Physica D 139 (2000) 276–300

Computing connectedness: disconnectedness and discreteness

V. Robins [a,*], J.D. Meiss [a], E. Bradley [b]

[a] *Department of Applied Mathematics, University of Colorado, Boulder, CO 80309, USA*

[b] *Department of Computer Science, University of Colorado, Boulder, CO 80309, USA*

Received 12 January 1999; received in revised form 15 October 1999; accepted 8 November 1999
Communicated by C.K.R.T. Jones

Abstract

The standard definition is that a set is *connected*

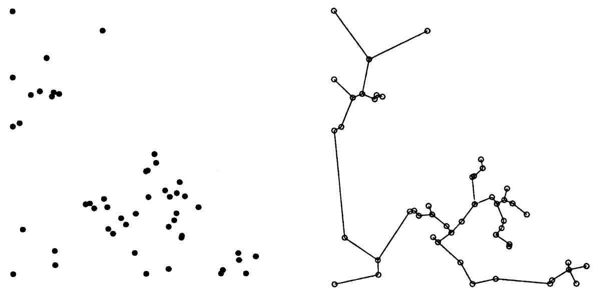


Fig. 1. A finite set of points and its minimal spanning tree. The weight of an edge is the Euclidean distance between the points it joins.

In order to compute [illegible] and [illegible] numerically, we need an appropriate way to organize the finite point-set data. The structure we use is a graph — the MST [26]. This choice was inspired by Yip's [27] work on computer recognition of orbit structures in two-dimensional area-preserving maps. Section 3.1 describes why the edge lengths of the MST naturally define the resolutions at which one should see a change in the number of components. In fact, *Conversely, if the MST of a finite set of points, X, has largest edge length* finding the diameter of that set of points. Finding the diameter of a set of N points in the plane is an order $O(N \log N)$ algorithm, since the computations can be restricted to points lying on the boundary of the convex hull. For subsets of higher-dimensional spaces, this restriction does not necessarily help and the algorithm we use is the brute-force comparison of distances between all pairs of points, which is $O(N^2)$ [26].

Finally, we must address the problem of how to determine the finest appropriate resolution, ϵ_0, as discussed at the beginning of Section 3. To do this, we examine how the number of isolated points in the ϵ-decomposition of the set X varies with resolution, i.e., the function $I(\epsilon)$. A point, x, is isolated at resolution ϵ if $d(x, X \setminus x) \geq \epsilon$. In terms of the MST, a point is isolated at resolution ϵ if all edges incident to it have length longer than ϵ. In all of the examples below, the underlying sets are perfect, so the finite point-set approximation is "bad" at any resolution for which there are isolated points. It follows that the resolution at which we start to see isolated points is one way to estimate ϵ_0, i.e., $\epsilon_0 = \inf\{\epsilon : I(\epsilon) = 0\}$. The validity of this approach is supported by the numerical evidence given in Section 4; the data for $C(\epsilon)$ and $D(\epsilon)$ blur at the resolution at which isolated points are first detected.

4. Examples

In this section we present some examples that illustrate the behavior of the number of ϵ-components, $C(\epsilon)$, the largest diameter, $D(\epsilon)$, and the number of isolated points, $I(\epsilon)$, for fractals with different topology. The goal is to show that these quantities give useful information about the topology. The first examples are relatives of the Sierpinski triangle. These sets are generated from a family of iterated function systems (IFS). The Hausdorff dimension of each set is identical, even though they have different topological structure, as the disconnectedness and discreteness indices highlight. For these fractals, we show that the cutoff resolution decreases when the number of data points is increased, which is not surprising since more points sampled from an attractor constitute a better approximation of the underlying set. We also vary the way in which the data cover the set and find that for a fixed number of points the cutoff resolution is a minimum when the data is uniformly distributed. Again, this is exactly what we expect, since isolated points appear at larger values of ϵ when points are not evenly spaced.

We next present a number of Cantor set examples to illustrate different types of scaling behavior in

