A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

LA-UR -85-4265

CONF-8509203--3

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

LA-UR--85-4265

DE86 004733

TITLE AN APPROACH TO ERROR-ESTIMATION IN THE APPLICATION OF DIMENSION ALGORITHMS

AUTHOR(S) Joachim Holzfuss and Gottfried Mayer-Kress

SUBMITTED TO CNLS Workshop "Dimensions and Entropies in Chaotic Systems" Proceedings held Sept. 11-16, 1985 at the Pecos River Conference Center

DISCLAIMER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

An Approach to Error-Estimation in the Application of Dimension Algorithms

Joachim Holzfuss* and Gottfried Mayer-Kress
Center for Nonlinear Studies, MS B258, Los Alamos National Laboratory
Los Alamos, N.M. 87545, USA

* Permanent address: Drittes Physikalisches Institut, Universitaet Goettingen, D-3400 Goettingen, Fed. Rep. of Germany

Three different methods for calculating the dimension of attractors are analyzed. An approach to error-estimation is presented and is used on various datasets. It is shown, that in some cases the errors can become very large.

1. Introduction

The dimension of attractors reconstructed from a time series [1] is of great physical interest especially in experimental situations [2,3]. It is a measure for the number of active modes modulating a physical process and therefore a measure of complexity. Many different methods [4,5,6] of calculating the dimension of attractors have been introduced. Very important questions are, how far these methods are reliable and how large the uncertainty of a calculated dimension is. Most of the algorithms used for dimension measurements of attractors, reconstructed from numerical and experimental data, average over certain variables, such as the number of nearest neighbors, the mass of a cube of a certain sidelength or pointwise dimensions from different reference points on the attractor. All these averages must be taken into account, if one wishes to determine a realistic error estimate of a fractal dimension. In the literature error estimates are mainly calculated by just averaging over some values of the scaling exponents obtained in different length scales (least squares fit). This method can truly underestimate existing errors. The errors can make it useless for experimentalists to deal with smaller/larger relationships between different definitions of fractal dimensions. The largest error source is the limitation in the number of data available to reconstruct an attractor from a time series. The number of data points necessary for filling this subset of a phase space with points to get the same probability measure as given by the attractor of the physical process might be very large and even increases exponentially with the dimension of the attractor. This also gives rise to the question, whether an analysis of high dimensional attractors is possible.

We first recall some different methods for calculating the dimension (II), and then introduce our approach to determine the error (III). After a short description of the computer programs and some ideas for automation of the

dimension calculation (IV), we analyze different data sets, including a 5-torus, gaussian noise and the Lorenz attractor (V).

2. Methods for Calculating the Dimension of an Attractor

To reconstruct an attractor from a time series of a single probe, we use the now classical method of time delay coordinates. In this method a vector $\vec{X}(t_k)$ in an n-dimensional phase space is constructed by taking delayed samples of the time series $x(t_k)$ as coordinates [1], such that

$$\vec{X}(t_k) = \left(x(t_k),\ x(t_k+T),\ x(t_k+2T),\ \cdots\ ,\ x(t_k+(n-1)T)\right) \qquad (1)$$

where t_k is the discrete time with k running from 1 to the number of data points and T is an 'arbitrary', but fixed time delay. The embedding dimension n is the number of coordinates of the embedding space. If T is chosen to be equal to the time delay, where there is a minimum in the mutual information between two measurements [7], a D-dimensional attractor is constructed best by taking the embedding dimension larger than $2D+1$. For signals with strong periodic contents this time delay is approximately equal to the first zero-crossing of the autocorrelation function.

We used three different methods for calculating the "fractal" dimension of an attractor.

1. The pointwise dimension (mass dimension) [4,8] can be defined as

$$D_2 = \lim_{n_{data}\to\infty} \lim_{r\to 0} \frac{\log \frac{1}{n_{data}} N_{\vec{X}_0}(r)}{\log r} \qquad (2)$$

and it consists of counting the number of data points $N_{\vec{X}_0}(r)$ within a cube of sidelength r centered at a point $\vec{X}_0$ on the attractor. Due to the fact, that one does not have an infinite amount of data points and also no infinite precision, as required in the definition, one has to average over several reference points $\vec{X}_0$. Fig. 1 shows the scaling behavior of $N_{\vec{X}_0}(r)$ of different reference points. We calculate the pointwise dimensions for 200 reference points and take the average, which yields a good estimate for the dimension of the attractor.

