A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

Optimal Embeddings of Chaotic Attractors from Topological Considerations

EUROPHYSICS LETTERS 15 March 1991

Europhys. Lett., 14 (6), pp. 521-526 (1991)

Optimal Embeddings of Chaotic Attractors from Topological Considerations.

W. LIEBERT (*) ($), K. PAWELZIK (*) and H. G. SCHUSTER (**)

(*) *Institut für Theoretische Physik, Universität Frankfurt*
Robert-Mayer-Strasse 8, D-6000 Frankfurt
(**) *Institut für Theoretische Physik und Sternwarte, Universität Kiel - D-2300 Kiel*

(received 9 July 1990; accepted in final form 8 January 1991)

PACS. 05.45 – Theory and models of chaotic systems.

Abstract. – Guided by topological considerations, a new method is introduced to obtain optimal delay coordinates for data from chaotic dynamic systems. By determining simultaneously the *minimal* necessary embedding dimension as well as the *proper* delay time we achieve optimal reconstructions of attractors. This can be demonstrated, *e.g.*, by reliable dimension estimations from limited data series.

During recent years much progress has been made in understanding and characterizing real dynamical systems which display chaotic time behaviour [1]. One of the main reasons for this progress was the important discovery that the strange attractor to which the trajectory $\boldsymbol{x}_i$ in phase space of a dissipative chaotic system becomes attracted in the course of time can be reconstructed from experimentally obtained scalar time series $\{x_i\} = x(i \cdot \delta t)_{i=1}^{N}$ using delay coordinates $\boldsymbol{x}_i = (x_i, x_{i+k}, \ldots, x_{i+(m-1)k})$ [2]. Here $\tau = k \cdot \delta t$ is the delay time and m is the embedding dimension. The well-known embedding theorems [3] ensure the possibility of a reconstruction using the delay coordinates for embedding dimensions $m \geqslant 2d + 1$ with d the capacity of the attractor.

While for infinitely long time series the choice of τ should be arbitrary, this is not the case for time series which contain a finite number of data points [4-6]. Since when fixing the observation scale the necessary amount N of data grows exponentially with the embedding dimension m as it does with other parameters [5] it is of great importance to find the minimal sufficient embedding which could be smaller than $2d + 1$. Therefore the proper choice of τ and m remains an important missing link between experimental data and theoretical description.

In this letter, we propose a new method based on topological considerations which determines unambiguously and simultaneously the proper delay time τ and the correct minimal embedding dimension m. Our method differs strongly from previous methods [5-7] which could be used so far and, in contrast to them, it deals τ and m at the same time using

($) Permanent address: Institut für Kernphysik, TH Darmstadt, Schlossgartenstrasse 9, D-6100 Darmstadt.

fewer data points. It can be easily implemented numerically and works also for small amounts of noisy data. These claims will be supported by numerical calculations.

We solve the problem of constructing optimal delay coordinates by reverting to the idea of the *embedding* itself. An embedding is a *topological* mapping, *i.e.* an invertible continuous, and bijective mapping. Roughly speaking for each neighbourhood of a point of the attractor itself as well as of the embedded attractor the following statement is valid: inner points remain inner points and points within the boundary which defines the neighbourhood remain boundary points. The properties of a topological mapping which preserves neighbourhood relations can be achieved by the iterated delay coordinate mapping only if m is big enough. The condition of continuity is already fulfilled by the delay coordinate mapping itself, but it does not strictly preserve the order of nearest neighbours. Our idea is to increase successively the embedding dimension m and to test the topological properties step by step. If preservation of neighbourhood relations is achieved by going from m to $m+1$ the embedding dimension m is sufficient. If the chosen m is too small, the mapping yields a projection of the attractor into a space of too low dimension and therefore violates the condition of injectivity (fig. 1). The possibility of finding many directions of projections is associated with a range of valid values for the delay time τ.

