ISSN 0002-9920

Notices

of the American Mathematical Society

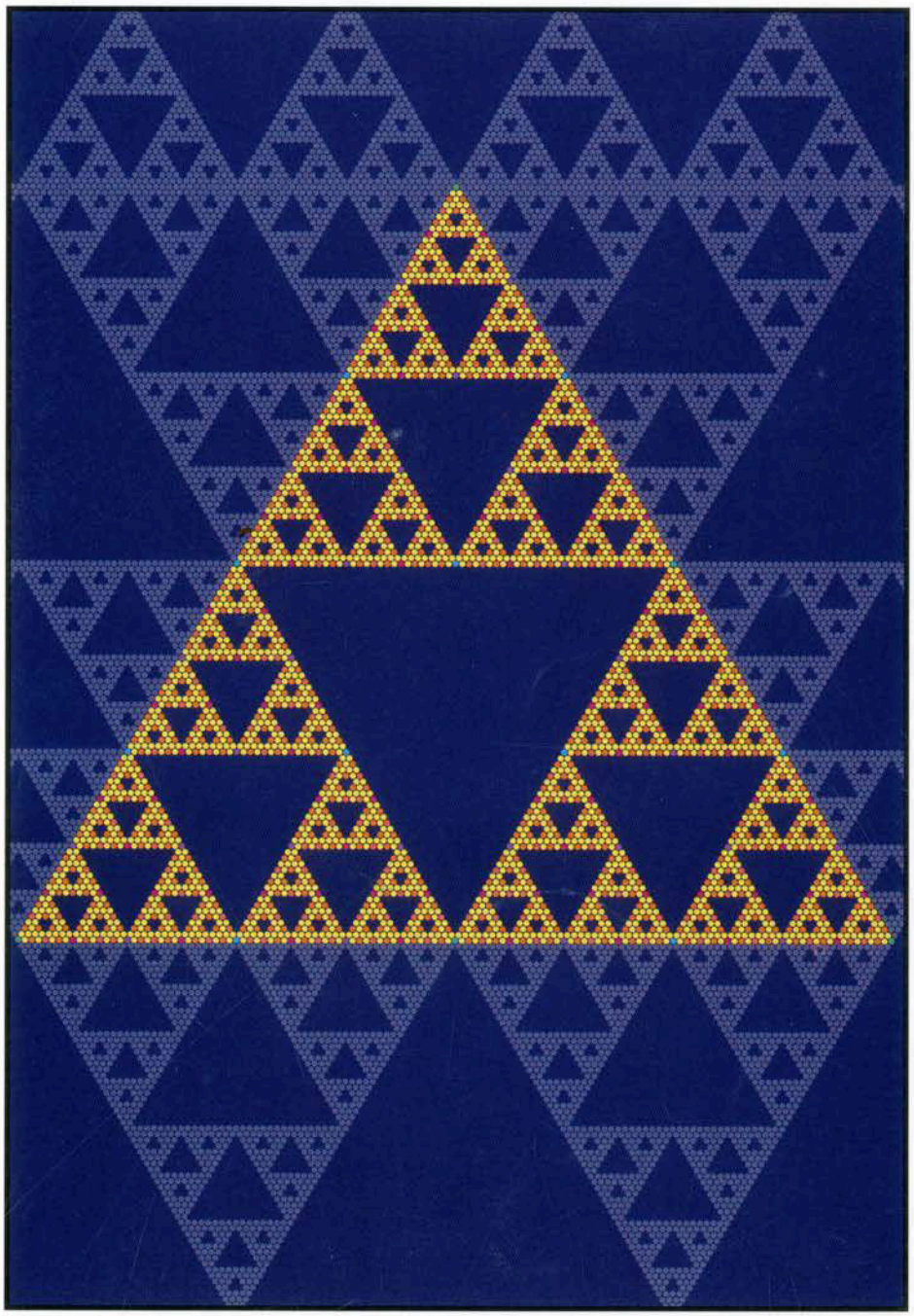
November 1999

Volume 46, Number 10

Analysis on Fractals
page 1199

Pseudorandomness
page 1209

Kerosinka: An Episode in the History of Soviet Mathematics
page 1217



The Sierpinski Gasket (see page 1208)

Analysis on Fractals

Robert S. Strichartz

From Manifolds to Fractals

Analysis on manifolds has been one of the central areas of mathematical research in the twentieth century. Rooted in the foundational work of the nineteenth century, with its rigorous theory of multidimensional calculus and the visionary ideas of Riemann, it has flowered into a richly layered mathematical tapestry. It has attracted mathematicians with diverse expertise and points of view, including topology, differential equations, differential geometry, functional and harmonic analysis, and probability theory. This heady mix of ideas has produced a vast body of work and a seemingly endless supply of challenging problems that should keep mathematicians busy well into the next century.

At the same time it has become apparent that many phenomena in the real world are best modeled by geometric structures that are much more irregular. The theory of fractals, as B. Mandelbrot [Ma] has so forcefully argued, seeks to provide the mathematical framework for such development. A theory of analysis on fractals is now emerging and is perhaps poised for the kind of explosive and multilayered expansion that has characterized analysis on manifolds. This article will explain some of what has been accomplished and where it might lead.

The central character in the theory of analysis on manifolds is the Laplacian. Thus the starting point for analysis on fractals will be the construction of an analogous operator on a class of fractals. This will not be a genuine differential operator, of course, but it will have quite a few of the features we have come to expect from anything labeled "Laplacian". It will be a local operator, and in fact $\Delta f(x)$ will be a limit in a suitable renormalized sense of the difference between an average value of f in a neighborhood of x and $f(x)$. We will be imitating the weak formulation of the Laplacian, so that $\Delta u = f$ will be interpreted to mean

$$\mathcal{E}(u, v) = -\int f v \, d\mu \tag{1}$$

for a suitable test class of functions v, vanishing on the boundary, where μ is a measure and $\mathcal{E}(u, v)$ is a bilinear form called a *Dirichlet form*. In the manifold case,

$$\begin{aligned} \mathcal{E}(u, v) &= \int \nabla u \cdot \nabla v \, d\mu = \\ &\int g^{jk}(x) \frac{\partial u}{\partial x^j} \frac{\partial v}{\partial x^k} \sqrt{g(x)} \, dx \end{aligned} \tag{2}$$

and $d\mu(x) = \sqrt{g(x)}\, dx$ in local coordinates, where $\{g_{jk}(x)\}$ is a given Riemannian metric, $g(x) = \det\{g_{jk}(x)\}$, and $\{g^{jk}(x)\}$ is the inverse of the matrix $\{g_{jk}(x)\}$. In the fractal case the Dirichlet form will come to play the leading role. There does not seem to be any canonical measure, and the measure on the right side of (1) may be different from the one on the right side of (2). There is certainly no analog of the Riemannian metric. It is interesting to ask in the manifold case whether or not the Dirichlet form determines the metric. The

Robert S. Strichartz is professor of mathematics at Cornell University. His e-mail address is `str@math.cornell.edu`.