2. Another method we used was the determination of the correlation dimension D_2 with the algorithm proposed by GRASSBERGER and PROCACCIA [5]. They showed the scaling of the correlation integral $C(r)$ for small r

$$C(r) \propto r^{D_2} \qquad (3)$$

with

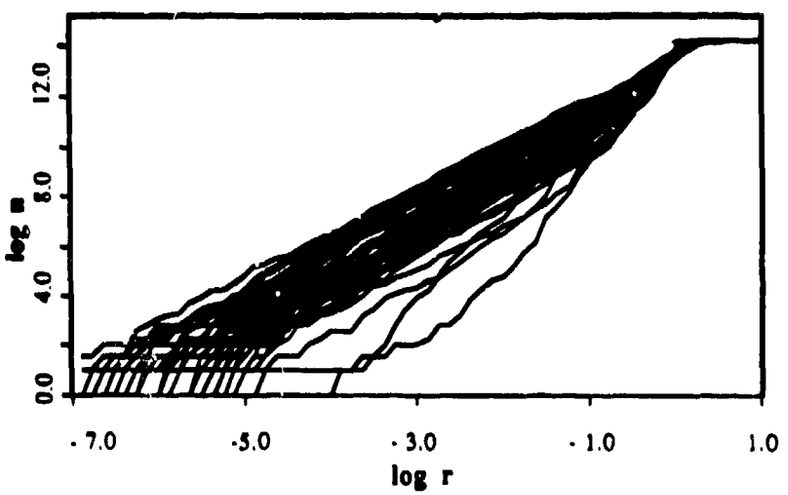

Fig 1: Scaling behavior of $N_{\vec{X}_0}(r)$ at 50 reference points $\vec{X}_0$ on the Lorenz attractor. The embedding dimension is 5.

$$C(r) = \lim_{n_{data}\to\infty} \frac{1}{n_{ref}} \sum_{j=1}^{n_{ref}} \frac{1}{n_{data}} \sum_{i=1}^{n_{data}} \Theta(r - |\vec{X}_i - \vec{X}_j|) \tag{4}$$

and Θ equal to 1 for positive and 0 for negative arguments.
Also

$$C(r) = \lim_{n_{data}\to\infty} \frac{1}{n_{ref}} \sum_{j=1}^{n_{ref}} \frac{1}{n_{data}} N_{\vec{X}_j}(r) \tag{5}$$

with $N_{\vec{X}_j}$ equal to the rightmost sum in eq. (4). $C(r)$ counts the number of points $N_{\vec{X}_j}(r)$ in a cube of fixed sidelength r, averages over all the cubes, that are centered at different reference points $\vec{X}_j$, and normalizes. An example shows, how the averaging is done.

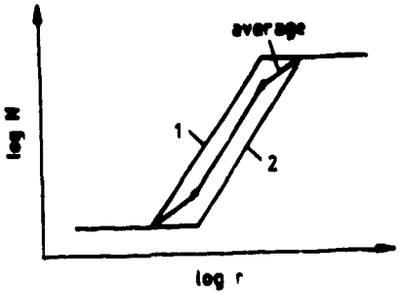

Fig. 2: sketch of the scaling behavior for 2 reference points with different scaling regions $[r_{min}, r_{max}]$ and their average (Grassberger and Procaccia method).

The scaling of the contents of the cubes with respect to r in two different regions on the attractor (fig. 2) is discribed by

$$N_1(r) = (\alpha_1 r)^D \quad \text{and} \quad N_2(r) = (\alpha_2 r)^D \quad \alpha_1, \alpha_2 = const. \tag{6}$$

The Grassberger and Procaccia algorithm averages over the N-values, while r is fixed:

$$C(r) = \frac{N_1(r) + N_2(r)}{2} = \frac{1}{2} \tag{7}$$

$$\bar{N}(r) = r^D \left(\frac{{\alpha_1}^D + {\alpha_2}^D}{2}\right) \tag{8}$$

For infinitely expanded scaling behavior we get the same scaling exponent D for the averaged values. Large differences in the α's and small scaling regions distort the scaling properties of $C(r)$.