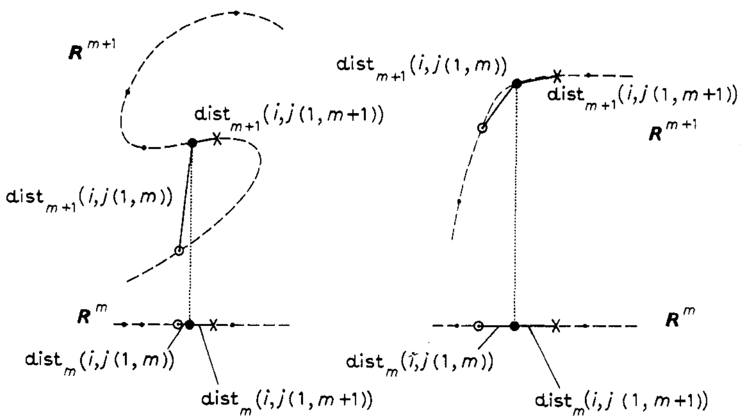

Fig. 1. – Illustration of the delay coordinate mapping as a projection of a $(m+1)$-dimensional reconstruction space onto a m-dimensional one. The change of nearest-neighbour order near a reference point $\boldsymbol{x}_i$ by transition from embedding dimension m to $m+1$ is visualized for *a*) not sufficient and *b*) sufficient embedding dimension. The crosses mark the next neighbour in $\boldsymbol{R}^{m+1}$ and the circles the next neighbour in $\boldsymbol{R}^m$, respectively. The relevant distances for obtaining the wavering product are shown.

The property that a mapping be topological is a local property, and therefore we can use the concept of nearest neighbours. Figure 1 shows how neighbourhood relations may change by going from m to $m+1$. The violation of topological properties is revealed if points closely located to the reference point in $\boldsymbol{R}^m$ become noticeably removed by the transition to $\boldsymbol{R}^{m+1}$. As a first step towards a quantitative measure of this effect we form (cf. fig. 1) the ratio Q_1 of the distance $\mathrm{dist}_{m+1}(i, j(k,m))$ measured in $\boldsymbol{R}^{m+1}$ between the reference point $\boldsymbol{x}_i$ and the point $\boldsymbol{x}_{j(k,m)}$ which was the k-th nearest neighbour in $\boldsymbol{R}^m$ (*old embedding*), and of the distance $\mathrm{dist}_{m+1}(i, j(k, m+1))$ measured in $\boldsymbol{R}^{m+1}$ between the same reference point and the point $\boldsymbol{x}_{j(k,m+1)}$ which is the k-th nearest neighbour in $\boldsymbol{R}^{m+1}$ (*new embedding*), respectively:

$$Q_1(i,k,m,\tau)=\frac{\mathrm{dist}^{\tau}_{m+1}(i,j(k,m))}{\mathrm{dist}^{\tau}_{m+1}(i,j(k,m+1))}, \tag{1}$$

where the index τ at the distance function dist indicates the expected dependence on the reconstruction parameter τ.

The denominator contains the *true* nearest neighbours in $\boldsymbol{R}^{m+1}$ and distances can only *increase* by going from m to $m+1$. Hence this quotient should have the following property: $Q_1(i,k,m,\tau)=1$, if the k-th nearest neighbour in the *new* embedding $\boldsymbol{R}^{m+1}$ was already the k-th nearest neighbour in the *old* embedding $\boldsymbol{R}^m$ and $Q_1(i,k,m,\tau)>1$, if this is not the case. $Q_1(i,k,m,\tau)>1$ measures the degree of violation of the topological properties. But, as mentioned before, the delay coordinate mapping does not preserve the order of nearest neighbours. Therefore the above considerations only hold strictly for the case $k=1$. Going from m to $m+1$ the distances of nearest neighbours can become larger in an unsystematic manner. A small shift within the ordering of distances is the consequence even in case of a sufficient value of the embedding dimension (cf. fig. 1*b*)). The effect of noise works in a similar manner. This *wavering* of points has to be suppressed to the greatest possible extent in order to differentiate this effect from the violation of topological properties.

The first refinement of eq. (1) is achieved by taking the product

$$P_i(m,\tau)=\left(\prod_{k=1}^{p} Q_1(i,k,m,\tau)\right)^{1/p} \tag{2}$$

over a neighbourhood of the reference point containing $p\geqslant 1$ nearest neighbours. The p-th root suppresses the p-dependence of the product of distances. (We choose, *e.g.*, $p=10$ in case of $N=10\,000$ data points.) Permutations within the order of nearest neighbours by going from m to $m+1$ due to the wavering of points cancel out in the product of distances. We have $P_i\geqslant 1$, where $P_i=1$ only if the p nearest neighbours remain the same p nearest neighbours when going from m to $m+1$ irrespective of their arrangement. But it turns out that this measure P_i is not yet sufficient because P_i does not come close to 1 even in case of sufficient embedding dimension.