Figures 3 and 5 were prepared using Mathematica. Figure 6 was produced by Kyallee Dalrymple.

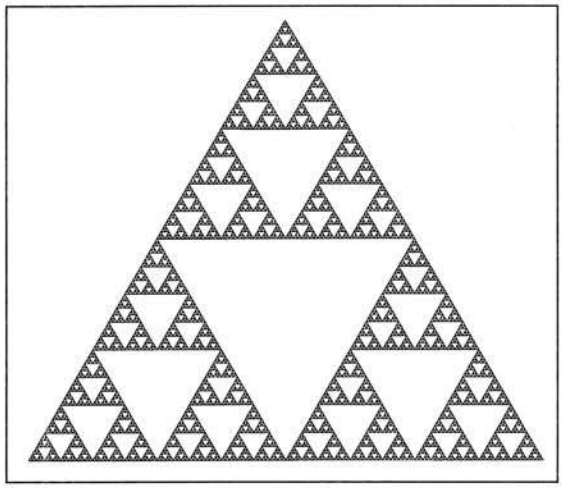


Figure 1. The Sierpinski gasket.

answer is yes when the dimension $n \neq 2$. When $n = 2$, conformal metrics yield the same Dirichlet form, and there is an obstruction in passing from a positive definite form in the gradient to a Riemannian metric, not present when $n \neq 2$.

This article will describe the approach introduced by J. Kigami, which is direct, constructive, and easy to explain. It is also possible to use probabilistic methods that will indirectly yield the same Laplacians. For this approach see M. Barlow [Ba], which is also a good source for references in this area. Later references to the probabilistic literature are found in [HK]. The forthcoming book [Ki3] will have an extensive bibliography. Other mathematical developments that might also be described as "analysis on fractals" are not described here; for example, the function spaces of A. Jonsson and H. Wallin [JW] and the work of J. Harrison [H] and U. Mosco [Mo].

Laplacian on the Sierpinski Gasket

In order to keep the discussion on a concrete level, I will concentrate on the construction of the fully symmetric Laplacian on the *Sierpinski gasket* SG (Figure 1), the familiar self-similar fractal generated by three contractions F_i in the plane with contraction ratio $1/2$ and fixed points at the vertices of an equilateral triangle. Its construction will be described in more detail below. This was the first example considered by J. Kigami [Ki1], who later extended the method to a class of fractals called "p.c.f." (post-critically finite) [Ki2]. The SG is typical for p.c.f. fractals, but this class is certainly very special, and other methods (nonconstructive) for producing Laplacians have been presented for other fractals.

Kigami's idea is to approximate the fractal from within by a sequence of finite graphs. The Laplacian on the fractal is then the renormalized limit of graph Laplacians. The same method works on the unit interval or the unit square, but we will see that there are features of the construction on SG that are more reminiscent of the interval than the square. In particular, points will have positive capacity.

In order to figure out the correct renormalization for the limit, we will construct the Dirichlet form first. The sequence of graphs $\{\Gamma_m\}$ is easy to describe inductively (see Figure 2). Γ_0 is just the complete graph on the three vertices V_0 of the triangle, and Γ_m with vertices V_m is obtained from Γ_{m-1} by applying the contractions F_i to the vertices V_{m-1}, with the edge relation $x \underset{m}{\sim} y$ holding if and only if x and y are vertices of the same cell $F_{i_1} \cdots F_{i_m}(SG)$ of order m. It is clear that the union of the vertices V_m is dense in SG, so a continuous function is determined by its restriction to vertex points. It turns out that we can work entirely within the class of continuous functions on SG.

The unit interval can also be obtained as a limit of graphs by taking V_m to be the dyadic points $\{j2^{-m} : 0 \leq j \leq 2^m\}$ with edges between consecutive points. The constructions that follow are closely modeled on this example.

On each graph Γ_m there is a naive Dirichlet form

$$E_m(f,g) = \sum_{x \underset{m}{\sim} y} (f(x) - f(y))(g(x) - g(y))$$

but these forms are not related to each other unless we multiply by the appropriate constants. That is, for

$$\mathcal{E}_m(f,g) = c_m \sum_{x \underset{m}{\sim} y} (f(x) - f(y))(g(x) - g(y)), \tag{3}$$

we would like the following consistency condition to hold: given a real-valued function f defined on V_{m-1}, we want

$$\mathcal{E}_{m-1}(f,f) = \mathcal{E}_m(\tilde{f},\tilde{f}), \tag{4}$$

where $\tilde{f}$ is the extension of f to V_m that minimizes $\mathcal{E}_m$. We call $\tilde{f}$ the *harmonic* extension. If we work this out for $m = 1$, we find that $c_1 = (5/3)c_0$, and the harmonic extension $\tilde{f}$ is given by the $(2/5, 1/5)$ law: the value of f at a vertex of $V_1 \setminus V_0$ is the weighted average of f at the vertices of V_0 with weights $2/5$ for the adjacent vertices and $1/5$ for the opposite vertex. In the general case the harmonic extension problem is seen to be local, so the same $(2/5, 1/5)$ law applies on each cell of order $m-1$, and the same relationship $c_m = (5/3)c_{m-1}$ holds for the renormalization constants. Thus we must have $c_m = (5/3)^m c_0$, and for simplicity we

