Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

Physics, Computer Science & Mathematics Division

To be submitted for publication

NUMERICAL ESTIMATES OF HAUSDORFF DIMENSION

Alexandre Joel Chorin

November 1981

RECEIVED
LAWRENCE
BERKELEY LABORATORY

FEB 22 1982

LIBRARY AND
DOCUMENTS SECTION

TWO-WEEK LOAN COPY

*This is a Library Circulating Copy
which may be borrowed for two weeks.
For a personal retention copy, call
Tech. Info. Division, Ext. 6782*

LBL-13642
c.2

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

NUMERICAL ESTIMATES OF HAUSDORFF DIMENSION*

Alexandre Joel Chorin

Department of Mathematics and Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

November 1981

* This work was supported in part by the Director, Office of Energy Research, Office of Basic Energy Sciences, Engineering, Mathematical, and Geosciences Division of the U.S. Department of Energy under contract W-7404-ENG-48. The work was performed while the author was a visiting member of the Courant Institute of Mathematical Sciences, whose hospitality is gratefully acknowledged.

ABSTRACT

Numerical methods for estimating Hausdorff dimension, useful in the analysis of turbulence, are explained and applied to a specific example. In particular, methods involving rescaling and approximation by Cantor sets are discussed.

§1. INTRODUCTION

The need to estimate Hausdorff dimension numerically has arisen in several problems connected with turbulence theory, in particular in the analysis of three dimensional vortex motion [2], [3], in the analysis of stochasticity in dynamical systems [1], [6], [13], in the theory of turbulent flames [4]. Related notions of dimension are also of significance ([6], [13]), and other applications are likely to appear (see, e.g., [11]). The methods used in the literature to estimate Hausdorff dimension include a straightforward application of the definition [2], or of a modified definition [6], [13], and a rescaling technique followed by an approximation by Cantor sets [3]. The purpose of the present paper is to explain these methods and validate them by applying them to a set whose dimension is known.

We begin by defining Hausdorff dimension. Consider a compact set C ; cover it by balls of radii $\rho_i \leq \rho$. Form the sum

$$S(D) = \sum \rho_i^D, \quad D = \text{positive number.}$$

Consider the quantity

$$h(D) = \lim_{\rho \rightarrow 0} \liminf S(D) \quad (1)$$

where the $\liminf$ is taken over all covers with $\rho_i \leq \rho$. $h(D)$ is the dimension of C in dimension D . The number

$$D^* = \begin{cases} \text{greatest lower bound of } D \text{ for which } h(D) = \infty, \\ \text{least upper bound of } D \text{ for which } h(D) = 0 \end{cases}$$

exists, and is the Hausdorff dimension of C . For a cube, $D^* = 3$; for a square, $D^* = 2$; for a segment, $D^* = 1$; according to [2], [3], the essential L_2 support of the vorticity in incompressible inviscid flow had dimension $D^* \sim 2.5$.

We shall apply our methods of estimation to the set Z of zeroes of Brownian motion. Let $x(t)$, $0 \leq t \leq 1$, be a realization of Brownian motion (for a definition, see, e.g., [9]). The set of its zeroes is the set of t 's such that $x(t) = 0$. This set has, with probability 1, Hausdorff dimension $D^* = 1/2$ ([8], [14]). We shall use the following properties of Brownian motion: (i) The interpolation property [10]: If $x(\cdot)$ is normalized so that $x(1)$ has variance $1/2$, and if $x(t_1)$, $x(t_2)$ are known, then the conditional distribution of $x(t)$, $t_1 \leq t \leq t_2$ is given by

$$x(t) = x(t_1) + (t - t_1)(x(t_2) - x(t_1))/(t_2 - t_1) + w((t_2 - t)(t - t_1)/(t_2 - t_1))^{1/2}, \quad (2)$$

where w is a Gaussian random variable with mean zero and variance $1/2$. (ii) Self similarity: $x(t)$, $0 \leq t \leq 1$, and $\sqrt{\alpha}x(t/\alpha)$, $\alpha > 0$, $0 \leq t \leq 1$, are equivalent processes.

