Fractals, similarity, intermediate asymptotics

Ya. B. Zel'dovich and D. D. Sokolov

S. I. Vavilov Institute of Physics Problems, Academy of Sciences of the USSR, Moscow, and M. V. Lomonosov State University, Moscow

Usp. Fiz. Nauk 146, 493–506 (July 1985)

Classical mathematical physics dealt only with perfectly smooth constructs; but today one works with curves having no tangent at any point, surfaces having no area, and other important objects once thought useful merely for tricky exam questions. Several aspects of these rough entities are described, such as the fractional-dimension concept, intermediate asymptotics, and the differential calculus of functions lacking derivatives, based on the Hölder index. Applications are possible in many branches of physics, including fluid dynamics, general relativity, and cosmology.

CONTENTS

1. INTRODUCTION

"The line is inconceivable," said the Skeptic philosopher Sextus Empiricus, "for the Geometers state that 'the line is length without breadth'; but we in our inquiry are unable to perceive length without breadth either in sensibles or in intelligibles; for whatever sensible length we perceive, we perceive as including a certain breadth."[1] So even in antiquity—for our extract dates from the end of the second century A.D.—people realized how limited the concept of dimension is when restricted to whole numbers. Little by little objects came to be envisaged that have more bulk than a line but nonetheless resemble it: a line with some breadth. Significantly, this feeling was long expressed in negative form, critical of ideas regarding one- and two-dimensional objects; indeed that is what the quote from Sextus conveys.

The concept of *line* has played a key role in the development of analytic geometry. One route here has been taken by differential geometry, a discipline concerned not only with lines and surfaces in our familiar environment of three-dimensional space. For progress in differential geometry soon gave rise to the idea of curved, non-Euclidean space, even multidimensional space, as well as a curved entity combining both space and time.

But perhaps a still larger contribution to the history of natural science has come from the invention of *analysis*—differential and integral calculus. Various formal, abstract definitions of derivatives and integrals were put forward; yet the very discovery of these principles was indissolubly linked with the concepts of motion and curves—the derivative as a rate of travel or a slope, the integral as a path length or an area. It is important to emphasize, though, that even motion and smooth curves were taken for granted. Newton and Leibnitz could hardly have arrived at the idea of derivatives by studying Brownian motion, tracing the velocities and trajectories of the microscopic particles that are caught up in thermal flight.

Thus Newton, Leibnitz, and one could add many other names right up to Euler, accepted continuity and the existence of derivatives as self-evident. Their silence on the matter did not remain unnoticed by succeeding generations. Textbooks abounded in such phrases as "the derivative, provided it exists." These notes of caution, like road signs warning the motorist of hazards, have a deep meaning. One can argue about just when certain reservations and complications ought to be introduced into teaching practice (the successful development of "nonstandard" analysis,[2,3] based on the concepts of infinitely large and small numbers, has shown that the critics have sometimes gone too far, and that the theory of limits can be handled in a spirit much closer to Newton and Leibnitz than is usually done), but the existence of unsmooth functions and the like is itself beyond any doubt.

In fact, in the latter half of the past century mathematicians of the school who criticized the foundations of analysis, above all Karl Weierstrass and Giuseppe Peano, devised functions that were continuous but nowhere had derivatives, as well as curves that everywhere densely filled a square. From the modern viewpoint the strange properties of these objects reflect their having been thought of as one-dimensional, whereas it would be more natural to regard them as objects of higher—including fractional—dimension, or in present nomenclature, *fractals*. But the preliminary steps in the study of fractals were not taken by design; the applications to physics and otherwise had not been foreseen, and the term "fractal" was coined by Benoit B. Mandelbrot[4] only in 1975.

 0038-5670/85/070608-09$01.80

Contemporaries of the critics often perceived their complaints as destructive to mathematics. In 1893 Charles Hermite wrote to Thomas Jan Stieltjes that he "turned away in fear and horror from this lamentable plague of functions with no derivatives."[5,4] Critics belonging to the next generation of mathematicians focused just as intently on the most fundamental aspects of their science: the concepts of set, natural number, proof, and so on. A resounding reaction to the earlier attitude was the celebrated utterance of David Hilbert (he himself was a prominent critic, although the dividing line between critics and noncritics is of course a very arbitrary one), who "refused to be driven out of the paradise which Georg Cantor had created" (see Hilbert's Weierstrass lecture[6]).

Starting from a different point, it was not a mathematician but an eminent theoretical physicist, the late Paul A. M. Dirac, who assimilated the nonclassical tenets in a positive, constructive way. His idea of the δ-function made a deep impression on mathematicians and physicists alike, compelling them to rethink the criticisms and to extract from them what might be of practical use. In speaking of how Dirac's approach influenced the development of fractal theory, we particularly have in mind the psychological effects, the change in the scientific atmosphere. For these questions are eternal riddles of science; they have to be examined anew at every stage of scientific progress, and the shifting viewpoints are of no less significance than would be the cultivation of some specific new problem.

By way of comparison, a similar change in attitude occurred among twentieth-century mathematicians of critical bent, when they found that the doctrines of the Scholastics and the ancient logicians, long since cast aside by the march of natural science, had unexpectedly become very relevant. A twentieth-century author can take issue both with his own contemporaries and with Aristotle[1)] (see Mikeladze's essay[8] for some examples). Actually the turnings of the tide extend over rather long intervals, with different views coexisting in contention. Thus, in a sense the physical applications of fractals began with the work of Albert Einstein and Marian Smoluchowski on Brownian motion at the start of our century. By the same token the dread of singularities in general relativity theory, a panic that held sway until just recently, was essentially an echo of the belief that undifferentiable objects are inadmissible.