3. The third method considered here, is the one proposed by TERMONIA and ALEXANDROWICZ [6]. It consists of averaging over the different radii of

cubes, which contain a fixed number of data points. They showed that N, the 'number of nearest neighbors', behaves like

$$N \cdot \bar{r}(N)^{D_F'}, \tag{9}$$

where $\bar{r}(n)$ is the average radius of the cubes containing N data points and D_F' the "fractal" dimension of the attractor. As an example we consider again the scaling behavior for two reference points, this time keeping N fixed and r variabel.

$$N = (\alpha_1 r_1)^D \qquad \text{and} \qquad N = (\alpha_2 r_2)^D \qquad \alpha_1, \alpha_2 = const. \tag{10}$$

Then

$$\frac{N^{\frac{1}{D}}}{\alpha_1} = r_1(N) \quad \text{and} \quad \frac{N^{\frac{1}{D}}}{\alpha_2} = r_2(N) \tag{11}$$

and

$$N^{\frac{1}{D}}\left(\frac{1}{2}\left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2}\right)\right) = \frac{r_1(N) + r_2(N)}{2} = \bar{r}(N) \tag{12}$$

gives

$$N = \bar{r}(N)^D \left(\frac{1}{2}\left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2}\right)\right)^{-D} \quad , \tag{13}$$

which results in the same scaling exponent D for the averaged values as for the single reference points.

All three methods use a different kind of averaging. The last two methods have one thing in common: they average over an ensemble of single values (N in case 2, r in case 3) and don't care about the orientation of the lines in fig. 1 i.e. the scaling behavior of the attractor at the different reference points. This may in some cases lead to a misinterpretation of the results, if the averaged values pretend a scaling behavior, that the values obtained from different reference points may have never had. In method 1 the additional information of the reference points is used when the averaged pointwise dimension is calculated.

3. Error Estimation

Each value of the pointwise dimension D_s was obtained by calculating the slope of a fitted straight line to each curve in the log r / log N plot using least squares fit. In order to determine the spread of values of the D_s we consider their standard deviation, given by

$$\Delta D = \sqrt{\overline{D^2} - \overline{D}^2} \tag{14}$$

which says, that 68.3 % of all values of a normal distribution lie in the interval $[D - \Delta D, D + \Delta D]$.

In case 2 and 3 straight lines are also fitted to the values in the doubly logarithmic plot using least squares fit. The least squares fit consists of finding a straight line

$$y_{ij} = a + bx_{ij} + z_{ij} \tag{15}$$

such that $\sum(z_{ij})^2$, the sum of squares of the errors z_{ij} is minimized. The index j denotes the different reference points and the index i their average value. It is also possible to use weighted least squares fit, which minimizes

$$\sum w_i z_i^2 = \sum w_i (y_i - a - bx_i)^2 \tag{16}$$

with

$$w_i = \frac{1}{\Delta^2 y_i} \tag{17}$$

and

$$y_i = \frac{1}{n_{ref}} \sum_{j=1}^{n_{ref}} y_{ij} \tag{18}$$

z_i and x_i have the same definition. Using weighted least squares fit reduces the weight of values, if their variance $\Delta^2 y_i$ is large. However, because of the finite size of the attractor, the variances become very small for large values of log r (saturation region) and therefore could give rise to a false fit. Therefore we have to exclude this possibility by e.g. imposing a lower bound for the slope, which should be fitted or by restricting the possible scaling ranges.

In terms of the Grassberger and Procaccia algorithm y_i is the average of the log N_{ij} values of a fixed radius r_i over all reference points $\vec{X}_j$ (fig. 3 a). In the Termonia and Alexandrowicz method y_i would be the average of the logarithms of the radii log r_{ij} with a fixed number of nearest neighbors N_i (fig. 3b) .

We use the average over the logarithmic values in determining $\Delta^2 y_i$ and minimize the errors z_{ij} of the straight line in the log r / log N plot. This gives a good approximation of the fully consistent way of averaging over the nonlogarithmic values of r or N, which would require a minimizing of

$$\sum_i (y_{ij} - ax_{ij}^b)^2 \tag{19}$$

In order to get the standard deviation of the slope, we have $\Delta y + \Delta z = (b\,\Delta x) + (\Delta a + \Delta b\,\Delta x)$. We exclude possible parallel shiftings of the lines obtained at the different reference points to be able to evaluate the largest possible error of the slope, because methods 2 and 3 yield no information about the orientation of these lines. This is done by setting Δa to zero. Now we get for the standard deviation of the slope

$$\Delta b = \frac{\Delta z}{\Delta x} \tag{20}$$

Δz is the standard deviation of the differences of the actual y-values to the y-values of the fitted line. Δx is the standard deviation of the x-values. From fig.4a,b we see, that Δb is the standard deviation of all the possible values of the slope b.