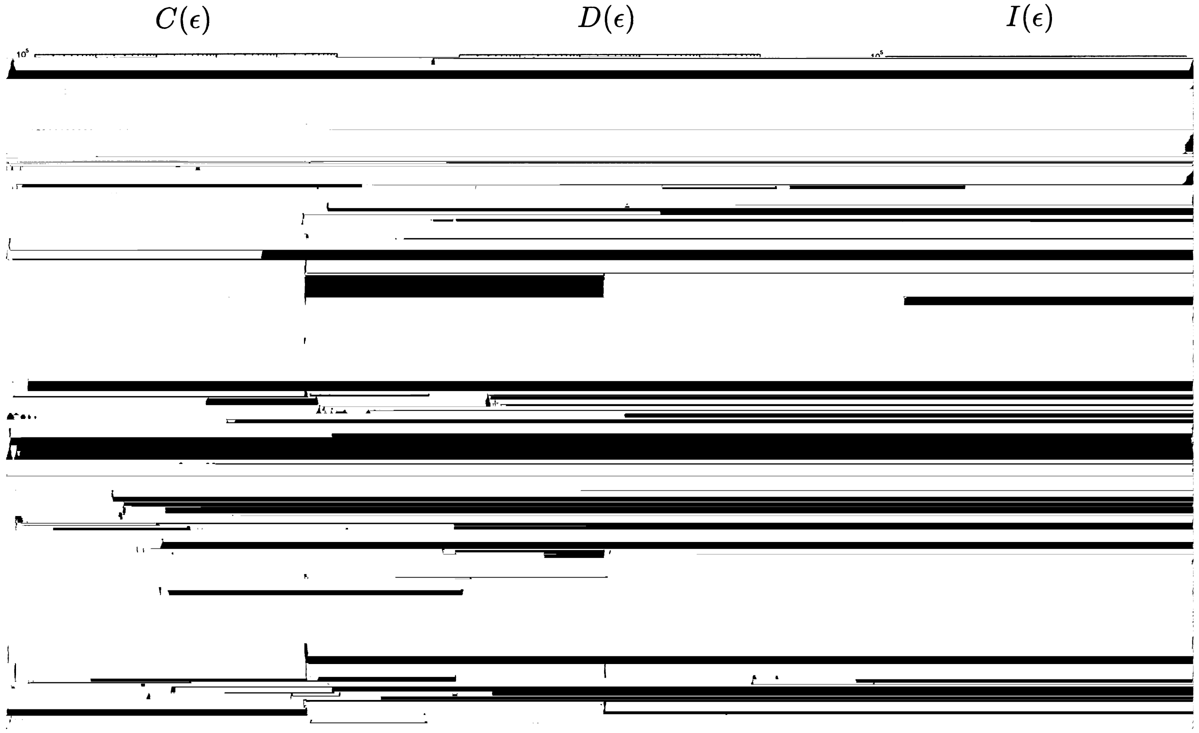


Fig. 4. $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ for the Sierpinski triangle. The top row gives results for 10^4 uniformly distributed points on the fractal and the bottom row for 10^5 points. All axes are logarithmic. The horizontal axis range is 10^{-5} 1. The solid lines represent $C(\epsilon)$ and $D(\epsilon)$ for ideal data; the dots are the computed values.

A finite point-set approximation to the triangle and the corresponding minimal spanning tree are shown in Fig. 3. The underlying set is connected and perfect, so we expect to see $C(\epsilon) = 1$, $D(\epsilon) = \sqrt{2}$, and $I(\epsilon) = 0$ for . This is reflected by the calculations of $C(\epsilon)$ and $D(\epsilon)$ for 10^4 and 10^5 point approximations to the triangle, as shown in Fig. 4. We see that for above a threshold value, the computed values of $C(\epsilon)$ and $D(\epsilon)$ are in exact agreement with our expectations. The point at which $C(\epsilon)$ and $D(\epsilon)$ deviate from the ideal values is the value of at which the number of isolated points, $I(\epsilon)$, becomes positive. This ϵ value is, of course, the cutoff resolution ϵ_c discussed in Section 3.2. At finer resolutions, i.e., ϵ_c, we see a sharp transition in the number of connected components from one to the number of points in the set; the diameters show a correspondingly sharp decrease. Both these effects are due to the narrow distribution of edge lengths of the MST. Clearly, the value of ϵ_c depends on the number of points, , covering the set. For the 10^4 point approximation, $\epsilon_c \approx 0\ 008$ and for $= 10^5$, $\epsilon_c \approx 0\ 0022$; We expect the relationship to be $\epsilon_c \approx 1/\sqrt{\ }$, since the data is homogeneously distributed on a subset of $\mathbb{R}^2$. This is supported by the data in Fig. 5a. Here, we plot cutoff resolution versus the number of points for $10^3 \leq \quad \leq 10^5$; the slope of the least-squares fit line is $-0\ 58$.

The results discussed so far are for uniformly distributed data; we now look at nonuniformly distributed data. As described earlier, we change the way an orbit covers the IFS attractor by choosing the functions $_1$, $_2$, and $_3$ with different probabilities. To generate Fig. 6a, we set $p_1 = 0\ 05$ and $p_2 = p_3 = 0\ 475$. This highly nonuniform distribution of points induces perceptible changes in the $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ data, Fig. 7, but the graphs remain qualitatively similar to those in Fig. 4. The cutoff resolution is significantly larger: $\epsilon_c \approx 0\ 04$ compared with $0\ 008$ for the uniform distribution with the same number of points. The growth in the number of -components for ϵ_c

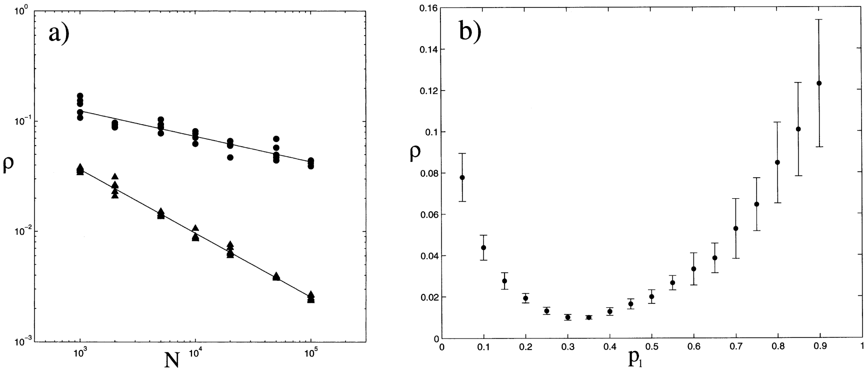


Fig. 5. (a) Cutoff resolution, ρ, as a function of the number of points, $10^3 \le N \le 10^5$, covering the Sierpinski triangle for two values of p_1: (●) marks data for the nonuniform distribution with $p_1 = 0.05$; (▲) marks data for $p_1 = \frac{1}{3}$, i.e., a uniform distribution, and (b) cutoff resolution as a function of p_1 for 10^4 data points on the Sierpinski triangle. The error bars are the standard deviation about the mean of 20 calculations of ρ for each value of p_1.

is also less rapid than that for the uniform data. Both these changes are due to a greater spread in the edge lengths of the MST. The geometry of the distribution is reflected in the graph of $D(\epsilon)$; the densely covered diagonal means $D(\epsilon) = \sqrt{2}$ for ϵ-values significantly less than ρ. We can lower the cutoff resolution by increasing the number of