Z is a random set. A comparison of our calculations here with the calculations in [2], [3] shows that Z is a much more irregular set than the sets of non-integer Hausdorff dimension generated by smooth differential equations, and the numerical estimation of the dimension of Z is not trivial. The methods used here are similar, but not identical, to the methods used in [2], [3]. Dimension is but one property of a set, and thought is required in each special case.

§2. A COVERING SCHEME

We begin by covering the set of zeroes of $x(\cdot)$ by segments of equal lengths (somewhat in the manner of the construction in [13]). Pick $x(1)$ by sampling the appropriate Gaussian distribution (an algorithm for sampling Gaussian distributions can be found, e.g., in [12]). By definition, $x(0) = 0$, and thus $[0, 1]$ contains at least one member of Z .

Divide $[0, 1]$ into $[0, 1/2]$, $[1/2, 1]$. A value of $x(1/2)$ can be found using (2): $x(1/2) = x(1)/2 + 2^{-3/2}w$. If $x(1/2)x(1) < 0$, $[1/2, 1]$ contains at least one zero.

More generally, define an iteration to be the following sequence of operations: Divide each one of the intervals from the preceding iteration into two halves; values of $x(\cdot)$ at the end points of the new intervals are either available from the preceding iteration or can be sampled by applying formula (2), which in this special case reads

$$x(\text{middle}) = \frac{1}{2}(x_- + x_+) + \frac{1}{2}\sqrt{\Delta} w, \quad (3)$$

where $x(\text{middle})$ is the value of $x(\cdot)$ in the middle of an old interval, x_- and x_+ are the values of $x(\cdot)$ at the left and right ends of an old interval, Δ is the length of a new interval, and w is defined as in (2).

Let x_- , x_+ now denote the values of $x(\cdot)$ at the ends of a new interval. If $x_-x_+ < 0$, the interval surely contains a zero of $x(\cdot)$. If $x_-x_+ > 0$, $|x_-| \geq K\sqrt{\Delta}$, $|x_+| \geq K\sqrt{\Delta}$, K large enough, there is a negligible probability that the interval contains a zero (see [10]) and the interval can be removed from further consideration. If neither of these conditions holds, the interval may or may not contain a zero.

We consider quantity $\lim_{\Delta \rightarrow 0} n\Delta^D$, where $\Delta = 2^{-i}$ is the length of the

intervals after the i -th iteration, and n is the number of intervals which are known to contain zeroes (they satisfy $x_-x_+ < 0$). The difference between this quantity and the quantity $h(D)$ in (1) lies in the fact that all the radii $\Delta/2$ are equal and also in the fact that the $\liminf$ operator in (1) has not been carried out. Thus $2^{-D}n\Delta^D$ is an upper bound on $h(D)$. The number

$$\tilde{D} = \begin{cases} \text{greatest lower bound of } D \text{ for which } n\Delta^D \rightarrow 0, \\ \text{least upper bound of } D \text{ for which } n\Delta^D \rightarrow \infty, \end{cases}$$

is an upper bound on the dimension D^* of Z . We shall be computing $\tilde{D}$, and we shall assume without proof that $\tilde{D} = D^*$. From the numerical point of view it does not matter whether this (highly plausible) assumption is correct, since we shall obtain estimates of $\tilde{D}$ with error estimates. In the calculations of [2], [3] the $\liminf$ in (1) is computed correctly. In [13], a quantity analogous to $\tilde{D}$ is evaluated.

In Table I we display values of $n\Delta^D$ for several values of D as functions of the iteration i . For $D < D^*$, $n\Delta^D$ should be increasing; for $D > D^*$, $n\Delta^D$ should be decreasing. We set $K = 4$ (we shall show below that this is a large enough value of K). The calculation must be stopped after a finite number of iterations because the number of intervals quickly overwhelms the available computer memory and because Δ shrinks to below the underflow limit of the computer arithmetic. A reasonable person looking at Table I could conclude that D^* lies somewhere between .45 and .60 -- not a dramatically accurate answer. In the next section we shall see how this estimate can be improved without catastrophic expense.

