FRACTALS AND BROWNIAN MOTION

Paul M. Covert

Undergraduate thesis (MATH 496)
submitted toward the degree
Bachelor of Science with College Honors

Department of Mathematics
University of Washington
Seattle, WA
June 1990
Prof. Krzysztof Burdzy, supervisor
I. BROWNIAN MOTION.

1. Random Walk.

As a warm-up to the Brownian motion process itself, let us consider first a less complicated process which has some of the same properties: the symmetric random walk on \( \mathbb{Z} \). This is constructed by starting at zero and repeatedly taking steps which move up one with probability 1/2 and down with probability 1/2. Or, more formally: let \( X_1, X_2, \ldots \) be an infinite sequence of independent, identically distributed random variables with 
\[
P\{X_k = 1\} = P\{X_k = -1\} = 1/2 \quad \text{for all } k \in \mathbb{Z}^+; \ 
\]
let \( Y_0 = 0 \), and for \( n \in \mathbb{Z}^+ \), let 
\[
Y_n = \sum_{k=1}^{n} X_k.
\]
Then the sequence \( Y_0, Y_1, Y_2, \ldots \) is called a symmetric random walk on the integers, or simply a “standard random walk”.

This process has what will turn out to be two of the three defining properties of Brownian motion: (1) for any positive integers \( k, n \), \( Y_{n+k} - Y_n = \sum_{j=n+1}^{n+k} X_j \) has the same distribution as \( Y_k = \sum_{j=1}^{k} X_j \); and (2) for any positive integers \( n_1 < n_2 < \ldots \), \( Y_{n_{j+1}} - Y_{n_j} \) is independent of \( Y_{n_{k+1}} - Y_{n_k} \) for all distinct \( j, k \). The latter property is due to the independence of the random variables \( X_i \), while the former follows from their identical distribution. (The third property of Brownian motion is continuity over the real numbers; we shall see later on that it is a limit of “increasingly continuous” random walks.)

Note also that, by the Central Limit Theorem, the distribution of \( Y_k / \sqrt{k} \) approaches a normal distribution as \( k \to \infty \); the normal distribution turns out to be another important characteristic of the Brownian motion process.

Let us now take a look at the zero set of the standard random walk, \( Z = \{ k \in \mathbb{Z}^+ : Y_k = 0 \} \). Since \( k \in Z \) implies that \( X_j = 1 \) for \( k/2 \) values of \( j \) in \( \{1, \ldots, k\} \) and \(-1\) for the other \( k/2 \), \( k \) must be an even number. Thus for odd \( k \), \( P[k \in Z] = 0 \). For \( n \in \mathbb{Z}^+ \),
\[
P[2n \in Z] = P[Y_{2n} = 0] = P[\{X_1, \ldots, X_{2n}\} \text{ contains } n "1"'s \text{ and } n "-1"'s] \\
= 2^{-2n} \binom{2n}{n} = (2n)!/(n!2^n)^2.
\]

The expected number of points in this set is then \( \sum_{1}^{\infty} (2n)!/(n!2^n)^2 \). We may judge whether this is finite or infinite using Stirling’s formula for the factorial, \( n! \approx \sqrt{2\pi n} n^{n+1/2} e^{-n} \),
which is very accurate for large \( n \). Using this, we have
\[
\sum_{1}^{\infty} 2^{-2n} (2n)!/(n!)^2 \approx \sum_{1}^{\infty} [2^{-2n} \sqrt{2\pi}(2n)^{2n+1/2} e^{-2n} / 2\pi n^{2n+1} e^{-2n}] = \frac{1}{\sqrt{2\pi}} \sum_{1}^{\infty} 1/\sqrt{n} = \infty.
\]
Thus the expected number of returns to zero is infinite. Indeed, the process will return infinitely many times with probability one.

We can also discuss multidimensional random walks, that is, ordered \( d \)-tuples of independent standard random walks. These will turn out to be analogous to multidimensional Brownian motion in various ways—for example, in their recurrence properties. The probability that a \( d \)-dimensional standard random walk will return to zero after \( 2n \) steps is the probability that all \( d \) of its components will return to zero independently, that is, \([2^{-2n} (2n)!/(n!)^2]^d\), which, by Stirling’s formula, is approximately \((1/\sqrt{n})^d\) for large \( n \). Now, \( \sum_{1}^{\infty} n^{d/2} \) diverges for \( d = 1 \) or 2, and converges for \( d > 2 \); thus we can expect infinitely many returns to zero in one and two dimensions, but not in three or more.

2. Brownian Motion as a Limit of Random Walks.

We define a Brownian motion process \( B(t) \) (with \( t \in [0, \infty) \)) as one for which (1) for \( t_2 > t_1 \), \( B(t_2) - B(t_1) \) is independent of \( B(t_1) \); (2) for \( t_0 > 0 \), \( B(t + t_0) - B(t) \equiv B(t_0) \) for all \( t \geq 0 \) (where “\( \equiv \)” denotes identical distribution); and (3) its sample paths are continuous and start at zero. (Another well-known property, the normal distribution of \( B(t) \) for each \( t > 0 \), follows from these three, and thus is not needed in the definition.)

Since, among these three properties, the standard random walk lacks only continuity, it seems as though what is needed is to take a random walk and make it continuous. In some sense, this actually works; the following explanation is based on that of Knight (1981).

First, we define a more general random walk, constructed so as to include a step duration \( t_0 \) and step size \( y_0 \), by letting each \( Y_j = \pm y_0 \) with probability \( 1/2 \) each, and \( X(kt_0) = \sum_{1}^{k} Y_j \). Furthermore, we can also extend this process to a continuous function on \([0, \infty)\) by defining \( X(t) \) linearly on each interval \([kt_0, (k+1)t_0]\) — “connecting the dots”, so to speak. (This will be called a “random walk function”.)
In approximating Brownian motion by random walks, what we want is a sequence of random walk functions with both step size and step duration decreasing to zero. Also, since for both the standard random walk, from which the process starts, and for Brownian motion itself, \( \text{Var} X(1) = \text{E}[X(1)]^2 = 1 \) (notation: "\text{Var}" = variance, "\text{E}" = expectation), we will require that \( \text{Var} X(1) = 1 \) for all stages in the sequence. This requirement has the effect of imposing a relationship between step size and step duration at each stage of the approximation; if we let \( t_0 = 1/k_0 \) for some \( k_0 \in \mathbb{Z}^+ \), then \( \text{Var} X(1) = \text{Var} X(k_0 t_0) = k_0 \text{Var}(t_0) = k_0 y_0^2 \) (the variance of a single step is just the step size squared) = \( y_0^2 / t_0 \); if \( \text{Var} X(1) = 1 \), then \( t_0 = y_0^2 \), i.e. the step duration is the square of the step size.