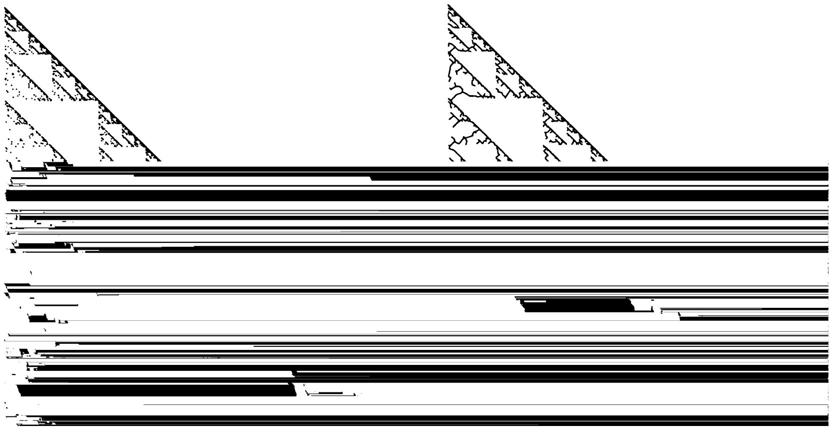


Fig. 6. (a) 10^4 points on the Sierpinski triangle generated by setting $_1 = 0.05$ and $_2 = {}_3 = 0.475$, and (b) the corresponding MST.

$$= \frac{0}{2}. \qquad C(\epsilon) = \begin{cases} 3 \;\; + 2 \cdot 3^{\;-1)/2} & \text{if } \;\text{is odd.} \\ 3 \;\; + 3^{\;/2} & \text{if } \;\text{is even.} \end{cases} \qquad D(\epsilon) = \frac{\;\;_0}{2^{\;-1}} \quad \text{for } \;\ge 3$$

These expressions give $\gamma = \log 3/\log 2 \approx 1.585$ and $= 1$.

We can see in Fig. 9 that the numerical calculations agree very well with the theory down to the cutoff resolution ≈ 0.003. When , the computed values of $C(\epsilon)$ are larger than the predicted values because isolated points are counted as extra components. For still smaller values of , every point is resolved as an isolated point and the $C(\epsilon)$ curve levels off. The meaningful portion of the data — between these extremes — shows a staircase periodicity about a linear trend. The slope of the linear trend is an estimate of γ. We determine γ numerically, using a least-squares fit, to be 1.41 ± 0.05. This is lower than the true limiting value given above (1.58) because of second-order effects at the relatively large values of for which the $C(\epsilon)$ data are valid. We estimate the slope of the true curve over the same range to be 1.48, which is closer to the value computed above.

The numerically calculated values of $D(\epsilon)$ also show a staircase periodicity about a linear trend. The data have a systematic bias for the jumps at slightly larger values of than predicted by theory. This is due to finite data

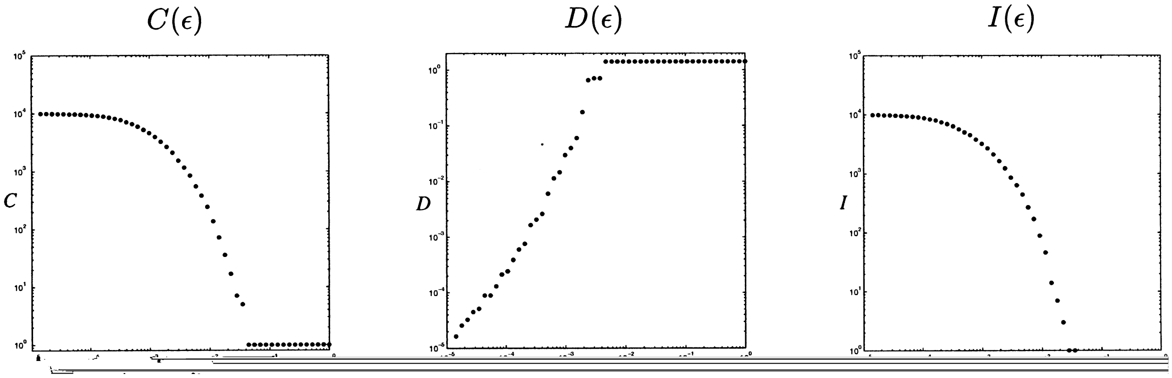


Fig. 7. $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ for the nonuniformly distributed data set in Fig. 6. All axes are logarithmic. The horizontal axis range is 10^{-5} 1.

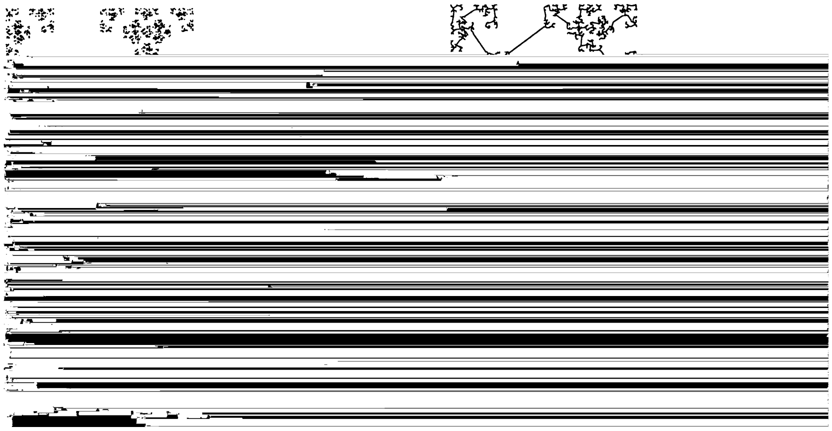


Fig. 8. (a) 10^4 points on the Cantor set generated by (2), and (b) the corresponding MST.

Table 1
This table summarizes values of γ and for Cantor subsets of the plane[a]

Data set	dim	γ	
Fig. 12a	1.262	1 23 ± 0 02	0 96 ± 0 04
Fig. 12b	1.126	1 11 ± 0 02	1 00 ± 0 03
Fig. 14a	1.131	1 13 ± 0 01	0 98 ± 0 03
Fig. 14b	2	0 80 ± 0 05	0 46 ± 0 05
Fig. 16	1 21 dim 1 34	1 36 ± 0 03	0 95 ± 0 05

[a] The numbers are estimated using a least-squares linear fit to logarithmic plots of and , respectively; the error margins are estimated by varying the scaling range; the second column gives the box-counting dimension, dim , for each set; these numbers are computed using formulas from Falconer [32].

data for a 10^5 point approximation, giving a value of $\gamma = 1\ 55 \pm 0\ 03$. This is in very close agreement with the theoretical value of $\gamma = \log 3 \triangleleft \log 2 \approx 1\ 585$.

