.093, 0.591, -0.647, 0.045, -0.322, 0.342, 0.025\ )$
1.926	$\vec{a_2} = (\ 0.701, -0.086, 0.046, 0.699, 0.055, -0.015, 0.088\ )$
1.705	$\vec{a_3} = (\ 0.010, 0.389, -0.279, -0.013, 0.624, -0.616, 0.043\ )$
1.013	$\vec{a_4} = (\ 0.024, -0.032, 0.020, -0.147, -0.096, -0.055, 0.982\ )$
0.236	$\vec{a_5} = (\ -0.047, 0.007, 0.029, -0.007, 0.698, 0.706, 0.107\ )$
0.072	$\vec{a_6} = (\ 0.384, 0.589, 0.620, -0.337, -0.051, 0.050, -0.056\ )$
0.068	$\vec{a_7} = (\ -0.591, 0.379, 0.342, 0.612, -0.068, 0.000, 0.105\ )$

With a_{ij} , the $j-th$ component of the $i-th$ eigen vector $\vec{a_i}$ belonging to the $i-th$ eigen value λ_i , the value of the $i-th$ principal component y_i is given by the following equations ($i = 1$ corresponds to the first (maximum) principal component) at every address.

$$y_i = \sum_{j=1}^{7} a_{ij} X_j \quad (i = 1, 2, \cdots, 7). \qquad (9)$$

Each X_j and therefore each y_i have 1332 addresses respectively. If the suffix i increases, the standard deviation of the principal components decreases. The square of the standard deviation of each principal component y_i is the corresponding eigen value ($\lambda_i = \overline{y_i{}^2}$).

4 Dynamical(Extended) PCA

In the dynamical PCA, we embed the seven principal components y_i $(i = 1, 2, \cdots, 7)$ of the descriptive PCA and the one additional quantity y_8 in the extended eight dimensional space. Then we standardize the variables in the same manner as the descriptive PCA.

$$Y_i = (y_i - \overline{y_i})/\delta_{y_i} \quad (i = 1, 2, \cdots, 8), \qquad (10)$$

where $\overline{y_i}$ and δ_{y_i} represent the mean values and standard deviations of y_i concerning 1332 addresses, respectively. Except $\overline{y_8}$ which is generally non-zero, $\overline{y_i}$ $(i = 1, 2, \cdots, 7)$ are zero because they are the principal components of the descriptive PCA.

Since we want to reproduce the equations (1),(2),(3) and (4), we adopt y_8 as the additional variable in the dynamical PCA, to be $y_8 = \Delta q_1$ in case 1, $y_8 = \Delta p_1$ in case 2, $y_8 = \Delta q_2$ in case 3, and $y_8 = \Delta p_2$ in case 4, respectively. The difference Δq_1 , etc. of the address k is created by the subtraction of the value of the adopted variable at the address $k-1$ from that of the address k $(k = 1, 2, \cdots, 1332)$. For the first address, the difference y_8 is approximatedly taken to be the same as the second address. The results are given by the following four cases.

$case1: \ y_8 \ = \ \Delta q_1$

λ_i

2.000 $\vec{b_1} = (\, 0.590, -0.074, 0.359, -0.019, 0.002, 0.113, 0.068, 0.707\,)$
1.000 $\vec{b_2} = (\, 0.202, -0.096, -0.061, -0.008, 0.516, -0.591, -0.575, 0.000\,)$
1.000 $\vec{b_3} = (\, 0.204, 0.738, -0.285, 0.190, 0.428, 0.336, 0.015, -0.000\,)$
1.000 $\vec{b_4} = (\, -0.074, 0.533, 0.380, 0.111, -0.542, -0.153, -0.486, -0.000\,)$
1.000 $\vec{b_5} = (\, 0.173, -0.270, -0.284, 0.868, -0.219, 0.042, -0.115, -0.000\,)$
1.000 $\vec{b_6} = (\, 0.254, -0.156, -0.499, -0.420, -0.291, 0.396, -0.495, -0.000\,)$
1.000 $\vec{b_7} = (\, 0.350, 0.232, -0.428, -0.145, -0.351, -0.575, 0.407, -0.000\,)$
0.000 $\vec{b_8} = (\, -0.590, 0.074, -0.359, 0.019, -0.002, -0.113, -0.068, 0.707\,)$

