Quantifying the closeness of fractal measures

Holger Kantz*
Department of Theoretical Physics, University of Wuppertal, D 42097 Wuppertal, Germany
(Received 24 November 1993; revised manuscript received 21 March 1994)

Numerically, fractal measures are represented by finite sets of discrete points, e.g., those generated by trajectories. We introduce a cross-correlation integral to evaluate the distinctness and overlap of two such sets, and present a criterion to determine up to what precision the two sets are realizations of the same measure.
PACS number(s): 05.45.+b

I. OUTLINE OF THE PROBLEM

Modern experiments in nonlinear dynamics can easily produce time series of 10^5 points (e.g., [1]), which by appropriate embedding procedures yield representations of the underlying attractors. Correspondingly, by performing numerical simulations of chaotic dynamical systems invariant measures on fractal objects can be approximated by huge numbers of points. Generally, however, no analytical way exists to derive or describe fractal measures of realistic systems. Each of the above sets is a finite sample drawn according to the underlying unknown distribution. Since such a set of points formally has dimension zero, two different realizations (e.g., two trajectories of the same system) are disjoint with probability one, if the attractor is different from a periodic orbit. In certain situations it is important to know whether two such sets represent the same fractal measure. If they do not, how different are they? In the following we will present a way to give a quantitative answer to both issues.

We generalize the correlation integral of Grassberger and Procaccia [2], which we call autocorrelation integral henceforth, to cross correlations between two sets. If the two sets form similar attractors, then on large scales the two autocorrelation integrals and the cross correlation will have comparable values. The difference between the two attractors will show up on small scales, where the cross correlations rapidly decay to zero. We base the definition of a distance between the two attractors on the length scale at which the transition between the two different behaviors occurs.

Before we enter the theoretical details in Sec. II and some numerical examples in Sec. III we want to further illustrate the problem and to review two approaches which are not appropriate for our purpose.

To have a guideline for the following considerations let us give one example where the closeness of sets is of particular interest. In nonlinear time series analysis [3] one deals with (in general univariate) experimental data $\{s_t\}$, representing the state of a chaotic dynamical system at successive times t, $t = 1, \ldots, T$. By the help of embeddings [4], a state space can be reconstructed, in which the data form a fractal set. However, due to the presence of noise and the finiteness of the data set the fractal structure is veiled on small scales. For different purposes (e.g., prediction [5], computation of Lyapunov exponents [6], dynamical noise reduction [7]) it is necessary to construct a dynamical system consistent with the data, i.e., to fit a function F which under iteration reproduces as accurate as possible a noise-free version of the attractor. For the above-mentioned applications it is essential to have a quantitative estimate of the deviation between the original and the synthetic attractor. Another issue in the same context is whether two different experimental data sets from the same experiment under slightly different conditions form the same attractor or by what amount their attractors differ. Finally, when performing nonlinear noise reduction in an inappropriate way the resulting cleaned attractor can be strongly deformed. Therefore one wants to check whether the outcome is compatible with the original data.

The well-known *numerical recipes* [8] suggests using the χ^2 test in order to decide whether two samples are consistent with the same distribution. Adopting this for our situation we should introduce a binning in the embedding space of the two sets (box counting) and compute $\chi^2 = \sum_i (p_i - q_i)^2 / \sqrt{p_i q_i}$ (p_i and q_i are the probabilities of finding a point of the corresponding set in box i). χ^2 varies between zero (maximal agreement) and infinity (disjoint supports of the two sets). Indeed, we could learn about the scale on which the two sets are indistinguishable, if we computed χ^2 as a function of the box size ϵ. For very large ϵ we expect χ^2 to be close to zero, and to increase below a length scale where the differences between the two sets become visible. However, we do not want to use this method, first since box counting becomes very inefficient for fractal sets embedded in high-dimensional spaces, and second since χ^2 does not yield any further characterization of the two sets.

A refinement of this approach was suggested by Wright and Schult [9], but their goal is different from ours. In particular, they are able to decide up to which degree a given signal is contained in the signal under investigation. They consider projections of the invariant measure onto at most two dimensions. This avoids the inefficiency of box counting, but it prohibits the development of their

*Electronic address: kantz@wpts0.physik.uni-wuppertal.de

method towards a reasonable definition of a distance.