From the standpoint of modern science a function with no derivative certainly is not just an abstract notion drawn from an arsenal of cunning questions set in exams on mathematical analysis, but the trajectory of a Brownian particle. So broken is the trajectory that it should be regarded as a "thick" line, a fractal. As we shall see, the description of fractals is itself very close to Weierstrass' example of a function nowhere differentiable. Weierstrass in effect was in possession of the fractal concept without suspecting it! Among the mathematician's tools at that time was analytic apparatus capable of describing such uneven objects. Taking the place of ordinary dimensions is the fractional dimensionality first introduced by Felix Hausdorff early in this century,[9] while the derivative is replaced by the so-called Hölder index or fractional derivative, a concept put forward by any number of mathematicians. These ideas will be discussed below, as well as the physical problems to which they pertain.

The fractional-dimension concept rests on analysis of conventional integer dimensions. We all have a feeling for what is meant by a line or circle being one-dimensional, a plane or sphere being two-dimensional, a ball or space being three-dimensional, and so on. Roughly speaking, we mean that the position of a point on a line can be specified by one coordinate; on a plane, by two; and in space, by three. That quantity—the number of coordinates—cannot be a fraction. To look for a way of introducing fractional dimensions, two steps are necessary: we have to find some relationship that characterizes dimensionality but does not rely on integers, and we need to pin down the weak point in our naïve ideas about dimensions, eliminating it so that we can ascribe to certain objects a fractional dimension.

The program just outlined can be put into practice as follows. With one-dimensional objects we associate the concept of length; with two-dimensional objects, area; and with three-dimensional ones, volume. Characterizing these concepts is a construct which by no accident similarly is called dimensionality: [cm], [cm^2], [cm^3]. From other branches of physics we know that these dimensions can also be fractional; in cgs units, for example, electric charge has dimensions of $[\text{g}^{1/2}\cdot\text{cm}^{3/2}\cdot\text{sec}^{-1}]$.

But in order to pull this fractional dimensionality along into fractal theory we have to look at the concept of coordinates somewhat more broadly. To any preassigned accuracy we can specify a point inside a square not just by a pair of coordinates but also by only one, provided we use a coordinate line that fills the square more and more densely. In fact a coordinate of that kind is by no means exotic. For instance, a person's address in a city could in principle be specified by giving the geographic coordinates of his residence; but we use a different procedure, first naming the street and then the building and apartment numbers (like a coordinate with integer and fractional parts). Indeed one could actually number all the city's dwellings in a single sequence (the Tokyo address system comes closest)—(transl. note: or the American nine-digit ZIP code). Now it no longer seems oversubtle to ask what the dimensions of a city are. If we think of it as an ensemble of streets, the city would be one-dimensional; as an area on the earth's surface, two-dimensional; if we allow for the heights of apartment houses, three-dimensional; and might there not be intermediate dimensions as well? In other words, how should we characterize a city: by the combined lengths of its streets (summing the first powers of the individual block lengths, from one intersection to the next), by its area (summing the squared block lengths, if the street grid is rectangular), or, for appropriate questions and types of cities, by summing some other power of the block lengths, say the three-halves power? We shall examine this approach more fully at the end of the next section.

The problem with the concept of fractional—indeed even integer—dimensions is that the route just outlined is not the only possible one; other approaches can be formulated as well, equally natural but leading to different results. We will begin by taking an approach less familiar to physicists but historically earlier—that of Hausdorff, who relied essentially on ideas that had been expressed by Euclid. Underlying the Hausdorff approach is the recognition that one- and two-dimensional structures are in effect three-dimensional portions of space, two or one of their characteristic scales being very small.

Before we turn specifically to fractals, let us recall two basic ideas on which that concept rests. When we speak of the unsmooth, broken-line trajectory of a Brownian particle or its infinitely high velocity, we naturally are idealizing the situation. On very tiny scales the finite mass of a Brownian particle and the finite intercollision time will manifest themselves, and the trajectory will become smooth. When we speak of a fractal surface we should think of a rough surface whose scale of irregularity gradually becomes smaller as the projected area of the irregularities diminishes. The irregularities should, however, still be much coarser than the interatomic distance scale; otherwise the concept of a boundary for the body would not apply at all. When we say that a lengthy molecule fills up a region of space (the corresponding mathematical image is called a Peano curve), we mean that from some filling factor onward we have to allow not only for the molecule's extension along a line but also for its thickness, and we can no longer describe the situation in terms of tangled lines. That is how in its most general form the fractal concept ties in with *intermediate asymptotics*[10]: although the scale of roughness is small, it remains much larger than something still smaller.

The second basic idea likewise is illustrated by the example of a rough surface. Imagine that on the surface is a whole hierarchy of mounds having a common base area but differing in height, so that the higher mounds occur much more rarely than the low ones. Clearly you can't describe a roughness of this kind by any single number, as a fractional dimension. Hence the fractal concept must build upon a premise which excludes such complications—upon the postulate that the corresponding power-spectrum expansions contain random phases. To deal with other situations one either has to treat singularities separately, as in δ-function theory, or one must introduce a hierarchy of dimensions peculiar to individual roughness systems (see Sec. 6).

On the whole the development of the fractal concept represents a typical story of the mathematically "impossible" being surmounted. Mandelbrot's book in its successive versions[4] has played a decisive role in the emergence of these ideas. In our country the first survey of fractals was published as a chapter in Barenblatt's book on similarity.[11]

2. FRACTALS AS THICK LINES

Suppose that we peer at a razor blade with a stronger and stronger microscope. First, to the unaided eye, it will look smooth. Then rough places and notches will appear; these will enlarge, and later we will see the object's crystalline structure. Its boundary will no longer have any definite meaning, and we will enter the quantum world. In order to characterize the intermediate-asymptotics region, where the body surface and the blade have become uneven although we are still far away from interatomic scale, we now need to introduce the concept of fractals (Fig. 1).