$case2: \ y_8 \ = \ \Delta p_1$

λ_i

2.000 $\vec{b_1} = (\, 0.086, 0.659, 0.009, 0.214, -0.004, 0.061, -0.096, -0.707\,)$
1.000 $\vec{b_2} = (\, 0.330, -0.008, 0.348, -0.026, -0.314, 0.567, 0.591, -0.000\,)$
1.000 $\vec{b_3} = (\, 0.359, -0.146, -0.210, 0.346, 0.787, 0.200, 0.162, -0.000\,)$
1.000 $\vec{b_4} = (\, 0.124, -0.192, 0.634, 0.606, -0.073, -0.413, -0.061, -0.000\,)$
1.000 $\vec{b_5} = (\, 0.107, 0.141, 0.616, -0.589, 0.444, -0.000, -0.211, 0.000\,)$
1.000 $\vec{b_6} = (\, 0.820, 0.002, -0.232, -0.236, -0.217, -0.409, -0.056, 0.000\,)$
1.000 $\vec{b_7} = (\, -0.217, 0.232, -0.019, -0.134, 0.180, -0.542, 0.745, 0.000\,)$
0.000 $\vec{b_8} = (\, 0.086, 0.659, 0.009, 0.214, -0.004, 0.061, -0.096, 0.707\,)$

$case3: \ y_8 \ = \ \Delta q_2$

λ_i

2.000 $\vec{b_1} = (\, 0.034, 0.686, -0.013, -0.104, -0.002, -0.062, 0.113, 0.707\,)$
1.000 $\vec{b_2} = (\, -0.783, 0.067, -0.029, -0.103, 0.152, -0.361, -0.466, -0.000\,)$
1.000 $\vec{b_3} = (\, 0.336, 0.028, -0.115, 0.240, 0.864, -0.208, -0.163, -0.000\,)$
1.000 $\vec{b_4} = (\, 0.012, 0.068, 0.666, 0.624, -0.173, -0.362, 0.032, 0.000\,)$
1.000 $\vec{b_5} = (\, 0.203, -0.108, 0.628, -0.707, 0.154, -0.155, -0.069, 0.000\,)$
1.000 $\vec{b_6} = (\, -0.340, 0.106, 0.382, 0.126, 0.339, 0.771, 0.048, -0.000\,)$
1.000 $\vec{b_7} = (\, -0.339, -0.158, -0.031, -0.071, 0.249, -0.264, 0.850, 0.000\,)$
0.000 $\vec{b_8} = (\, 0.034, 0.686, -0.013, -0.104, -0.002, -0.062, 0.113, -0.707\,)$

$case4: \; y_8 = \Delta p_2$

λ_i	
2.000	$\vec{b_1} = (\, 0.693, -0.037, 0.058, -0.015, -0.010, -0.107, -0.053, 0.707 \,)$
1.000	$\vec{b_2} = (\, -0.104, -0.034, 0.327, 0.003, -0.376, 0.591, 0.625, -0.000 \,)$
1.000	$\vec{b_3} = (\, 0.141, 0.492, -0.291, 0.278, 0.617, 0.425, 0.124, -0.000 \,)$
1.000	$\vec{b_4} = (\, -0.001, 0.827, 0.082, -0.051, -0.499, -0.077, -0.226, 0.000 \,)$
1.000	$\vec{b_5} = (\, -0.081, 0.054, 0.818, 0.468, 0.260, -0.117, -0.147, -0.000 \,)$
1.000	$\vec{b_6} = (\, 0.019, -0.109, -0.353, 0.827, -0.356, -0.186, 0.134, 0.000 \,)$
1.000	$\vec{b_7} = (\, -0.033, 0.235, 0.034, -0.129, 0.184, -0.627, 0.706, 0.000 \,)$
0.000	$\vec{b_8} = (\, 0.693, -0.037, 0.058, -0.015, -0.010, -0.107, -0.053, -0.707 \,)$

In the similar manner as the descriptive PCA, the principal components z_i of the dynamical PCA can be given by the following equations.

$$z_i = \sum_{j=1}^{8} b_{ij} Y_j \quad (i = 1, 2, \cdots, 8), \qquad (11)$$

where b_{ij} indicates the $j-th$ component of the $i-th$ eigen vector $\vec{b_i}$ corresponding to the $i-th$ eigen value of the dynamical PCA. We have not seven but eight principal components becuase we have embedded the quatities in 8-dimenional space in the dynamical PCA.