We need a second term which compensates the fact that some reconstruction points near the boundary of the neighbourhood of a reference point leave this neighbourhood by going from m to $m+1$ due to the delay coordinate mapping itself or due to added noise. For that reason we introduce a second ratio

$$Q_2(i,k,m,\tau)=\frac{\mathrm{dist}_m^\tau(i,j(k,m))}{\mathrm{dist}_m^\tau(i,j(k,m+1))} \tag{3}$$

analogous to the first one (eq. (1)), but here the distances are calculated in the *old embedding* $\boldsymbol{R}^m$. Generally, Q_1 exceeds the reciprocal of Q_2 in case of sufficient embedding dimension because of the continuity of the delay coordinate mapping([1]). The topological argument is that the distances of points to the reference point, which are small in $\boldsymbol{R}^{m+1}$, remain small when projected down to $\boldsymbol{R}^m$.

We define a product which compensates the wavering by

$$W_i(m,\tau)=\left(\prod_{k=1}^{p} Q_1(i,k,m,\tau)\,Q_2(i,k,m,\tau)\right)^{1/2p}, \tag{4}$$

and we call it the *wavering product*. If the wavering product is $W_i(m,\tau)\approx 1$, then the

---

([1]) Only in the irrelevant case of the transition from $\boldsymbol{R}$ to $\boldsymbol{R}^2$ can almost all points leave the neighbourhood of the reference point in $\boldsymbol{R}$ in such a manner that they have comparable distances in $\boldsymbol{R}^2$. This would result in a product $Q_1\cdot Q_2$ much smaller than 1.

topological properties are achieved locally in good approximation by the mapping into $\boldsymbol{R}^m$. Now we have to average the wavering product over a sufficiently large number of reference points, in order to get a measure for the whole accessible attractor. Since we are only interested in deviations of $W_i(m, \tau) \approx 1$, we introduce $\overline{W}(m, \tau) = \ln \langle W_i(m, \tau) \rangle_i$, where $\langle .. \rangle_i$ denotes the mean value of W_i (2). Furthermore we normalize the averaged wavering product and divide $\overline{W}$ by τ in order to eliminate the explicit τ-dependence of the wavering product, which occurs in case of sufficient embedding dimension (3).

The averaged wavering product $\overline{W}$ can easily be calculated numerically from scalar time series (4). We discuss some model system calculations (5). The treatment of the Rössler system [8] by most methods is very cumbersome because it is almost two-dimensional. For the most part algorithms calculating the minimal embedding dimension fail either by obtaining the underestimated value of $m = 2$ or by missing the convergence of the algorithm at $m = 3$. Figure 2 shows results for the Rössler system. A good convergence of $\overline{W}/\tau$ *vs.* m

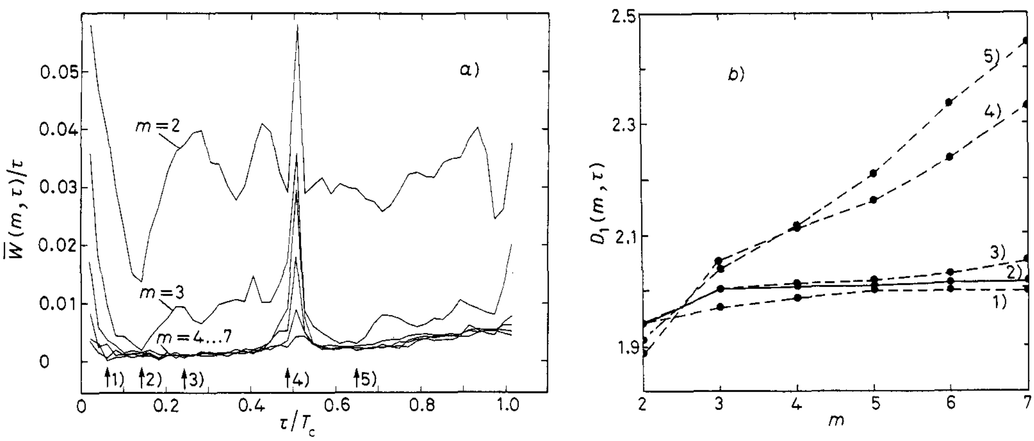

Fig. 2. – Numerical results for the Rössler system (see ref. [8]) with $a = 0.15$, $b = 0.2$, $c = 10$, $\delta t = \pi/25$ using $N = 10\,000$ data points. *a*) Averaged and normalized wavering product $\overline{W}/\tau$ for a range of embedding dimensions m and delay times τ in terms of the first recurrence time T_c. *b*) Information dimension *vs.* embedding dimension m for some values of delay time τ, as marked by numbered arrows in *a*).

(2) The average is taken over about 10% of the data points as reference points.