To find a sequence of random walks which approximate a Brownian motion, we shall let \( R_0(t) \) be the function constructed by connecting the dots on the standard random walk, and for \( n \in \mathbb{Z}^+ \), let \( R_n(t) \) be a similar random walk function of step duration \( 2^{-2n} \) and step size \( 2^{-n} \). (Thus for each \( n \), where \( R'_n(t) \) exists, it equals \( 2^n \) or \( -2^n \); letting \( n \to \infty \), it can be guessed that the limiting case, Brownian motion, will be nowhere differentiable, although everywhere continuous.)

To ensure that the functions \( R_n \) do indeed converge to a Brownian motion, we need also to require a certain relationship between \( R_n \) and \( R_{n+1} \) for each \( n \), namely, that (with the steps of \( R_n \) being twice the size of those of \( R_{n+1} \)), \( R_n \) steps up each time \( R_{n+1} \) moves two steps up, and down whenever \( R_{n+1} \) moves down twice. This can perhaps be made more clear graphically, as follows.

Fig. 1(a) shows a the typical path of a typical random walk \( R_3[0,1] \), determined by 64 coin flips. The step size here is \( 1/8 \), and the step duration \( 1/64 \). On this, we note the first point taking the value \( +1/4 \) or \( -1/4 \) (here it is \( +1/4 \)); the first point after that taking 0 or \( +1/2 \) (here, 0); and so on. This should give about 16 points (14 in this case); we use these to trace out the beginning of \( R_2(t) \), as shown in Fig. 1(b). Continuing in the same way, we find \( R_1 \) (Fig. 1(c)), and \( R_0 \) (Fig. 1(d)).

What we are doing here is ensuring that the sample paths of the random walks \( R_n(t) \) approximate the same Brownian motion sample path, with the approximation becoming increasingly close as \( n \) increases. The way we are doing this, if we have \( R_n[0, \infty) \), then
with probability 1, we uniquely determine \( R_{n-1}[0, \infty) \) — and, indeed, \( R_m \) for all \( m < n \). We say that \( R_{n-1} = M_{n-1}(R_n) \), and, for \( m < n \), \( R_m = M_m(M_{m+1}(\ldots(M_{n-1}(R_n))\ldots)) \).

Now, let \( \Omega_\infty \) be the set of sequences of random walk functions \((R_0, R_1, \ldots)\) such that \( R_{n-1} = M_{n-1}(R_n) \) for all \( n \). For \((R_0, R_1, \ldots) \in \Omega_\infty \), \( t \in [0, \infty) \), let \( R(t) = \lim_{n \to \infty} R_n(t) \), with probabilities defined as one would expect: \( P[R(t_0) > a] = \lim_{n \to \infty} P[R_n(t_0)] \). Knight (1981) proves that \( R_n \) converges uniformly to \( R \) with probability 1. The proof is beyond the scope of this paper.

The uniform convergence makes things very nice here; since each \( R_n \) is continuous, \( R \) must be continuous also (and obviously \( R(0) = 0 \)). Regarding the properties of homogeneity (that, given \( t_0 > 0 \), \( R(t + t_0) - R(t) \) is identically distributed for all \( t \geq 0 \)), and independence of increments (that, for all \( t_n > \ldots > t_1 > 0 \), \( R(t_1), R(t_2) - R(t_1), \ldots \), and \( R(t_n) - R(t_{n-1}) \) are all independent), we first check that they hold on the dyadic rationals (\( \{j2^{-n}, j, n \in \mathbb{Z}^+\} \)). The increments are clearly independent on this set, since for dyadic rationals \( t_1 = j_12^{m_1}, \ldots, t_n = j_n2^{m_n} \), we let \( M = \max\{m_1, \ldots, m_n\} \); then the increments in question are clearly independent for \( R_m \) with \( m > M \), and thus for \( R \) also. Similarly, homogeneity holds on the dyadic rationals also, since for \( t_0 = j2^M \), \( R_m(t + t_0) - R_m(t) \) is distributed as \( R_m(t_0) \) for all \( t = k2^m \), for all \( m > M \). Since these properties hold on the dyadic rationals, which are dense in \( \mathbb{R} \), and since \( R(t) \) is continuous, they must hold for \( R(t) \) on all of \( \mathbb{R} \).

We have shown that \( R(t) = \lim_{n \to \infty} R_n(t) \), where the \( R_n(t) \)'s are random walks of step size \( 2^{-n} \) and step duration \( 2^{-2n} \) such that \( R_{n-1} = M_{n-1}(R_n) \) for all \( n \), is indeed a Brownian motion process, which we shall henceforth denote \( B(t) \).

### 3. Properties of Brownian motion.

(i) **Normality.** In examining the distributions of the increments of \( B(t) \), we first recall that, for all \( n \), \( R_n(1) \) is a random variable with expectation 0 and variance 1. Since \( R_n(1) \) is a sum of \( 2^{2n} \) independent random variables which are equal to \( 2^{-n} \) with probability \( 1/2 \) and \( -2^{-n} \) with probability \( 1/2 \), its distribution approaches the normal distribution as \( n \to \infty \); thus \( B(1) \equiv N(0, 1) \).
Similarly, for positive integers \( n_0, n \geq n_0 \), \( R_n(2^{-n_0}) \) has expectation 0 and variance \( 2^{-n_0} \); again, \( R_n(2^{-n_0}) \to N(0, 2^{-n_0}) \) as \( n \to \infty \), so \( B(2^{-n}) \equiv N(0, 2^{-n}) \) for all \( n \in \mathbb{Z}^+ \). Finally, any \( t > 0 \) can be expressed as a sum \( t = n + \sum_{j=1}^{\infty} a_j 2^{-j} \), where \( n \) is a nonnegative integer and each \( a_j \) is 0 or 1; due to the independence and homogeneity properties, \( B(t) \) is a sum of independent normals whose expectations are zero and whose variances add to \( t \). Therefore, for all \( t > 0 \), \( B(t) \equiv N(0, t) \).

