J. Indian Inst. Sci., July–August 1998, **78**, 257–265.

A new method to embed time-series data and for parameter identification

PRABHAKAR G. VAIDYA*
CSIR Center for Mathematical Modelling and Computer Simulation (C-MMACS), Bangalore 560 037, India.

Received on August 2, 1998.

Abstract

This paper develops a scheme to estimate the derivatives of a discretely sampled data more accurately than traditional methods. The method is based on applying singular-value decomposition to Taylor's expansion but in a matrix form. The validity and robustness to noise of the proposed scheme are illustrated by numerical experiments.

1. Introduction

'A picture is worth a thousand words'; so is an equation which is worth a million words. The truth of this statement becomes self-evident when one deals with a large amount of data. It has been the dream of every analyst of observational data since Keppler to find an equation which will fit the data.

Both these ideas have been beautifully expressed by Packard *et al.*[1] in a fascinating paper appropriately titled 'geometry from a time series', later formalized by Takens.[2]

Experimentally or numerically generated data often consists of a measurement of a single scalar variable at different times. For example, we might measure a variable X at different times leading to a sequence of scalar values: $X(t_{-1})$, $X(t_{-2})$, $X(t_{-3})$... at times t_{-1}, t_{-2}, t_{-3},... If a model equation is to be developed from the sequence and expect it to be of an order equal to N, where, let us assume, N is higher than 1 but less than, say 20, we would need to generate a set of vectors from these scalar data and then define the dynamics of the data with the help of this set, either in terms of difference of differential (or more complex) equations.

In the Packard–Taken's method, the vectors are constructed from the sequence of X itself. Thus, the first vector would have the first N data points as its components. This way we embed an R_n space in an R_1 sequence.

The method has been discussed extensively in literature. Its main drawback is its vulnerability to experimental noise. We have an alternative method which is far more intuitive and seems to offer several advantages. It is based on a technique of state-space vectors which is quite well established. It is common to plot velocity vs displacement in mechanics. In higher dimensions, higher derivatives become higher components of the state space. The main draw-

*On leave from the School of Mechanical and Materials Engineering, Washington State University, Pullman, Washington 99164, USA. email: prabhakargv@hotmail.com

back is the accurate determination of derivatives which has led to the Packard alternative. In what follows, we would demonstrate a singular-value decomposition technique[3] to accurately determine even very high-order derivatives.

Once this is achieved, we would show how to obtain state-space portraits from experimental data. This method is robust as it can tolerate and, in fact, suppress a small amount of random noise. This is followed by an illustration of how to find unknown parameters (if the form of equations is known) using this method.

2. Numerical differentiation

The formulae for numerical differentiation are quite well known. For example, if the samples are at a uniform interval h, the central-difference formula gives:

$$dX/dt = (X(t+h) - X(t-h))/2h.$$

This formula is a little more accurate than just a forward or backward difference formula. An improvement involving $t+2h$ and $t-2h$ terms can be easily derived using Taylor's series. However, even *Numerical recipes*[3] claims that the algebra gets 'increasingly tedious' with additional terms.

This objection, however, can now be easily overcome given our ability to invert and find singular-value decomposition of quite large matrices. We would show below how this can be done and how, in principle, extremely accurate determination to a very high degree of accuracy for a very large order of derivative is possible, provided certain underlying assumptions about reasonable smoothness, etc. are valid.

This procedure works in the presence of a small amount of noise also. In fact, the very first output of the procedure is that it yields a filtered version of the X series!

2.1. *Taylor's expansion in a matrix form*

The Taylor series expansion is given by

$$f(t+h) = f(t) + \left(h\frac{d}{dt}f + \frac{h^2}{2!}\frac{d^2}{dt^2}f \right) + \text{higher terms},$$

or in a compact notation

$$X(t+h) = X(t) + \sum_{n=1}^{N} \frac{h^n}{n!}\frac{d^n}{dt^n}X(t).$$

If we take a series of samples, not necessarily at a uniform interval

$$S_n = X(t+h_n) = X(t) + \sum_{m=1}^{M} \frac{h_n}{m!}\frac{d^m}{dt^m}X(t).$$

We can write this in a matrix form, if we define

$$A_{n,m} = \frac{(h_n)^m}{m!}$$

and

$$F_0 = X(t)$$

$$F_m = \frac{d^m}{dt^m} X(t).$$

Then,

$$S_n = A_n, \Sigma_m \cdot F_m$$

or

$$S = AF.$$

The problem now boils down to finding F given S and A. And since, in most cases, A is not square, and there are errors (noise) present, we would do this by means of Moore–Penrose generalized inverse procedure.

