Restoration of Missing Data and Reconstruction of Dynamical Systems

Manabu YUASA, Nawo YAMAMOTO, Masafumi UMETANI
RIST, Kinki University, Higashi-Osaka
577-8502, JAPAN
and
M.K.DAS
Institute of Informatics & Communication,
University of Delhi South Campus,
Benitojuarez Road, New Delhi 110021, INDIA

(Received 26 December, 2005)

Abstract

The eliminated data in Lorentz dynamical system is restored by using the generalized Principal Component Analysis(PCA). The restored data and the original data which has been eliminated, are compared. Both data have good coincidence. Adopting the original data only (case 1) and the data which includes partly the restored one (case 2), the reconstruction of Lorentz dynamical system, i.e. the system of the original differential equations, is examined respectively by the method of repeated PCA.

Key words: Restored data, Principal Component Analysis, Reconstruction of dynamical systems

1 Introduction

Recently, with the rapid progress of the electronic technology, the beginning of the operation of the newly constructed many large telescopes, and the increase of the opportunity of observations in the space, a large number of huge data base have been constructing in the field of observational astronomy. In these observational data base the small partial lack of the data is common. If we make an analysis of the data statistically, we should adopt the data as much as possible regarding not only physical quantities but also the number of observed stars in order to produce the increased statistical accuracy. For example, if there is one star which has all observational data of the physical quantities except for one quantity, it is too wasteful to exclude all the data of such a star. From this point of view, the imperfect data should be included in the analysis if the imperfectness is small and the adjusted values are restored by some valid method.

In this paper, we simulate the restoration of the eliminated data produced with the numerical integration of the differential equations in Lorentz dynamical system (Lorentz 1963) by the method of generalized PCA (Unno and Yuasa 1992) which gives the most probable adjusted values to the imperfect data (Singh and Yuasa et al.). After the restoration, the restored data and the original one are compared. The restoration error is small and both data have rather good coincidence.

In addition, we examine the reconstruction of Lorentz dynamical system, i.e. the system of the differential equations from which we have produced the original data with the numerical integration. The reconstruction is performed by using the method of repeated PCA (Unno and Yuasa 2000). Two cases are examined, namely, one case uses the original data only and another case uses the data which includes partly the restored one. The accuracy of the reconstruction in the case of including the restored data somewhat decreases in comparison with the case of using the original data only. But in some case, the data partly including restored one gives the similar reconstruction accuracy to the case of using the original data only.

2 Restoration of the Eliminated Data

We have produced the original data of Lorenz dynamical system with the numerical integration of the following equations:

$$\frac{dx}{dt} = -ax + ay, \quad (1)$$

$$\frac{dy}{dt} = -xz + rx - y, \quad (2)$$

$$\frac{dz}{dt} = xy - bz, \quad (3)$$

where $a = 10, b = \frac{8}{3}$, and $y = 28$.

The numerical integration has been performed by Runge-Kutta method, adopting $x(0) = 0, y(0) = 1, z(0) = 0$ as the initial condition and the step width 0.0002. Then we have adopted every 10 step values between $t = 0$ and $t = 1.6$ as the original data $x(i), y(i)$ and $z(i) \quad (i = 1, \cdots, 800)$ for the following analysis.

Using the method of the restoration (Unno and Yuasa 1992) which gives the most probable adjusted values to the missing data, we have examined the simulation of the restoration of the data by eliminating one value of the specified varaible for each i of the original data. The procedure is mentioned in the following. Embedding the prepared data, namely $Q_1^{(i)} = x(i), Q_2^{(i)} = y(i), Q_3^{(i)} = z(i), Q_4^{(i)} = x(i)y(i)$, and $Q_5^{(i)} = z(i)x(i) \quad (i = 1, \cdots, 800)$, in the 5-dimensional space, we eliminate one of the data, for example $Q_1^{(s)}$.

The normalized data $q_j^{(i)} \quad (j = 1, \cdots, 5; i = 1, \cdots, 800)$ for applying PCA is introduced by

$$q_j^{(i)} = \frac{Q_j^{(i)} - < Q_j >}{\sigma_j}, \quad (4)$$

where $< Q_j >$ and σ_j represent the mean value and the standard deviation of the quantity Q_j respectively. If we introduce the weight $w_j^{(i)}$ for $Q_j^{(i)}$ and the another weight $v_j^{(i)} = 1 - w_j^{(i)}$ for the virtual added data $x_j^{(i)}$, the elimination of the data $Q_1^{(s)}$ corresponds to putting $w_1^{(s)} = 0$ and all other weights $w_j^{(i)}$ except for $w_1^{(s)}$ equal to 1.