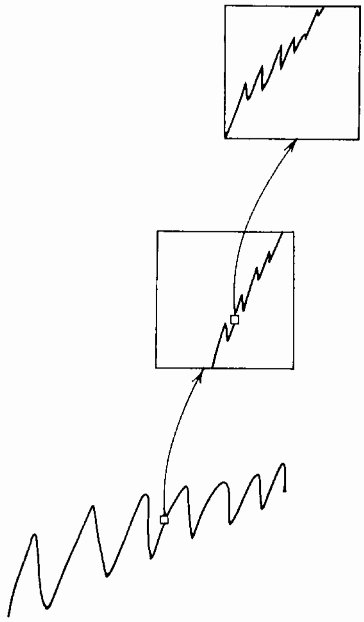

FIG. 1. A fractal curve has a uniform structure over a wide range of scales.

As mentioned in the Introduction, one of the first actual mathematical examples of a fractal was devised by Weierstrass[12] to illustrate a function without a derivative. Geometrically, his function is a curve $y = y(x)$ with a single-valued projection on the x axis and specified by the series

$$y(x) = \sum_{n=0}^{\infty} A^n \cos(B^n \pi x),$$

where $0 < A < 1$, while the product AB is suitably large (Weierstrass originally took $AB > 1 + 3/2$, but we shall understand $AB > 1$, a condition established by Hardy[13]). It will be fitting for us at the outset to consider the more general case of functions

$$y(x) = \sum_{n=0}^{\infty} a(k_n) \cos(k_n x + \varphi_n), \tag{1}$$

where the $a(k_n) \sim k_n^{-\alpha}$, with the $k_n \sim n \to \infty$.

Although Weierstrass' example can readily be put into the form (1), there are wide gaps in the summation and all the phases φ_n vanish. It is well understood (see, for instance, Zygmund[14]) that the rate α at which the Fourier spectrum falls off corresponds to the number of derivatives of the function $y(x)$. In fact, on differentiating the series (1) term by term we find

$$\frac{d^j}{dx^j} y(x) = \sum a(k_n)(k_n)^j \cos(k_n x + \psi_n).$$

For the original series to converge we must have $a_n \sim k^{-\alpha}$, with $\alpha > 0$. Clearly the series for the jth derivative will converge only if $\alpha > j$.

It is important to emphasize that all the arguments to follow rely on the random-phase approximation; in other words, they apply to a representative function of the form

$$y(x) = \sum a(k_n)\cos(k_n x + \varphi_n),$$

where the φ_n constitute a sequence of independent random numbers distributed uniformly between 0 and 2π. For any individual function the slowness in the decline of the spectrum will be related, and simply so, to the roughness at some one point. The δ-function, for example, has a flat spectrum, $a(k) \equiv \text{const}$; for the θ-function, $a(k) \sim 1/k$, but these functions are discontinuous only at the origin of coordinates. Fractal curves, on the other hand, are so constructed that in a statistical sense all their points are alike.

We would also point out that the k_n have not been identified with the integers n themselves; we merely stipulate that $k_n \sim n$. What is important for our purposes is that not only the phases but also the frequencies of the summed harmonics are random in nature. Otherwise if α is small enough ($\alpha < 1/2$) the curve $y(x)$ can develop randomly distributed peaks going off to infinity. Suggestive of that possibility is the fact that a series of the form $y = \Sigma(\sin nx/n^\alpha)$ is unbounded near the points $x = 0$ and $x = 2\pi$; indeed in this example

$$\int_0^{2\pi} y^2(x)\,dx = \infty,$$

and the series convergence tests fail to ensure continuity of the summed series at just those points where $\sin(x/2) = 0$.

From our vantage point today, Weierstrass' most significant result was to verify that the function he had constructed did satisfy the criterion of being "in general position." But the general-position argument itself was only formulated later, after Cantor had developed the theory of sets. This consideration strongly affected the mathematician's outlook: in the nineteenth century people spent years in proving, say, that the number π is transcendental, while now we are usually satisfied to say that almost all numbers are transcendental.

Thus according to the theory of trigonometric series, if the exponent α is in the range $0 < \alpha < 1$ then the function $y(x)$ will be continuous but not differentiable. The curve corresponding to it will be a fractal. If $1 < \alpha < 2$, the "trajectory" $y = y(x)$ will be smooth, but motion along it will take place at fractal velocities. And if we want all the derivatives to be nonfractal [$y(x) = 1/(1+x^2)$ or e^{-x^2}], then the spectrum $a(k)$ should fall off faster than any power of k (that is, $\sim e^{-k}, e^{-k^2}$).

We have been making negative, limiting statements: to a function whose spectrum does not fall off fast enough, the concepts of analysis are not fully applicable. By introducing Hölder indices we can give these assertions a positive character, saying that the function $y(x)$ does not have one derivative but a smaller number, α, of derivatives. If the fundamental formula of "smooth" analysis is written as

$$\Delta y = \mu\,\Delta x,$$

we may write for our function (see Fig. 2)

$$\Delta y = \mu_H (\Delta x)^\alpha,$$

where μ, μ_H denote respectively the ordinary and Hölder derivatives (properly speaking, left and right derivatives

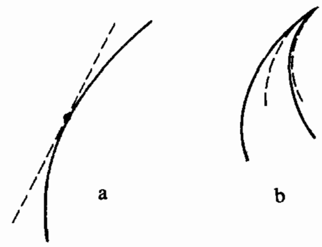

FIG. 2. The structure of smooth and fractal curves in the neighborhood of a point. a) A smooth curve has a tangent (dashed line); b) a fractal curve has a curvilinear cone $\Delta y \sim |\Delta x|^\alpha$ (dashed curve) serving as a tangent.

μ_{H_-}, μ_{H_+} will arise for $\Delta x < 0$, $\Delta x > 0$). With this convention we can work with $y(x)$ in much the same way as with a smooth function.