It should be noted that if the intervals are uniform (or of fixed ratio with one another) a transformation of the type

$$T = \frac{t}{h}$$

and

$$X(T+1) = X(T) + \sum_{m=1}^{M} \frac{1}{m!} \frac{d^m}{dT^m} X$$

makes A independent of h.

2.2. *Finding derivatives accurately*

To illustrate the procedure, consider a data sequence (generated from the Lorenz's equation, although in this section this knowledge is not assumed) represented in Fig. 1.

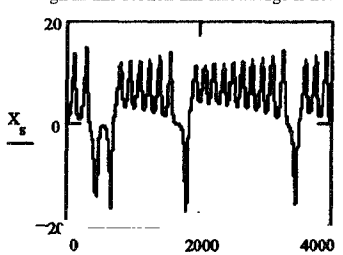


FIG. 1. A sample data sequence.

At this stage all we need to know is the value of the samples and the step size h. In this example, the step size is 1/200. Let us denote the pth sample of X by X-p. At this point, we calculate various derivatives of X. Since the sampling is uniform, one way to do this is to choose a set of $N+1$ samples, with X-p at their centre. If we account for M derivatives in Taylor's expansion, using the rescaling with the sampling interval h, we get

$$m := 0..M$$

$$n := 0..N$$

$$A_{n,m} := \frac{\left(n - \frac{N}{2}\right)^m}{m!}.$$

In this case, we choose

$$N := 50$$

and

$$M := 18$$

$$h := \frac{1}{200}$$

Then, A would have $N+1$ rows and $M+1$ columns. As described in the previous section, we need to find the generalized inverse of A. Many standard routines would do this. For example, using Mathcad, we get a matrix.

$$SV := \text{svd}(A)$$

which has $(N+1)+(M+1)$ rows (in our case, 70) and $(M+1)$ (our case 19) columns.

This is to be separated into two matrices:

U: = submatrix x(SV, 0, N, 0, M)

V: = submatrix x(SV, $N+1$, $N+M+1$, 0, M)

We also form a diagonal matrix W from the singular values given by

SVS: = svds (A)

W: = diag(SVS)

Now the singular-value procedure ensures that $U \times W \times V^T$ will be equal to A. In the numerical example, we can check this by seeing that

$$\max((U \cdot W \cdot VT - A) = 9.536743 \times 10^{-7}$$

is indeed very small.

In practice, we often need to modify W to take care of very small eigenvalues by an interesting procedure. (see, for example, *Numerical recipes*[3] which includes a delightful comment: "it

is not very often that one gets to set infinity = 0!"). For the sake of easy reference, a sample algorithm would be

$$WI:=\left|\begin{array}{l}\text{for } m\in 0..r\,\mathrm{SVS}-1\\ \left(\mathrm{WID}_m\leftarrow 0\right)\text{ if }\mathrm{SVS}_m<1\\ \left(\mathrm{WID}_m\leftarrow\dfrac{1}{\mathrm{SVS}_m}\right)\text{ otherwise}\\ \mathrm{WI}\leftarrow\mathrm{diag}(\mathrm{WID})\end{array}\right.$$

where the value of 1 was chosen by observing the spectrum of the singular values. Now the generalized inverse is given by

$$AI:=\mathrm{V}\cdot\mathrm{WI}\cdot\mathrm{UT}.$$

We now do this at each p and get the F matrix which would have estimates of the M derivatives. For the rest of the paper, we would use the following notation for some of the early derivatives:

$P = X$

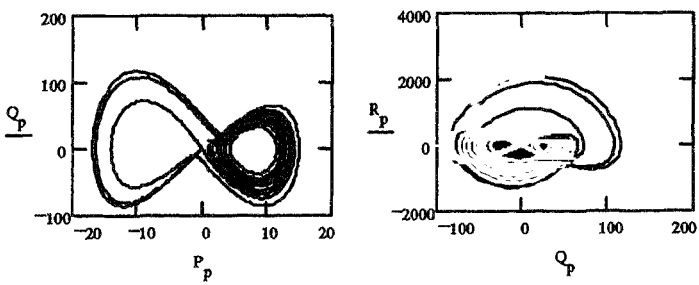
$Q = dX/dt$

$R = dQ/dt$

$S = dR/dt$

With this notation, we can plot the results for the first 2,000 ps (Fig. 2).