In this simple case, the virtual added data $x_1^{(s)}$ becomes the restored data of $Q_1^{(s)}$ (Yuasa et al. 2005). The value of $x_1^{(s)}$ is given by the solution of the following separated simultaneous algebraic equations:

$$\Big(\sum_{l=1}^{n} \frac{\mu_{l1}^2}{\lambda_l}\Big) x_1^{(s)} + \sum_{l=1}^{n} \frac{\mu_{l1}}{\lambda_l} \Big(\sum_{k=2}^{n} \mu_{lk} q_k^{(s)}\Big) = 0$$
$$(s = 1, \cdots, 800), \quad (5)$$

where λ_l is the l-th eigen value of PCA and μ_{lj} represents the j -th component of the l-th eigen vector of PCA.

By changing the columns of original data $Q_1^{(i)}$ and $Q_2^{(i)}$, we can compute the restored value $x_2^{(s)}$ for supplementing the eliminated value $q_2^{(s)}$. In the same manner, we can get the restored values $x_3^{(s)}, x_4^{(s)}$ and $x_5^{(s)}$ for the eliminated value $q_3^{(s)}, q_4^{(s)}$ and $q_5(s)$ respectively. The restoration error, namely the difference between the restored data and the original data is shown in Fig.1 $\sim$ Fig.3 for the case of the elimination and restoration of $x(s), y(s)$ and $z(s) \quad (s = 1, \cdots, 800)$ respectively. In each Fig. the horizontal axis is the difference between the normalized original vaiable $q_j^{(s)}$ (mean value is 0 and the standard deviation is 1) and the restored value $x_j^{(s)}$ and the vertical axis represents the frequency distribution of the corresponding data. These Figs. show the restoration error is small, though there is a little remarkable restoration error in the variable z. We can conclude the restoration is successfully performed.

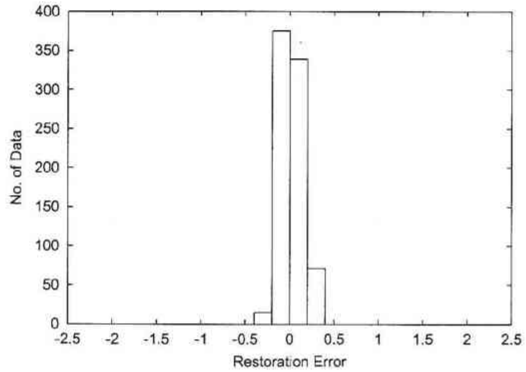

Fig.1. The frequency distribution of restored data is shown with respect to the restoration error, i.e. $q_1^{(s)}$(eliminated value of x)-$x_1^{(s)}$(restored value of x) for $s = 1, \cdots, 800$.

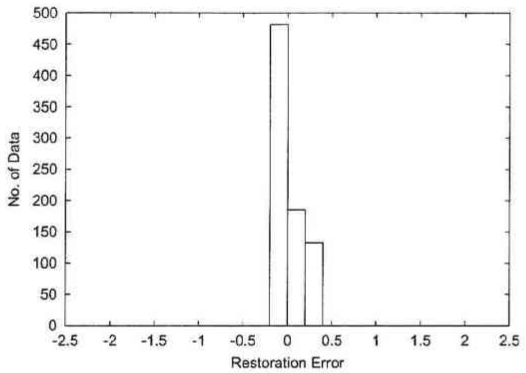

Fig.2. The frequency distribution of restored data is shown with respect to the restoration error, i.e. $q_2^{(s)}$(eliminated value of y)-$x_2^{(s)}$(restored value of y) for $s = 1, \cdots, 800$.

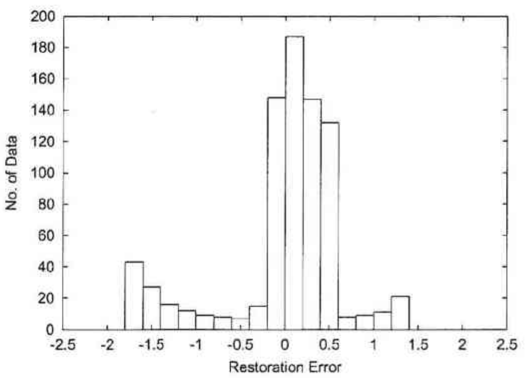

Fig.3. The frequency distribution of restored data is shown with respect to the restoration error, i.e. $q_3^{(s)}$(eliminated value of z)-$x_3^{(s)}$(restored value of z) for $s = 1, \cdots, 800$.