(ii) **Uniqueness.** The Brownian motion process \( B(t) \) is unique up to multiplication by a constant scale factor and addition of a constant multiple of \( t \); that is, any process \( Y(t) \) satisfying the three axioms of Brownian motion satisfies \( Y(t) \equiv aB(t) + bt \), where \( a \) and \( b \) are real constants.

This property is related to the normal distribution of \( Y(1) \). Since \( Y \) must be homogeneous with independent increments, the fact that, for any \( n \), \( Y(1) = [Y(1/n) - Y(0)] + \ldots + [Y(1) - Y((n - 1)/n)] \) makes \( Y(1) \) the sum of \( n \) independent and identically distributed random variables distributed as \( Y(1/n) \). This becomes normal as \( n \to \infty \) as long as the \( Y(1/n) \)'s are reasonably well behaved, which they will be in this case; and so \( Y(1) \equiv N(b, a) \) for some \( a, b \). In fact, the same argument shows that \( Y(t) \) is normally distributed for all \( t \), by looking at sums of \( Y(t/n) \).

Now, this process has the property that mean and variance are linear in \( t \); that is, there exist constants \( a, b \in \mathbb{R} \) such that \( E[Y(t)] = bt \), \( Var[Y(t)] = at \) for all \( t \). Thus, for a continuous, homogeneous \( Y(t) \) with independent increments, there exist \( a, b \) such that \( Y(t) \equiv N(bt, at) \) for all \( t \); that is, \( Y(t) \equiv aX(t) + bt \).

Conversely, any \( Y \) such that \( Y(t) \equiv aX(t) + bt \) clearly satisfies the axioms of Brownian motion. Thus \( Y(t) \equiv aX(t) + bt \) is necessary and sufficient for continuity, homogeneity, and independent increments, and we have shown the uniqueness property.

(iii) **Invariance properties.** The simplest of the "self-similarity" properties of Brownian motion is the Markov property— i.e., for \( t_0 > 0 \), \( B(t + t_0) - B(t) \equiv B(t_0) \). This is, of course, just a restatement of the homogeneity property, but it is important to note that, at any time on a Brownian motion, we can "stop and start over", and the path after starting over will still be a Brownian motion path.
The scaling property of Brownian motion is that, for \( c > 0 \), \( B(t) \equiv \sqrt{c}B(t/c) \). This can be seen to be true since \( \sqrt{c}B(t/c) \) clearly has continuous paths, is homogeneous, and has independent increments, and \( \text{E}[\sqrt{c}B(t/c)] = 0, \text{Var}[\sqrt{c}B(t/c)] = (\sqrt{c})^2 \cdot (1/c) = 1 \). This implies that Brownian motion on any \([0, c]\) will be just like that on \([0,1]\), multiplied by the scale factor \( \sqrt{c} \).

Another invariance property is that \( B(t) \equiv tB(1/t) \). This is shown to be true by the fact that \( tB(t) \) is normally distributed for all \( t \) with covariances matching those of \( B(t) \). For \( t_2 > t_1 > 0 \), \( \text{Cov}[B(t_1), B(t_2)] = \text{E}[B(t_1)B(t_2)] = \text{E}[B(t_1)(B(t_1) + (B(t_2) - B(t_1)))] = \text{E}[B(t_1)]^2 + \text{E}[B(t_1)(B(t_2) - B(t_1))] = \text{Var}B(t_1) + 0 = t_1 = \min\{t_1, t_2\} \); also, \( \text{Cov}[t_1B(1/t_1), t_2B(1/t_2)] = t_1t_2\text{Cov}[B(1/t_1), B(1/t_2)] = t_1t_2\min\{1/t_1, 1/t_2\} = t_1t_2/t_2 = t_1 = \text{Cov}[B(t_1), B(t_2)] \). The significance of this property is that it produces an equivalence between “local” properties, as \( t \to 0 \), and “global” properties, as \( t \to \infty \). In particular, the distribution of zeroes of the sample paths \( B(t) \) on \((0, \infty)\) matches that of \( tB(1/t) \), and thus of \( B(1/t) \).

(iv) The Law of the Iterated Logarithm. An important result which has implications about the zero set of \( B(t) \) is the Law of the Iterated Logarithm: that, with probability 1,

\[
\limsup_{t \to 0} \frac{B(t)}{\sqrt{2t \log \log (1/t)}} = 1.
\]

By the symmetry \( B(t) \equiv tB(1/t) \), this is equivalent to

\[
\limsup_{t \to \infty} \frac{B(t)}{\sqrt{2t \log \log t}} = 1.
\]

The proof of the Law is omitted here, but it should be noted that the standard deviation of \( B(t) \) is \( \sqrt{t} \), and so what this means is that as \( t \) grows large, \( \sqrt{2t \log \log t} \) is a very slowly increasing number of standard deviations of \( B(t) \); for example, \( \sqrt{2 \log \log 10^6} \approx 2.29 \) and \( \sqrt{2 \log \log 10^{100}} \approx 3.30 \).

(v) Recurrence. The Law of the Iterated Logarithm can also be combined with the symmetry \( B(t) \equiv -B(t) \) to give

\[
\liminf_{t \to 0} \frac{B(t)}{\sqrt{2t \log \log (1/t)}} = -1
\]
with probability 1. Thus, in every $(0, \varepsilon)$, the sample path $B(t)$ oscillates infinitely many times between $+\sqrt{2t \log \log (1/t)}$ and $-\sqrt{2t \log \log (1/t)}$ (since, for each time $B(t)$ approaches $+\sqrt{2t \log \log (1/t)}$, there is an approach to $-\sqrt{2t \log \log (1/t)}$ closer to zero, and vice versa). This then implies that, with probability 1, the sample paths $B(t)$ have infinitely many zeroes in $(0, \varepsilon)$ for any $\varepsilon > 0$; furthermore, in any interval, there will be infinitely many local maxima and minima. Also, by the local-global symmetry noted earlier, there are infinitely many zeroes in each $(1/\varepsilon, \infty)$, implying that a Brownian motion path never “escapes”: for any $c > 0$, $B(t) = 0$ for infinitely many $t > c$ with probability 1.

