10

Volume 133, number 6 PHYSICS LETTERS A 21 November 1988

INTRINSIC LIMITS ON DIMENSION CALCULATIONS

Leonard A. SMITH
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, UK

Received 15 June 1988; accepted for publication 2 September 1988
Communicated by A.P. Fordy

The combined influences of boundary effects at large scales and nonzero nearest neighbor separations at small scales are used to compute intrinsic limits on the minimum size of a data set required for calculation of scaling exponents. A lower bound on the number of points required for a reliable estimation of the correlation exponent is given in terms of the dimension of the object and the desired accuracy. A method of estimating the correlation integral computed from a finite sample of a white noise signal is given.

Since Richardson's early attempts to quantify the complexity of coastlines and natural boundaries [1], a variety of methods have been devised for computing scaling exponents to describe complex distributions. The method most commonly employed today is the Grassberger–Procaccia algorithm (GPA) [2,3], which estimates the correlation exponent, ν, from a data set. In this Letter, we determine an intrinsic limit to the accuracy of calculations which consider the global scaling of a statistic on an object [#1]. Specifically, the number of points required to estimate the correlation exponent of a nonlacunar set to within 5% of its true value increases at least as fast as

$$N_{\min} \geqslant 42^M , \tag{1}$$

where M is the greatest integer less than the dimension of the set. It is stressed that $N_{\min}$ is a necessary bound on the amount of data required for a reliable calculation of ν. Many data sets in published work do not satisfy even this requirement, and, in many circumstances, more data is required.

Correlation exponent. First we consider the problem of determining the correlation exponent of an object in $\mathbb{R}^M$. Consider a set of points x_i ($i=1, 2, 3, \ldots, N$) distributed over the object. The correlation integral, $C_2(l)$, is defined as [2,3]

$$C_2(l) = \lim_{N\to\infty} (1/N^2) \sum_{i=1}^{N} \sum_{j=1}^{N} \Theta(l - |\boldsymbol{x}_{ij}|) , \tag{2}$$

where N is the number of points in the set, $\boldsymbol{x}_{ij}$ is the vector separating the ith and jth points, and Θ is the Heaviside function which is equal to zero for negative argument and one otherwise. For convenience, we shall use the supremum norm, so that the magnitude of a vector is the maximum of its cartesian components. This choice of norm is also convenient in numerical computations. As l approaches zero, we expect the correlation integral to have the form [4,5]

$$C_2(l) = \chi(l) l^{\nu} , \tag{3}$$

where ν is called the correlation exponent and is equal to the generalized Renyi dimension d_2 [4,21] and $\chi(l)$ is a possibly oscillatory O(1) function of l which reflects the lacunarity of the set [5,18–20]. The structure of $\chi(l)$ is generated by sparse or empty regions (lacunae) in the object. In the correlation integral, this oscillatory structure is most clearly expressed in strictly self-similar sets where $\chi(l)$ is strictly periodic in $\log(l)$.

Operationally, it is useful to define a function $\hat{\nu}(l)$ through

[#1] Throughout this Letter the term object will signify the underlying geometric object (e.g. strange attractor or physical boundary) while a data set consisting of a finite collection of points which approximate an object will be referred to as a set or a reconstruction.

$$\hat{\nu}(l) = \frac{\mathrm{d}\log C_2(l)}{\mathrm{d}\log l}, \tag{4}$$

where $C_2(l)$, and thus $\hat{\nu}(l)$, is computed from a data set consisting of a finite number of points. The structure of $\hat{\nu}(l)$ is determined by both the structure of the object and observational limits such as the noise level and the amount of data. When calculated from a large but finite number of data points, $\hat{\nu}(l)$ displays four regions of different behavior [8,9] as shown in fig. 1. At length scales less than the mean nearest neighbor distance, l_{nn}, the set scales as a collection of isolated points and $\hat{\nu}(l)$ tends to zero (region I). Next in increasing l, is a region dominated by any noise in the data (region II). Here $\hat{\nu}(l)$ approaches (but is less than) the dimension of the embedding space, independent of the true value of ν of the object. It is in the next region, region III (often called the plateau), that $\hat{\nu}(l)$ approximates ν. In general, $\hat{\nu}(l)$ fluctuates about ν due to $\chi(l)$; if $\chi(l)$ is constant (i.e. the set is nonlacunar), then there may exist a plateau of slowly decreasing $\hat{\nu}(l)$. The