3 Reconstruction of the Dynamical System

We reconstruct Lorentz dynamical system by the method of repeated PCA (Unno and Yuasa 2000), using the original data only (case 1) and the data which includes the restored data partly (case 2). We mention case 1 first. At the early stage of this study we have prepared 5000 data (from $t = 0$ to $t = 10$). We have examined the reconstruction by using several kinds of data number. Even if we adopt 2000 or 5000 data the results have been same as 700 data at least 2 main significant digits. If we adopt 800 data (from $t = 0$ to $t = 1.6$, time interval is 0.002) and embed them in the 9-dimensional space, i.e. $x, y, z, x^2, y^2, z^2, xy, yz$ and zx, we have got the following reconstructed equations:

$$
\begin{aligned}
\frac{dx}{dt} &= -10.389x + 10.115y - 0.001z \\
&\quad -0.000x^2 - 0.000y^2 + 0.000z^2 \\
&\quad +0.000xy - 0.000yz + 0.010zx \\
&\quad -0.001, \qquad (6) \\
\frac{dy}{dt} &= +29.138x - 1.698y + 0.142z \\
&\quad +0.012x^2 - 0.002y^2 - 0.003z^2 \\
&\quad -0.002xy + 0.027yz - 1.034zx \\
&\quad +0.179, \qquad (7) \\
\frac{dz}{dt} &= +0.017x - 0.008y - 2.739z \\
&\quad -0.007x^2 - 0.025y^2 + 0.003z^2 \\
&\quad +1.038xy - 0.001yz + 0.004zx \\
&\quad +0.007. \qquad (8)
\end{aligned}
$$

The difference between these equations (6) $\sim$ (8) and the original Lorentz equations (1) $\sim$ (3) is rather small. We have succeeded in the reconstruction of the Lorentz dynamical system within the accuracy of almost 2 significant digits for each coefficient of variables in the right hand side of the corresponding differential equations. Next we show the results of case 2. If we adopt 740 original data (from 1 to 740) and 10 restored data (from 741 to 750) whose restoration error seems to be rather large (case2-A), the reconstructed equations are as follows:

$$\begin{aligned}
\frac{dx}{dt} &= -14.074x + 11.902y - 0.703z \\
&\quad -0.139x^2 + 0.038y^2 + 0.031z^2 \\
&\quad +0.006xy - 0.018yz + 0.084zx \\
&\quad -0.457, \qquad (9) \\
\frac{dy}{dt} &= +29.264x - 1.757y + 0.164z \\
&\quad +0.016x^2 - 0.004y^2 - 0.004z^2 \\
&\quad -0.002xy + 0.027yz - 1.037zx \\
&\quad +0.196, \qquad (10) \\
\frac{dz}{dt} &= -0.587x + 0.276y - 2.849z \\
&\quad -0.030x^2 - 0.018y^2 + 0.008z^2 \\
&\quad +1.039xy - 0.004yz + 0.016zx \\
&\quad -0.073. \qquad (11)
\end{aligned}$$

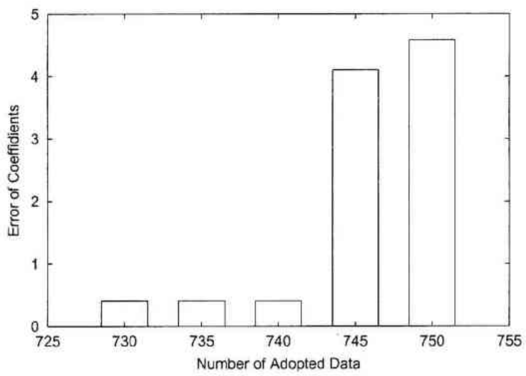

Fig.4. The variation of the error of the coefficients, Δc, of $\frac{dx}{dt}$ involved in including the restored data partly. At 745 and 750 on the horizontal axis, 5 and 10 restored data are used respectively.