<u>i</u>	<u>D = .35</u>	<u>D = .40</u>	<u>D = .45</u>	<u>D = .50</u>	<u>D = .55</u>	<u>D = .60</u>	<u>D = .65</u>	<u>D = .70</u>
1	.35	.24	.17	.12	.088	.062	.044	.031
2	.34	.23	.16	.11	.075	.051	.035	.024
3	.27	.17	.11	.078	.051	.034	.022	.014
4	.51	.32	.20	.13	.084	.053	.034	.021
5	.46	.28	.17	.10	.067	.041	.025	.015
6	.68	.40	.24	.14	.085	.050	.030	.017
7	.74	.42	.29	.14	.080	.046	.026	.015
8	.97	.53	.29	.16	.091	.051	.028	.015
9	1.04	.55	.29	.16	.085	.045	.24	.015
10	1.06	.55	.28	.14	.079	.039	.020	.010

Table I.

Values of $n\Delta^D$ as Functions of i and D

§3. RESCALING

Suppose the number of segments in the calculation (both the segments known to contain zeroes and those which may turn out to contain zeroes) exceeds a preset number n_0 (we usually set $n_0 = 100$). Then rescale n , i.e., throw away half the segments in such a way that the expected number of segments which contain zeroes in the rejected half equals the number of such segments in the retained half. We shall call such a rejection "unbiased." An unbiased rejection can be easily achieved; for example, each time a segment is halved, store the parameters relating to one half of the old segment in the array location in which the old segment was remembered and store the parameters which describe the other half at the end of the array. A rejection of the first or the second half of the resulting array is unbiased. Perform this rescaling at each iteration if needed. Segments which are not likely to contain zeroes may, in the interest of efficiency, be thrown out as soon as they are generated.

If Δ becomes smaller than a preset Δ_0 (we usually set $\Delta_0 = .01$ or $\Delta_0 = .001$), rescale Δ , i.e., double Δ , and, in order to leave Z invariant, multiply all values x_- , x_+ , by $\sqrt{2}$ (see the self similarity property (ii) above).

If there are n segments with a sure zero in the computer at the end of an iteration, there would have been $\sim Nn$ segments in the computer if rescaling had not been used. $N = 2^{\ell_1}$, ℓ_1 = number of rescalings of n . If Δ is the length of the segments as stored in the computer, the real length should be Δ/M , $M = 2^{\ell_2}$, ℓ_2 = number of rescalings of Δ . Thus $D < D^*$ if $NM^{-D}n\Delta^D \rightarrow \infty$, $D > D^*$ if $NM^{-D}n\Delta^D \rightarrow 0$. We could again estimate D^* by following trends in the evolution of $NM^{-D}n\Delta^D$ for several values of D .

However, a sharper estimate is available. By construction, $1 \leq n \leq 2n_0$ and $\Delta_0 \leq \Delta \leq 2\Delta_0$. Thus $n\Delta^D$ is, for each D , a positive quantity bounded from above and bounded away from zero. Thus $NM^{-D}n\Delta^D$ tends to zero or infinity if NM^{-D} does. Therefore, as $N \rightarrow \infty$ and $M \rightarrow \infty$, we must have $M^{-D^*} = O(N)$ and $D^* \rightarrow \log N / \log M$. In Table II we display values of $\log N / \log M$ obtained after 60 iterations in different runs, each run using a different sequence of random numbers to generate Z . In 60 iterations n is typically rescaled 6 to 7 times and Δ about 40 times. In Table III we display estimates of D^* obtained in a similar fashion with 600 iterations ($n_0 = 100$). These iterations are very inexpensive (each one takes a second or two on a VAX); after 600 iterations, the error in the estimate of D^* is under 1%; without rescaling, the same accuracy would have presumably required about $2^{400} \cong 10^{100}$ segments stored in the computer -- an unimaginably expensive enterprise.

Finally, we can check how large K must be before we are reasonably sure that an interval contains no zeroes. Numerical experiment shows that the estimates of D^* are independent of K as long as $K \geq 2$.

§4. APPROXIMATION BY CANTOR SETS

We now describe a method for estimating the dimension of a set by approximating the given set by a suitable Cantor set.