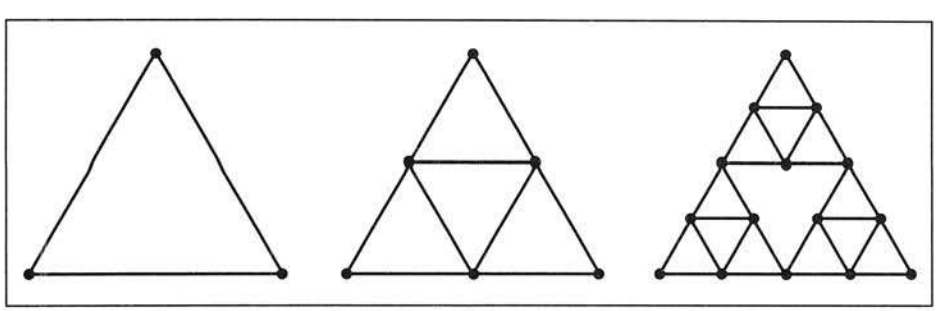


Figure 2. The first three graphs $\Gamma_0, \Gamma_1, \Gamma_2$ approximating the Sierpinski gasket.

take $c_0 = 1$. Note that the local harmonic extension law is the analog of the fact that a linear function on an interval takes on the average value of its endpoint values at the midpoint; no such result holds for harmonic functions on a square.

The consistency condition (4) means that for any continuous function f on SG, $\mathcal{E}_m(f,f)$ is a monotone increasing function of m, so

$$\mathcal{E}(f,f) = \lim_{m\to\infty} \mathcal{E}_m(f,f) \tag{5}$$

is always defined (in $[0,\infty]$). We let $\operatorname{dom}\mathcal{E}$ be the set of functions f for which $\mathcal{E}(f,f)$ is finite. The constants, and only the constants, have zero energy, and it is not hard to see that $\operatorname{dom}\mathcal{E}$ modulo constants is a Hilbert space with inner product $\mathcal{E}(f,g)$ defined by the same sort of limit as in (5). The energy form satisfies the self-similar identity

$$\mathcal{E}(f,g) = \sum_{i=1}^{3} (5/3)\mathcal{E}(f\circ F_i, g\circ F_i) \tag{6}$$

and is symmetric under the 6-element symmetry group of the equilateral triangle.

The three vertices in V_0 are, by definition, the *boundary* of SG. Note that every nonboundary vertex in V_m has exactly four neighbors in V_m. We define a *harmonic function* on Γ_m to be one that assumes the value at a nonboundary vertex x equal to the average of the values at the neighboring vertices. A *harmonic function* on SG is just a continuous function whose restrictions to Γ_m are all harmonic. It is uniquely determined by its values on the boundary, and its values on V_m are obtained from its values on V_{m-1} by the harmonic extension $(2/5, 1/5)$ law. Thus the space of harmonic functions is 3-dimensional. Figure 3 shows the graph of a harmonic function.

To define a Laplacian from the Dirichlet form via the weak formulation (1) requires that we choose a measure μ for the right side of (1). There is certainly a natural measure on SG, namely the self-similar probability measure satisfying

$$\mu = \sum_{i=1}^{3} \frac{1}{3}\mu\circ F_i^{-1}, \tag{7}$$

or equivalently

$$\int_{\mathrm{SG}} f\,d\mu = \sum_{i=1}^{3} \frac{1}{3}\int_{\mathrm{SG}} f\circ F_i\,d\mu. \tag{8}$$

In fact, such an identity determines μ, and μ is the normalized Hausdorff measure of dimension $\log 3/\log 2$ restricted to SG. The measure of each of the 3^m cells of order m is just 3^{-m}. The identity (8) makes it possible to evaluate many integrals exactly (for example, inner products of harmonic functions). With this choice of μ we define $u \in \operatorname{dom}\Delta$ and $\Delta u = f$ if $u \in \operatorname{dom}\mathcal{E}$, f is continuous, and (1) holds for all $v \in \operatorname{dom}\mathcal{E}$ with v vanishing on the boundary V_0. This is in fact a useful

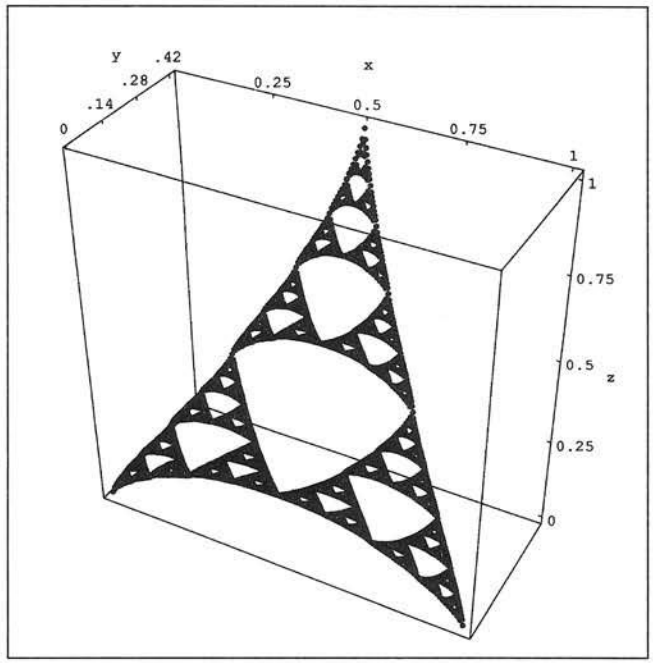


Figure 3. The graph of a harmonic function on the Sierpinski gasket, with boundary values $0, 0, 1$ at the vertices $(0,0), (1,0), (1/2, \sqrt{3}/4)$. The function was actually plotted just at the points of V_7.

definition for proving theorems, but a more explicit definition is the pointwise formula

$$\Delta u(x) = \lim_{m\to\infty} (3/2)5^m \sum_{y\underset{m}{\sim}x} (u(y) - u(x)) \tag{9}$$

for any nonboundary vertex point x. The renormalization factor 5^m in (9) is just the product of $(5/3)^m$ in (3) divided by the factor $(1/3)^m$ from the measure μ. The relatively unimportant constant $3/2$ arises because vertices in V_m do not correspond exactly to sets of measure $(1/3)^m$. The exact theorem [Ki1] is that $u \in \operatorname{dom}\Delta$ if and only if the limit in (9) exists uniformly. Note that (9) is, as promised, a renormalized limit of graph Laplacians, but the renormalization constant 5^m cannot be explained by any superficial dimension arguments. Of course (9) exhibits Δu as a limit of difference quotients and shows the local nature of this Laplacian: $\Delta u(x)$ depends only on the values of u in any neighborhood of x. But (9) is valid only for vertex points, and although these points are dense in SG, they are a set of measure zero for μ and are far from being typical points. From (9) it is not hard to show that a function h is harmonic if and only if $\Delta h = 0$.