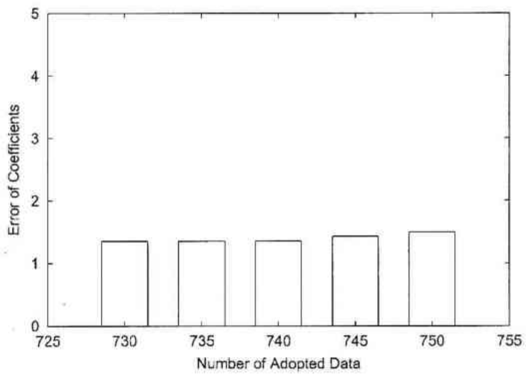

Fig.5. The variation of the error of the coefficients, Δc, of $\frac{dy}{dt}$ involved in including the restored data partly. At 745 and 750 on the horizontal axis, 5 and 10 restored data are used respectively.

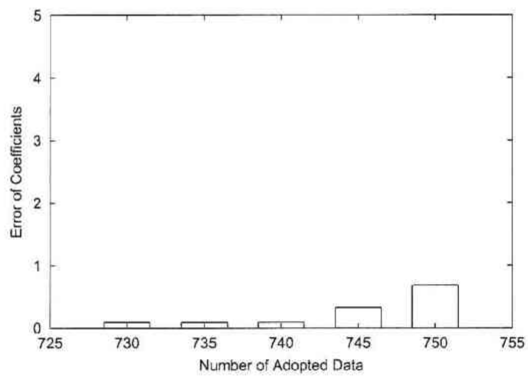

Fig.6. The variation of the error of the coefficients, Δc, of $\frac{dz}{dt}$ involved in including the restored data partly. At 745 and 750 on the horizontal axis, 5 and 10 restored data are used respectively.

Equations (9)$\sim$(11) indicate the accuracy of the reconstruction of the differential equations somewhat decreases in comparison with the case of using the original data only (equations (6)$\sim$(8), case 1).
In Fig. (4)$\sim$(6), we show the error of the coefficients, which is computed with the formula, $\Delta c = \sqrt{\sum_{i=0}^{9}(c_i - c_i(o))^2}$, where c_i and $c_i(o)$ $(i = 1, \cdots, 9)$ represent each coefficient of the variables $x, y, z, x^2, y^2, z^2, xy, yz$ and zx in the right hand side of the corresponding reconstructed and original differential equations respectively, and c_0 and $c_0(o)$ are the constant terms in the same meaning.
In each Fig. the error of the coefficients, Δc, is computed by adopting the original data only up to 740 on the horizontal axis. At 745 and 750 on the horizontal axis, we have included the restored data by 5 (from 741 to 745) and 10 (from 741 to 750) respectively. We can see from these Figs Δc is somewhat increasing after 740 except for $\frac{dy}{dt}$. The restored and added data does not seem to be so useful.

And if we adopt 700 original data and 10 restored data whose restoration error seems to be rather small (case 2-B), the reconstructed equations are as follows:

$$
\begin{aligned}
\frac{dx}{dt} &= -10.631x + 10.236y - 0.084z \\
&\quad -0.011x^2 + 0.003y^2 + 0.003z^2 \\
&\quad +0.001xy - 0.001yz + 0.014zx \\
&\quad -0.042 \qquad (12)
\end{aligned}
$$

$$
\begin{aligned}
\frac{dy}{dt} &= +29.130x - 1.694y + 0.138z \\
&\quad +0.011x^2 - 0.002y^2 - 0.003z^2 \\
&\quad -0.002xy + 0.027yz - 1.034zx \\
&\quad +0.178 \qquad (13)
\end{aligned}
$$

$$
\begin{aligned}
\frac{dz}{dt} &= +0.041x - 0.019y - 2.729z \\
&\quad -0.006x^2 - 0.025y^2 + 0.003z^2 \\
&\quad +1.038xy - 0.001yz + 0.003zx \\
&\quad +0.011 \qquad (14)
\end{aligned}
$$