4.2. Cantor sets in the plane

One of our objectives is to use our techniques to identify and characterize phase space structures in dynamical systems. Cantor sets are often present in chaotic dynamical systems, so it is useful to examine some simple Cantor set examples to gain a better understanding of the different types of scaling that can occur in the and graphs. In Figs. 12, 14 and 16, we show five Cantor sets in the plane. In each case, the orbit has 50 000 points. Four of these have zero Lebesgue measure and one (Fig. 14b) has positive measure, so it is termed a fat Cantor set (this is analogous to the term “fat fractal” for fractals with positive measure [31]). All are attractors of iterated function systems of the form

$$= [\ ] = {}_1[\ ] \cup {}_2[\ ] \cup {}_3[\ ] \cup {}_4[\ ]$$

The generating functions, , become increasingly complex in this series of examples. The three simplest involve only affine transformations and another uses conformal functions; the functions that generate the fat Cantor set cannot be written in closed form. The geometric structure of each set is reflected in the type of staircase seen in the graphs of and . For the four examples with zero Lebesgue measure, we expect to see $\gamma = \dim$, the box-counting dimension; this is supported by our results, summarized in Table 1. For the Cantor set with positive measure, the value of γ is significantly different from the dimension. We again observe, for all of the examples, that the cutoff resolution, , is well thefl7Tjfl /F10 1 Tffl 0.64920 TDfl [(,-)25(alue)-360.5(weer)emal functions;56functions;56function

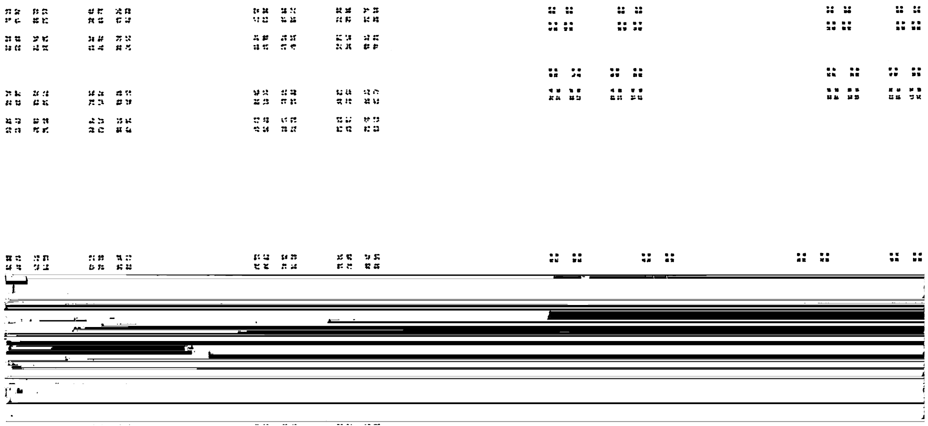


Fig. 12. Cantor sets generated by iterated function systems of four similarity transformations. Both sets have 50 000 points: (a) similarities with contraction ratio $\frac{1}{3}$, and (b) the upper two similarities have ratio $\frac{1}{4}$ and the lower two have ratio $\frac{1}{3}$.

integers and , where is one of the two longest edges. Values of γ and , presented in Table 1, are again very close to the expected values.

To generate the set in Fig. 14a, more general affine transformations are used, each contracting by $\frac{1}{3}$ horizontally and $\frac{1}{4}$ vertically. The corresponding graphs of (and (in Fig. 15 show the now familiar staircase scaling

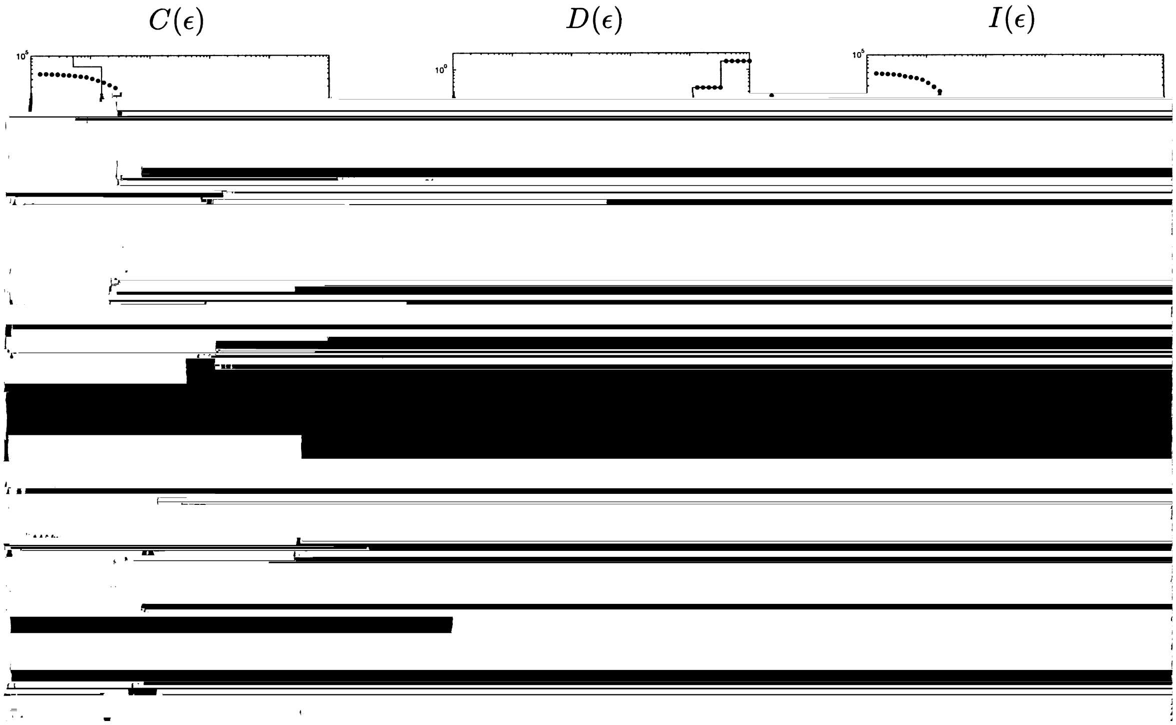


Fig. 13. ((and (for the Cantor sets in Fig. 12. The top row is data for Fig. 12a; the second row is for Fig. 12b. All axes are logarithmic. The horizontal axis range is 10^{-5} 1. The solid lines represent (and (for ideal data; the dots are the computed values.