There is a version of the Gauss-Green formula valid for this Laplacian. Not only is this an interesting result in itself, but also it is an important technical tool. To state the result we need to define normal derivatives at the boundary points $x \in V_0$:

$$\partial_n f(x) = \lim_{m\to\infty} (5/3)^m \sum_{y\underset{m}{\sim}x} (f(y) - f(x)). \tag{10}$$

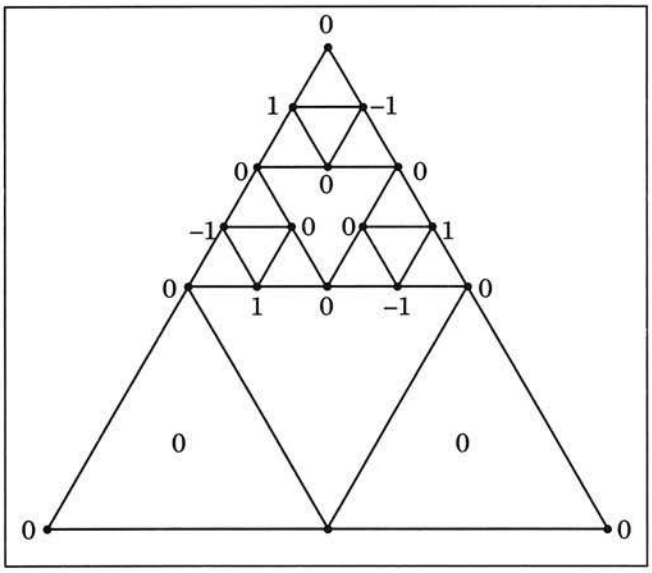


Figure 4. The values of an eigenfunction on Γ_3 with eigenvalue $\lambda_3 = 5$, identically zero on two of the three cells of order 1.

Note that there are exactly two neighboring points when $x \in V_0$. The limit exists when f is in $\operatorname{dom}\Delta$.

Theorem 1. (Gauss-Green Formula). *If u and v are in* $\operatorname{dom}\Delta$,

$$\int_{SG}(u\Delta v - v\Delta u)\,d\mu = \sum_{x\in V_0}\left(u(x)\partial_n v(x) - v(x)\partial_n u(x)\right). \tag{11}$$

There is also a Green's function $G(x, y)$ for solving

$$\Delta u = f, \quad u\big|_{V_0} = 0 \tag{12}$$

uniquely via

$$u(x) = \int_{SG} G(x, y)f(y)\,d\mu(y). \tag{13}$$

The Green's function is continuous, symptomatic of the fact that points have positive capacity. There is an explicit formula for G that we omit here.

Eigenfunctions and Spectral Decimation

With the definition of the Laplacian in place, it is possible to consider analogs of the classical equations involving the Laplacian. We have already mentioned harmonic functions, for which there exists a simple and effective local extension algorithm. The harmonic functions are the analogs of linear functions on an interval. Similarly, the eigenfunctions of the Laplacian

$$-\Delta f = \lambda f \tag{14}$$

are the analogs of sines, cosines, and exponentials. By imposing either Dirichlet ($f|_{V_0} = 0$) or Neumann ($\partial_n f|_{V_0} = 0$) boundary conditions on solutions of (14), we obtain a discrete family of *Dirichlet* (or Neumann) *eigenfunctions*, with eigenvalues forming a discrete *Dirichlet* (or Neumann) *spectrum*. Moreover, arbitrary functions may be expanded in infinite series of either type of eigenfunctions, giving the analog of Fourier sine and cosine series. The spectrum was first studied by the physicists R. Rammal and G. Toulouse in 1983 and 1984. In 1992 M. Fukushima and T. Shima gave a mathematical description of eigenvalues and eigenfunctions. Many illustrations can be found in [DSV].

The eigenfunctions may in fact be computed via an effective local extension algorithm. To see what this should be like, we observe that the function $\sin \pi kx$ on the unit interval is an eigenfunction not only of the differential operator d^2/dx^2 but also of the symmetric second difference operator

$$\Delta_h^2 f(x) = h^{-2}(f(x+h) + f(x-h) - 2f(x)), \tag{15}$$

with eigenvalue $2h^{-2}(\cos \pi kh - 1)$ that tends to $-(\pi k)^2$ as $h \to 0$. Thus, if we let V_m denote the dyadic points $j2^{-m}$ in the unit interval, the eigenfunctions of Δ_h^2 with $h = 2^{-m}$ on V_m are restrictions to V_m of eigenfunctions of d^2/dx^2, and we obtain the whole Dirichlet spectrum (with the correct eigenvalues) in the limit as $m \to \infty$. Moreover, there is a bifurcation of eigenfunctions as we extend from V_m to V_{m+1}. If we start with $f(x) = \sin \pi kx$ on V_m with $1 \le k < 2^m - 1$, we can extend f to V_{m+1} as either $\sin \pi kx$ or $\sin \pi(k + 2^m)x$, giving different eigenfunctions with different eigenvalues. Here we can begin the whole process with the single Dirichlet eigenfunction $\sin \pi x$ on V_1.