(3) We have introduced this τ rescaling in order to allow an appropriate comparison of values of $\overline{W}(m, \tau)$ for just sufficient embedding dimension with values of $\overline{W}(m, \tau)$ for not sufficient m using small τ-values. Otherwise one could go wrong in the determination of the optimal embedding parameter m because the τ-dependence of $\overline{W}(m, \tau)$ causes a relative flattening of $\overline{W}(m, \tau)$ for small τ-values.

(4) The nearest-neighbour algorithms can be accelerated by taking into account the structure of the delay coordinate mapping and that we are only interested in a limited number of nearest neighbours. Then we can avoid the calculation of the distances and the sorting for all points and for all embeddings by fixing a boundary distance—only a fraction part of the extension of the attractor—only beneath which we look for nearest neighbours. If this procedure fails for some reference points at higher embedding dimensions, we can use an algorithm suggested by FREEDMAN J. H. *et al.*, *IEEE Trans. Comp.* (1975), 1000. Using the maximum norm instead of Euclidean distances facilitates the computations.

(5) For the simple geometrical example of a 2-torus we get the correct result which is a minimal sufficient embedding $m = 3$ (cf. ref. [11]).

occurs at the correct value of the embedding dimension ($m = 3$) and for a delay time $\tau = 0.143\,T_c$ with T_c being the first recurrence time of the system. For this proper value of τ the averaged and normalized wavering product has its minimum. The information dimensions D_1 [9] show a clear convergence to the correct value slightly lying above 2.0 at the minimal embedding dimension for the optimal value of τ (full line in fig. 2*b*)). As expected, it turns out that, using almost all other values of τ, for the number of data points N and reference points chosen, the dimension estimation does not saturate even for $m \geqslant 2d + 1$ (cf. ref. [5], footnote 1). Specifically, the calculated dimensions converge using τ-values in the left neighbourhood of the optimal τ-value by enlarging the embedding dimension from $m = 4$ up to $m = 6$ such that $m \geqslant 2d + 1$ but they diverge for τ-values greater than the optimal one.

The Mackey-Glass system formally has an infinite number of degrees of freedom but its strange attractor has a finite dimension [10]. Therefore it mimicks a typical experimental situation. Figure 3 shows that the attractor of the Mackey-Glass system should be embeddable in a phase space of dimension $m = 4$ at a value of $\tau = 0.3\Gamma$, with $\Gamma = 30$ the internal time delay of the system. Note that larger values of τ only seem to be similarly appropriate, due to our τ rescaling of the wavering product.

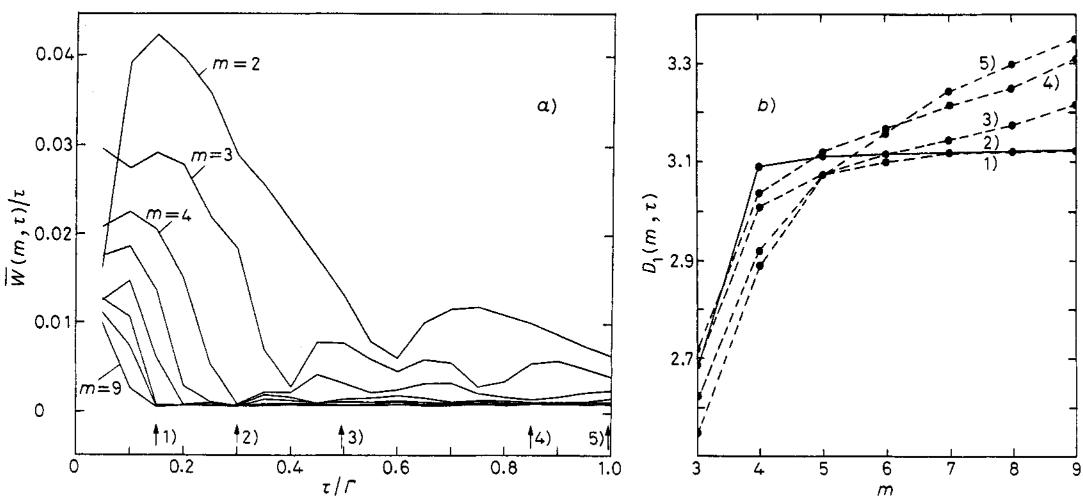

Fig. 3. – Numerical results for the Mackey-Glass equation (see ref. [10]) with $a = 0.2$, $b = 0.1$, $c = 10$, $\Gamma = 30$ using $N = 9000$ data points. Averaged and normalized wavering product and information dimension as depicted in fig. 2. The delay times are given in terms of the internal time constant Γ.