5 Reproduction of Differential Equations

The eighth (minimum) eigen values of the dynamical PCA are nearly zero ($< 10^{-5}$) in each case. Then the eighth principal components in each case can be put equal to zero. These relations give the following equation in each case.

$$z_8 = \sum_{j=1}^{8} b_{8j} Y_j = 0 \qquad (12)$$

We have calculated the value of eighth eigen vector $\vec{b_8}$ of dynamical PCA in section 4. On the other hand the standardized variables $Y_j \; (j = 1, \cdots, 7)$ of dynamical PCA can be transformed to the original variables x_j and Y_8 can be transformed to the difference Δx of the focused variable x. In $case1 \sim case4$ Δx represents $\Delta q_1, \Delta p_1, \Delta q_2$ and Δp_2 respectively. With the help of the equation (7),(8),(9),(10), we can solve the equation (12) with respect to the difference of the focused variable as follows:

$$\Delta x = c_0 + \sum_{j=1}^{7} c_j x_j \qquad (13)$$

The difference Δx has been created by the subtraction of successive x values. The interval of time Δt of successive values of x is 0.03 in our arranged data, so we have to divide both sides of the equation (13) by Δt to get the differential equations concerning q_1, p_1, q_2 and p_2.

The results are as follows:

$$\begin{aligned} case1 \quad \frac{dq_1}{dt} &= 1.000 p_1 \\ &+0.015 q_1 - 0.000 q_2 - 0.000 p_2 - 0.000 q_1^2 \\ &-0.000 q_2^2 + 0.030 q_1 q_2 - 0.000 \end{aligned} \qquad (14)$$

$$\begin{aligned} case2 \quad \frac{dp_1}{dt} &= -1.000 q_1 - 1.999 q_1 q_2 \\ &+0.021 p_1 + 0.005 q_2 - 0.000 p_2 + 0.024 q_1^2 \\ &-0.024 q_2^2 - 0.000 \end{aligned} \qquad (15)$$

$$\begin{aligned} case3 \quad \frac{dq_2}{dt} &= 1.000 p_2 \\ &+0.000 q_1 - 0.000 p_1 + 0.015 q_2 + 0.015 q_1^2 \\ &-0.015 q_2^2 + 0.000 q_1 q_2 + 0.000 \end{aligned} \qquad (16)$$

$$case4 \quad \frac{dp_2}{dt} = -1.000q_2 - 1.000q_1^2 + 0.999q_2^2 - 0.048q_1q_2 -0.005q_1 + 0.000p_1 + 0.021p_2 -0.000 \tag{17}$$

Above results show Hénon-Heiles' original differential equations (1)-(4) are successfully reproduced by our method almost up to three significant figures concerning the coefficients in the case of adopted data($H = 0.0940$; non-chaotic).

6 Accuracy of the Reproduction and Chaos

To see the influence of chaotic motion on the accuracy of the reproduction of the differential equations, we have examined the reproduction of Hénon-Heiles' system using the data arranged from the various values of Hamiltonian. The numerical value of Hamiltonian can be physically regarded as the total energy of the system, so we hereafter use total energy E to represent the numerical value of Hamiltonian H.

Using the method of Poincaré mapping, Hénon and Heiles(1964) analyzed the dynamical system given by the equations (1)-(5). Along each trajectory, they investigated the successive mapping of the point satisfying the conditions $q_1 = 0$ and $p_1 > 0$ on the $q_2 - p_2$ plane.

As an important result, they found the qualitative difference concerning the behavior of the successive points between the trajectories having the total energy $E \leq \frac{1}{9}$ and $\frac{1}{9} < E < \frac{1}{6}$. In the case of $E \leq \frac{1}{9}$, each trajectory gives the succesive points lying on the smooth curve like fig.3(a). On the contrary, in the case of $\frac{1}{9} \leq E \leq \frac{1}{6}$, some trajectories still produce the smooth curve but some other initial conditions give the random successive points without any regularity. In fig.3(b) the random dispersed points indicate the Poincaré mapping produced from a unique trajectory.