There are two familiar examples of physical objects having such a low smoothness. One is Norbert Wiener's process, a mathematical abstraction of Brownian motion in which the mass of the Brownian particle and the intercollision time both tend to zero. It is the displacement of the Brownian particle during time t that is called the Wiener process: $x(t) = w_t$ (here letter w stands for Wiener, rather than for probability, the custom in many branches of physics). At any given time the particle will undergo an acceleration analogous to a δ-function; its correlator will have a flat spectrum. Hence the sum of numerous instantaneous impulses will cause a displacement $w_{t+dt} - w_t = w_{dt} = dw_t \sim (\Delta t)^{1/2}$, and the Hölder index will be 1/2 (the Brownian-particle trajectory will have a half-derivative).

We cannot go into detail here about the mathematical analysis applicable to Brownian-motion–type functions, but we should like to mention a remarkable formula of Kiyosi Itô[15] for calculating the differential of the function $F(w_t)$, where F is a smooth function. In this case dF, the increment to a quantity of order dt, is expressed not by $F'dw_t$ but by $dF = F'dw_t + 1/2F''(dw_t)^2$. It is the second term in Itô's formula which enables one not only to describe diffusion from an analytic point of view in diffusion-equation language (à la Langevin, Smoluchowski, and Wiener) but later on to obtain formulas such as that of Kac and Feynman for solving evolutionary equations by means of Wiener (or, in quantum mechanics, Feynman) integrals. Further information on this analytic technique and its applications to describe thermal conductivity and diffusion will be found in the review literature.[16,17]

The other example is Kolmogorov turbulence. If we discount the fact that at very short distances (or equivalently, for very large wave numbers) viscosity comes into play, then the velocity field will experience variations of order $\delta v \sim (\delta r)^{1/3}$. Accordingly the velocity field of Kolmogorov turbulence will be continuous, but it will have only a one-third derivative. Of course we are dealing with a more complicated entity here—a rough vector field.

Now let us see what the dimensions of the resultant fractal curves are. Hausdorff's approach was to define fractional dimensionality as follows. Around every point of our set we construct a circle of radius $\varepsilon \to 0$, and we add up the area $S(\varepsilon)$ of all the circles combined (actually it suffices to construct a finite number of circles, of order $1/\varepsilon$). Some of the circles will overlap and naturally in computing $S(\varepsilon)$ we count the overlap areas only once. It is the rate at which $S(\varepsilon)$ diminishes toward smaller ε that determines the dimension. In fact, for a smooth curve the area $S(\varepsilon) \sim \varepsilon L$, where L is the curve length. If the curve degenerates to a point, then $S(\varepsilon) \sim \varepsilon^2$; for a plane domain, $S(\varepsilon) \sim \varepsilon^0$. In the case of our fractal we can estimate the area in the following way (Fig. 3). Let $k = 1/\varepsilon$ denote the wave vector; then individual sine-wave periods are contained within an ε-circle, and if the quantity $a(k_n) \equiv a(1/\varepsilon)$ falls off more slowly than ε as $\varepsilon \to 0$, the band of circles will have a width not of order ε but of order $a(1/\varepsilon)$. If instead $a(1/\varepsilon) < \varepsilon$, the band of circles will be able to track all the kinks in the curve, which will therefore be a smooth rather than a fractal curve.

We now return to the fractal curve. Let its projection have a length of order unity. The combined area of all the ε-circles will be $S(\varepsilon) \sim a(1/\varepsilon) \sim \varepsilon^\alpha$. If $0 < \alpha < 1$, the area $S(\varepsilon)$ will diminish more slowly than for a smooth curve, and our fractal will occupy a place intermediate between a line and a plane region. Hausdorff proposed a definition whereby an object γ such as this fractal would have a dimension

$$\dim_{\text{ext}} \gamma = 2 - \alpha.$$

The subscript, standing for "exterior," means that to evaluate this quantity we have to go beyond the bounds of the curve itself.

This example has a serious shortcoming: the x axis has been singled out. But we can readily introduce a modification to make the axes equivalent. To this end we regard the curve γ as expressed parametrically:

$$x = \sum a(k_n) \cos(k_n t + \varphi_{1n}),$$
$$y = \sum b(k_n) \cos(k_n t + \varphi_{2n}).$$

In order for the two directions to be equivalent, the coefficients a, b have to fall off by the same law:

$$a \sim b \sim k^{-\alpha}, \quad 0 < \alpha < 1.$$

If we estimate the area of an ε-neighborhood of our curve in two dimensions, we will again obtain ε^α, as can easily be seen (rather than what one might at first have expected, $\varepsilon^{2\alpha}$, because we now have to consider the departure of the roughness from some average state). We still have $\dim_{\text{ext}} \gamma = 2 - \alpha$.

However, the fractal dimension of our curve can be defined in a different way by utilizing only concepts associated with the curve itself, not with how the curve γ is situated in the plane. We have alluded to this approach in the Introduction. Let us assign a parameter t along our curve. We subdivide the t axis into segments of length ε and compute the length of γ, taking into account only those kinks for which t changes by at least ε. In this fashion we find that the segments will have a combined length of order $a(1/\varepsilon)(1/\varepsilon) \sim \varepsilon^{\alpha-1}$, becoming infinite as ε decreases. Notice that the only thing of importance to us in this argument is our measurement of the distance along the curve on a definite scale of t-variation, not the placement of the curve in the plane.