The story is quite similar on SG. Aside from a few complications, every eigenfunction

$$-\Delta_m f_m = \lambda_m f_m \tag{16}$$

of the graph Laplacian

$$\Delta_m f_m(x) = \sum_{y \underset{m}{\sim} x}(f(y) - f(x)) \tag{17}$$

on Γ_m can be extended by an explicit local algorithm in two distinct ways to an eigenfunction f_{m+1} on Γ_{m+1} with eigenvalue λ_{m+1}, where the eigenvalues are related by the quadratic equation

$$\lambda_m = \lambda_{m+1}(5 - \lambda_{m+1}). \tag{18}$$

Furthermore, eigenfunctions of Δ on SG arise by taking limits as $m \to \infty$. What is new is that the process needs to be started at different levels. Eigenvalues can have high multiplicity (much higher than is suggested by the 6-element symmetry group), and the eigenfunctions can be completely localized. For example, Figure 4 shows the values of an eigenfunction on Γ_3 with eigenvalue $\lambda_3 = 5$. By extending this function and passing to the limit, we obtain a Dirichlet eigenfunction on SG that vanishes identically on two of the three cells of order 1. This means that there are solutions to the heat equation or wave equation on SG for which the heat

(or vibration) never escapes from a small cell. This is perhaps not so surprising in view of the topological structure of SG, where single junction points (the vertices in V_m) control the connection between neighboring cells, and any signal with an odd symmetry in a neighborhood of such a junction point will not get past. Another peculiar feature of the eigenfunctions on SG is that they may assume a constant value along a line segment. For example, the ground state Dirichlet eigenfunction (Figure 5) assumes its maximum value along the whole inverted triangle connecting the nonboundary vertices in V_1.

As a consequence of this description of eigenvalues called the *spectral decimation method*, the spectrum (Dirichlet or Neumann) satisfies a Weyl asymptotic law with a dimension $d_S = 2\log 3/\log 5$ known as the *spectral dimension*. The example of SG is a bit special, since the spectral decimation method works only for a very limited class of fractals. J. Kigami and M. Lapidus established the Weyl asymptotic law in full generality using different methods in 1993.

Heat and Wave Equation

Using the Laplacian Δ on SG for the space part, we can consider space-time heat and wave equations $u_t = \Delta u$ and $u_{tt} = \Delta u$ for $u(x,t)$ a function on $\mathrm{SG} \times [0,\infty)$. The heat equation may also be considered as the diffusion equation for a stochastic process that may be described as Brownian motion on SG, with Δ as its infinitesimal generator. It is possible to give an independent description of this Brownian motion, and in fact this predates the explicit constuction of the Laplacian we have described. The heat kernel then describes the transition probabilities for the process.

Estimates for the heat kernel of the expected Gaussian type have been obtained by B. Hambly and T. Kumagai [HK], and there are some interesting features. The dimension that appears in the on-diagonal estimates is the spectral dimension d_S. The distance that appears in the off-diagonal estimates is not the Euclidean distance. This should not be too surprising, since the geometry that comes from the embedding of SG in the plane plays no role in the construction of the Laplacian. The distance that is relevant is the *intrinsic resistance metric*. If we regard the graphs Γ_m as electric circuits, with the edges consisting of resistors whose resistance is the reciprocal of the conductance constant c_m, then the effective resistance between vertices is independent of m and by continuity yields a metric on SG. Another way to describe this metric $d_R(x,y)$ is as the infimum of the energy of a function f that satisfies $f(x) = 0$ and $f(y) = 1$. The intrinsic resistance metric is not exactly self-similar, but asymptotically it scales by a ratio of $3/5$ with each contraction. Thus cells of order m have diameter approximately $(3/5)^m$.

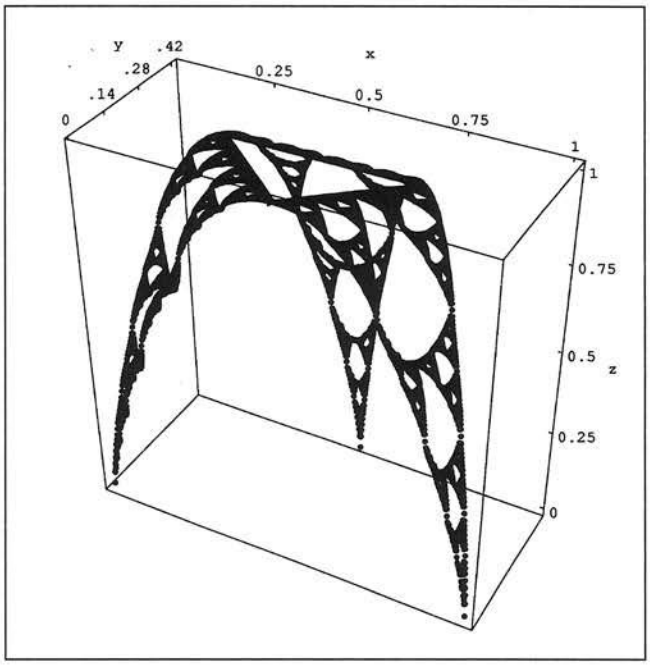


Figure 5. The graph of the ground-state Dirichlet eigenfunction on the Sierpinski gasket. This function is strictly positive on the interior and achieves its maximum value of 1 everywhere on the principal interior triangle. The function has $\lambda_1 = 2$. For $m \geq 1$, λ_{m+1} is $(1/2)(5 - \sqrt{25 - 4\lambda_m})$, the negative square root being chosen at each stage.

An interesting connection between the intrinsic resistance metric and the spectral dimension is the formula

$$d_S = 2d_H/(d_H + 1),$$

where d_H denotes the Hausdorff dimension in this metric.

The wave equation presents another surprise: there is no finite propagation speed. This can be explained in terms of a mismatch between the scaling properties of the second time derivative (factor of 4) and the Laplacian (factor of 5). If we were to construct a "Sierpinski harp" by wiring strings along the edges of Γ_m and coupling the strings appropriately at the vertices in V_m, we would need to increase the tension on the strings as m increases in order to obtain in the limit a model for wave propagation on SG. As every musician knows, increasing tension increases pitch, but the reason behind the increase in pitch is that the speed of propagation increases. Thus, for large m a vibration can travel along the edge of the harp at high speed. However, most of the energy of the vibration will not discover this potential superhighway but instead will get snarled in the local traffic of the convoluted connections in the graph Γ_m. In the limit, small amounts of energy can travel at arbitrarily large speed. From a scaling perspective, vibrations appear to travel faster on a smaller scale.