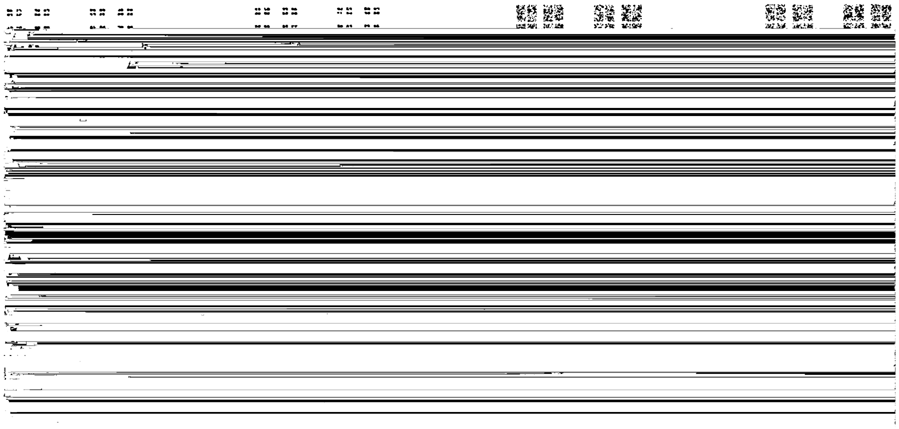


Fig. 14. Two Cantor sets with largest gaps of $\frac{1}{2}$ and $\frac{1}{3}$: (a) a set generated by an IFS of four affine transformations with horizontal contraction of $\frac{1}{3}$ and vertical contraction of $\frac{1}{4}$, and (b) a fat Cantor set, generated as the cross product of two Cantor sets of positive measure in the real line.

pattern. Compared to the second Cantor set example, the larger steps in these graphs reflect the more regular geometric structure of the set.

The fourth example is a Cantor set with positive Lebesgue measure and therefore a dimension of two. It is possible to represent this set as the attractor of an iterated function system of the general form above. The functions involved,

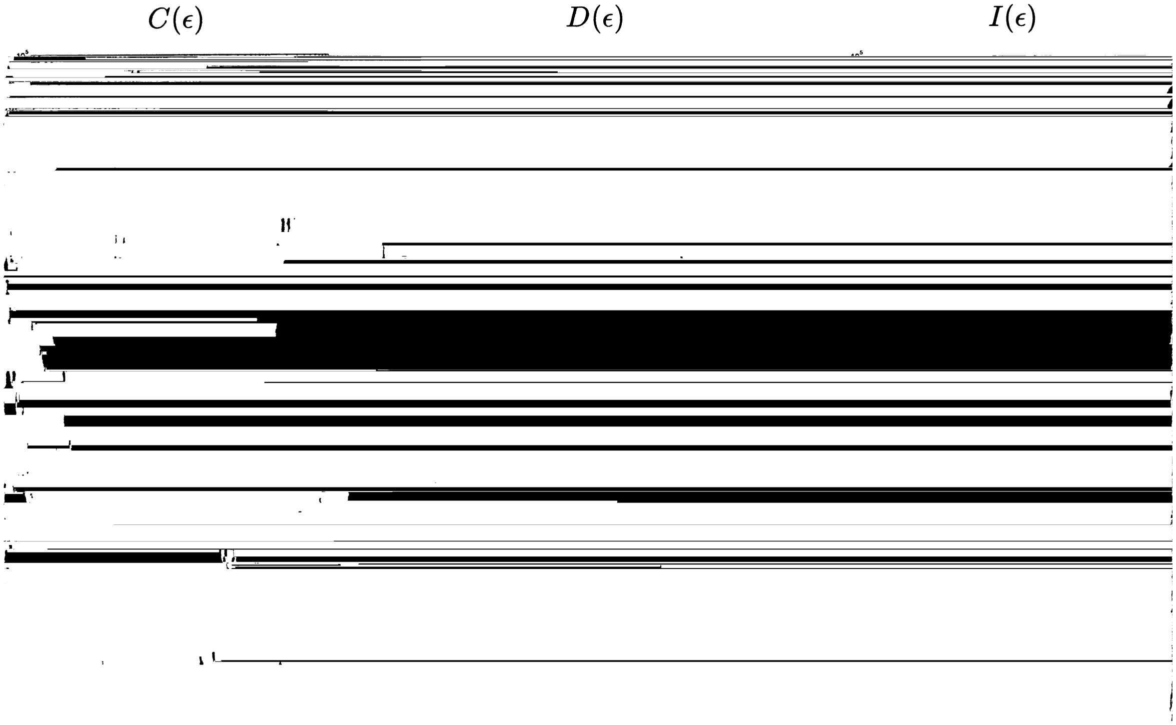


Fig. 15. [illegible] and [illegible] for the 2D Cantor sets in Fig. 14. The top row is data for Fig. 14a; the second row for Fig. 14b, the fat Cantor set. All axes are logarithmic. The horizontal axis range is 10^{-5} 1.