The Hausdorff distance is a mathematical concept used to characterize the distinctness of two sets [10]. If one defines the distance between a single point $\vec{x}$ and a set Y by

$$d(\vec{x}, Y) = \min_{\vec{y}\in Y}(||\vec{x} - \vec{y}||), \tag{1}$$

where $||\,||$ denotes some norm, then the Hausdorff distance between the sets X and Y is

$$h(X, Y) = \max\{\, \sup_{\vec{x}\in X} d(\vec{x}, Y)\,;\, \sup_{\vec{y}\in Y} d(\vec{y}, X)\}. \tag{2}$$

The Hausdorff distance has very nice properties, which make it valuable for mathematical purposes [10]. Unfortunately, for our purpose it is not appropriate, since it relies on the points of each set which are farthest away from the other set. Consider the situation that the two sets are realizations of the same Gaussian distribution. The largest distances $d(x, Y)$ will occur for x in the tails, where the densities are smallest. Obviously, this is not a good criterion for our purposes. This is closely related to the fact that the analog of Eq. (2) for measures (instead of finite sets) is sensitive only to the supports of the two measures, not to the measures themselves.

II. THE CROSS-CORRELATION INTEGRAL: THEORETICAL CONSIDERATIONS

Grassberger and Procaccia [2] introduced the (auto) correlation sum of a set X

$$C_{XX}(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - |\vec{x}_i - \vec{x}_j|), \tag{3}$$

where Θ is the Heaviside step function and $\vec{x}_i$ are elements of the set representing the invariant measure. For univariate time series $\vec{x}_i$ usually is a delay vector of sufficient dimension m. It is important [13] to exclude the pairs $i = j$. For numerical ease we will use the sup-norm, as usual.

If a given data set possesses a sufficiently strong self-similarity, there exists a range of ϵ where $C_{XX}(\epsilon)$ exhibits the scaling behavior $C_{XX}(\epsilon) \propto \epsilon^{D_2}$. The exponent D_2 is called the correlation dimension of the set X. D_2 is only one quantity out of a hierarchy of generalized dimensions D_q [11, 12] which characterize a fractal object. However, even if all D_q of two sets X and Y coincide, this does not guarantee that the two fractal measures are identical. Thus the mere computation of C_{XX} and C_{YY} does not suffice, especially not for the case where no clear scaling behavior is present and D_2 is undefined.

To compare two sets X and Y, we introduce the cross-correlation sum

$$C_{XY}(\epsilon) = \frac{1}{|X||Y|} \sum_{\vec{x}\in X} \sum_{\vec{y}\in Y} \Theta(\epsilon - |\vec{x} - \vec{y}|) \tag{4}$$

together with the autocorrelation sums $C_{XX}(\epsilon)$ and $C_{YY}(\epsilon)$. Note that the cross correlation, Eq. (4), is symmetric in X and Y. $|X|$ and $|Y|$ denotes the number of elements in each set. The autocorrelation sum, Eq. (3), thus is a special case of Eq. (4). In the remainder of this paper we will show how the knowledge of these three sums characterizes the relationship between the two finite sets.

To establish some rigorous results we will need two variants of the sums in Eqs. (3) and (4). On the one hand, if the number of points in each set goes to infinity, and μ and ν are the underlying measures, then the sums converge to integrals:

$$\begin{aligned}\tilde{C}_{\mu\nu}(\epsilon) &:= \lim_{|X|,|Y|\to\infty} C_{XY}(\epsilon) \\ &= \int \mu(\vec{x})\, \nu(\vec{y})\, \Theta(\epsilon - |\vec{x} - \vec{y}|)\, d^m x\, d^m y.\end{aligned} \tag{5}$$

On the other hand, for theoretical considerations sometimes it will be more convenient to deal with the box-counting equivalents of Eqs. (3) and (4). Consider a partition $\mathcal{P}_\epsilon$ in the (reconstructed) state space. Then p_i (q_i) is the probability to find a point of X (Y) in the ith box of $\mathcal{P}_\epsilon$, and

$$B_{XY}(\epsilon) = \sum_{\mathcal{P}_\epsilon} p_i q_i \propto C_{XY}(\epsilon). \tag{6}$$

In particular, $B_{XX}(\epsilon)$ has the same scaling properties as C_{XX} for self-similar sets [2]. Note, however, that we do not recommend the use of B_{XY} for numerical evaluations; it will be employed here only to establish some relations which are much harder to show for the correlation sums.