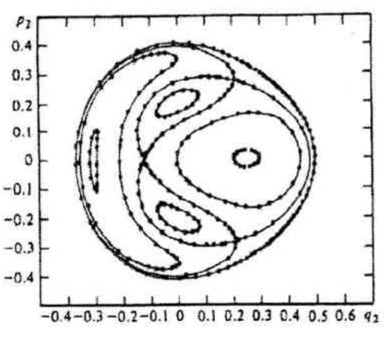

Fig.3(a). Poincaré mapping(E=0.083), Hénon & Heiles 1964.

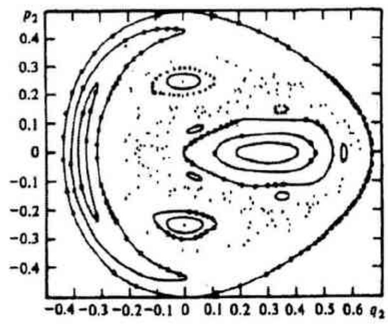

Fig.3(b). Poincaré mapping(E=0.125), Hénon & Heiles 1964.

If the total energy E is given, the area which the successive points can move in $q_2 - p_2$ plane is given by the inequality $p_2^2 + q_2^2 - 2/3q_2^3 < 2E$. The outer most curve in fig.3(a) and 3(b) indicates the boundary of this movable area, of which the inside is covered by the smooth curves and/or the random points. The relative area covered by smooth curves in the movable area increases with the increase of total energy (fig 4). The trajectory which produces the dispersed points corresponds to the chaotic motion. The chaotic motion comes into appearance at the critical total energy $E = \frac{1}{9}$ and the area of the chaotic motion dominates with the increase of the total energy excess over the critical value.

The chaotic motion is to have a higher complexity than the non-chaotic motion. Namely, in the chaotic motion, two very similar initial conditions give the quite different trajectories as time goes by. Anticipating some effects of chaos on our reproduction method, we intend to search for the influence of the chaotic motion on our method by comparing various total energy cases. In this analysis we adopt the same reproduction method as the preceding sections, embedding the variables $q_1, p_1, q_2, p_2, q_1^2, q_2^2$ and q_1q_2 in the 7-dimensional space. The adopted total energy values for data1~data10 are 0.094, 0.103, 0.110, 0.119, 0.128, 0.144, 0.152, 0.159, 0.165, and 0.167 respectively.

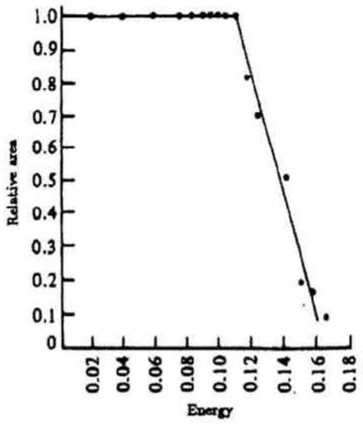

Fig.4. Relative area covered by the smooth curves as a function of energy, Hénon & Heiles 1964.

In each data the number of the address is fixed as 332, then the total energy is considered to be a single parameter. The dependence of the accuracy of the reproduction on the total energy is investigated by using the minimum eigen value, Λ_{min}, in the dynamical PCA. The errors of the constructed dynamical system, $\triangle$, is given by the following equation(Unno, Yuasa 2000):

$$\triangle = \frac{1-c^2}{\sqrt{2}}, \tag{18}$$

where c is the sum of the correlation between each principal component in the descriptive PCA and the additional quantity y_8. On the other side, the relation of c and Λ_{min} is given by the equation(Unno, Yuasa 2000):

$$\Lambda_{min} = 1 - c, \tag{19}$$

By the substitution of the equation (18) into the equation (19), the error is expressed with Λ_{min} as

$$\triangle = \frac{\Lambda_{min}(2-\Lambda_{min})}{\sqrt{2}}, \tag{20}$$

If the minimum eigen value, Λ_{min}, is very small($\Lambda_{min} \ll 1$), the equation (20) is approximately written as