Let us try to interpret the divergence we have found in the cumulative length as ε diminishes. Suppose that we have erred in evaluating the dimension of our object γ and that we actually are not studying a curve after all, but are seeking to parametrize a region with a single parameter. The parametrization will be very poor, of course: the coordinate line will grow denser and denser, crossing itself as it fills up the region, and forming a sort of lattice there (Fig. 4). Successive strips in the lattice will be separated by a distance of order ε, while the number of quadrangles will be $\sim 1/\varepsilon^2$ (each strip corresponds to an individual rough spot on the curve). If we sum the lengths of the segments in this broken line we will obtain a quantity $\sim \varepsilon \cdot 1/\varepsilon^2 \to \infty$. Actually the region is two-dimensional, so what we need is not its length but its area; that is, we should sum not the lengths of the sides, but the squared lengths.

If the dimension is fractional, however, we must add up not the first or second powers but the μth powers of the lengths, for some value of μ. In order to keep the resulting sums finite, which will be true if we have chosen the dimension correctly (as in our example involving a region), we should set $\mu = 1/\alpha$. The sum will then have dimension $[\text{cm}^{1/\alpha}]$, the number $1/\alpha$ itself evidently representing the dimension. Accordingly it is natural to adopt for the interior dimension of a fractal γ the number

$$\dim_{\text{int}} \gamma = \frac{1}{\alpha}.$$

This expression also holds, of course, for curves in a space of arbitrary number of dimensions.

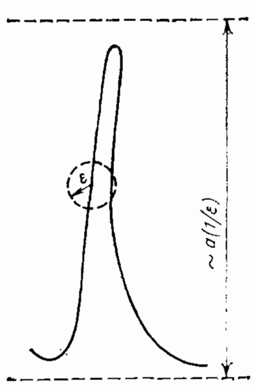

FIG. 3. How the exterior dimension of a fractal is defined. The width of a neighborhood of the fractal curve is determined not by the radius of the small circle but by the size of the curve's kinks.

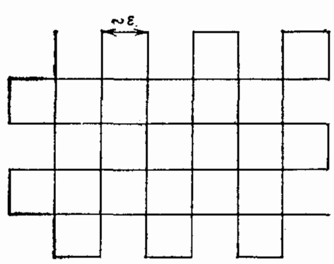

FIG. 4. How the interior dimension of a fractal is defined. The links in the broken line that densely covers the plane region are short, but there are a great many of them. Even so, the sum of the squared link lengths remains finite.

In a plane, a fractal-type curve can have an exterior dimension ranging from 1 to 2 (the dimension of the space containing the curve), and an interior dimension from 1 to ∞. The two will coincide only for the trivial case of a smooth curve. When dealing with various physical problems one has to use different definitions of fractal dimensionality. For example, if we are interested in the problem of adsorption on a thin fiber, we will need to know how many atoms can fit next to the fiber, that is, its exterior dimension. If instead we want to estimate the filament's weight, the significant dimension will be the interior one.

To compute the exterior dimension of a fractal curve in 3-space, we write

$$x_i = \sum a_i(k_n)\cos(k_n t + \varphi_{in}), \quad a_i(k) \sim k^{-\alpha}.$$

One can easily estimate the volume of an ε-neighborhood of the curve: $V(\varepsilon) \sim \varepsilon^{2\alpha}$. Then by analogy with the two-dimensional case,

$$\dim_{\rm ext}\gamma = 3 - 2\alpha.$$

More generally, in n-space,

$$\dim_{\rm ext}\gamma = n - (n-1)\alpha.$$

We see, then that the exterior dimension of a fractal curve lies between the dimension of a smooth object and the dimension of the surrounding space, while the interior dimension can take any value between the smooth-object dimension and infinity.

3. FRACTALS AS THICK SURFACES

A rough surface can be assigned a fractal dimension in a similar way. Suppose that our surface Φ is specified by the parametric equations

$$x_i(u, v) = \sum\sum a_i(k) e^{i(k_u u + k_v v)},$$
$$a_i \sim |k|^{-1-\alpha}.$$

We are now working with a double Fourier series, and in order for it to be summable the exponent α has to exceed 1. Consequently in the fractal interval of interest to us, $0<\alpha<1$, the double series diverges in the conventional sense, and we can sum it only by taking principal values—that is, first the harmonics within a sphere of fixed radius, and then along the radii.

Let us break down the (u,v) parameter domain into triangles of diameter ε. Although a subdivision into squares would be more easily visualized, a pattern of triangles (a triangulation) is much simpler technically, because unlike a quadrangle a triangle is a rigid figure, being determined solely by its side lengths.

We add up the areas of the triangles tesselating the surface, again taking into account only harmonics no shorter than ε. The sum will be of order $[a(1/\varepsilon)(1/\varepsilon)]^2(1/\varepsilon^2)$, and for α in the range $0<\alpha<1$ it will increase without bound as ε diminishes. To keep the sum finite we have to sum the $(1/\alpha)$-th powers of the areas; thus to obtain the correct power of the unit of length we should take the dimension of the object to be

$$\dim_{\rm int}\Phi = \frac{2}{\alpha}, \quad 0<\alpha<1.$$

As for the exterior dimension of a fractal surface, it will have the Hausdorff value

$$\dim_{\rm ext}\Phi = 3-\alpha, \quad 0<\alpha<1,$$

since an ε-neighborhood of the surface has a volume of order $a(1/\varepsilon)(1/\varepsilon)$.

The noncoincidence of the interior and exterior dimensions of a fractal surface means that unlike the case for regular surfaces differentiable repeatedly, for a fractal surface the relationship between the interior and exterior properties is violated. In other words, we can compute, say, the curvature of a regular surface in two ways, either from the Riemann curvature tensor (as in general relativity) or from the curvatures of surface sections, and the results will coincide; but generally speaking this will not be true for a fractal surface. Surprisingly, for smooth but irregular surfaces—surfaces that have a tangent plane but no higher derivatives—we actually encounter a fractal-like behavior, even though their dimension is exactly 2. This interesting phenomenon, discovered by the American mathematician John Nash, comes about as follows.