(vi) Nowhere-differentiability. With probability 1, Brownian sample paths are nowhere differentiable. (Not just on a set of measure zero, but literally nowhere.) This can perhaps be explained by the fact that differentiability at any point requires that a function be “approximately linear” over some interval, however small; this must contain an open interval with rational endpoints, of which there are only countably many in $(0, \infty)$. In any such interval, $B(t)$ will be “approximately linear” with probability zero; since there are only countably many of them, the probability of approximate linearity on any one of them is zero, and thus $B(t)$, although continuous, is nowhere differentiable.

4. Multidimensional Brownian Motion.

By a “$d$-dimensional Brownian motion”, we mean an ordered $d$-tuple of independent standard Brownian motions; we denote this $B_d(t)$. One important property of $B_d(t)$ is its spherical symmetry; that is, for any rotation $T$ of $\mathbb{R}^d$, $TB_d(t) \equiv B_d(t)$. This spherical symmetry turns out to be necessary for extending the uniqueness property of Brownian motion to several dimensions. Given the usual three properties plus spherical symmetry, multidimensional Brownian motion is again unique up to multiplication by a constant and addition of a constant multiple of time— that is, only processes of the form $aB_d(t) + bt$, $a \in \mathbb{R}$, $b \in \mathbb{R}^d$, are spherically symmetric and satisfy continuity of paths, homogeneity, and independence of increments.

The answer to the question of whether there is (pointwise) recurrence on the paths $B_d(t)$ is that for $d \geq 2$, $P[B_d(t) = 0$ for some $t > 0] = 0$, and, similarly, for $t_0 > 0$,
P[B_d(t) = B_d(t_0) for some t > t_0] = 0. In other words, there is no pointwise recurrence. The difference between the one-dimensional case and the others is that, in more than one dimension, the path can move around the origin without actually hitting it.

However, we may also define a weaker type of recurrence, known as “neighborhood recurrence” (as contrasted with the previously discussed “pointwise recurrence”). A process $X(t)$ starting at 0 is said to be neighborhood recurrent if, for any $\varepsilon > 0$ and any $N > 0$, $P[|X(t)| < \varepsilon$ for some $t > N] = 1$; that is, if it can always be expected to return arbitrarily close to zero as $t \to \infty$. If it is not neighborhood recurrent, it will wander off to infinity with probability 1, and is said to be transient. Note that neighborhood recurrence is equivalent to the property that $\liminf_{t \to \infty} |X(t)| = 0$; if $X(t)$ is transient, then $\liminf_{t \to \infty} |X(t)| = \infty$.

For multidimensional Brownian motion, it turns out that $B_2(t)$ is neighborhood recurrent, but that for $d \geq 3$, $B_d(t)$ is transient. While this is difficult to prove, it can be explained intuitively by noting that it is just what should have been expected based on the recurrence properties of multi-dimensional random walks, which always return to zero in one and two dimensions, but not in three or more.

This further implies that, for any neighborhood $N_\varepsilon(x)$ of radius $\varepsilon > 0$ about a point $x \in \mathbb{R}^2$, the path $B_2(t)$ will intersect $N_\varepsilon(x)$ infinitely many times with probability 1. (Certainly the probability of hitting it once is positive, and, since $B_2(t)$ always returns near zero, there will be infinitely many chances to use this positive probability, so to speak.) Thus, although the trail of the path $B_2(t), 0 \leq t < \infty$ has Lebesgue measure zero in $\mathbb{R}^2$ (since any given point is hit with probability zero), it is still dense in $\mathbb{R}^2$.

II. FRACTALS

1. Topological and Hausdorff Dimension.

Two introductory writings on fractals are Mandelbrot’s well-known *The Fractal Geometry of Nature* (1982), and an article by Taylor (1986). While Taylor proposes a fine-tuning
of Mandelbrot's definition, we shall stick to the earlier and simpler definition of Mandelbrot for purposes of this paper. We shall define as a fractal any set for which the Hausdorff dimension exceeds the topological dimension. (Taylor's definition also uses the "packing dimension", which will not be discussed here.)

For the sets of Euclidean geometry, the concept of dimension is simple: the dimension is zero for an isolated point, one for a line or curve, two for a plane or a surface, and three for a solid. Likewise, in linear algebra, an $n$-dimensional vector space is one which requires $n$ independent vectors to form a basis.

In such cases, it might seem clear that an $n$-dimensional set is one on which any point can be described by $n$ independent parameters. However, this idea is shown to be useless by such counterexamples as the "plane-filling curve" of Peano: a continuous mapping from the (clearly 1-dimensional) interval $[0, 1]$ to the (clearly 2-dimensional) unit square $[0, 1]^2$. Obviously, a better way to define the concept of dimension is needed. We shall consider two such definitions: the topological dimension and the Hausdorff dimension.

The topological dimension is the easier to explain of the two; it is more intuitive, and it has the nice property of only taking on integer values. A collection of disconnected points (i.e., a set of which the only connected subsets are single points) is said to have topological dimension zero, and for positive integers $n$, a set has topological dimension $n$ if (1) it does not have dimension less than $n$, but (2) any connected subset of it can be disconnected by the removal of some $(n - 1)$-dimensional set. This gives the familiar results for the sets of Euclidean geometry (a curve can be disconnected by removing points, etc.) and vector spaces, and, for the more complicated sets associated with Brownian motion, it gives the expected results of dimension zero for the zero set of a sample path $B(t)$ and one for the trail of a path $B_d(t)$.

Hausdorff dimension (also known as Hausdorff-Besicovitch dimension) is more complicated; it is based on the Hausdorff measure with respect to the functions $\phi(s) = s^\alpha$ for $\alpha > 0$, which is defined as

$$
\phi - m(E) = \liminf_{\delta \to 0} \inf_{\sum_j \phi(\text{diam } E_j) : E \subset \bigcup_j E_j; \text{diam } E_j < \delta, \forall j} \{ \sum_j \phi(\text{diam } E_j) : E \subset \bigcup_j E_j; \text{diam } E_j < \delta, \forall j \}.
$$
For example, for any \( n \), the unit interval \([0,1]\) can be covered by \( n \) intervals of diameter (length, in this case) \( 1/n \), as if \( \phi(s) = s \), then \( \sum \phi(\text{diam } E_j) \) is just the sum of the lengths of the intervals, which will give an infimum of 1 for any \( \delta \); thus \( s^1 - m[0,1] = 1 \). However, suppose \( \phi = s^2 \); then, as \( \delta \to 0, n \to \infty \) if we are to keep \( 1/n < \delta \), and we are looking at \( \lim_{n \to \infty} \sum_{j=1}^n (1/n)^2 = \lim_{n \to \infty} (1/n) = 0 \); likewise, since \( \lim_{n \to \infty} \sum_{j=1}^n (1/n)^{1/2} = \lim_{n \to \infty} \sqrt{n} = \infty, s^{1/2} - m[0,1] = \infty \). In this way, we can see that \( s^\alpha - m[0,1] = \infty, \alpha < 1; 1, \alpha = 1; 0, \alpha > 1 \).