Fig. 16. A Cantor set generated by an IFS consisting of four nonlinear affine transformations, each mapping the unit circle into a circle of radius $\frac{1}{3}$: (a) the data set with circle boundaries, and (b) a close-up of one of the four clusters.

however, are limits of piecewise linear approximations and it is not possible to write them in closed form. Instead, we generate the set as the cross product of two positive measure Cantor subsets of the unit interval. These sets are constructed as follows: at each level, ≥ 1. $2^{\ -1}$ gaps of length $<2^{\ -1}$ are removed from the center of an interval remaining from level -1. The sum of the gap lengths is $2^{\ -1} - 1$; choosing and to make this length less than one ensures the Cantor set has positive measure. It is easy to recursively generate the end points of the gaps (down to some level) and these points are used as the finite point-set approximation. For the set in

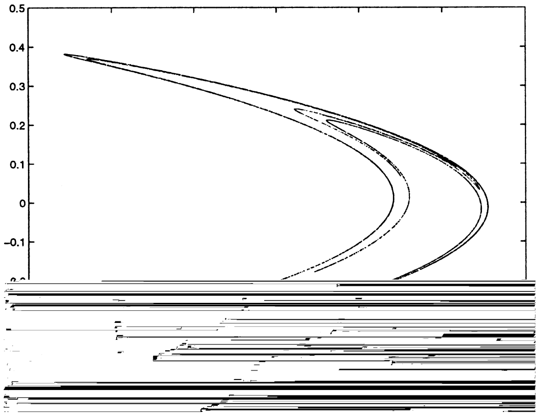


Fig. 18. An orbit on the Hénon attractor.

of the form $f_i(z) = \frac{1}{3}z^2 + c_i$, where $z = x + iy$, and the translations, c_i for $i = 1, \ldots, 4$, take the values $\{\pm\frac{1}{2}, \pm\frac{1}{2}i\}$. Notice that although we choose f_i with equal probability, the nonlinearity introduces a nonuniformity to the distribution of points over the Cantor set. The cutoff resolution $\epsilon_c \approx 5 \times 10^{-4}$ is nevertheless comparable to the previous examples with uniformly distributed data. Scaling in the graphs of $D(\epsilon)$ and $I(\epsilon)$ occurs in two distinct ϵ intervals, see Fig. 17. For $0.005 < \epsilon < 1$, there are three shallow steps reflecting the large-scale structure that is visible in Fig. 16a. The second portion of the data, for $\epsilon_c < \epsilon < 0.005$, has a steeper slope, corresponding to the limiting small-scale structure of the set. The values of γ and k given in Table 1 are slopes of the $D(\epsilon)$ and $I(\epsilon)$ over the interval $\epsilon_c < \epsilon < 0.005$. We find, as for the previous zero-measure Cantor sets, that γ is close to the box-counting dimension and $k \approx 1$.

4.3. Examples from dynamical systems

Our last two examples are Cantor sets from iterated maps.

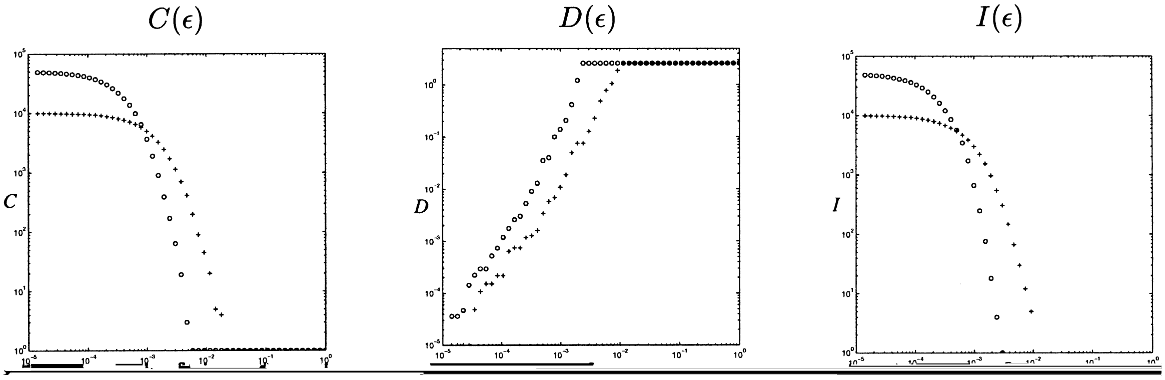


Fig. 19. $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ data for two orbits on the Hénon attractor. The crosses (+) represent calculations for the orbit of 10^4 iterates and the circles (○) are for an orbit with 5×10^4 points. All axes are logarithmic. The horizontal axis range is $10^{-5} < \epsilon < 1$.

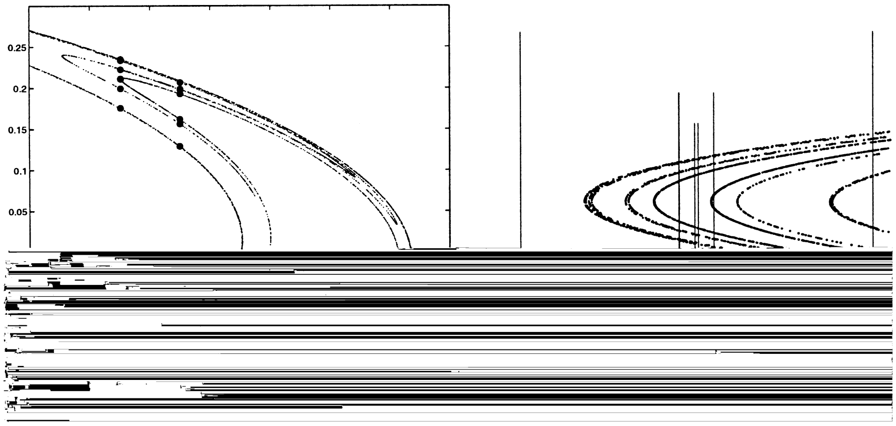


Fig. 20. (a) A close-up of the Hénon attractor. The dark spots are points in the three cross-sections considered in the text: slices at , = 0 302435., = 0 5 and ˷ = 0, and (b) a small part of the slice at , = 0 302435, ˷ = 0 22 that shows the folding of the attractor. The pairs of vertical lines are the boundaries of the different subslices of widths 2×10^{-5}. 2×10^{-6} and 2×10^{-7}.