We would like to mention here that the above-calculated values for the optimal delay time are about 10% to 15% smaller than those we had obtained by another method based on the correlation integral [5] improving the information theoretical approach [6]. Furthermore, for the value of $\tau \approx 0.23\,T_c$ suggested in [6] for the Rössler system, calculated from the mutual information the convergence of D_1 *vs.* m, is obviously not as convincing so for the above-discussed optimal one (cf. fig. 2*b*)). Using the information theoretical approach of ref. [6] the underestimated value of $m = 2$ as an appropriate embedding dimension is suggested. In fact, the topological approach is superior to the information theoretical one.

We tested our algorithm also for noisy and short time series ($N \approx 2000$), for the Mackey-Glass system with $\Gamma > 30$ (resulting in larger embedding and fractal dimensions) and for further model system data (Lorenz, Duffing). We always obtain reliable values for the wavering product and for the calculated dimensions [11]. The wavering product has also been used successfully for experimental data [12].

To conclude, we have presented a method for solving simultaneously the problem of the proper choice of the delay time τ and of determining the minimal sufficient embedding dimension m for topologically correct attractor reconstructions, using finite scalar time series. This provides a solid basis for calculating characteristic quantities of experimental systems which display chaotic motion. Our method also yields an upper limit for the *number of degrees of freedom* of the dynamical system in its long-time behaviour and therefore a best estimate for the minimal number of equations needed for modelling the underlying dynamics of the attractor. From that we can improve algorithms calculating Lyapunov exponents by eliminating parasitic exponents [13]. Furthermore, determining m and τ means evaluating two important parameters for algorithms predicting dynamical systems [14].

* * *

This work was supported by the DFG-Sonderforschungsbereich *Nicthlineare Dynamik* Frankfurt/Darmstadt.

REFERENCES

[1] Eckmann J.-P. and Ruelle D., *Rev. Mod. Phys.*, **57** (1985) 617; Schuster H. G., *Deterministic Chaos*, 2nd edition (Physik-Verlag, Weinheim) 1988.
[2] First mentioned in Packard N. H., Crutchfield J. P., Farmer J. D. and Shaw R. S., *Phys. Rev. Lett.*, **45** (1980) 712.
[3] Mañé R., in *Springer Lecture Notes in Mathematics*, Vol. **898**, edited by D. A. Rand and L. S. Young (1981), p. 320; Takens F., in *Springer Lecture Notes in Mathematics*, edited by D. A. Rand and L. S. Young (1981), p. 366.
[4] Destexhe A., Sepulchre J. A. and Babloyantz A., *Phys. Lett. A*, **132** (1988) 101; Havstad J. W. and Ehlers C. L., *Phys. Rev. A*, **39** (1989) 845.
[5] Liebert W. and Schuster H. G., *Phys. Lett. A*, **142** (1989) 107.
[6] Fraser A. M. and Swinney H. L., *Phys. Rev. A*, **33** (1986) 1134; *IEEE Trans. Inf. Theory*, **35** (1989) 245.
[7] Froehling H., Crutchfield J. P., Farmer D., Packard N. H. and Shaw R., *Physica D*, **3** (1981) 605; Cremers J. and Hübler A., *Z. Naturforsch. A*, **42** (1987) 797; Cenys A. and Pyragas K., *Phys. Lett. A*, **129** (1988) 227.
[8] Rössler O. E., *Phys. Lett. A*, **57** (1976) 397.
[9] Grassberger P., *Phys. Lett. A*, **107** (1985) 101; Hentschel H. G. and Procaccia I., *Physica D*, **8** (1983) 435.
[10] Mackey M. C. and Glass L., *Science*, **197** (1977) 287; Farmer D., *Physica D*, **4** (1982) 366.
[11] Liebert W., *Chaos und Herzdynamik* (Harri Deutsch Verlag, Frankfurt am Main) 1990.
[12] Chennaoui A., Pawelzik K., Liebert W., Schuster H. G. and Pfister G., *Phys. Rev. A*, **41** (1990) 4151.
[13] Eckmann J.-P., Kamphorst S. O., Ruelle D. and Ciliberto S., *Phys. Rev. A*, **34** (1986) 4971; Liebert W., *Lyapunov Spectrum from Scalar Chaotic Time Series Revisited*, to be published.
[14] Pawelzik K. and Schuster H. G., *Unstable Periodic Orbits and Prediction*, in *Phys. Rev. A* (in press).