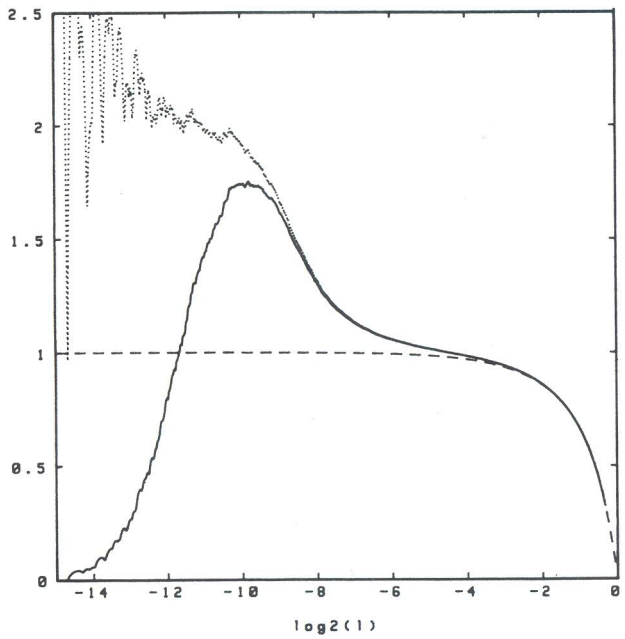

Fig. 1. Theoretical and observed correlation exponent functions computed with a data set describing a line ($\nu=1$) consisting of $N=2^{14}$ points where pseudo-random noise (uniformly distributed on $\pm 2^{-9}$) has been added to each point. In this case $M=2$, $r=1.5$; $\hat{\nu}(l)$ is shown with the self-pairs included (solid curve) and omitted (dotted), the dashed curve represents $\bar{\nu}(l)$. All logarithms are taken base 2.

estimated correlation exponent, ν_{est}, is computed from $\hat{\nu}(l)$ in this region; for cases in which we are interested in determining the largest value of ν_{est} which should be expected for a given N, ν_{est} will be taken as the maximum of $\hat{\nu}(l)$ in this region. Region III is followed by a more rapid transition as $\hat{\nu}(l)$ decreases to zero for l greater than the diameter of the set (region IV). The unit of length is chosen such that the diameter of the set is equal to one. By estimating the width of the outermost and innermost transition zones we compute a limit on the number of points required for region III to be observable.

For finite data sets, it is essential that either the double sum of the correlation integral is defined by eq. (1) including the $i=j$ terms (hereafter called self-pairs) or l_{nn} is calculated explicitly. This is required in order to distinguish the scaling of true noise at small l from fluctuations due to finite N. When the self-pair terms are omitted, $\log C_2(l)$ is not bounded as l goes to zero; fluctuations in $\log C_2(l)$ arise due to the sparsity of points with neighbors closer than l_{nn}. Steep descents in $\log C_2(l)$ result in large values of $\hat{\nu}(l)$ (which may exceed M) and may be misinterpreted as being characteristic of the noise region. Recently, Grassberger [22] has determined finite sample corrections which extract information from these length scales when computing the spectrum of Renyi dimensions. Alternatively, when the $i=j$ terms are included in computing d_2 the nearest neighbor statistics are reflected in the transition from region I to region II (computed below).

Hypercubes. In the case of an ideal, noise free data set, $\hat{\nu}(l)$ will differ from ν due to the lacunarity of the object and the finite diameter of the object (edge effects). Guckenheimer [10] noted that edge effects might bias the calculation of scaling exponents; we isolate this effect by considering a uniformly covered hyper-cube of dimension M. Hyper-cubes are nonlacunar ($\chi(l)$ is constant), and the correlation integral may be computed in the limit of infinite N. From this computation we determine a function $\bar{\nu}(l)$ which characterizes the behavior of $\hat{\nu}(l)$ at large scales.