First consider a regular surface whose area is $\sim S$. If the surface is closed, one would naturally expect a body inscribed in the surface to have a volume of order $S^{3/2}$. The precise coefficient will of course depend on the particular form of the surface—on the distribution of its Gaussian curvature, but the dimensional factors will always be the same. What Nash found[18] is that if the surface has a smoothness between 1 and 2, so that an extrinsic curvature can no longer be calculated from the standard formulas, then the area of the surface will be quite unrelated to the enclosed volume: within an arbitrarily small sphere one can inscribe a globe so severely wrinkled (although in such a way as to preserve the metric, that is, all distances between points along the surface) that its area is as large as desired. It is presently believed[19] that this phenomenon will dispapear when the smoothness reaches a limit somewhere between 1.07 and 1.7. After a fashion, the effect displayed by Nash surfaces is reminiscent of that forecast by the Czech novelist Jaroslav Hašek, who has one of his heroes avow that "inside the terrestrial globe there's another globe, a good deal bigger than the one outside."

The simplest concept of dimensionality, the one familiar to all, is widely used in biology to interpret the relative populations of different animals. Perhaps for some species it might also be possible to discern certain properties inherent in Nash-type boundaries. Fractal dimensions have in fact been invoked to model the absorptivity of lungs.

4. FRACTALS AS SPACETIME FOAM

The interior dimension of a surface carries over without significant change to the case of curved spaces in general relativity theory (space being subdivided, of course, not into triangles but into their four-dimensional analogs, simplexes). If spacetime on microscales resembles a foam, as Wheeler first suggested,[21] the dimension perceived by a macroscopic observer could be very different from the microscopic dimension. These questions, however, still have received very little attention from mathematicians, although certain assessments are now feasible.

In cosmology one often employs a "flat" spectrum of primordial fluctuations, in which all harmonics of the perturbations in the density of a Friedmann model are equally probable. Plainly, this flat spectrum can only be viewed as an intermediate asymptotic. At one time such a spectrum was regarded as a convenient, reasonable premise to cover a situation about which we lack more definite knowledge[22]; but it has now become clear, in the context of inflationary cosmology,[23] that theory can indeed predict a model with critical density (that is, comoving Euclidean space) and a flat density-perturbation spectrum.[24] The curvature of the model then will also have a flat spectrum, by analogy to the second derivative of a Wiener process. It would therefore be natural to suppose that some analogy exists between a Wiener process and the metric itself. That would imply, as we have seen, that in the intermediate-asymptotics region the dimension of comoving space would not be 3, but rather more than 3.

At a later stage the fractal character of space will have been lost, giving way to a cell-lattice structure[25] that represents the structure in the distribution of matter, not of space. To be sure, in light of Einstein's equations wherever the density becomes infinite in this structure, a singularity will also occur in the metric. It will resemble the singularity at the place where two non-coaxial cylinders cross each other.[26] But the corresponding power spectrum of the metric will be strongly phased—no longer a fractal spectrum.

In attempts to describe the development of small initial fluctuations, their growth rate, and the rate of cosmological expansion, an allowance for fractal dimensionaltiy seems unlikely to alter the prevailing estimates, but may be useful for resolving the dilemmas posed[27] by the small amplitude of the angular irregularities in the cosmic microwave background.

It is interesting to consider Nash surfaces in multidimensional space as well. A multidimensional surface embedded in a space of high enough dimension may exhibit Nash properties, even being differentiable an arbitrary number of times.[28] For a four-dimensional surface, the dimension of the space that would envelop such crumpled Nash surfaces is estimated[29] to lie somewhere between 11 and 29. In this regard it is pertinent to recall an idea remote from fractal theory: the multidimensional Kaluza-Klein generalizations of field theory. Space has a higher dimension on microscales in these models than it does on macroscales, because the extra dimensions turn out to be periodic in coordinates whose period is vanishingly small.

Another way to lower the dimension as one goes over to macroscales has also been proposed.[30] One can suppose that a potential develops in multidimensional spacetime analogous to the potential which keeps quarks within nucleons; but instead of preventing the escape of quarks it would inhibit the free existence of particles outside a four-dimensional spacetime surface. It would seem that the presence of crumpled Nash surfaces in the enveloping space ought to render this picture inadmissible, for the surrounding space would have a dimension ranging from 11 to 29. The lower limit of this dimension interval is in curious coincidence with the dimension of enveloping space that is inferred from particle theory.[31]

5. FRACTALS AS CONTOUR LINES

In physics the fractal concept was originally put forward to cope with a single salient example: measuring the length of the coast of Great Britain. A shoreline, one will recall, can be defined as follows. In a plane, introduce a function $h(x,y)$ to represent the elevation above sea level; then the shoreline will comprise the solution of the equation $h(x,y) = 0$.

It was found that the larger the scale the map was drawn on, the longer the coast seemed to be. Thus a coastline turns out to be a fractal, whose length ought to be measured by introducing a quasilength with dimensions of centimeters to some power. Even though this argument has long been recognized by specialists it has not yet come into cartographic practice. To illustrate, the administration of the National Park of Lithuania reports the area, depth, and shore length of all the lakes in the park, although the shorelines can hardly have been rigorously measured, whereas the diameter of the lakes (a perfectly definite quantity) is not quoted. For practical purposes, of course, one can accept a shoreline if one knows on what scale it was measured; but that is just the information which tends to be omitted.