$$\triangle \sim \sqrt{2}\Lambda_{min}(\times 100\%). \tag{21}$$

Then we compute the minimum eigen value, Λ_{min}, for 4 reproduced equations $\frac{dq_1}{dt}\frac{dp_1}{dt}\frac{dq_2}{dt}$ and $\frac{dp_2}{dt}$ corresponding to the 10 data sets respectively. The result is shown in fig.5(a), where the horizontal axis is energy and the vertical axis is minimum eigen value Λ_{min}. In fig.5(a), we can see the minimum eigen value begins to increase at the vicinity of the critical energy $E = \frac{1}{9} = 0.111$ for the appeareance of the chaotic motion, and continues to increase with the increase of total energy except for the largest 2 energy case in the equation $\frac{dp_1}{dt}$.

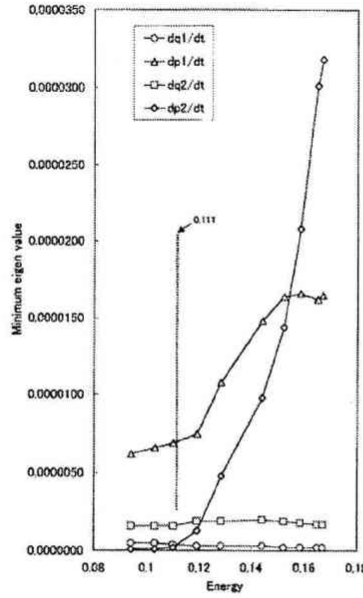

Fig.5(a). Minimum eigen value of each reproduced equation, as a function of energy.

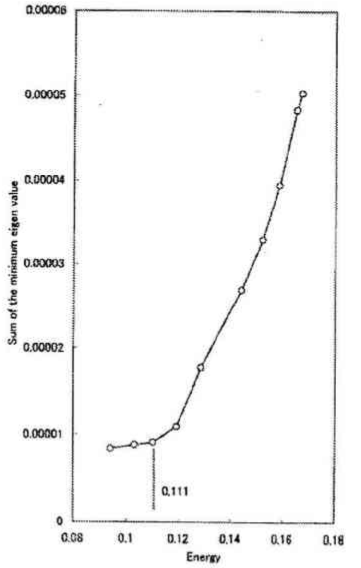

Fig.5(b). Sum of the minimum eigen value of the reproduced 4 equations, as a function of energy.

As the sum of the minimum eigen values corresponding to the 4 equations $\frac{dq_1}{dt}\frac{dp_1}{dt}\frac{dq_2}{dt}$ and $\frac{dp_2}{dt}$ reproduced from the unique data should give the accuracy of the whole system, we have calculated the sum of the minimum eigen values and the result is shown in fig.5(b) whose horizontal axis is the energy.

Fig.5(b) shows that the sum of the minimum eigen value begins to increase at the vicinity of the critical energy and continues to increase with the increase of the total energy. This fact can be interpreted as the reflection of fig.4. Namely, the accuracy of our reproduction method based on the principal component analysis has a clear inverse correlation with the relative area covered by the smooth curves in fig.4. The appearance of the chaotic motion reflects on the accuracy of our reproduction method for dynamical systems.

7 Effects of Noise for Reproduction

The observed data inevitably contain the observational errors. On the other hand, the adopted data in this analysis so far, are arranged from the numerical integration under the condition to conserve the total energy up to 10 figures which probably brings the accuracy of the adopted 7 variables $q_1, p_1, q_2, p_2, q_1^2, q_2^2$ and q_1q_2 almost up to the similar figures.

For the purpose of clarifying the limitation of applying our method to the observed data sets with observational errors, we have investigated the effects of white noise overlapped on the arranged data from the numerical integration. The results are shown in fig.6(a),6(b) (E=0.110; non-chaotic) and 6(c),6(d) (E=0.144; chaotic). In the four figures, the vertical axis is the minimum eigen value, and the horizontal axis is the relative magnitude of the overlapped white noise. The typical absolute value of the variables q_1, p_1, q_2, p_2 is roughly regarded as 10^{-1} and the overlapped white noise is examined in the magnitude of $10^{-5}, 10^{-4}$ and 10^{-3} respectively, or, the examined relative magnitude of the overlapped white noise is 0.01%, 0.1%, and 1%, respectively. Fig. 6(a), 6(b), 6(c), and 6(d) show clearly that reproduced differential equations lose rapidly their accuracy under the condition of the overlapped white noise of the magnitude of 10^{-3}(= 1% relative noise). The errors mainly originate from non-linear terms in the differential equations. If the overlapped white noise becomes 10^{-2}(= 10% relative noise) the reproduced differential equations have no resemblance to the original equations at all. Then we conclude that observational data are required to have at least 3 significant figures(= 0.1% relative observational errors) for the rigorous reproduction of the differential equations.