The unbounded propagation speed seems to defy basic physical principles, but the resolution of this problem is simply that true fractals do not exist in nature. Once one gets to the molecular level, a different model is needed. Nevertheless, objects that have a fractal structure at several scales may often still be profitably modeled by fractals.

Numerical Analysis

Where exact solutions to fractal differential equations are unavailable, a number of techniques for finding approximate solutions are known. One rather obvious method, the analog of the finite-difference method, is to use $5^m\Delta_m$ on Γ_m as an approximation to Δ on SG. This method was used for the space components in the heat and wave equations in [DSV]. It is also possible to develop an analog of the finite-element method by constructing spline spaces [SU]. The analog of the space of polynomials of degree at most $2j+1$ on an interval is the space $\mathcal{H}_j$ of multiharmonic functions satisfying $\Delta^{j+1}u = 0$. By using the Green's function representation (13) it is possible to give effective local algorithms for computing multiharmonic functions, starting from the boundary data

$$(19) \qquad \begin{cases} \Delta^k f\big|_{V_0} & k \le j/2, \\ \partial_n \Delta^k f\big|_{V_0} & k < j/2. \end{cases}$$

We then define spline spaces $S_m\mathcal{H}_j$ of functions that are piecewise in $\mathcal{H}_j$ on each of the 3^m cells of order m and that satisfy matching conditions at the nonboundary vertices in V_m corresponding to the data (19). The spline space $S_m\mathcal{H}_0$ is just the space of continuous piecewise harmonic functions at level m. By taking higher values of j we allow the splines to be "smoother".

To find the finite-element approximation to the solution of (12), for example, we choose values for j and m and take the subspace of $S_m\mathcal{H}_j$ satisfying the boundary condition $u|_{V_0} = 0$. The approximate solution is the function in this space satisfying the integrated equation (1) for all v in this space. By choosing a natural basis for this space we obtain a sparse system of linear equations. The values of $\mathcal{E}(u, v)$ can be computed theoretically, but the right side of (1) requires numerical integration. The spline spaces may be used also to develop efficient numerical integration methods analogous to Simpson's method. They are also useful for cutting and pasting operations on functions. A version of the finite-element method implemented by M. Gibbons and A. Raj may be found at `http://mathlab.cit.cornell.edu/~gibbons/`.

Taylor Approximations

It would be difficult to convince a calculus student that the second derivative is the more basic concept and the first derivative is a subordinate notion. Yet that is the situation we are in at this stage of the development of calculus on SG. In fact, there is no completely satisfactory analog of the gradient, although I will describe two distinct approaches to the problem. On an interval the tangent line is defined by the local approximation of a differentiable function by linear functions. This is the first of the sequence of Taylor polynomial approximations, involving higher derivatives, for functions with greater smoothness. The first derivative and also the higher derivatives appear as coefficients of the Taylor polynomials. On SG the analogs of linear functions are harmonic functions, and the analogs of polynomials are the multiharmonic functions. These are the functions that should serve as local approximations to a general "differentiable" function, and the coefficients identifying the approximation should serve as components of various derivatives of the function. With some luck, one may also compute the derivatives as limits of difference quotients.

The situation turns out to be more complicated for three reasons. First, the results are different at the special vertex points in V_m and at generic points, where the theory is somewhat incomplete. Second, at a vertex point the approximation rate that characterizes the Taylor approximation must be described by two different estimates, a faster rate for the odd part and a slower rate for the even part (hence overall). Third, the region where the approximation takes place is limited by the geometry of the point regardless of the function.

To describe the situation in more detail, we begin by defining a tangential derivative $\partial_T f(x)$ at boundary points $x \in V_0$ to go along with the normal derivative defined by (10):

$$(20) \qquad \partial_T f(x) = \lim_{m\to\infty} 5^m(f(a_m) - f(b_m)),$$

where a_m and b_m are the two neighbors of x in V_m. The limit exists if f is in $\operatorname{dom}\Delta$. Next we localize both derivatives to the cells of order m whose boundary points are the vertices in V_m. Each nonboundary vertex x_0 is a boundary point for two such cells, and so there are four derivatives defined at x_0. However, under reasonable assumptions, such as $f \in \operatorname{dom}\Delta$, the two normal derivatives are related: in fact, they sum to 0. The two tangential derivatives are independent however. So we can define a gradient of f at x_0 consisting of three component derivatives. This leads to an embarrassing dimensional miscount, since we have four numbers (the value of the function at x_0 plus the three derivatives) to match a harmonic function, and the space of harmonic functions is only 3-dimensional. The resolution of this paradox is that we should try to match only a local harmonic function, not a global harmonic function. The point x_0 has a natural system of neighborhoods $U_m(x_0)$ consisting of pairs of adjacent cells of order m meeting at x_0.

Here we must require that m be sufficiently large so that $x_0 \in V_m$. Each of these neighborhoods has exactly four boundary points, and there is a 4-dimensional space of local harmonic functions on $U_m(x_0)$. The first Taylor approximation to f at x_0 is defined to be the local harmonic function h that matches the values of f and its three derivatives at x_0. Note that the domain of definition of h is limited by the geometry of x_0 alone.

Next we describe the local approximation properties of h. We define a reflection symmetry R in $U_m(x_0)$ that fixes x_0 and reflects each cell to itself through the angle bisector of x_0. (Here R does not permute the two cells.) The overall estimate takes the form

$$(21) \qquad |f(x) - h(x)| = o((3/5)^m) \quad \text{for } x \in U_m(x_0),$$

while the estimate for the odd part is

$$(22) \qquad |(f(x) - f(Rx)) - (h(x) - h(Rx))| = o((1/5)^m)$$

for $x \in U_m(x_0)$. It is not hard to show that at most one local harmonic function can satisfy (21) and (22) and that if such a local harmonic function exists, then its derivatives at x_0 must match the corresponding derivatives of f at x_0. In addition, $f(x_0) = h(x_0)$. The following existence theorem is proved in [S]:

Theorem 2. *Suppose f is in* $\mathrm{dom}(\Delta)$ *and Δf satisfies a Hölder condition of any positive order. Then for each vertex point x_0 there exists a local harmonic function h satisfying (21) and (22).*

There are analogous statements involving higher-order Taylor approximations by local multiharmonic functions with better estimates, under assumptions that f belongs to the domain of a power of Δ. The occurrence of a power of $3/5$ in (21) and (10) is a consequence of the fact that $c_m = (5/3)^m$ in (3). However, the power of 5 that appears in (22) and (20) is just coincidentally the same as the power of 5 that appears in (9). This coincidence is related to the additional hypothesis of Hölder continuity required in Theorem 2.