Fig. 4a : standard deviation of x-values and of the errors z of the y-values
b : the possible changes of the slope b

In case 2 the errors are symmetric around the average slope. In case 3, where we calculate $r(N)$ and therefore get an error for $\frac{1}{D}$, they get asymmetric when solving for the slope D.

To modify the definition into a better computable form we have

$$\Delta^2 b = \frac{\Delta^2 z}{\Delta^2 x} = \frac{\overline{z^2} - \overline{z}^2}{\overline{x^2} - \overline{x}^2} \tag{21}$$

Due to the least squares fit $\overline{z} \approx 0$. Taking the average between the differences of all the log r or log N values and the values of the straight line, which is fitted to the logarithms of the averages of the nonlogarithmic r or N-values implies a very small correction $\overline{z} \; | = 0$, which can be neglected. Therefore we get

$$\Delta^2 b = \frac{\sum_{i=1}^{n_{rad}} \sum_{j=1}^{n_{ref}} (y_{ij} - a - bx_{ij})^2}{\sum_{i=1}^{n_{rad}} \sum_{j=1}^{n_{ref}} (x_{ij} - \overline{x})^2} \tag{22}$$

where n_{ref} is the number of reference points and n_{rad} is the number of the averaged values of distances r sub i. With the average over all reference points (17), its variance is given by:

$$\Delta^2 y_i = \sum_{j=1}^{n_{ref}} \frac{(y_{ij})^2}{n_{ref}} - \left(\sum_{j=1}^{n_{ref}} \frac{y_{ij}}{n_{ref}} \right)^2 \qquad (23)$$

and with

$$\sum_{i=1}^{n_{rad}} \sum_{j=1}^{n_{ref}} (x_{ij} - \bar{x})^2 = n_{ref} \sum_{i=1}^{n_{rad}} (x_i - \bar{x})^2 \qquad (24)$$

$$(x_{ij} = x_i \ f.a. \ j)$$

we get

$$\Delta^2 b = \frac{\sum_{i=1}^{n_{rad}} \Delta^2 y_i + \sum_{i=1}^{n_{rad}} (y_i - a - b x_i)^2}{\sum_{i=1}^{n_{rad}} (x_i - \bar{x})^2} \qquad (25)$$

This expression allows to compute the variances of all the y_i's first and to get the variance of the slope when fitting over a certain number of y_i's. Expression (22) can also be modified by using

$$a = \bar{y} - b\bar{x} \qquad (26)$$

and

$$b = \frac{\sum_{i=1}^{n_{rad}} \sum_{j=1}^{n_{ref}} \frac{x_{ij}\, y_{ij}}{n_{rad} * n_{ref}} - \bar{x}\,\bar{y}}{\Delta^2 x} \qquad (27)$$

with $\bar{x}$ and $\bar{y}$ equal to the grand mean into

$$\Delta^2 b = \frac{\Delta^2 y}{\Delta^2 x} - b^2 \qquad (28)$$

The variances are taken over all $n_{rad} * n_{ref}$ points.

4. Description of the Programs

The programs we use are designed for a fully automatic analysis of the data sets. First of all the program has to find the scaling region. This is done by looking for the interval of a given length in the log r / log N plot, where the root-mean-square error of the deviation from a fitted straight line, which is defined as (see also eqn. (16))

$$\sqrt{\frac{1}{n_{rad}} \sum_{i=1}^{n_{rad}} w_i z_i^2} \quad , \qquad 29$$

is the smallest. In case 1, where we consider the averaged pointwise dimension, the weights w_i are all 1. Here the program calculates the slope of a fitted line at each reference point. Then it averages over just 20% of the obtained values of the pointwise dimensions, neglecting all the curves, that don't show a scaling behavior over the entire range of r-values. After calculating the average and the standard deviation it repeats these steps in a different embedding dimension. In case 2 and 3 the procedure is about the same, except there is just one line to be fitted. Also the weights are set to their respective values (eqn. 17). The procedure also features a "self-blowup" of the length of the fitted line: If the total length of the scaling region is unknown, the program starts with a given short interval length and finds the scaling region by minimizing the RMS error (eqn. 29). Then it repeats this step with an enlarged interval until a certain threshold value of the RMS error is reached, thus indicating, that the fitted interval length exceeds the length of the scaling region. In all cases we find that 0.05 is a "good" threshold value.