These are not the usual state space pictures of the Lorenz's attractor. However, they are just as fascinating. The original goal of Packard of arriving at the geometry from data has been achieved.



(a) (b)

FIG. 2. State–space portraits

2.3. *Finding Y and Z*

Now, let us, for the first time in the numerical example, assume that we know that our data is from the Lorenz's equation:

$$\begin{bmatrix} \frac{d}{dt}X \\ \frac{d}{dt}Y \\ \frac{d}{dt}Z \end{bmatrix} = \begin{bmatrix} \sigma \cdot (Y - X) \\ \rho \cdot X - Y - X \cdot Z \\ X \cdot Y - \beta \cdot Z \end{bmatrix}$$

As a first step, let us for now assume that we do know the parameters of the equation. In that case, it is easy to show that, if

$$S := \frac{1}{\sigma}$$

the best estimates of X, Y, Z would be

$$X2_p := P_p$$

$$Y2_p := Q_p \cdot s + P_p$$

$$Z2_p := \rho - 1 + \frac{\left[-R_p \cdot s - Q_p \cdot (1+s)\right]}{P_p}$$

we can see that these do indeed turn out to be very good estimates (Fig. 3).

Now, we can plot the familiar state–space pictures (Fig. 4).

2.4. *Noise reduction*

One of the main problems with the Packard–Taken's procedure is that it is sensitive to the errors caused to extraneous noise, which, at least in experimental situation is inevitable. Once again choosing the Lorenz's equation, and using a finer step size of 1/2,000, a uniform random noise of range of +0.001 was added to X.

The same procedure was followed with a high N of 200 and somewhat lower M of 10 (one can justify both these numerical observations by means of some reasonable ad hoc arguments). The results can be seen in Fig. 5.

In this figure, P is the estimate of X at the centre of the block, $X2$ is the value obtained after adding the random noise. It is easily seen that the procedure gives an estimate, which is much closer to the actual X, thus removing a lot of noise.

2.5. *Finding parameters and initial conditions*

Another recent development deals with the possibility of rapidly identifying parameters of the Lorenz's equation.[4] It has been shown that following the above notation, we get the following relationship

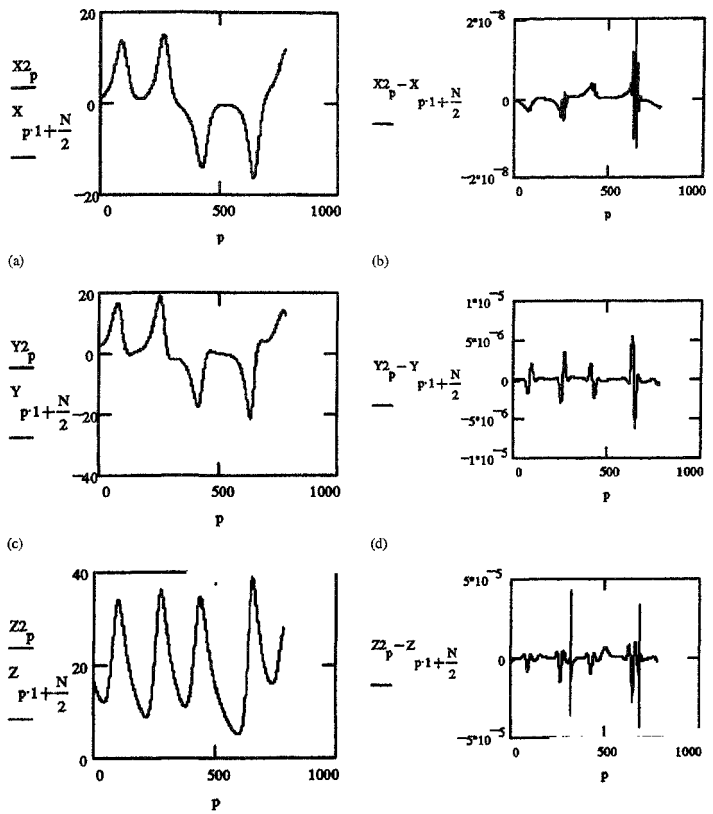
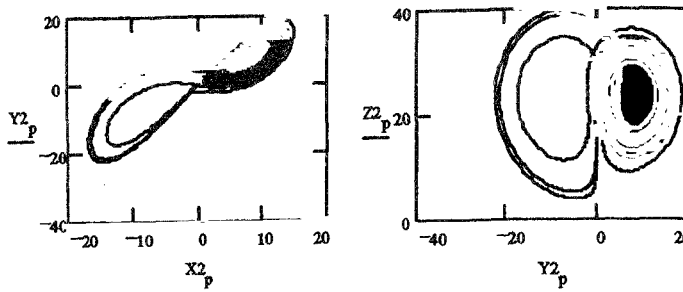