The fractal definition that we have given does not fully correspond to this classical problem. When studying a coast in detail one encounters along with the main shoreline a system of reefs or islets; a network of fine contour lines will demarcate bodies of water and islands that lie next to the shore, as suggested in Fig. 5. By contrast, our parametrized fractal structures can have crossing points (such as in two-dimensional Brownian-type motion), whereas a contour

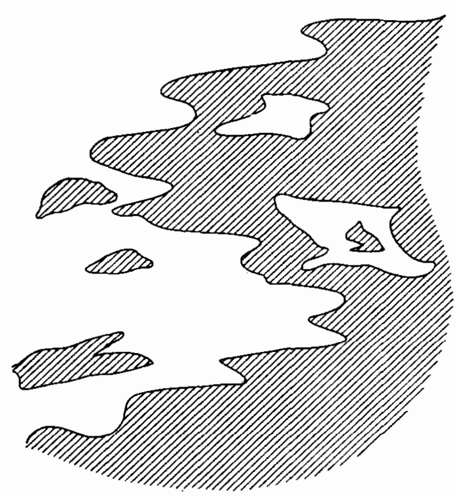

FIG. 5. The structure of the contour lines for a random function. Shaded areas represent water-filled depressions. Dry land is studded with lakes; the adjacent sea, with islets.

line, with probability unity, will experience no such self-crossings (for that to happen, a random function and its gradient would both have to become zero, giving an indeterminate system of equations).

To define the fractal dimension of a contour line by a random function we must again determine how many Hölder derivatives it has. That number, α, is specified by the behavior of the spatial correlation function over short distances: for a smooth random field,

$$\langle \varphi(x)\, \varphi(x+r) \rangle = 1 - Ar^2$$

as $\mathbf{r} \to 0$, and for a fractal field,

$$\langle \varphi(x)\varphi(x+r) \rangle = 1 - Ar^{\alpha}.$$

The exponent α expresses the fractal dimension of the boundary line.

6. FRACTALS AS DENSE POINT SETS

Up to this stage the fractal dimension has always been fairly sizable. To obtain objects of small dimension, say 1/10, we have to consider classes of numerous (in the limit, infinitely many) points which densely fill a fractal formation. We can evaluate the Hausdorff dimension of such a fractal in the plane by surrounding each of its points with a little circle of radius ε, adding up the area of all the circles, and establishing how that area depends on ε. In mathematics such fractals have long been known—an example is the Cantor set, a discontinuum (an everywhere unconnected set with the power of the continuum); but these objects were thought to be no more than a playground for clever mathematicians. Today we know that Cantor-type sets are more than toys; they exemplify *strange attractors*,[32] such as the fractal set of zeros of a one-dimensional Wiener process (which has dimension 1/2). It has lately been proposed,[33,34] for instance, that fractals of this sort might serve to describe the intermittency of turbulence.

Indeed, as we mentioned in Sec. 2, in Kolmogorov turbulence the velocity is continuous but has only a one-third derivative. The situation can be depicted graphically as follows. Along one axis let us plot the number σ of derivatives of the velocity, and along the other axis the Hausdorff dimension d of the set comprising the points at which the velocity does *not* have σ derivatives (Fig. 6). Then in the case of Kolmogorov turbulence d will be given by a step function of the type

$$d(\sigma) = 3\theta\left(-{}^1/_3 + \sigma\right),$$

so that if $\sigma < 1/3$ the differentiability property will exist almost everywhere, but if $\sigma > 1/3$, hardly anywhere. Such a function $d(\sigma)$ corresponds to a Guassian velocity field, devoid of intermittency.

If structures are present in the turbulence and it is decidedly non-Gaussian, we would expect $d(\sigma)$ to vary gradually from 3 to 0. In experimental practice it is more convenient to determine some relationship for the dependence of the higher moments of the velocity field on distance and moment number.[34] Such measurements have been performed[35] and they enable functions $d(\sigma)$ to be constructed

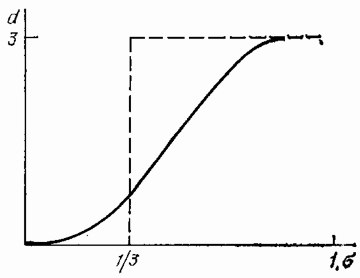

FIG. 6. How the intermittency of turbulence can be described by fractional dimensions. *Dashed line*, dimensionality of Kolmogorov turbulence; *solid curve*, real turbulence.

for real turbulence (curve in Fig. 6). Thus far $d(\sigma)$ is none too accurate but the curves definitely indicate departures from a Gaussian state. The $d(\sigma)$ curve can be described verbally by saying that the velocity field has about a 1/10 derivative almost everywhere, but a 9/10 derivative hardly anywhere; in between the dimension of the point set undergoes a smooth change.

To close our account of fractal structures, we would ask the reader to keep in mind that in the wake of the nineteenth-century era when fractal theory was born, the era of mathematical criticism (the twentieth century) has given us a great many other strange creatures with which, unlike fractals, we still don't quite know what to do. Monsters have been identified such as extraordinary sets that are their own elements, non-Aristotelian logics, intuitional mathematical analysis. In time perhaps these too will find their place in physics. To those who might wish to contemplate this subject we would commend the book by Fraenkel *et al.*,[36] one of the few available monographs written in language comprehensible to the nonspecialist.

Think of the structure of space and of fields once quantum zero-point vibrations have been introduced, but remember as well how particles are left unscattered by those structures.