The story for local approximation at a generic point x_0 is quite different. If we specifically assume that x_0 does not belong to any V_m, then x_0 belongs to a unique cell $U'_m(x_0)$ of order m, and this gives a natural system of neighborhoods of x_0. Each cell has three boundary points, and local harmonic functions on the cell are determined by the values at these boundary points. In fact, each local harmonic function is the restriction of a unique global harmonic function, and the extension and restriction are easily described in terms of three matrices M_i and their inverses M_i^{-1}. The restriction from SG to $F_i(SG)$ of a harmonic h is given by

$$(23) \qquad h \circ F_i \Big|_{V_0} = M_i h \Big|_{V_0},$$

where

$$(24) \qquad M_1 = \begin{pmatrix} 1 & 0 & 0 \\ 2/5 & 2/5 & 1/5 \\ 2/5 & 1/5 & 2/5 \end{pmatrix}$$

and M_2 and M_3 are obtained from M_1 by cyclic permutations of indices. The restriction to the cell $F_{i_1} \cdots F_{i_m}(SG)$ is then given by

$$(25) \qquad h \circ (F_{i_1} \cdots F_{i_m}) \Big|_{V_0} = M_{i_m} \cdots M_{i_1} h \Big|_{V_0},$$

and the extension is the inverse relation

$$(26) \qquad h \Big|_{V_0} = M_{i_1}^{-1} \cdots M_{i_m}^{-1} h \circ (F_{i_1} \cdots F_{i_m}) \Big|_{V_0}.$$

Now if a nonvertex point x_0 is given, the neighborhood system $U'_m(x_0)$ corresponds to a unique sequence $\{i_j\}$. We let h_m denote the harmonic function that matches f at the three boundary points of $U'_m(x_0)$. By (26) this means

$$(27) \qquad h_m \Big|_{V_0} = M_{i_1}^{-1} \cdots M_{i_m}^{-1} f \circ (F_{i_1} \cdots F_{i_m}) \Big|_{V_0}.$$

If the limit exists as $m \to \infty$, we call the harmonic function $h = \lim_{m \to \infty} h_m$ the first-order Taylor approximation to f at x_0. Using H. Furstenberg's 1963 theory of products of random matrices, [S] shows that for μ-almost every point x_0, the first-order Taylor approximations at x_0 exist for every $f \in \mathrm{dom}\,\Delta$, and the estimate

$$(28) \qquad |f(x) - h(x)| = O(\beta^m) \quad \text{for } x \in U'_m(x_0)$$

holds for $\beta > \beta_0$. Moreover, the estimate (28) uniquely characterizes the harmonic function h. The value of the constant β_0 in (28) can only be estimated.

Energy Measures

Since the Dirichlet form $\mathcal{E}(u, v)$ is the analog of $\int \nabla u \cdot \nabla v \, d\mu$, it is tempting to look for the analogs of both $\nabla u \cdot \nabla v$ and dx within it. In fact, there is a standard procedure for associating a measure $\nu_{u,v}$ to every $u, v \in \mathrm{dom}\,\mathcal{E}$ (positive when $u = v$) such that

$$(29) \qquad \mathcal{E}(u, v) = \int d\nu_{u,v}.$$

In this case the simplest way to describe $\nu_{u,v}$ is to take for $\nu_{u,v}(A)$, when A is any simple set, the same limit that defines $\mathcal{E}(u, v)$, but restricting the sums in $\mathcal{E}_m(u, v)$ to points in A. If we can find a positive measure ν with the property that $\nu_{u,v}$ is absolutely continuous with respect to ν for all $u, v \in \mathrm{dom}\,\mathcal{E}$, then we can write

$$(30) \qquad \nu_{u,v} = \Gamma(u, v)\, d\nu$$

for the appropriate Radon-Nikodym derivative $\Gamma(u, v)$. Here Γ is called the *carré du champs* operator. S. Kusuoka in 1989 showed that (30) holds

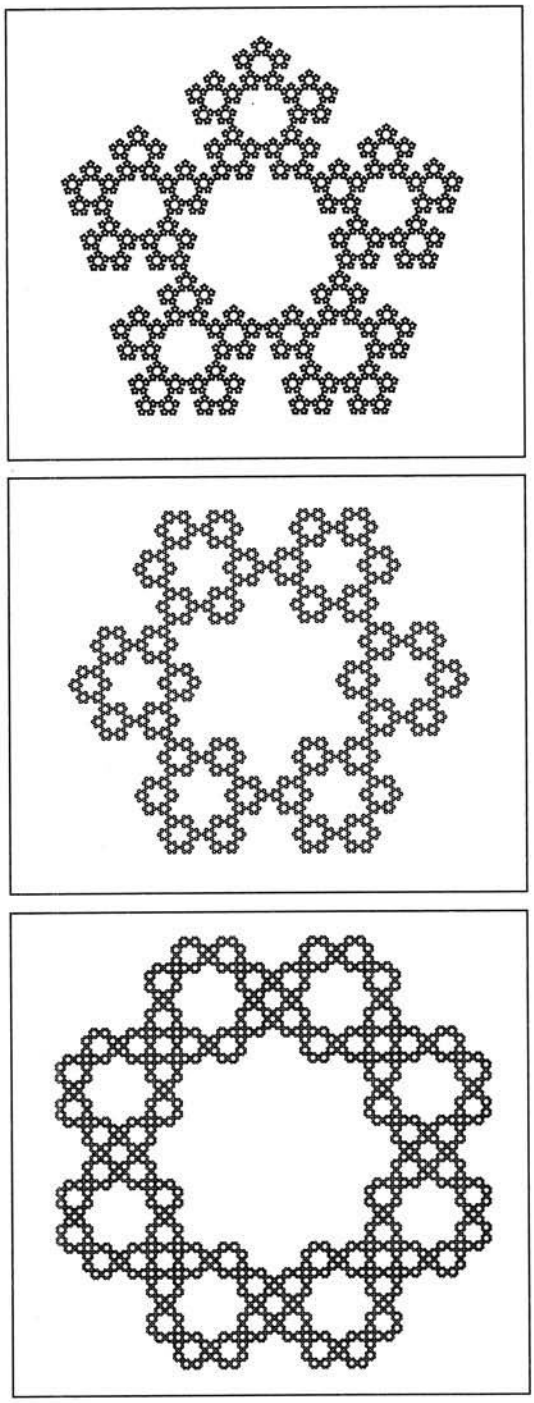


Figure 6. The pentagasket (top), hexagasket (middle), and octagasket (bottom).