It turns out to be true that for any set \( E \) in any Euclidean space \( \mathbb{R}^d \), there exists an \( \alpha_0 \) such that \( s^\alpha - m(E) \) is infinite for \( \alpha < \alpha_0 \) and zero for \( \alpha > \alpha_0 \). This \( \alpha_0 \) is called the \textbf{Hausdorff dimension} of \( E \). For notational purposes, this will be called "dim \( E \)"; the topological dimension of \( E \) will be "top dim \( E \)". For any \( E \subset \mathbb{R}^d \), dim \( E \geq \text{top dim } E \); if equality does not hold, we say that \( E \) is a \textbf{fractal}.

This again gives the familiar results for points, curves, surfaces, and solids. For example, the unit square \( E = [0,1]^2 \subset \mathbb{R}^2 \) can be covered by \( n^2 \) squares of side length \( 1/n \) and thus diameter \( \sqrt{2}/n \); we guess that \( E \) has Hausdorff \( s^2 \) measure \( \lim_{n \to \infty} \sum_{j=1}^n (\sqrt{2}/n)^2 = 2 \), and, since this is neither zero nor infinite, dim \( E = 2 \).

However, the Hausdorff dimension has the interesting property that it need not always be integer-valued. For example, the triadic Cantor set, \([0,1] - (1/3, 2/3) - (1/9, 2/9) - (7/9, 8/9) - \ldots \), can be covered with \( 2^n \) intervals of length \( 3^{-n} \) for nonnegative integers \( n \). Now, \( \lim_{n \to \infty} \sum_{j=1}^{2^n} (3^{-n})^{\log 2/\log 3} = \lim_{n \to \infty} 2^n [2^{\log 2/\log 3}]^{-n} = \lim_{n \to \infty} 2^n (2^{-n}) = 1 \); thus the Cantor set has Hausdorff dimension \( \log 2/\log 3 \approx 0.6309 \). Clearly, the Cantor set has topological dimension \( 0 \) (it is completely disconnected); thus it is our first example of a fractal.

It should be noted that any countable set \( E \) has dim \( E = 0 \), as we may take \( \{E_j\}_{j=1}^\infty \) to be an enumeration of the single-point sets comprising \( E \); then, for \( \alpha > 0 \), \( \sum_{j=1}^\infty (\text{diam } E_j)^\alpha = \sum 0 = 0 \).

It may be noticed that, when \( \alpha \) is an integer, the Hausdorff \( s^\alpha \) measure on \( \mathbb{R}^\alpha \) is very similar to \( \alpha \)-dimensional Lebesgue measure. (They are equivalent measures; that is, their ratio is bounded by constants.) Thus the Hausdorff \( s^\alpha \)-measure, where \( \alpha \) is not necessarily
an integer, may be thought of as an $\alpha$-dimensional analog of Lebesgue measure.

Now, for "fat Cantor sets" of positive Lebesgue measure on $\mathbb{R}$, constructed by removing intervals which decrease in length relative to the intervals from which they are removed (for example, removing the middle third at the first stage of the construction, then middle ninths, middle 27ths, etc.), we arrive at a set $E$ with $\dim E = 1$, since positive Lebesgue measure implies positive Hausdorff $s^1$-measure. However, these sets are totally disconnected (since any interval has a sub-interval taken out of it), and so have top dim 0; thus they are still fractals. Note that this means that a fractal set $E$ need not have a non-integer-valued $\dim E$.

It can be seen that a fractal set is generally a figure which is infinitely "squiggly"; a fractal curve $E(\text{top dim } E = 1)$ must have infinite length in any neighborhood of any of its points, a fractal surface has infinite area in any such neighborhood, etc.

2. Self-similarity.

One important property common to many (though by no means all) fractals is that of self-similarity. In general, a set $E \in \mathbb{R}^d$ is called self-similar (with respect to ratio $r$ and integer $N$) if it can be partitioned into subsets $E_1, ..., E_N$ which are congruent to each other and to the dilation of the original set by a scaling constant $r < 1$. This implies that each $E_i$ is also self-similar, that the pieces into which those pieces are decomposed are self-similar, and so on out to infinity. Thus a self-similar set can be decomposed into arbitrarily many subsets which are arbitrarily small dilations of itself.

We may deduce from this that a self-similar set will be either perfectly "smooth" or infinitely "detailed", but nowhere in between. For example, the unit interval $[0, 1]$ can, for any positive integer $n$, be decomposed into $n$ subsets which are congruent to $(1/n)[0, 1]$, so it is self-similar with respect to any $((1/n), n)$. Similarly, the triadic Cantor set $C$ has pieces $[0, 1/3] \cap C, [2/3, 1] \cap C$ which are both congruent to $(1/3)C$, as they are constructed from intervals of length $1/3$ by having middle thirds of intervals removed out to infinity; thus $C$ is self-similar w.r.t. $1/3$ and 2. However, none of the intermediate stages in the construction of $C$ — for example, $[0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$, is self-similar;
although they are composed of several smaller pieces, the pieces are not congruent to
dilations of the original set.

It is the infinitely detailed type of self-similar set which is likely to be a fractal, as
infinite detail is necessary to get the Hausdorff dimension above the topological dimension.
The triadic Cantor set is such a fractal, and we shall be able to use self-similarity to
construct many more self-similar fractals, such as the Koch snowflake, which will be seen
in the next section.