First let us show the validity of a more rigid version of the Schwarzian inequality for the three correlation sums. From $B_{XX} + B_{YY} = \sum_i (p_i - q_i)^2 + 2p_i q_i$ the inequality

$$B_{XY}(\epsilon) \le \frac{1}{2}[B_{XX}(\epsilon) + B_{YY}(\epsilon)] \tag{7}$$

follows immediately. Correspondingly, for the correlation integrals we have to evaluate

$$\begin{aligned}\tilde{C}_{\mu\mu} + \tilde{C}_{\nu\nu} - 2\tilde{C}_{\mu\nu} = \int\int \Theta(\epsilon - |\vec{x} - \vec{y}|)[\nu(\vec{x}) - \mu(\vec{x})] \\ \times[\nu(\vec{y}) - \mu(\vec{y})]\, d^m x\, d^m y.\end{aligned} \tag{8}$$

From that we would like to conclude

$$\tilde{C}_{\mu\nu}(\epsilon) \le \frac{1}{2}[\tilde{C}_{\mu\mu}(\epsilon) + \tilde{C}_{\nu\nu}(\epsilon)]. \tag{9}$$

The right-hand side of Eq. (8) is positive definite for sufficiently small ϵ. For smooth measures this can be seen easily, since (in one dimension) it is (apart from the normalization) nothing but the integral over the autocorrelation function of $f(x) = \nu(x) - \mu(x)$ from $-\epsilon$ to ϵ. Due to the assumed smoothness there has to be a finite interval around zero where the autocorrelation function is positive, hence Eq. (9) is valid for small ϵ.

Unfortunately, for the correlation sums the inequality is valid only on average, and may be violated in certain situations. One such example is a set Y which consists of the N points of X shifted by a very small amount, δ,

which is smaller than the minimal distance ϵ_0 between two points of X. Then the three correlation sums practically coincide for large ϵ. For $\delta < \epsilon < \epsilon_0$, $C_{XY} = 1/N$ remains constant, whereas $C_{XX} = C_{YY} = 0$, thus violating the inequality Eq. (9). Only for $\epsilon < \delta$ is it valid again. This could be remedied by adding the "diagonal" pairs $\vec{x}_i = \vec{x}_j$ to the autocorrelation sums, Eq. (3); but this destroys the scaling behavior of C on small scales. If instead these two sets X and Y represented two slightly different periodic orbits of the same period, then the measures μ and ν would consist of δ functions at each point of the sets X and Y, respectively. For this situation Eq. (9) is fulfilled for any ϵ.

The inequalities Eqs. (7) and (9) are of certain importance, since from them one immediately derives that the underlying measures are identical if the three correlations coincide on all scales:

$$\tilde{C}_{\mu\nu}(\epsilon) = \tilde{C}_{\mu\mu}(\epsilon) = \tilde{C}_{\nu\nu}(\epsilon),$$

$$\forall\epsilon \implies \mu(\vec{x}) = \nu(\vec{x}) \text{ almost everywhere.} \quad (10)$$

As we said, equal scaling properties of the two autocorrelation integrals only mean that the two fractal objects have equal correlation dimensions, while an agreement of the three integrals in the limit of large scales is trivially guaranteed by the normalization. The converse of our proposition is true due to the definition of $\tilde{C}_{\mu\nu}$.

If the box-counting versions of the correlation sums attain the same values for all ϵ, then $B_{XX}+B_{YY}-2B_{XY} = 0$ identically. Together with Eq. (7) it follows immediately that $p_i = q_i \forall i$ and for all partitions $\mathcal{P}_\epsilon$. Writing down the corresponding expression for the correlation integrals, $\tilde{C}_{\mu\mu}(\epsilon) + \tilde{C}_{\nu\nu}(\epsilon) - 2\tilde{C}_{\mu\nu}(\epsilon) = 0$ for all ϵ yields $\mu(\vec{x}) = \nu(\vec{x})$, since in the limit of small ϵ Eq.(9) is valid. Thus in both situations it is important that the correlations agree on small scales, which automatically induces an agreement on all larger scales. A mere agreement on large scales does not suffice.

Let us now evaluate $\tilde{C}_{\mu\nu}$ for measures which have different support. If the supports are disjoint, the cross-correlation integral is zero for sufficiently small ϵ. If instead the two supports intersect each other transversely, we claim that $\tilde{C}_{\mu\nu}(\epsilon) \propto \epsilon^m$, if $m \in \mathbf{N}$ is the minimal dimension needed to embed the union of the two measures. Due to the assumed independence of μ and ν, the dependence on ϵ remains unchanged, if one of the two measures is replaced by the Lebesgue measure on a smooth manifold, in which the corresponding measure is embedded. If the other measure is embedded in the same manifold, this yields the exponent m, as we claimed. If the other measure is not embedded in this manifold, one has to increase its dimension, otherwise this reasoning does not yield an answer [14]. As a test, one can easily evaluate Eq. (5) for two intersecting lines, say, the x and y axes. Choosing the points $(x, 0)$ of X as reference points, the contribution of the set Y to $C(X, Y)$ for given x is the part of the y axis falling into a square of length 2ϵ centered on $(x, 0)$. Thus we find