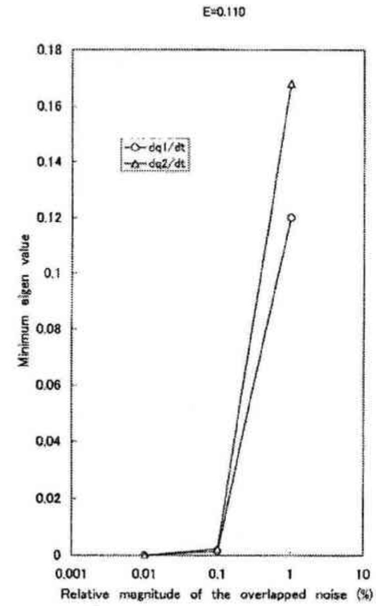

Fig.6(a). Minimum eigen value of the reproduced 2 equations, as a function of the relative magnitude of the overlapped noise(E=0.110).

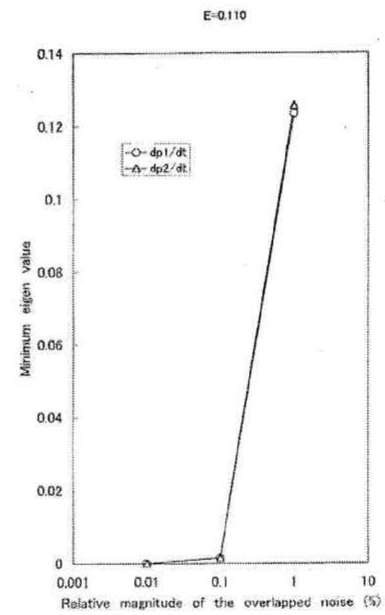

Fig.6(b). Minimum eigen value of the reproduced 2 equations, as a function of the relative magnitude of the overlapped noise(E=0.110).

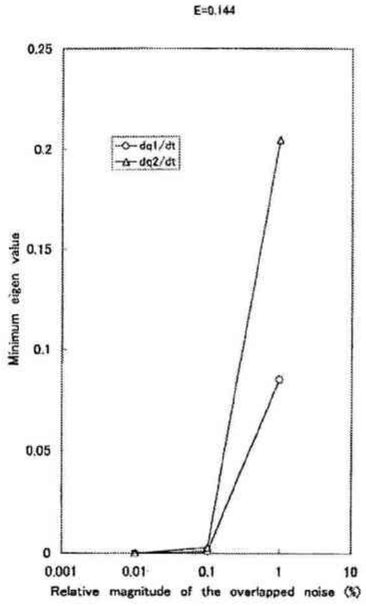

Fig.6(c). Minimum eigen value of the reproduced 2 equations, as a function of the relative magnitude of the overlapped noise(E=0.144).

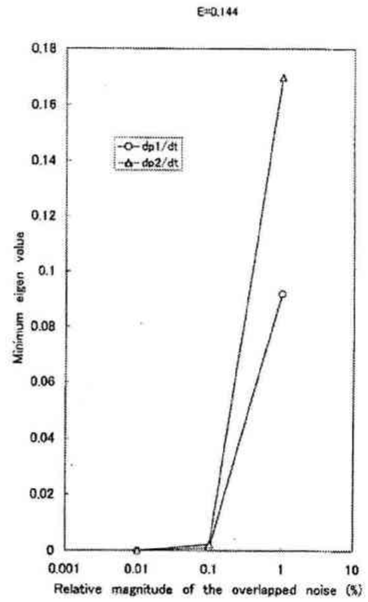

Fig.6(d). Minimum eigen value of the reproduced 2 equations, as a function of the relative magnitude of the overlapped noise(E=0.144).