[1)]Yet another parallel is usually omitted from current discussions: paradoxes such as that of Cesare Burali-Forti (how big is the biggest set?) and the set-theoretic methods for resolving them bear some resemblance to the problems encountered in analyzing the concept of the supreme good as presented in the *Nicomachean Ethics* of Aristotle.[7]

[1]W. G. Bury (tr.), Sextus Empiricus [Against the Physicists **1**, 390–391], Harvard Univ. Press (1936), Vol. 3, p. 187 [Russ. Transl., **1**, Mysl, M., 1975].
[2]P. Cartier, "Perturbations singulières des équations différentielles ordinaires et analyse non-standard," Séminaire Bourbaki 1981/82, Soc. Math. France, Paris (1982), pp. 21–44 (No. 580) [Usp. Mat. Nauk **39**, No. 2, 57 (No. 236) (1984)].
[3]A. K. Zvonkin and M. A. Shubin, Usp. Mat. Nauk **39**, No. 2, 77 (1984) [Russ. Math. Surveys **39**, No. 2, 69 (1985)].
[4]B. B. Mandelbrot, Les Objects Fractals: Forme, Hasard et Dimension, Flammarion, Paris (1975); Fractals: Form, Chance, and Dimension, Freeman, San Francisco (1977); The Fractal Geometry of Nature, Freeman (1982, 1983), 468 pp.
[5]C. Hermite, in: Correspondance d'Hermite et de Stieltjes, ed. S. Baillaud and H. Bourget, Gauthier-Villars, Paris (1905), **2**, 318.
[6]D. Hilbert, Math. Annalen **95**, 161 (1925).
[7]Aristotle, Works [Russ. Transl. **2**, Mysl, M., 1978].
[8]Z. Mikeladze, "Foundations of Aristotelian Logic" [in Russian], *ibid.*, p. 5.

[9]F. Hausdorff, Math. Annalen **79**, 157 (1918).
[10]G. I. Barenblatt and Ya. B. Zel'dovich, Usp. Mat. Nauk **26**, No. 2, 115 (No. 158) (1971) [Russ. Math. Surveys **26**, No. 2, 45 (1972)].
[11]G. I. Barenblatt, Podobie, avtomodel'nost', promezhutochnaya asimptotika Gidrometeoizdat, Leningrad, 1978 [Engl. Transl. Similarity, Self-Similarity, and Intermediate Asymptotics, Plenum (1980)]; 2nd ed., Theory and Applications to Geophysical Hydrodynamics [in Russian], Leningrad (1982).
[12]K. Weirstrass, Abhandlungen aus der Funktionenlehre, Springer, Berlin (1886).
[13]G. H. Hardy, Trans. Am. Math. Soc. **17**, 301 (1916).
[14]A. Zygmund, Trigonometrical Series, Stechert, New York (1935) [Dover (1955)]; 2nd ed., Cambridge Univ. Press (1959) [Russ. Transl. **1**, Mir, M., 1965].
[15]K. Itô, Proc. Japan Acad. **22**, Nos. 1–4, 32–35 (1946).
[16]H. P. McKean, Stochastic Integrals, Academic Press (1969) [Russ. Transl., Mir, M., 1973].
[17]S. A. Molchanov, A. A. Ruzmaĭkin, and D. D. Sokolov, Usp. Fiz. Nauk **145**, 593 (1985) [Sov. Phys. Usp. **28**, 307 (1985)].
[18]J. Nash, Ann. Math. **60**, 383 (1954).
[19]I. Ya. Bakel'man, A. L. Verner, and B. E. Kantor, Vvedenie v differentsial'nuyn geometriyn "v tselom." (An Introduction to Differential Geometry as a Whole), Nauka, M., 1973.
[20]G. I. Barenblatt and A. S. Monin, Proc. Nat. Acad. Sci. USA **80**, 3540 (1983).
[21]J. A. Wheeler, "Neutrinos, gravitation, and geometry," Rend. Scuola Intl. Fis. Enrico Fermi **11** (July 1959), Bologna (1960); Geometrodynamics (Topics Mod. Phys. **1**), Academic Press (1962), pp. 1–130 [Russ. Transl. IIL, M., (1962)].
[22]Ya. B. Zel'dovich and I. D. Novikov, Relyativistskaya astrofizika, Nauka, M., 1967 [Engl. Transl., Relativistic Astrophysics, Univ. Chicago Press (1971); 2nd ed., **2**, The Structure and Evolution of the Universe, Nauka (1975) [Univ. Chicago Press (1983)].
[23]A. A. Starobinskiĭ, Phys. Lett. B **117**, 175 (1982).
[24]J. J. Halliwell and S. W. Hawking, Phys. Rev. D. **31**, 1777 (1985).
[25]S. F. Shandarin and Ya. B. Zel'dovich, Phys. Rev. Lett. **52**, 1488 (1984).
[26]Ya. B. Zel'dovich, D. D. Sokolov, and A. A. Starobinskiĭ, in: 150 let geometrii Lobachevskogo (150 Years of Lobachevskian Geometry) All-Union Inst. Sci. Tech. Inform [VINITI], M., (1977), p. 271.
[27]A. A. Starobinskiĭ, Pis'ma Astron. Zh. **9**, 579 (1983) [Sov. Astron. Lett. **9**, 302 (1983)].
[28]J. Nash, Ann. Math. **63**, 20 (1956).
[29]M. L. Gromov, Dokl. Akad. Nauk SSSR **192**, 1206 (1970) [Sov. Math. Dokl. **11**, 794 (1971)].
[30]D. W. Joseph, Phys. Rev. **126**, 319 (1962).
[31]E. Cremmer, B. Julia, and J. Scherk, Phys. Lett. B **76**, 409 (1978).
[32]M. I. Rabinovich, Usp. Fiz. Nauk **125**, 123 (1978) [Sov. Phys. Usp. **21**, 443 (1979)].
[33]U. Frisch, in: Chaotic Behavior of Deterministic Systems (Les Houches Sess. 36, July 1981), North-Holland (1983), p. 665.
[34]G. Parisi, in: Recent Advances in Field Theory and Statistical Mechanics (Les Houches Sess. 39, Aug. 1982), North-Holland (1984), p. 473.
[35]F. Anselmet, Y. Gagne, E. J. Hopfinger, and R. A. Antonia, J. Fluid Mech. **140**, 63 (1984).
[36]A. A. Fraenkel, Y. Bar-Hillel, and A. Levy, Foundations of Set Theory, 2nd ed., North-Holland, Amsterdam, 1973, [Russ. Transl. of earlier ed. Mir, M., 1966].

Translated by R. B. Rodman