5. Analysis of Different Datasets

In the analysis the total number of data points is always 20,000. We average over 200 reference points. The time delay T for the reconstruction of the attractor is chosen according to reference 7 by calculating the mutual information. As the first example we consider a 5-torus, constructed from a time series with a Fourier spectrum of 5 incommensurate frequencies. We analyze the data with all three methods (fig. 5a,b,c). In all three figures we see, that the calculated dimension of the attractor converges with increasing embedding dimension, and it has the value $D = 5$ from about $2D + 1 = 11$. The standard deviations of the attractor dimension, described by the error bars, converge also to certain values. For the averaged pointwise dimension we get $5 +/- 0.3$, for the correlation dimension $5 +/- 0.5$ and for the dimension obtained by the TERMONIA/ALEXANDROWICZ [6] method $5 + 0.8 / - 0.7$.

The calculation of the dimension of the Lorenz attractor (fig 6a,b,c) shows also good convergence at about $2D+1$. The values for D obtained by the different methods were $2 +/- 0.15$ in the case of the averaged pointwise dimension, $2 +/- 0.6$ in case 2 and $2 + 0.6 / - 0.4$ in case 3. The growing of the standard deviation in case 2 and 3 is due to successive shortening of the scaling region in higher embedding dimensions because of geometrical effects. Smaller scaling regions with constant variance at each average value in the $\log r$ / $\log N$ diagram result in a larger variance of the slope of the fitted line. The average of the pointwise dimension is not affected by this, when the length of the fitted line equals the length of the scaling region.

In the third example we analyzed gaussian noise, to have an easy example of higher dimensional attractors (fig. 7a,b,c). Noise is considered to be spacefilling, i.e. each phase space of every embedding dimension is filled. Each one of the three methods couldn't produce this result. This is truly seen in the

deviation from the 45° line. The deviation is due to the increasing amount of data necessary for calculating higher dimensions. Also it can be extracted, that the error increases almost linearly with the attractor dimension. Compared to the averaged pointwise dimension, the dimension calculated with the two other methods showed a very large possible error.

6. Conclusions

Three different methods of calculating the dimension of attractors were analyzed. To each of those an approach to estimate the error was presented, which was based on calculating standard deviations for certain variables. Examples of analyzed data sets showed, that the averaged pointwise dimension provided the smallest possible error in the calculated dimension. It also seemed to be closer to the real value. It was shown, that the possible errors in the methods of GRASSBERGER/ PROCACCIA [5] and TERMONIA/ ALEXANDROWICZ [6] can be very large. High dimensional analysis is shown to be very difficult, because of the linear growth of the errors with the embedding dimension.

Acknowledgements

J.H. wants to thank the Center for Nonlinear Studies for the hospitality and financial support and W. Lauterborn for very useful discussions. We are deeply indepted to Erica Jen for permission of using versions of her codes which are the basis of parts of our numerical calculations. We also appreciate very helpful discussions with Erica Jen and J. Doyne Farmer. All computations were done in Los Alamos on CRAY 1 computers.

References:

1 Packard, N. H.,Crutchfield, J. P., Farmer, J. D., Shaw, R. S.: Geometry from a Time Series. Phys. Rev. Lett. 45, 9 (1980) 712

2 Brandstaetter, A., Swift, J.,Swinney, H. L., Wolf, A., Farmer, J. D., Jen, E., Crutchfield, P. J.: Low-Dimensional Chaos in a Hydrodynamic System. Phys. Rev. Lett. 51 ,16 (1983) 1442

3 Lauterborn, W., Holzfuss, J.: Evidence for a Low-Dimensional Strange Attractor in Acoustic Turbulence. Preprint Gottingen

4 Farmer, J. D., Ott, E. and Yorke, J. A.: The Dimension of Chaotic Attractors. Physica 7D (1983) North Holland 153-180

5 Grassberger, P. and Procaccia, I.: Characterization of Strange Attractors. Phys. Rev. Lett. 50, 5 (1983) 346

6 Termonia, Y. and Alexandrowicz, Z.: Fractal Dimension of Strange Attractors. Phys. Rev. Lett. 51, 14 (1983) 1265

7 A. Frazer, this volume

8 Mandelbrot, B.: The Fractal Geometry of Nature (W. H. Freeman and Co., San Francisco, 1982).