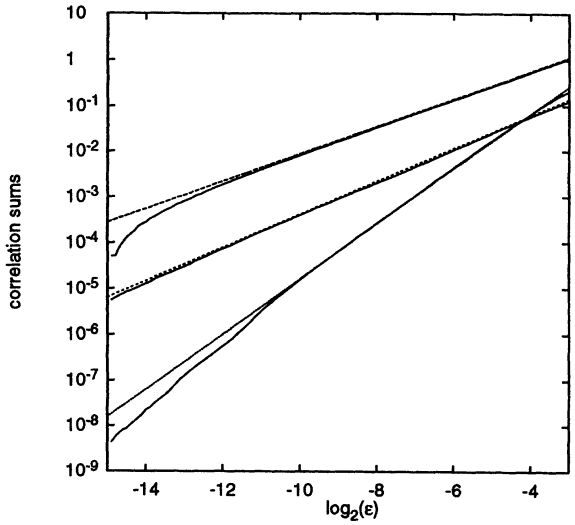

FIG. 1. The sums C_{XX} (top curve), C_{YY} (middle), and C_{XY} (bottom curve) as a function of ϵ, where set X consists of 5000 points representing the bisectrix in the plane and set Y consists of 5000 points of the Hénon attractor ($a = 1.4$, $b = 0.3$). The straight lines represent ϵ^α, where $\alpha = 1$, $\alpha = 1.2$, and $\alpha = 2$, respectively.

$$C_{\text{line-line}}(\epsilon) = \int dx \int dy\, \Theta(\epsilon - |(x,0) - (0,y)|_{\text{sup}}) = 4\epsilon^2 \quad (11)$$

independent of the dimension of the space in which the lines are embedded. The intersection of two planes yields exponents 2, 3 or 4, depending on whether the planes are identical, intersect themselves on a line, or (in a more than three-dimensional space) intersect in a point. Furthermore, we computed $C_{XY}(\epsilon)$ numerically for the intersection of the Hénon attractor, which has a correlation dimension of about 1.2, with the diagonal in two dimensions (2D). The result yields very precisely $C_{\text{Hénon, line}} \propto \epsilon^2$ (see Fig. 1).

III. NUMERICAL RESULTS

After these theoretical considerations we want to pass to the numerical computation of the cross correlations for different finite sets. We begin with different versions of the Hénon attractor. Since it can be embedded in two dimensions, we can visualize the attractors, which facilitates the interpretation of the results. In practice, however, the considerations presented in this letter become more relevant in higher dimensions, where visual inspection can be applied only to projections of the sets and thus is less reliable. In such cases we suggest computing the correlation integrals for one sufficiently large value of the embedding dimension.

In our first example, set X is a trajectory of 5000 points (after discarding transients) of the Hénon system, $x_{t+1} = 1 - ax_t^2 + bx_{t-1}$ with the standard parameters [15]. The set Y consists of the same points, but disturbed by 3% additive uniformly distributed noise. The two data sets are shown in Fig. 2 for a two-dimensional embedding.

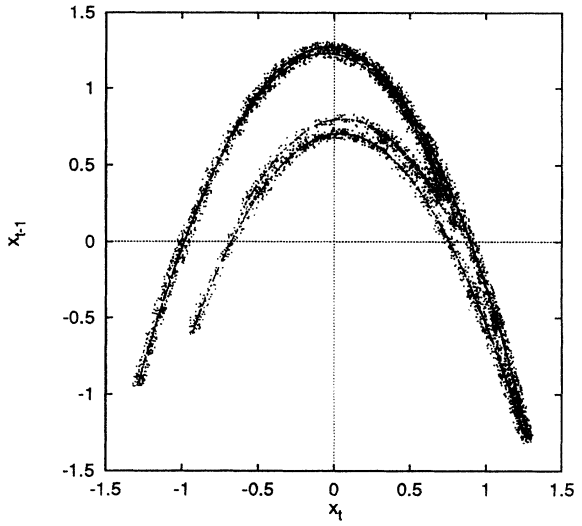

FIG. 2. 5000 points of a clean Hénon trajectory (a=1.4, b=0.3) and the same data contaminated by 3% additive noise.

In Figs. 3 (a) and 3(b) we present the three correlation sums and the local scaling exponents, $d \ln C / d \ln \epsilon$.