In the limit of infinite N, the correlation integral is equal to the probability that the distance separating two randomly chosen elements, x and y, is less than l. For the one dimensional hyper-cube (the unit interval) this probability is

$$P(|x-y|<l)=l(2-l)\,. \tag{5}$$

For an M-dimensional cube, each of the M elements of the separation vector $|\boldsymbol{x}-\boldsymbol{y}|$ are independent, hence

$$P(|\boldsymbol{x}-\boldsymbol{y}|<l)=[P(|x-y|<l)]^M\,. \tag{6}$$

Thus the correlation integral of an M-dimensional cube is

$$\bar{C}_2(l)=[l(2-l)]^M \tag{7}$$

and

$$\bar{\nu}(l)=M[1-l/(2-l)]\,. \tag{8}$$

Note that $\bar{\nu}(l)$ decreases with increasing l and always underestimates ν.

We now relate the accuracy of an estimate of ν (which should, of course, equal M) to the number of points in the data set by defining a quality factor, Q, and requiring

$$\bar{\nu}(l)\geqslant Q\nu, \quad 0\leqslant Q\leqslant 1\,. \tag{9}$$

Solving eq. (8) for l then provides an outer limit $l_{\max}$, such that for $l>l_{\max}$, $\hat{\nu}(l)$ does not provide a sufficiently accurate estimate of ν.

Estimating ν from $C_2(l)$ requires a knowledge of $C_2(l)$ over a range of $l<l_{\max}$, denote this range by R (where $R\equiv l_{\max}/l_{\min}\geqslant 1$). Three factors govern the choice of R. The first factor is the smallest range sufficient to estimate the slope of $\log C_2(l)$, call this range r. The second is the minimum acceptable width of region III (the "plateau"); if R is equal to one this region consists of only a point. Finally, in lacunar sets there is no flat plateau, the value of $\hat{\nu}(l)$ will fluctuate and R must be chosen sufficiently large to identify this behavior. Bearing these constraints in mind, we solve for $l_{\min}$

$$l_{\min}=\frac{2(1-Q)}{R(2-Q)}\,. \tag{10}$$

Resolving an M-dimensional hyper-cube to a scale $l_{\min}$ requires at least

$$M_{\min}\geqslant(1/l_{\min})^M \tag{11}$$

points. Thus a lower bound on the number of points required for an estimate of ν to exceed a fraction Q of the true value is

$$N_{\min}\geqslant\left(\frac{R(2-Q)}{2(1-Q)}\right)^M\,. \tag{12}$$

Assuming $Q=0.95$ and $R=4$ yields eq. (1).

When a fixed number of pseudo-random points are embedded in spaces of increasing M, the observed decrease in the maximum value of $\hat{\nu}(l)$ is faster than that predicted from eqs. (8) and (11). One reason for this is the nonzero separation of nearest neighbors; considering the transition from point-like scaling provides an estimate of $\hat{\nu}(l)$ at small length scales, the function $\underline{\nu}(l)$, which we now calculate under the assumption of a uniform density.

At small scales, the number of points within a distance l of a given point $\boldsymbol{X}$ (far from any boundary) is

$$k(l)=\rho V(l)+1\,, \tag{13}$$

where ρ is the density of points, $V(l)$ is the volume within a distance l and the addend of unity represents the point at $\boldsymbol{X}$. At length scales smaller than l_{nn}, the quantization of $k(l)$ becomes important. Assuming a uniform density, $\rho=N$, we have