The Cantor sets we shall use are constructed as follows: Consider the interval $[0, 1]$, and divide it into $m \geq 2$ segments of length $1/m$. Keep $s \leq m$ of these segments, $s > 1$, and throw out the others. Divide each one of the remaining segments into m pieces of equal lengths and throw out all but s of these, etc. The remaining set has Hausdorff

.49	.45
.46	.50
.54	.52
.53	.54
.51	.48

Table II.

Estimates of D^* with 60 Iterations

=====

.506 , .500 , .503 , .496

Table III.

Estimates of D^* with 600 Iterations

dimension $D^* = \log s / \log m$. This can be easily seen if one assumes that the measure of the remainder in dimension D^* is positive: The Hausdorff measure of disjoint sets is additive, and if a set B is similar to a set A with a similarity ratio L , the ratio of their measure in dimension D is L^D . Thus, if $h(D^*)$ is the measure of the remainder set in dimension D^* , we have

$$h(D^*) = (s/m^{D^*})h(D^*),$$

i.e.,

$$D^* = \log s / \log m.$$

For details, see [7], [11].

Consider again the set Z . Let ℓ_1 be an integer. Divide $[0, 1]$ into ℓ_1 segments of equal lengths. Find the values of a realization of $x(\cdot)$ at the end points of those intervals, using the interpolation formula (2). This is done with ease if $\ell_1 = 2^{\ell_2}$, ℓ_2 integer, since then formula (3) can be used recursively. Delete the intervals which are not likely to contain zeroes, i.e., those characterized by $x_- x_+ > 0$, $|x_-| \geq 2\sqrt{\Delta}$, $|x_+| \geq 2\sqrt{\Delta}$, where as before Δ is the length of the interval and x_- , x_+ the values of $x(\cdot)$ at its ends. Suppose there are $\ell_3^{(1)}$ intervals left. Define $\bar{D}_1 = \log \ell_3^{(1)} / \log \ell_1$.

Throw away all but ℓ_4 of the remaining intervals. Divide each one into $\ell_1 = 2^{\ell_2}$ pieces, decide which pieces are unlikely to contain zeroes and throw them out. Suppose there are $\ell_3^{(2)}$ pieces left; there are $\ell_3^{(2)} / \ell_4$ such pieces in each of the starting ℓ_4 intervals. Let $\bar{D}_2 = \log(\ell_3^{(2)} / \ell_4) / \log \ell_1$.

Keep iterating in this manner: Start with l_4 pieces, divide each one into l_1 smaller pieces, compute the number $l_3^{(1)}$ of pieces left, and let $\bar{D}_1 = \log(l_3^{(1)}/l_4)/\log l_1$. Whenever the intervals become too small for convenient computation, rescale them, i.e., multiply their lengths Δ by a suitable factor A and multiply the x_- , x_+ by $\sqrt{A}$, as was done above.

If all the $l_3^{(i)}$ were equal to a fixed integer l_3 , independently of the choice of intervals to subdivide, all the $\bar{D}_i$ would be equal, and the set remaining after an infinite sequence of rejections would be a Cantor set of Hausdorff dimension $\bar{D} = \bar{D}_i$ for all i . This set would be larger than Z because at each step we keep intervals which may contain zeroes, but in fact will turn out to contain none. Thus $\bar{D} > D$. If the $l_3^{(i)}$ are not equal, we can view the numbers $\bar{D}_i$ as estimates of $\bar{D}$. The first 10 estimates of $\bar{D}$ are listed in Table IV. The averages of the $\bar{D}_i$ after 50 steps, with $l_2 = 5$, $l_4 = 4$, in one particular run, was .56 with a standard deviation .025. Thus $.56 \pm .025$ is an estimate of an upper bound $\bar{D}$ on D^* .

We now produce a construction which will provide an estimate of a lower bound $\underline{D}$ of D^* . Proceed exactly as in the preceding construction but change the criteria for rejecting intervals. Retain each interval which surely contains a zero ($x_-x_+ < 0$). Reject each interval not likely to contain a zero (defined as in the preceding section). Consider the remaining segments whose fate is uncertain (and which were retained in the preceding construction). Given an interval of length Δ , with values x_- , x_+ , of $x(\cdot)$ at its extremities, let $p(\Delta, x_-, x_+)$ be the probability that it contains a zero. Construct an algorithm which retains the interval with probability p (and rejects it with probability $(1 - p)$).