The correlation sum of the clean data exhibits a scaling range $C_{XX}(\epsilon) \propto \epsilon^{D_2}$, where $D_2 \approx 1.2$ is the correlation dimension of the Hénon attractor. The noisy data show the same behavior on large scales, but on smaller scales the correlation sum decays much faster with decreasing ϵ, representing the fact that the data points are scattered in the plane. Consequently, the scaling exponent is 2 in our two-dimensional embedding. The cross correlation C_{XY} shows exactly the same behavior as C_{YY}. The interpretation is simple: since the clean data are embedded in the noisy ones, taking them as reference points for the correlation sum is essentially the same as taking a corresponding subset of Y as reference points, which leads to C_{YY}.

Next we compare two clean Hénon attractors for slightly different parameter values (Fig. 4). Again, we show the correlation sums together with the local scaling exponents (Fig. 5). The two autocorrelation sums show that the two slightly different attractors have nearly the same correlation dimension, $D_X \approx 1.2$ and $D_Y \approx 1.15$. The cross-correlation sum essentially coincides with them for $\epsilon > \epsilon_s = 2^{-6}$. Below we will use this coincidence to define the closeness of the two attractors. Only for scales below ϵ_s does it become evident that the two sets of points do not represent the same measure, and, in particular, have a different support. Below this scale, the scaling behavior of C_{XY} is determined by the intersections of the two sets. Thus the number of pairs with a distance smaller than ϵ decays rapidly with ϵ. As we claimed before, this decay again is characterized by a power law, and the slope is close to $m = 2$ [note that the fluctuations in Fig. 5(b) are anticorrelated and one should fit a straight line to the curves in panel (a) rather than averaging over the fluctuations in panel (b)]. The slight violation of our prediction $m = 2$ presumably arises from the fact that some intersections are close to tangential

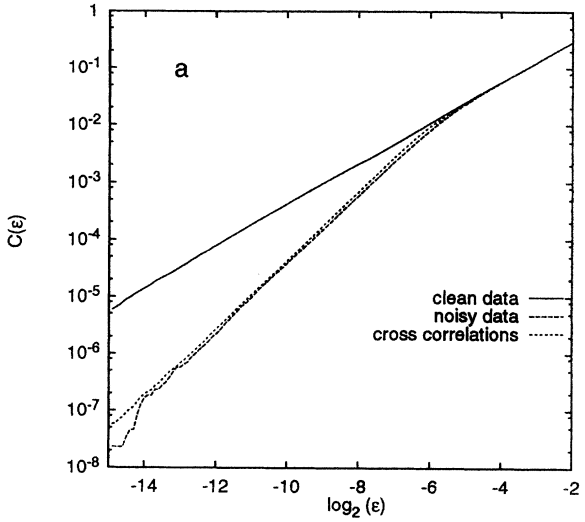

FIG. 3. The three correlation sums (a) and the local scaling exponents (b) for the two data sets shown in Fig. 2 in a two-dimensional embedding.

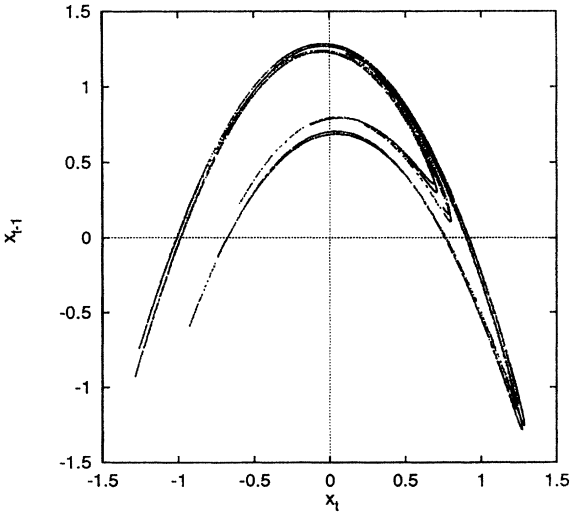

FIG. 4. Two Hénon trajectories of length 5000 each, a=1.4, b=0.3 and a=1.35, b=0.31, respectively.

and that due to the small number of pairs contributing to C_{XY} an overestimation of the scaling exponent is quite likely [16].

If we enlarge the embedding dimension, generically there will be no true intersections left, since almost any intersection in two dimensions will be transformed into a disjoint passage. This results in a much faster decay of $C_{XY}(\epsilon)$ for $\epsilon < 2^{-6}$ with an exponent between 3 and 4, as shown by the dotted curve in Fig. 5.

Finally, if the two sets already had no intersections in two dimensions due to some shift, then the cross-correlation sum would again contain a part above some value ϵ_s, where it is almost identical to the two correlation integrals, and below ϵ_s it would rapidly decay to zero.