4.3.1. The Hénon attractor

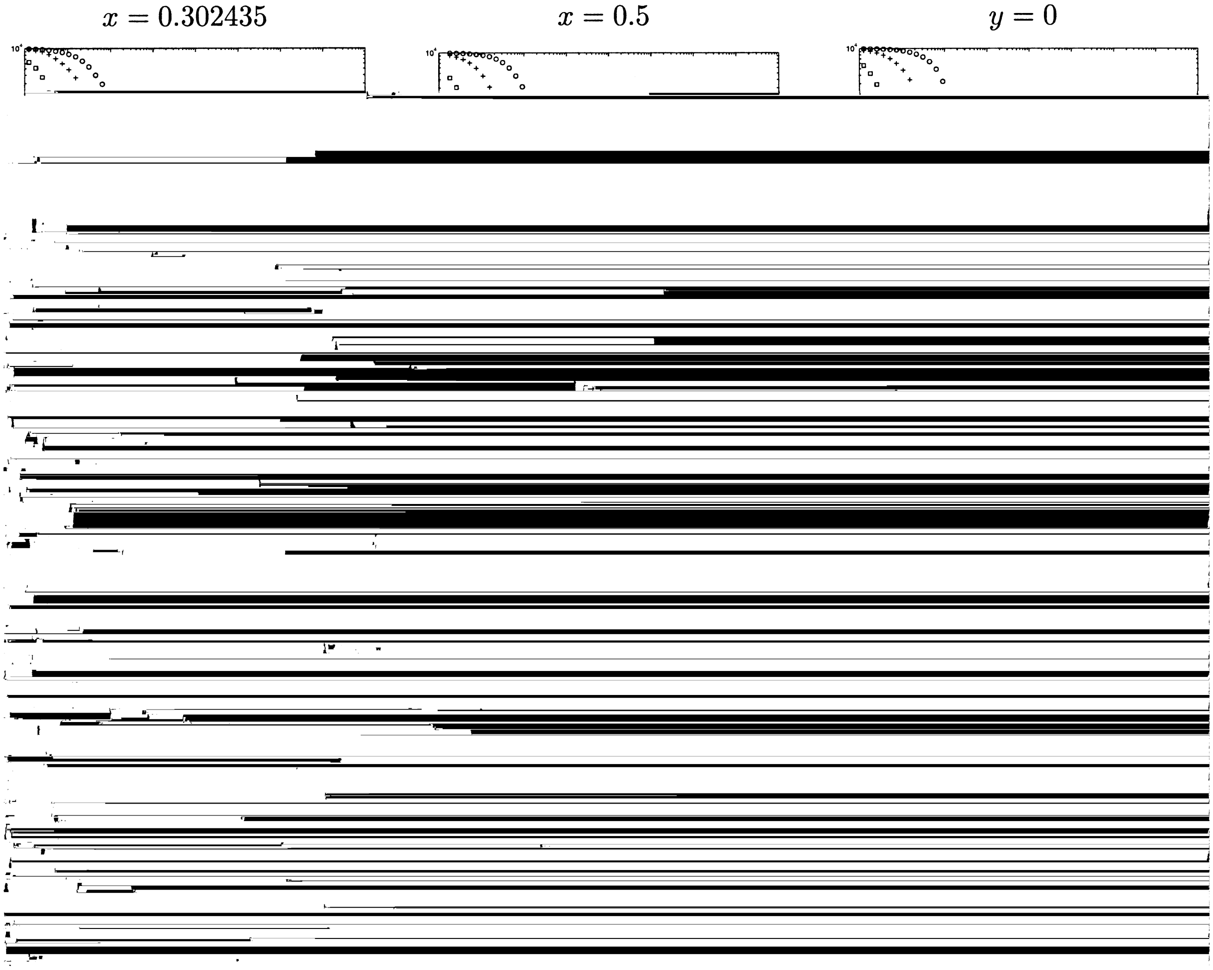


Fig. 21. Top row: (, middle (, and bottom: (data for three sections of the Hénon attractor. The circles ($\bigcirc$) represent calculations for a section of width 2×10^{-4}, the crosses (+) for one of width 2×10^{-5}, the squares ($\Box$) of width 2×10^{-6}, and the stars ($*$) width 2×10^{-7}. All axes are logarithmic. The horizontal axis range is 10^{-8} 1.

cross-section can therefore not be a Cantor set; though of course, removing any fold-tangency points does leave a Cantor set.

The sections at $x = 0.5$ and $y = 0$ have simpler structure. The graphs of (show the now familiar staircase structure of a Cantor set. The flat segments in each graph of (are due to the finite width of each slice, making the data appear like a Cantor set of lines. Values of γ and are calculated from the (and (data for the thinnest slice at each section. The results are summarized in Table 2. The multifractal nature of the Hénon attractor [34] means that we expect to see the value of the dimension vary for different cross-sections. For the three examples given here, though, the variation is not significant.

The above results give strong numerical support for the common belief that cross-sections of the Hénon attractor are Cantor sets. The box-counting dimension of the Hénon attractor is estimated to be about 1.27 [34,35]. Results about the dimension of intersections of sets [32] imply that the dimension of a cross-section through the Hénon

Table 2
Values of γ and for the three sections of the Hénon attractor shown in Fig. 20

Section	γ	
$= 0\ 302435$	$0\ 25 \pm 0\ 01$	$0\ 9 \pm 0\ 1$
$= 0\ 5$	$0\ 26 \pm 0\ 01$	$0\ 85 \pm 0\ 04$
$= 0$	$0\ 27 \pm 0\ 01$	$0\ 88 \pm 0\ 02$

attractor should be $1\ 27 - 1 = 0\ 27$. The values of γ given in Table 2 are in close agreement with this value, providing further support for our conjecture that Cantor sets of zero-measure have γ equal to the box-counting dimension.