Furthermore, we can use similarity considerations to estimate the Hausdorff dimension
of self-similar fractals, since Hausdorff dimension is based on the relationship between sizes
of small sets and numbers of them needed to cover the whole set. We note that, if a set $E$
is self-similar with respect to ratio $r$ and integer $N$, i.e. composed of $N$ pieces congruent
to $rE$, then those $N$ pieces will each be composed of $N$ subsets congruent to $r(rE) = r^2 E$,
and thus $E$ is self-similar w.r.t. $r^2$ and $N^2$, and, continuing this, with respect to $r^m$ and
$N^m$ for any positive integer $m$. Thus, for all possible such pairs $r, N$, the logarithmic ratio
$\log N / \log(1/r)$ is preserved. This constant is called the similarity dimension of $E$. In
many cases, this turns out to be equal to the Hausdorff dimension, thus giving us a way to
get at dim $E$ without having to worry about coverings of $E$ by sets of diameter less than $\delta$.

It is clear that the similarity dimension gives the usual results for the trivial cases of
the unit interval ($N$ segments of length $1/N$), the unit square ($N^2$ squares of side length
$1/N$), and the unit cube ($N^3$ cubes of side length $1/N$). Also, the Cantor set, composed
of 2 pieces congruent to $(1/3)C$, has similarity dimension $\log 2 / \log 3$, again equal to the
Hausdorff dimension.

The property of self-similarity can be further extended to unbounded sets $E \in \mathbb{R}^d$, which are considered self-similar if, for some $r$, $rE = E$. This is essentially the case as above
with $N = 1$, in that $E$ is "partitioned" into one (improper) subset which is congruent to
$rE$. However, we cannot define a similarity dimension in this way, as we would be dividing
by $\log 1$, that is, zero. Constructing a dimension from $r$ alone will not work, either, as $r$ is
not unique ($r^m E = E$ for all $m$). Thus, to estimate dim $E$ through similarity in this case,
we need to look at self-similar bounded subsets of $E$, if it has any.

3. Some examples of fractals.

The concept of self-similarity gives us a way to construct self-similar fractals using "initiators" and "generators". The triadic construction of $C$ provides a well-known example of this method. We start with its initiator, the unit interval $[0,1]$; then, at each stage of the construction, each straight line segment is replaced by a figure with the same endpoints, but with the shape of its generator,

The Cantor set is then the limit as the number of steps of the construction goes to infinity.

Another well-known self-similar fractal constructed in this way is the Koch snowflake. Its initiator is an equilateral triangle (Fig. 2(a)), and, at the first step of the construction, we replace each straight line segment by the generator (Fig. 2(b)), producing the Star of David (Fig. 2(c)). After two steps, we have Fig. 2(d); the limit at infinity has about that shape, but is infinitely detailed.

Technically, the snowflake is not itself self-similar, according to our definition introduced in the previous section; it is a closed curve which is not composed of several smaller closed curves. However, if we take a subset of it between the endpoints of one side of the original triangle—that is, the limit of

we find that that set is self-similar; since each line segment is replaced by four segments of
one-third its length, the similarity dimension is \( \log 4 / \log 3 \approx 1.2618 \); this is its Hausdorff
dimension also, and thus the dimension of the entire snowflake.

(It is interesting to note in passing that the Hausdorff dimension of the intersection
between the Koch snowflake and any of the stages in its construction is \( \log 2 / \log 3 \approx 0.6309 \). Why? Because, for any segment at any step, its middle third is pushed away
at the next step, and thus, at the limit at infinity, a triadic Cantor dust is left from the
segment.)

Another interesting fractal is the Sierpiński carpet. The construction of this is initiated
by taking the unit square, and generated by dividing the square into 9 squares of side length
1/3, removing the middle one, and keeping the outer eight. Thus the Sierpiński carpet is
the limit of

and has similarity (and Hausdorff) dimension \( \log 8 / \log 3 \approx 1.893 \). Note that this set is
equivalent to \( \{(x, y) \in \mathbb{R}^2 : x \in C \text{ or } y \in C \} \).

There also exists a self-similar fractal in \( \mathbb{R}^3 \) which is similarly related to the Cantor set:
the Menger sponge. Here, the initiator is the unit cube, and the generator is the cube minus
the 1/3-size cube at the center and the 1/3-size cubes in the middles of the six faces, leaving
the 20 1/3-size cubes on the corners and edges. Its dimension is \( \log 20 / \log 3 \approx 2.7268 \).
Note that, on each of the faces, the set of points which remain in the Menger sponge \( M \)
is precisely the Sierpiński carpet \( S \), and that \( M = \{(x, y, z) : (x, y), (x, z), (y, z) \in S \} = \{(x, y, z) : \text{two of } x, y, z \text{ are in } C \} \).
III. FRACTALS RELATED TO BROWNIAN MOTION

1. The Brownian trail.

So far, we have seen several self-similar fractals, all of which are non-random. Indeed, the definition of self-similarity which we have been using does not permit randomness. However, there do exist random fractals which are, in an important sense, self-similar; in order to discuss “random self-similarity”, we will need a new definition.

A bounded random set $E$ is statistically self-similar if, for some $r < 1$, $N \in \mathbb{Z}^+$, $E$ can be partitioned into $N$ subsets $E_1, ..., E_N$, all of which have the same distribution as $rE$. Similarly, an unbounded random $E$ is statistically self-similar if it has the same distribution as $rE$ for some $r < 1$. We may simply call such sets “self-similar” if there is no possibility of confusion.

As an example of a (statistically) self-similar random set, let us recall the trail of multi-dimensional Brownian motion, $B_d[0, \infty) = \{x \in \mathbb{R}^d : B_d(t) = x \text{ for some } t \in [0, \infty)\}$, with $d \geq 2$. $(B_1[0, \infty) = \mathbb{R}$ almost surely, so the case $d=1$ is not very interesting.) These Brownian trails turn out to be statistically self-similar random fractals, with probability 1.

First, we note that the scaling property $\sqrt{c}B_d(t/c) \equiv B_d(t)$ continues to hold for $d > 1$. $(B_d(t)$ is a d-tuple of independent 1-dimensional Brownian motions, for which we know $B(t) \equiv \sqrt{c}B(t/c)$; a d-tuple of these, independent, is $\sqrt{c}B_d(t/c)$.) Note also that, for positive real $c$, $\{t/c : t \in [0, \infty)\} = [0, \infty)$, so $B_d[0, \infty) \equiv \sqrt{c}B_d[0/c, \infty/c) \equiv \sqrt{c}B_d[0, \infty)$; thus, we have shown that the trail of $d$-dimensional Brownian motion is statistically self-similar (with respect to any $r = c < 1$, in fact).