IV. THE DISTANCE BETWEEN FINITE SETS

As we have seen, by the help of the cross-correlation sum together with the standard correlation sums one can characterize the relationship between two sets. Now we want to summarize the information contained in plots like Figs. 3 and 5 in a single number, a characteristic length scale. Having this in mind, a natural definition of a distance between two sets would be the value of ϵ, below which the cross correlations start to differ significantly from the autocorrelations. However, we do not like the word *distance* for this length scale, since this fits badly to intuition in the case that one set is embedded in the support of the other, as in Fig. 2. Therefore we coin the term C_2 *similarity*. We call two sets C_2 similar above the value of ϵ, when $C_{XY} \simeq C_{XX} \simeq C_{YY}$. For a precise definition we have to introduce a tolerance ρ, such that we can define the C_2-similarity scale ϵ_s of two sets:

$$\begin{aligned} &\ln C_{XX} - \ln C_{XY} < \rho \wedge \ln C_{YY} - \ln C_{XY} < \rho \\ &\qquad\qquad \text{if } \epsilon > \epsilon_s, \\ &\ln C_{XX} - \ln C_{XY} > \rho \vee \ln C_{YY} - \ln C_{XY} > \rho \\ &\qquad\qquad \text{if } \epsilon < \epsilon_s, \end{aligned} \tag{12}$$

where $\wedge$ and $\vee$ denote the logical "and" and "or," respectively. This means that for $\epsilon > \epsilon_s$ the two sets are indistinguishable (up to the arbitrarily fixed precision ρ) when viewed by C_{XY}. For the data shown in Fig. 2 this is $\epsilon_s \approx 2^{-4}$ (the variance of the noise is about one-sixteenth of the attractor), and for those in Fig. 4 it is $\epsilon_s \approx 2^{-6}$ with a very weak dependence on ρ. Note that in the latter example C_{XX} and C_{YY} alone do not suffice to determine the length scale below which the differences of the attractors are dominant.

Finally we want to demonstrate that with the help of C_2 similarity two different trajectories of the same system are clearly identified to represent the same measure. In this case ϵ_s should be of the same order of magnitude as the lower end of the scaling regions of the autocorrelation sums (which is about T^{-1/D_2}, T the trajectory length). For smaller ϵ all three curves will fluctuate due to a lack of neighbors. Nevertheless, when interpreting the range of these fluctuations as statistical errors, all three correlation sums should agree also on the small scales.

We want to present two examples of our method applied to experimental data. The first consists of two very short parts of an experimental data set obtained from the Zürich NMR laser [1]. The first and the last 300 data points out of a time series of length 40 000 are taken as two different trajectories. The embedding dimension is $m = 4$. From a complete data analysis [18] it is known that the data are subject to noise of about 1.8%. The correlation dimension of the clean attractor (taking the whole data set and after noise reduction) is about 1.5 [18]. Due to our criterion developed above from Fig. 6 one would clearly decide that the two trajectories do represent the same attractor. Of course the agreement becomes even better if we take longer trajectories.

The same goes for the data shown in Fig. 7. The first and last 500 points from 10 000 points recorded from a far-infrared- (FIR) laser experiment [17] are taken as two sets. They are embedded in four dimensions. The data contain 0.4% noise [19], including the discretization error due to an 8 bit analog-to-digital converter (the latter leading to the steps in Fig. 7). The attractor dimension is slightly above 2.