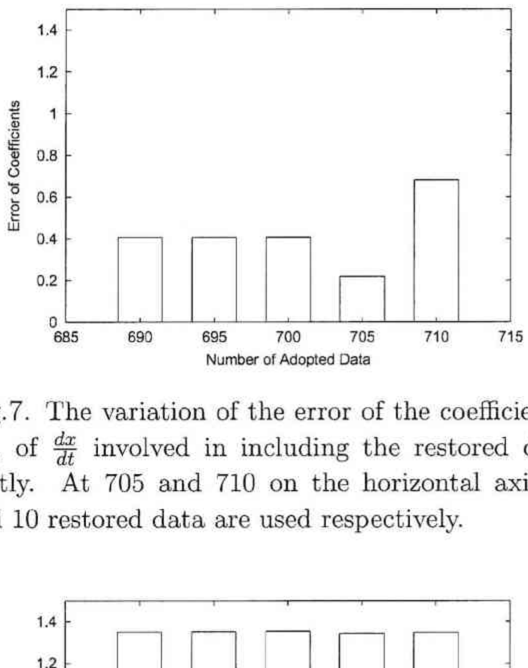

Fig.7. The variation of the error of the coefficients, Δc, of $\frac{dx}{dt}$ involved in including the restored data partly. At 705 and 710 on the horizontal axis, 5 and 10 restored data are used respectively.

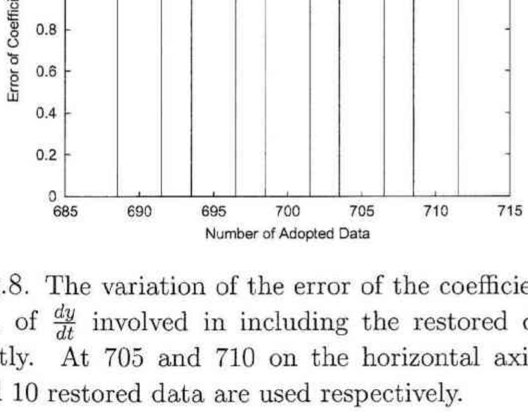

Fig.8. The variation of the error of the coefficients, Δc, of $\frac{dy}{dt}$ involved in including the restored data partly. At 705 and 710 on the horizontal axis, 5 and 10 restored data are used respectively.

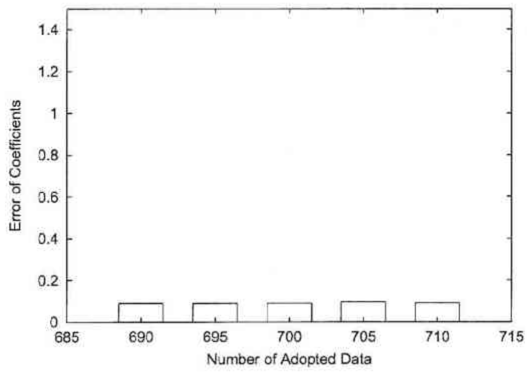

Fig.9. The variation of the error of the coefficients, Δc, of $\frac{dz}{dt}$ involved in including the restored data partly. At 705 and 710 on the horizontal axis, 5 and 10 restored data are used respectively.

In both of case2-A and case2-B, the restored data is used with respect to the variable x. If we compare the results of case 1 and case 2, we can see the variation of the coefficients is large in the equations (9) and (12). The reason is interpreted as we have used the difference of x in the determination of the reconstructed equations (9) and (12). So the two equations are effected directly and other equations are effected indirectly by the difference of the restored variable x.
In each Fig. (7)~(9), the error of the coefficients, Δc, is computed by adopting the original data only up to 700 on the horizontal axis. At 705 and 710 on the horizontal axis, we have included the restored data by 5 (from 701 to 705) and 10 (from 701 to 710) respectively. We can see from these Figs Δc is almost similar before and after 700 on the horizontal axis. So, we can conclude the restored data is useful and should be included in the statistical analysis to improve the accuracy. In case 2-A we can not say such a recommendation. In the real data analysis, case 2-A and case 2-B are to be mixed. We will investigate further examples to clarify the meaningfulness of including the restored data in the statistical analysis.

Acknowledgement

We are grateful to Emeritus Prof. W. Unno of the University of Tokyo for valuable discussions. M. Yuasa would like to thank JSPS (Japan Society for Promotion of Science) and DST (Department of Science & Technology, India) for a financial support.

References

[1] Lorentz E.N. 1963, J. Atmos. Sci. 20, 130

[2] Lorentz E.N. 1963, J. Atmos. Sci. 20, 448

[3] Singh H., Yuasa M., Yamamoto N. and Gupta R. 2006, PASJ 58, in press

[4] Unno W. and Yuasa M. 1992, Ap&SS 189, 271

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

[6] Yuasa M., Umetani M., Yamamoto N. and Das M.K. 2005, Sci. & Tech. Kinki Univ. No.17,1