We may also examine bounded subsets $B_d[0, t_0)$ of this to check for similarity properties. The set $B_d[0, t_0)$ is equal to $B_d[0, t_0/n) \cup ... \cup B_d[t_0(n-1)/n, t_0)$ for any $n$. Now, all $n$ of these are congruent (in distribution) to $B_d[0, t_0/n) \equiv \sqrt{1/n}B_d[0, t_0)$. Thus $B_d[0, t_0)$ is the union of $n$ sets with the same distribution as $\sqrt{1/n}B_d[0, t_0)$, so it is statistically self-similar (for any $t_0 > 0$).

Furthermore, we can also get a similarity dimension from this: as, for each number of pieces $n$, the scaling ratio is $\sqrt{1/n}$, the similarity dimension (using the same definition as
in the non-random case) is \( \log n / \log(1/\sqrt{1/n}) = \log n / \log n^{1/2} = 2 \). Thus we may guess (correctly) that the trail of Brownian motion in two or more dimensions has Hausdorff dimension 2, although its topological dimension is 1.

Another way of looking at the dimension of the Brownian trail follows. The trail of \( B_2(t) \) in \( \mathbb{R}^2 \) is a connected set which is dense in the plane, so it should have Hausdorff dimension 2 even though its 2-dimensional Lebesgue (and thus Hausdorff) measure is zero (since the probability of hitting any given point is zero). For \( d \geq 3 \), the projection on any plane of the trail of \( B_d(t) \) is the trail of a 2-dimensional Brownian path, which again has Lebesgue measure zero. Clearly \( \dim B_d[0, \infty) \geq 2 \), since projection into a smaller space cannot increase dimension; however, we would expect that the projection into \( \mathbb{R}^2 \) of a set with \( \dim \geq 2 \) should, in most cases, have positive 2-dimensional Lebesgue measure, which it does not. Thus \( \dim B_d[0, \infty) = 2 \).

2. The Brownian zero set.

The next set we shall examine is the set of zeros of (one-dimensional) Brownian paths, \( Z = \{ t \in [0, \infty) : B(t) = 0 \} \). We have shown previously that \( Z \) is infinite; importantly, it turns out to be uncountable also. (For example, if we didn’t know that the set of rationals in \([0,1)\) were countable, similarity considerations might lead us to believe its Hausdorff dimension to be 1.)

The unbounded random set \( Z \) is statistically self-similar with respect to any ratio \( r < 1 \). By one of the scaling properties of Brownian motion, \( B(t) \equiv \sqrt{1/r}B(rt) \), so the zeroes of \( B(t) \) are distributed as the zeroes of \( \sqrt{1/r}B(rt) \), which of course are just the zeroes of \( B(rt) \). Thus \( Z \equiv rZ \) for all \( r < 1 \). However, bounded subsets of \( Z \) are not statistically self-similar (for example, although \( Z[0, 1/n] \) is equivalent to \((1/n)Z[0, 1] \), \( Z[1/n, 2/n] \) is not); thus we cannot estimate \( \dim Z \) by a similarity dimension. Another method is needed.

The method used is based on the distribution of the lengths of the gaps between points of linear dusts, that is, sets of disconnected points on a line, such as \( Z \) or the Cantor set \( C \). We recall that, for a linear dust \( E \), \( \dim E = D \) if the number of intervals
of length \( r \) required to cover \( E \) will be proportional to \( r^{-D} \). In the self-similar case, 
\[
D = \log N / \log(1/r); \quad N = r^{-D}.
\]
For more general dusts, we cannot talk about subsets which are smaller images of the original set; we can, however, deal with clusters of points ("tremas") contained in intervals between gaps of at least a given length \( u \). For example, \( C \) has 4 tremas between gaps of length \( \geq 1/9 \), 8 between gaps of 1/27 or more, etc.

What we need to require of the set \( E \) in order to use this to estimate \( \dim E \) is that the lengths of the tremas are proportional to the lengths of the gaps. (Clearly the Hausdorff dimension is based on length and number of tremas of length less than \( \delta \) as \( \delta \to 0 \); the number of tremas will, of course, be the same as the number of gaps, and if the lengths of the tremas are about the same as the lengths of the gaps, we can get a Hausdorff dimension from properties of the gaps.) Then, if there is a \( D > 0 \) such that, for \( u_2 > u_1 > 0 \), the gap lengths \( U \) have the property that 
\[
P[U \geq u_2 | U \geq u_1] = (u_1/u_2)^D
\]
(at least in the limit as \( u_1, u_2 \to 0 \)), the constant \( D \) will be equal to \( \dim E \).

This clearly works for the Cantor set, for which any two consecutive gaps of length \( \geq 3^{-n} \) are separated by a trema of length \( 3^{-n} \). Here, for \( u_1 = 3^{-n_1}, u_2 = 3^{-n_2}, n_2 > n_1 > 0 \), there are \( 2^{n_1} \) gaps of length at least \( u_1 \), and \( 2^{n_2} \) of length \( \geq u_2 \); as \( u_1, u_2 \to 0, n_1 \) and \( n_2 \) go to infinity, and
\[
P[U > u_2 | U > u_1] = ((2^{n_2} - 1)/(2^{n_1} - 1)) \approx 2^{-n_1}/2^{-n_2} = (3^{-n_1}/3^{-n_2})^{\log 2/\log 3} = (u_1/u_2)^{\log 2/\log 3}.
\]
Thus this comfortingly agrees with what we already know— that \( \dim C = \log 2/\log 3 \).

This, then, is the method used to find the dimension of \( Z \). It turns out that the tremas have lengths which, in expectation, are proportional to the lengths of the gaps, and that, for \( u_2 > u_1 > 0 \), the probability that a gap length exceeds \( u_2 \) given that it exceeds \( u_1 \) is 
\[
(u_1/u_2)^{1/2} \quad \text{(according to Mandelbrot (1982), p.236)},
\]
leading us to conclude that \( \dim Z = 1/2 \).

In fact, \( Z \) turns out to be the 1/2-dimensional case of what is called "Levy dust"— a self-similar fractal on the line which has the Markov property (that the "future", \( t > t_0 \), is independent of the "past", \( t < t_0 \), for any \( t_0 \)). A Levy dust \( L^D \) can be specified for any \( 0 < D < 1 \); here, we have \( Z \equiv L^{1/2} \).
3. The graph of Brownian motion.