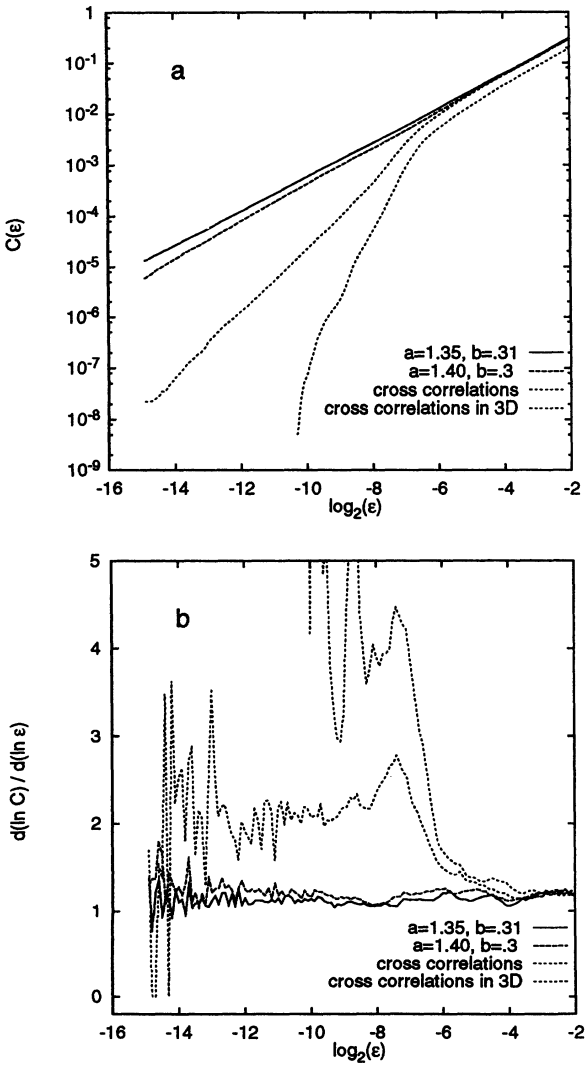

FIG. 5. The correlation sums (a) and local scaling exponents (b) for the data of Fig. 4.

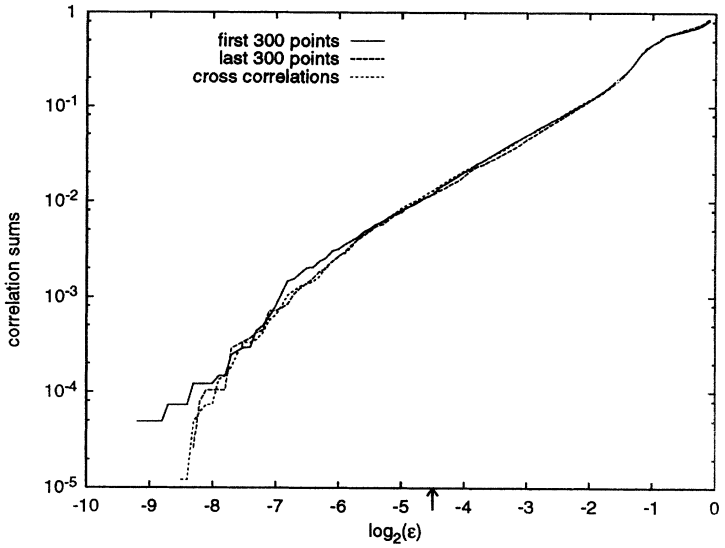

FIG. 6. The three correlation sums for the first and last 300 points of an experimental time series of length 40 000, taken from the Zürich NMR-laser experiment [1], embedded in four dimensions. The arrow indicates $\epsilon = T^{-1/D_2}$.

Thus we conclude that two sets represent the same fractal measure, if the C_2-similarity scale is of the order of its natural minimum and if in addition the three sums agree on smaller scales within their assumed statistical errors. The minimum of ϵ_s is determined by the onset of statistical fluctuations in the correlation sums, and in the presence of a clear scaling behavior it is of the order of $\epsilon_s \approx T^{-1/D_2}$. As we have shown, even very short experimental data sets can yield stable results. Comparing two parts of a data set with a large temporal distance in this way can also serve as a simple test for stationarity.

To anticipate misunderstandings let us repeat that the characterization of two sets in terms of the cross-correlation sum and in particular the definition of ϵ_s by no means depends on the existence of a scaling range in any of the three correlation sums. Therefore the C_2-similarity scale ϵ_s is a useful characteristic even in cases where the correlation integrals do not lead to a reliable estimate of the attractor dimensions.

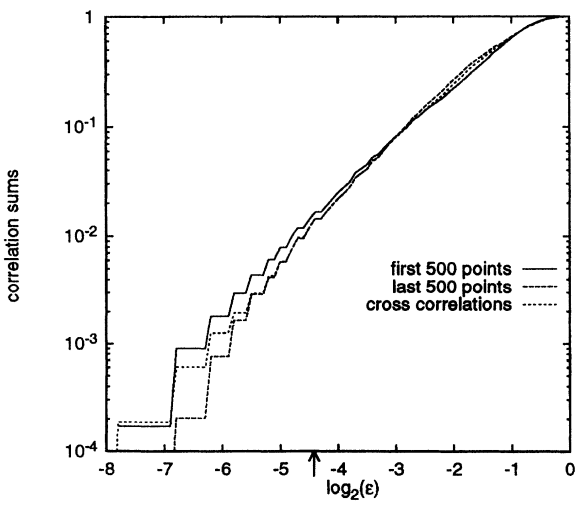

FIG. 7. The three correlation sums for the first and the last 500 points from a data set of length 10 000, obtained from a FIR-laser experiment in Braunschweig [17], embedded in four dimensions. The arrow indicates $\epsilon = T^{-1/D_2}$.