FIG. 3. Comparison of estimates of derivatives.

$$\left[-P\cdot Q+\beta\cdot\frac{(-R-Q)}{P}+\frac{\left(-P\cdot(S+R)+Q\cdot(R+Q)\right)}{p^2}\right]\cdot s+\beta\left(\rho-1-\frac{Q}{P}\right)+\frac{\left(-P\cdot R+Q^2\right)}{p^2}-P^2=0.$$

Since we can calculate P, Q, R, etc. with a high level of accuracy, we can evaluate them at consecutive points and just with a few points arrive at the following set of equations which can



(a) (b)

FIG. 4. State-space portraits.

be solved. The following worked-out example shows how this can be done with just a few points from the X set, and substituting the values P, Q, R and S at each point (again, using Mathcad):

Given

$$(-3451.6 - 229.29\beta)\cdot s + \beta\cdot(\rho + 22.179) + 281.19 = 0 \tag{1}$$

$$(-4701.7 - 270.11\cdot\beta)\cdot s + \beta\cdot(\rho + 25.588) + 408.04 = 0 \tag{2}$$

$$(-7075.5 - 328.24\beta)\cdot s + \beta\cdot(\rho + 30.755) + 647.19 = 0 \tag{3}$$

$$(-6123.6 + 145.39\beta)\cdot s + \beta\cdot(\rho - 20.968) + 563.52 = 0 \tag{4}$$

$$(-12334. - 421.49\beta)\cdot s + \beta\cdot(\rho + 39.463) + 1174.7 = 0 \tag{5}$$

find the parameters with a minimum error:

$$\min \mathrm{err}(\beta, \rho, s) = \begin{pmatrix} 2.657 \\ 24.837 \\ 0.1 \end{pmatrix}$$

These turn out to be quite close to the ones used.

3. Conclusions

We have demonstrated in this paper how to calculate the derivatives of a discretely sampled data, far more accurately than the usual central difference formulas do. This is accomplished by setting the Taylor's series formulation in a matrix form and then using singular-value decomposition. These derivatives are used to generate a set of vectors from scalar data. This method is clearly seen to have several advantages over the conventional method of embedding data. Using the example of a numerically generated data, it has been shown how to obtain state–

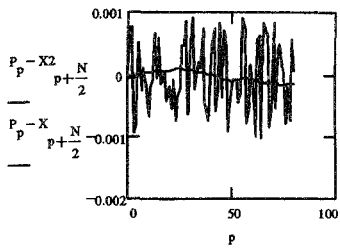


FIG. 5. Example of a noisy data sequence.

space portraits from sampled data. It is also shown that this method is robust in the sense that it can tolerate and, in fact, suppress a small amount of random noise. This is followed by an illustration of how to find unknown parameters, if the form of equations is known, using this method.

Acknowledgments

Thanks are due to Prof. V. K. Gaur and Dr. R. N. Singh of C-MMACS for many helpful discussions.

References

1. PACKARD, N. H. *et al.* Geometry from a time series, *Phys. Rev. Lett.*, 1980, **45**, 712–715.

2. TAKENS, F. *Dynamical systems and turbulence, Lecture Notes in Mathematics*, Vol. 898, D. A. Rand and L. S. Young (eds), Springer–Verlag, 1981.

3. PRESS, W. H. *Numerical recipes in C*, Second edition, Cambridge University Press, 1992, p.59ff.

4. VAIDYA, P. G. *Parametric identification for the Lorenz's equation*, C-MMACS, PD 9803.