$$k(l)=N(2l)^M+1, \quad l_{\mathrm{nn}}\approx l\ll 1\,. \tag{14}$$

The correlation integral is the normalized sum over all points, or

$$\underline{C}_2(l)=\frac{1}{N^2}\,[N^2(2l)^M+N]\,, \tag{15}$$

which implies

$$\underline{\nu}(l)=\frac{M}{1+N^2(2l)^M}\,N^2(2l)^M\,. \tag{16}$$

$\underline{\nu}(l)$ is an upper bound on $\hat{\nu}(l)$ in the sense that it is limited only by the nonzero l_{nn} and is not sensitive to edge effects. Now as $\overline{\nu}(l)$ assumes a zero l_{nn} and places an upper bound on $\hat{\nu}(l)$ due to the presence of boundaries, and these functions are, respectively, nondecreasing and nonincreasing, an upper bound on the observable values of $\hat{\nu}(l)$ occurs at their intersection. With the above assumption of a uniform density, a stronger bound on the number of points needed to observe $\hat{\nu}(l)$ greater than $Q\nu$ is found by requiring that the intersection of $\overline{\nu}(l)$ and $\underline{\nu}(l)$ occur at $l=l_{\min}$. The agreement between $\underline{\nu}(l)$ and $\hat{\nu}(l)$ at small l may be improved by defining $\underline{\nu}(l)$ as a centered finite difference of $\log \underline{C}_2(l)$ over a step

in $\log l$ equal to $\log R$. Simultaneous plots of $\bar{\nu}(l)$, $\underline{\nu}(l)$ and $\hat{\nu}(l)$ for low dimensional examples (see fig. 2), show that behavior of $\hat{\nu}(l)$ is well described; calculation of the value of the intersection of $\bar{\nu}(l)$ and $\underline{\nu}(l)$ provides a quick method of estimating the maximum value of $\hat{\nu}(l)$ observed for embedded white noise as a function of M, N, and R.

Time series. When analyzing a data set which describes a physical distribution or a solution to a system of ODEs, the dimension of the embedding space is known. When applied to time series data, the GPA computes the correlation integral of the data set reconstructed in successively higher embedding [3,6,7]. For each embedding dimension, M, an estimate of the correlation exponent, ν_{est}, is determined from $\hat{\nu}(l)$. As N approaches infinity, ν_{est} approaches M whenever $\nu \geqslant M$. For finite data sets, this is no longer the case. For a given N, solving eq. (12) for M provides a method for selecting the highest embedding dimension for which we can expect accurate computations. If ν_{est} for this embedding is not significantly less than QM, the deviation of ν_{est} from M calculated in higher dimensional embeddings does not imply a fractal structure. Indeed such a deviation is expected due to the increasing volume of higher dimensional hypercubes (in which a fixed number of points are embedded). A data set of 2^K randomly distributed data points covers an M-dimensional square region down to a scale of $l_{\text{nn}} \approx 2^{-K/M}$. For $M=2$ and $K=10$, $l_{\text{nn}}=0.01$. The same set embedded in a five-dimensional space yields an $l_{\text{nn}}=0.25$. As l_{nn} increases, so does the difference between ν_{est} and M (see eq. (8)). Embedding a fixed number of points in higher dimensional reconstructions forces the apparent convergence of the correlation exponent.

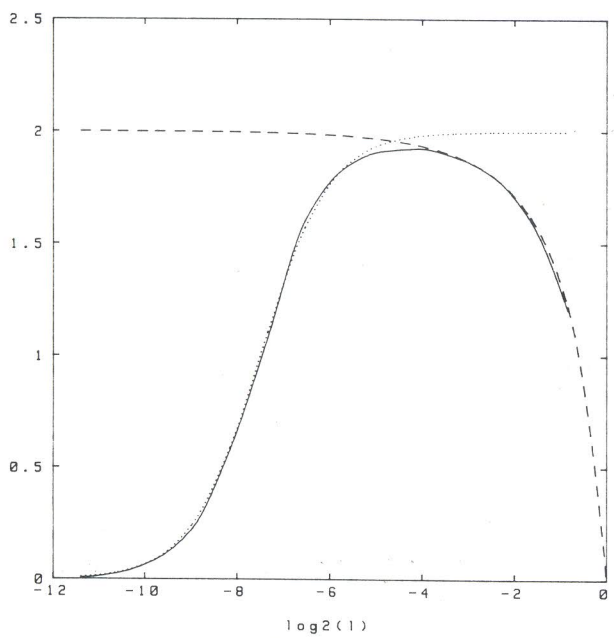

Fig. 2. Theoretical and observed correlation exponent functions computed from a data series generated by the cat map ($\nu=2$) with $M=2$, $N=2^{13}$, $r=2.0$. Curves of $\hat{\nu}(l)$ (solid), $\bar{\nu}(l)$ (dashed), and $\underline{\nu}(l)$ (dotted) are shown.