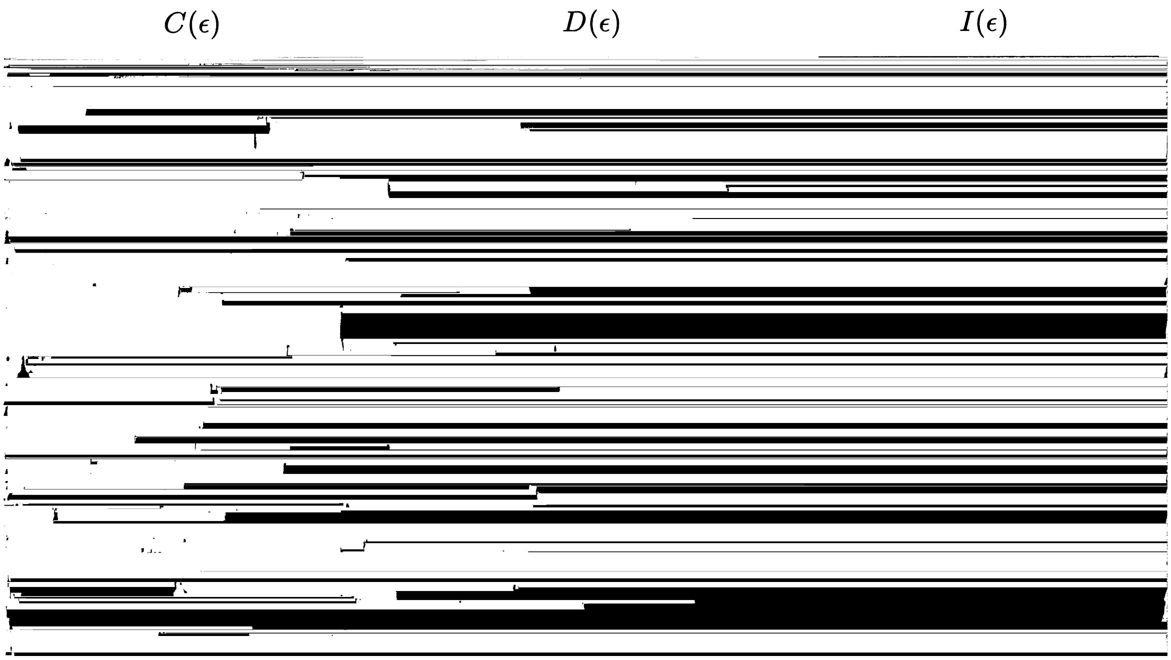
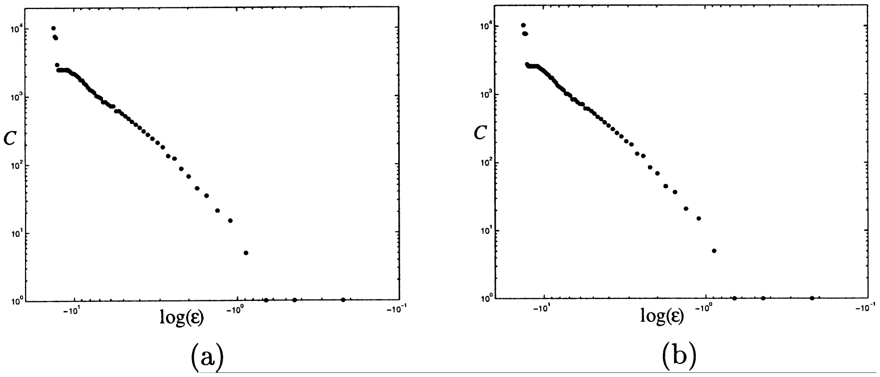


Fig. 23. $C(\epsilon)$, $D(\epsilon)$ and $I(\epsilon)$ data for the two cantori: (top row) data for the cantorus in Fig. 22a, and (bottom row) data for the cantorus in Fig. 22b. All axes are logarithmic. The horizontal axis range is $10^{-15} \le \epsilon \le 1$.

$$> \approx \begin{pmatrix} 1.9152 & -2.0358 \\ 0.5214 & 0.0847 \end{pmatrix}$$

The graphs of $D(\epsilon)$ in Fig. 23 are similar to those for previous Cantor sets. These graphs tell us that the cantori are totally disconnected, because $D(\epsilon) \to 0$. Again, we estimate to be very close to one: for the cantorus of Fig. 22a $= 1.09 \pm 0.05$; for that of Fig. 22b $= 1.01 \pm 0.02$. The graphs of $C(\epsilon)$,



(a) (b)

Fig. 24. $C(\epsilon)$ versus $\log \epsilon$ for the two cantori of Fig. 22a and 22b, respectively. All axes are logarithmic. The horizontal range is $-20 \le \log \epsilon \le -0.1$.

-V.00018is possibly logarithmic, rather than polynomial, in-V.00028. To

Acknowledgements

Thanks to A. Hogan for assistance with programming and the binary tree algorithm. JDM was supported in part by NSF grant number DMS-9971760. EB and VR were supported by NSF National Young Investigator award #CCR-9357740; EB was also supported by ONR #N00014-96-1-0720 and a Packard Fellowship in Science and Engineering.

References

[1] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Geometry from a time series, Phys. Rev. Lett. 45 (1980) 712–716.
[2] T. Sauer, J.A. Yorke, M. Casdagli, Embedology, J. Stat. Phys. 65 (1991) 95–116.
[3] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.S. Young (Eds.), Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Vol. 898, Springer, Berlin, 1981, pp. 366–381.
[4] P. Grassberger, I. Procaccia, Characterization of strange attractors, Phys. Rev. Lett. 50 (1983) 346–349.
[5] A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time series, Physica D 16 (1985) 285–317.
[6] R.W. Ghrist, P.J. Holmes, M.C. Sullivan, Knots and links in three-dimensional flows, Lecture Notes in Mathematics, Vol. 1654, Springer, Berlin, 1997.
[7] G. Mindlin, R. Gilmore, Topological analysis and synthesis of chaotic time series, Physica D 58 (1992) 229–242.
[8] N.B. Tufillaro, P. Wyckoff, R. Brown, T. Schreiber, T. Molteno, Topological time-series analysis of a string experiment and its synchronized model, Phys. Rev. E 51 (1) (1995) 164–174.
[9] M.R. Muldoon, R.S. MacKay, J.P. Huke, D.S. Broomhead, Topology from a time series, Physica D 65 (1993) 1–16.