We now examine the graph in \( \mathbb{R}^2 \) of a one-dimensional Brownian motion path against its time coordinate, that is, \( G = \{(x, t) \in \mathbb{R}^2 : t > 0, B(t) = x\} \). Note that \( G \) is not self-similar; since \( \sqrt{c}B(t/c) \equiv B(t) \), dilation on the time axis by \( r > 1 \) requires a dilation of the space variable by \( \sqrt{r} \) to retain statistical congruence to a dilation of the original graph. (This property of retaining congruence to the original figure after multiplications of different coordinates by different constants is called “self-affinity”).

The Hausdorff dimension of \( G \) is 3/2; the following explanation is taken from Mandelbrot (1982). In a small interval \((t_0, t_0 + \epsilon)\), the difference between the maximum and the minimum of \( B(t) \) is expected to be on the order of \( \sqrt{\epsilon} \), the standard deviation of \( B(t_0 + \epsilon) - B(t_0) \); thus it can be covered by a rectangle of height \( \approx \sqrt{\epsilon} \) and width \( \epsilon \), or by about \( \epsilon^{-1/2} \) squares of side length \( \epsilon \). (Remember that \( \epsilon \) is small, and so \( \epsilon^{-1/2} \) is large.) Now, let \( G[0,1] \) be the restriction of \( G \) to the time interval \([0,1]\); clearly \( \dim G[0,1] = \dim G \).

For large \( N \), on each time interval \([j/N,(j+1)/N]\) for \( j < N \), the graph can be covered by about \( (1/N)^{-1/2} = \sqrt{N} \) squares of side length \( 1/N \). There are, of course, \( N \) such time intervals in \([0,1]\), so \( G[0,1] \) should be covered by approximately \( N^{3/2} \) squares of side length \( 1/N \), and so \( \dim G = 3/2 \).

This gives us a new perspective on the zero set \( Z \) seen in the last section, from which we obtain a simpler explanation of why \( \dim Z = 1/2 \). The zero set is simply the restriction of \( G \) to the line \( \{x = 0\} \). Since all of the level sets \( \{x = c\} \) have the same distribution (except that they have different “starting points”), and they are created by the intersection of a set in the plane with a line, which is essentially a plane squeezed down onto one less dimension, the restriction of the plane set to the line should have dimension one less than that of the plane set. Thus \( \dim Z = 3/2 - 1 = 1/2 \), agreeing which our previous result.

4. Self-avoiding Brownian motion.

The term “self-avoiding Brownian motion” was coined by Mandelbrot due to the resemblance between the the set it describes and the “self-avoiding random walk”, a 2-dimensional standard random walk conditioned never to hit the same point twice. The
self-avoiding Brownian motion (SABM) is related to the "Brownian bridge" $T[0,1] = \{(x,y) : B_2(t) - tB_2(1) = (x,y) \text{ for some } t \in [0,1] \}$, a plane curve which is the trail of Brownian motion conditioned to have $B(0) = B(1) = 0$.

The SABM is defined to be the outer boundary of $T[0,1]$, or, more formally, the boundary of the unbounded connected component of $\mathbb{R}^2 - T[0,1]$; we denote this set by $F$. The exact Hausdorff dimension of $F$ is unknown; Mandelbrot conjectured it to be $4/3$, based on computer simulations of it as a limit of self-avoiding random walks, and Burdzy and Lawler (1990) have proven that $1 \leq \dim F \leq 3/2 - 1/4\pi^2 \approx 1.475$.

However, there are some topological properties of $F$ that can be examined, particularly in response to the question, "Does the self-avoiding Brownian motion really avoid itself?" For example, while the figure-eight is a closed curve, its outer boundary clearly does not avoid itself, but rather comes back to the middle point after it has already been there. To answer this question for the SABM set $F$, we define the terms "cut point" and "double cut point". A point $x \in \mathbb{R}^2$ is said to be a cut point of a Brownian path $B_2[0,1]$ if its "past" does not intersect its "future", that is, if for some $t \in (0,1)$, $B(t) = x$ and $B(0,t)$ is disjoint from $B(t,1)$; $x$ is a double cut point if the path $B[0,1]$ intersects it twice, and the path in between the hitting times is disjoint from the rest of the path, that is, $B(t_1) = B(t_2) = x$ and $B(0,t_1) \cup B(t_2,1)$ is disjoint from $B(t_1,t_2)$. If the path is topologically a figure-eight, the point of intersection will be a double cut point.

By "self-avoiding", we mean that $F$ is topologically equivalent to a circle; then, and only then, is it a closed curve which hits no point twice (excepting the point of beginning and end, which we are not considering). For this to be the case, we need to have no double cut points on $F$. (Actually, a double cut point may just set off a loop on the interior of the curve, and not affect the shape of the outer boundary; however, if a double cut point exists, the probability that it will set off an exterior loop is positive, thus making the outer boundary non-self-avoiding.) However, Burdzy and Lawler (1990) proved that, with probability 1, 2-dimensional Brownian paths have no double cut points; thus the "self-avoiding Brownian motion" is indeed self-avoiding.

It is interesting to note that, if we do not require that $B(1) = B(0)$, then the outer
boundary of $B[0,1]$ need not be self-avoiding; this is because a cut point of a non-closed Brownian path $B[0,1]$ will, by definition, separate its past from its future, and may (with positive probability) separate sets on which the outer boundary has closed loops (Fig. 3); thus the outer boundary need not be a simple closed curve, that is, self-avoiding. It was shown by Burdzy (1987) that a 2-dimensional Brownian path must (with probability 1) contain cut points; thus $B(1) = B(0)$ is needed to make $F$ necessarily self-avoiding.

REFERENCES


Figure 1(a).

Random walk with step size $1/8$

and step duration $1/64$. 
Figure 1(b).

Random walk with step size 1/4 and step duration 1/16.
Figure 1(c).

Random walk with step size $1/2$

and step duration $1/4$. 
Figure 1(d).

Random walk with step size 1 and step duration 1.
Figure 2(a)

Initiator of the Koch snowflake
Figure 2(b)

Generator of the Koch snowflake
Figure 2(c)
First stage of the construction of the Koch snowflake.
Figure 2(d)
Second stage of the construction of the Koch snowflake.
Figure 3. Cut point on Brownian path.