Note that a data set consisting of N_{min} points may not be sufficient even in the case of a nonlacunar object. For example, when reconstructing a data series from the cat map [11] with $M=3$ the data lie on a finite number of two-dimensional sheets; to observe the asymptotic behavior of $\hat{\nu}(l)$, length scales smaller than the separation of the nearest sheets must be sampled. (A similar behavior occurs with multiplicative congruential pseudo-random number generators for arbitrary M.) Even when the embedding dimension is known, strong inhomogeneities in the observed density may result in biased estimates of ν [12].

It remains to relate N_{min} to the number of points in a time series of length T. The hyper-cube argument above assumed that the points in the set were randomly distributed on the object. The points in a data set constructed from a time series are not, in general, uniformly distributed over the object. More importantly, these points are provided by single trajectory and thus are dynamically related [#2]. For chaotic phenomena, this effect is thought to be minor when the series is long relative to the decay of correlations. (For quasi-periodic signals, this difficulty remains.) For both types of systems, the number of points to be compared with N_{min} is determined by the dynamical time scale of the system, not a time scale of the experimental apparatus.

At least three time scales are relevant when embedding a time series. The sampling time, τ_s, determined by the experimental sampling rate, the de-

[#2] This ordering may in fact provide additional information useful in determining the statistics of the set [13,14]. In this paper, only the geometry of the point set is considered; if additional information encoded in the ordering of the points can be used to extend the series, the limits determined here would apply to the extended record.

lay time, τ_i, determined by the rate at which new information appears in the system [15], and the reconstruction time, τ_r, which reflects the separation of the consecutive base points used in the reconstruction. The reconstruction time is often chosen to be equal to the sampling time; this may bias $\hat{\nu}(l)$ due to the one-dimensional nature of the trajectory. While the effect of dynamically close points (points contributed from the same stretch of time series) may be removed [16], the effect is still present from data points in regions where the trajectory passes near itself. One method to limit this effect globally is to choose τ_r based on a statistic of the nearest neighbor separation of dynamically distant points (points constructed from portions of the time series many τ_i apart). A second method is to allow τ_r to vary and record points separated by a given arc-length along the trajectory. In either case, calculating the typical (dynamically distant) nearest neighbor separation provides a direct estimate of l_{min} of the data set; the choice of a τ_r which results in a decrease in the average nearest neighbor separation below this value does not significantly improve the covering of the object.

The assumption of randomly placed points implies that the number of points in a reconstruction for the GPA is $N_0 = T/\tau_i$. In practice, additional points may be used as long as dynamically nearby points do not dominate the scaling. Specifically, when the series is reconstructed with $\tau_r = \tau_i$, dynamically closest points are separated, on average, by the mean distance between two randomly chosen points (in M dimensions) $\sigma(M)$, while $l_{nn} \approx (1/N_0)^{1/M}$. The number of points may be increased (i.e. τ_r decreased) by a factor of up to $\sigma(M)/l_{nn}$ without introducing dynamically consecutive points as nearest neighbors. The number of points in the reconstruction is then

$$N_1 = \sigma(M)(T/\tau_i)^{(M+1)/M}, \tag{17}$$

where

$$\sigma(M) = 1 - \frac{4^M M!}{(2M+1)!}. \tag{18}$$

Note that $\sigma(1) = \frac{1}{3}$ and $\sigma(M)$ approaches one as M increases.

Basically, this procedure samples the trajectory down to the nearest neighbor scale determined by N_0. For small M (and most test cases), short recurrence times allow frequent sampling of the series, while as M increases, N_1 approaches N_0.