.35

.74

.74

.38

.16

.52

.69

.35

.68

.61

Table IV.

First 10 Estimates of $\bar{D}$, $l_2 = 5$, $l_3 = 4$

This is easily done: Let k be an integer. Divide Δ into k subintervals of lengths Δ/k , and use (2) again to construct a Brownian arc connecting x_- and x_+ . If the resulting arc crosses the t -axis, we keep the segment, and if it does not, we reject the segment. For k large enough (in practice, $k > 64$), we are keeping the segment with approximately probability p .

As before, if the $\ell_3^{(i)}$ were equal we would have a set with Hausdorff dimension $\underline{D} = \underline{D}_1 \equiv \log(\ell_3^{(i)}/\ell_4)/\log(\ell_1)$, where the $\ell_3^{(i)}$ are the numbers of segments retained according to the new criteria, while ℓ_1, ℓ_4 are defined as before. $\underline{D}$ is a lower bound on D^* because the remaining set is too small -- each doubtful segment has a probability p of being rejected, but if it is kept in one iteration it may be rejected later because the applications of the interpolation formula (2) are independent. However, the $\underline{D}_i$ are not equal and are only estimates of $\underline{D}$. In Table V we list the first 10 estimates of $\underline{D}_i$ of $\underline{D}$, with $\ell_2 = 5$, $\ell_4 = 4$; the average of the $\underline{D}_i$ after 50 steps was $.48 \pm .015$. Thus,

$$\underline{D} < D^* < \bar{D},$$

and we have found

$$\bar{D} = .56 \pm .025$$

$$\underline{D} = .48 \pm .015.$$

$\underline{D} - \bar{D} \approx 0$ because Z is a random set, unlike the sets encountered in the applications to differential equations. We have obtained fairly sharp estimates of $\bar{D}$ and $\underline{D}$ but a rather poor estimate of D^* . The best estimate of D^* was obtained by the rescaling procedure of the preceding section.

.38

.48

.51

.40

.36

.44

.56

.29

.55

.64

Table V.

First 10 Estimates of $\underline{D}$, $l_2 = 5$, $l_3 = 4$

BIBLIOGRAPHY

- [1] P. Bernard & T. Ratiu, (Eds.), Berkeley Turbulence Siminar, Springer (1977).
- [2] A.J. Chorin, Estimates of Intermittency, Spectra and Blow-up in Developed Turbulence, to appear in Comm. Pure Appl. Math.
- [3] A.J. Chorin, The Evolution of a Turbulent Vortex, to appear in Comm. Math. Physics.
- [4] A.F. Ghoneim, A.J. Chorin & A.K. Oppenheim, Numerical Modeling of Turbulence in a Combustion Tunnel, to appear in Phil. Trans. Roy. Soc. London.
- [5] F. Hausdorff, Dimension and Ausseres Mass, Math. Annalen (1919).
- [6] P. Fredericksen, J.L. Kaplan, E.D. Yorke & J.A. Yorke, The Liapunov Dimension of Strange Attractors, manuscript, (1981).
- [7] J.P. Kahane & R. Salem, "Ensembles Parfaits et Séries Trigonometriques," Hermann, Paris, (1963).
- [8] R. Kohn, Hausdorff Dimension Properties of Stochastic Processes, Lecture Notes, Princeton (1977).
- [9] J. Lamperti, "Probability Theory," Benjamin, N.Y. (1966).
- [10] P. Levy, "Le Movement Brownien," Mem. Sc. Math., fasc. 126, Gauthiers Villars, Paris (1954).
- [11] B. Mandelbrot, "Fractals: Form, Chance & Dimension," Freeman, San Francisco (1976).
- [12] R.E.A.C. Paley & N. Wiener, "Fourier Transforms in the Complex Domain," AMS Colloquium Publications, Providence (1934).
- [13] F. Takens, Detecting Strange Attractors in Turbulence, manuscript (1981).
- [14] S.J. Taylor, Proc. Camb. Phil. Soc., 51, 265 (1955).

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720