8 Discussion

The final results (14),(15),(16),(17) in section 5 show that the coefficients coincide with those of the original equations (1),(2),(3),(4) almost up to the three figures. The accuracy results mainly from the time interval of the arranged data sets.

In the preceeding sections, the dimension m of the embedded space has been chosen as 7. If we choose $m = 10$ and the number of arranged data as 1332 at every $t = 0.03$ between $t = 0$ and $t = 40$, we obtain the differential equations similar to the equations (14),(15),(16),(17).

In the case of $m = 10$, the variables of embeded space are chosen as $x_1 = q_1$, $x_2 = p_1$, $x_3 = q_2$, $x_4 = p_2$, $x_5 = q_1^2$, $x_6 = q_2^2$, $x_7 = q_1 q_2$, $x_8 = p_1^2$, $x_9 = p_2^2$, $x_{10} = p_1 p_2$.

The final results are as follows:

$$\begin{aligned} \frac{dq_1}{dt} &= 1.000p_1 \\ &\quad +0.015q_1 - 0.000q_2 - 0.000p_2 - 0.000q_1^2 - 0.000q_2^2 \\ &\quad +0.030q_1q_2 - 0.000p_1^2 - 0.000p_2^2 + 0.000p_1p_2 + 0.000 \end{aligned} \tag{22}$$

$$\begin{aligned} \frac{dp_1}{dt} &= -1.000q_1 - 2.000q_1q_2 \\ &\quad +0.015p_1 + 0.001q_2 + 0.000p_2 + 0.017q_1^2 - 0.014q_2^2 \\ &\quad -0.012p_1^2 + 0.015p_2^2 - 0.002p_1p_2 - 0.000 \end{aligned} \tag{23}$$

$$\begin{aligned} \frac{dq_2}{dt} &= 1.000p_2 \\ &\quad +0.000q_1 + 0.000p_1 + 0.015q_2 + 0.015q_1^2 - 0.016q_2^2 \\ &\quad +0.000q_1q_2 - 0.000p_1^2 - 0.001p_2^2 + 0.000p_1p_2 + 0.000 \end{aligned} \tag{24}$$

$$\begin{aligned} \frac{dp_2}{dt} &= -1.000q_2 - 0.999q_1^2 + 1.000q_2^2 \\ &\quad -0.001q_1 - 0.032q_1q_2 - 0.000p_1 + 0.015p_2 \\ &\quad +0.001p_1^2 + 0.002p_2^2 + 0.028p_1p_2 - 0.000 \end{aligned} \tag{25}$$

The method which we have adopted is proved concretely capable of the empirical construction of dynamical systems from observational data sets. This method is applicable to many research fields and we expect the method will be a strong tool for analyzing data of complex systems.

We have found a strong correlation between the accuracy of the reproduction and the chaotic motion. Since the orbit in the chaotic motion is considered to shift more easily to the neighbouring orbit than in the non-chaotic motion, this process may probably bring the lowering of the accuracy in the reproduction of the chaotic case. To obtain more detailed circumstances, further study including some other chaotic system, for example Lorenz system, will be useful and our future task.

On the effects of noise, we have obtained a conclusion that the reproduction of the differential equations requires the data of at least 3 significant figures(= 0.1% relative observational errors). This conclusion however is for the rigorous reproduction of the coefficient of the differential equations and in the case of a qualitative rough analysis for the observed data somewhat smaller significant figures may actually be available.

The author is grateful to Emeritus Prof. W. Unno of the University of Tokyo for valuable discussions and the constant encouragements.

References

[1] Hénon M., Heiles C. 1964, AJ 69, 73

[2] Unno W. 1995, Science & Technology, Kinki Univ. 7, 1

[3] Unno W., Yuasa M., Onishi T. 1996, Science & Technology, Kinki Univ. 8, 61

[4] Yuasa M., Unno W. 1996, Science & Technology, Kinki Univ. 8, 67

[5] Yuasa M., Unno W., Onishi T. 1997, Science & Technology, Kinki Univ. 9, 19

[6] Unno W., Onishi T., Yuasa M. 1997, Viva Origno 25, 139

[7] Yuasa M., Unno W., Magono S. 1999, PASJ 51, 197

[8] Unno W., Yuasa M. 2000, PASJ 52, 127