Conclusions. The arguments above demonstrate that a large quantity of data is required to accurately estimate the correlation exponent. Specifically, to observe a value of $\hat{\nu}(l)$ within 5% of the embedding dimension, M, with $M=6$, $R=4$ (i.e. $\hat{\nu}(l) > 5.7$) requires the analysis of a minimum of 5489031774 (uncorrelated) data points. Embedding "long" data records from a truly random sequence in higher dimensional spaces will yield estimates of the correlation exponent which are less than M; this is not evidence for low dimensional behavior. For example, if a white noise signal of this length were embedded with $M=20$, $R=1$, the value of $\hat{\nu}(l_{nn})$ would be significantly less than 16.

The rapid increase in computer time required for the GPA with increases in the number of data points currently limits the accessible exponents to about 5 or 6. In cases where this amount of data is available, long period oscillations in $\chi(l)$ may remain undetected, in which case $\hat{\nu}(l)$ will still provide a poor estimate of ν. A computationally tractable alternative is to compute the local scaling (and estimate the point-wise dimension) about singular points in the distribution (e.g. unstable periodic orbits in a chaotic mapping where a local lacunarity oscillation is often visible [8]). Such calculations also require large data sets and provide only local scaling information; but the computational cost increases linearly in the number of points and, when a periodic oscillation is present, they provide a meaningful estimate of the remaining uncertainty in the point-wise dimension.

If a short time series from a high dimensional system is embedded into increasingly higher dimensional embedding spaces, ν_{est} will appear to converge due to the effects discussed above. The correlation integral of low dimensional signals can be measured with "small" data sets [17] (the data set is still large compared to N_{min}), but the analysis of a sample from a high dimensional signal with the same number of points will also yield $\nu_{est} \ll M$. It is necessary to have sufficient data to distinguish these two behaviors.

I have benefited from many conversations with B. Mestel, C. Sparrow, E.A. Spiegel and A. Wolf, and from the hospitality of the King's College Research

Center. This work was supported by NASA Cooperative Agreement NCC 5-29 and the Science and Engineering Research Council.

References

[1] L.F. Richardson, in: General systems yearbook, Vol. 6 (1961) p. 139.
[2] P. Grassberger and I. Procaccia, Rev. Lett. 50 (1983) 346.
[3] P. Grassberger and I. Procaccia, Physica D 9 (1983) 189.
[4] H.G.E. Hentschel and I. Procaccia, Physica D 8 (1983) 435.
[5] L.A. Smith, J.-D. Fourier and E.A. Spiegel, Phys. Lett. A 114 (1986) 465.
[6] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, Phys. Rev. Lett. 45 (1980) 712.
[7] D. Ruelle and F. Takens, Commun. Math. Phys. 20 (1971) 167.
[8] L.A. Smith, Lacunarity and chaos in nature, Ph.D. Thesis, Columbia University (1987).
[9] A. Brandstater and H. Swinney, Phys. Rev. A 35 (1987) 2207.
[10] J. Guckenheimer, Contemp. Math. 28 (1984) 357.
[11] V.I. Arnold and A. Avez, Ergodic problems of classical mechanics (Addison-Wesley, Reading, 1968).
[12] L.A. Smith and E.A. Spiegel, Strange accumulators, in: Ann. N.Y. Acad. Sci., Vol. 497. Chaotic phenomena in astrophysics (1987).
[13] M. Casdagli, Nonlinear prediction of chaotic time series, submitted to Physica D.
[14] J.D. Farmer and J.J. Sidorowich, Phys. Rev. Lett. 59 (1987) 845.
[15] A.M. Fraser and H.L. Swinney, Phys. Rev. A 33 (1986) 1134.
[16] J. Theiler, Phys. Rev. A 34 (1986) 2427.
[17] N.B. Abraham, A.M. Albano, B. Das, G. De Guzman, S. Yong, R.S. Gioggia, G.P. Puccioni and J.R. Tredicce, Phys. Lett. A 114 (1986) 217.
[18] A. Arneodo et al., Fractal dimensions and the $f(\alpha)$ spectrum of the Hénon attractor, Preprint.
[19] R. Badii and A. Politi, Phys. Lett. A 104 (1984) 303.
[20] A.P. Mulhaupt, Phys. Lett. A 112 (1987) 403.
[21] A. Rényi, Probability theory (North-Holland, Amsterdam, 1970).
[22] P. Grassberger, Phys. Lett. A 128 (1988) 369.