Finally, the C_2-similarity scale can be used to compare more than two sets. We propose that the triangular inequality holds:

$$\epsilon_s(XZ) \leq \epsilon_s(XY) + \epsilon_s(YZ). \tag{13}$$

Intuitively this is clear, since C_2 similarity means that an ϵ_s tube around the one set covers an essential part of the other set, such that Eq. (13) leads back to the corresponding inequality for the norm used in Eq. (4).

V. CONCLUSIONS

We have introduced the cross-correlation sum to characterize the relation between two finite sets. Based on this, we have defined the C_2-similarity scale, which is the length scale above which the cross correlations between the sets are essentially identical to their autocorrelations. The C_2-similarity scale has the properties of a distance and thus can be used to compare more than two sets. In numerical examples we have demonstrated the usefulness of this concept and shown that in fact different short parts of a long experimental data set will be recognized as belonging to the same attractor.

Nevertheless, one problem remains. The cross-correlation sum (as any other real space method like the χ^2 test and the Hausdorff distance) sensitively depends on any translation or rescaling of one data set with respect to the other. In cases where no intrinsic scale is given one should define the cross-correlation sum as the maximum over all possible translations, rotations, and (possibly nonlinear) rescalings of one of the two sets. The maximum of C_{XY} is obtained when the overlap of X and Y is largest, such that the C_2-similarity scale ϵ_s is minimized.

For the applications we have in mind this problem will be of minor importance. With respect to nonlinear time

series analysis one task is the comparison of a given data set and an attractor generated after processing the data (e.g., with data from the fitted dynamics or after nonlinear noise reduction). In this case a natural range for the synthetic data exists and it is exactly the deviation of the two sets within this reference frame which one wants to measure. If, instead, two experimental time series are to be compared, then the problem of a relative scale and offset may be important. Still, if one can assume that the data are related by linear transformations, normalizing the two scalar distributions of the data will suffice.

ACKNOWLEDGMENTS

I want to thank P. Grassberger and T. Schreiber for fruitful discussions. I am particularly grateful to G. Mantica, from whom I learned about the Hausdorff distance between sets. Furthermore, I owe thanks to R. Brown for a careful reading of the manuscript.

[1] P. Bösinger, E. Brun, and D. Meier, Phys. Rev. Lett. **38**, 602 (1977).
[2] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).
[3] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos **1**, 521 (1991).
[4] T. Sauer, J.A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[5] *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, SFI Studies in the Sciences of Complexity Proc. Vol. XV (Addison-Wesley, Reading, MA, 1993).
[6] J.-P. Eckmann, S. Oliffson–Kamphorst, D. Ruelle, and S. Ciliberto, Phys. Rev. **A 34**, 4971 (1986).
[7] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, Chaos **3**, 127 (1993).
[8] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, England, 1992), Sec. 14.3.
[9] J. Wright and R.L. Schult, Chaos **3**, 295 (1993).
[10] M. Barnsley, *Fractals Everywhere* (Academic Press, San Diego, 1988).
[11] P. Grassberger, Phys. Lett. **97A**, 227 (1983).
[12] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman, Phys. Rev. A **33**, 1141 (1986).
[13] P. Grassberger, Phys. Lett. A **128**, 369 (1988).
[14] Another argument is as follows: Consider the box-counting equivalence B_{XY}, Eq. (6). The number of boxes, for which both p_i and q_i are different from zero, scales ϵ^{-D_s}, where $D_s = D_X + D_Y - m$ is the dimension of the intersection of the two sets (m is the embedding dimension). The weights of nonempty boxes scale as $p_i \approx \epsilon^{D_X}$ and $q_i \approx \epsilon^{D_Y}$, respectively, such that we expect that $B_{XY} \propto \epsilon^m$ for $D_X + D_Y \geq m$. Note that this reasoning in particular contains the situations discussed above (line-line, line-plane), and furthermore it correctly predicts the behavior of B_{XY} in the case where the two sets are identical, since then $D_s = D_X = D_Y$, such that $B_{XY}(\epsilon) \propto \epsilon^{-D_s+D_X+D_Y} = \epsilon^{D_X}$.
[15] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[16] J.B. Ramsey and H.-J. Yuan, Nonlinearity **3**, 155 (1990).
[17] U. Hübner, N.B. Abraham, and C.O. Weiss, Phys. Rev. A **40**, 6354 (1989).
[18] H. Kantz, T. Schreiber, I. Hoffmann, T. Buzug, G. Pfister, L.G. Flepp, J. Simonet, R. Badii, and E. Brun, Phys. Rev. E **48**, 1529 (1993).
[19] H. Kantz, in *Time Series Prediction: Forecasting the Future and Understanding the Past* (Ref. [5]).