with the choice of $\nu = \nu_{h_1,h_1} + \nu_{h_2,h_2}$, where $\{h_1, h_2\}$ is an orthonormal basis for the harmonic functions mod constants in the energy inner product. We refer to $\nu_{f,g}$ as *energy measures* and ν as the *Kusuoka measure.* A perhaps surprising observation is that the Kusuoka measure and hence all the energy measures are singular with respect to the self-similar measure μ defined by (7) and (8) that we used in the construction of the Laplacian! We can think of $\Gamma(u, v)$ as the analog of $\nabla u \cdot \nabla v$, but we must keep in mind that $\Gamma(u, v)$ is defined only almost everywhere with respect to ν. It is not clear whether there is any meaningful way to define ∇u so that $\Gamma(u, v) = \nabla u \cdot \nabla v$.

The singularity of the Kusuoka measure has another disquieting consequence: multiplication is forbidden in dom(Δ). Specifically, if $u \in \mathrm{dom}(\Delta)$ is not constant, then u^2 is not in dom(Δ). This can be explained by the putative identity

$$\Delta u^2 = 2u\Delta u + \nabla u \cdot \nabla u, \tag{31}$$

which can be interpreted correctly only in terms of measures [BST]. A different explanation can be based on the incommensurability of the different approximation rates in the local Taylor approximation. If f is in dom(Δ) and x_0 is a vertex point where $\partial_n f(x_0) \neq 0$, then the rate of convergence of $f(x)$ to $f(x_0)$ is bounded above and below by a multiple of $(3/5)^m$ on $U_m(x_0) \setminus U_{m+1}(x_0)$. Then $(f(x) - f(x_0))^2$ converges to 0 too rapidly to have nonzero normal derivative. But it also converges to 0 too slowly to have normal derivative equal to 0, for that implies a rate of at least $m(1/5)^m$.

The impossibility of multiplication is a serious obstacle to the interpretation of the domains of Δ and powers of Δ as spaces of smooth functions. Perhaps that is just the nature of things. Another possible response is to study a different Laplacian, constructed by taking the Kusuoka measure on the right side of (1). This eliminates the problem of multiplication and enables us to make perfect sense out of (31). But it has the disadvantage that ν is not self-similar, so computations with this Laplacian will not be independent of scale. Certainly when dealing with physical models, one will not have the luxury of choosing a measure at will if the measure is to have the interpretation of mass distribution.

Here is an entertaining diversion concerning the Kusuoka measure. Suppose we carry out the same procedure for the standard Dirichlet form $\int |\nabla u|^2\, dx$ on the unit disc. In this case there is an infinite orthonormal basis $\{h_i\}$ of harmonic functions modulo constants in this inner product, and we would take

$$\nu = \sum_{i=1}^{\infty} |\nabla h_i(x)|^2\, dx, \tag{32}$$

the sum being independent of the choice of orthonormal basis. The computation of ν is straightforward but lengthy, and the result is a multiple of the Riemannian measure associated to the hyperbolic metric on the disc. This should come as no surprise, since we have already observed that the Dirichlet form in two dimensions is a conformal invariant, so we might as well start out by working in the hyperbolic metric. Then the Möbius transformations of the disc are isometries, and ν must be Möbius invariant (because composing the orthonormal basis with a Möbius transform produces another orthonormal basis). Up to a constant multiple, there is a unique σ-finite Möbius invariant measure.

P.C.F. Self-Similar Fractals

Kigami [Ki2] has described a class of fractals called *post-critically finite* (p.c.f.), for which a similar theory of Dirichlet forms and Laplacians may be constructed, provided a certain algebraic problem can be solved. In the interest of simplicity I will describe a more limited class of fractals that seems to contain all the interesting examples. The key property of SG that we want to maintain is that it is connected, but just barely: the removal of a finite number of points makes it disconnected, so these junction points control all access from one point of the set to another. These fractals are often referred to as *finitely ramified.*

We will work within the class of self-similar sets in $\mathbb{R}^n$. A self-similar set is defined to be the unique nonempty compact set K satisfying

$$K = \bigcup_{i=1}^{N} F_i K \tag{33}$$

for a family $\{F_i\}$ of contractive similarities. The *critical set* C is defined to be the set of all intersection points $F_i K \cap F_j K$ for $i \neq j$, and the *post-critical set* P is defined to be the pre-images of C under the mappings F_i and their iterates. The p.c.f. assumption we make is that P is a finite set, and by definition it is the *boundary* of K.

To get a feeling for what is and is not in a p.c.f. fractal, consider the class of polygaskets, which are constructed from a regular N-gon in the same way that SG is constructed from a triangle. We adjust the contraction ratio for F_i (with fixed points the vertices of the N-gon) so that the images just touch. When N is not divisible by 4, the image N-gons touch at single vertices, and we obtain a p.c.f. fractal with P equal to the N vertices of the original N-gon. But when 4 divides N, the intersections are infinite (for $N = 4$ we obtain a square). Figure 6 shows the pentagasket and hexagasket, which are p.c.f., and the octagasket, which is not.

For a p.c.f. fractal K we define cells of order m to be images $F_{i_1} \cdots F_{i_m} K$, and we define graphs Γ_m by taking $V_0 = P$,

$$V_m = \bigcup_{i=1}^{N} F_i V_{m-1},$$

and the edge relation $x \underset{m}{\sim} y$ if x and y belong to the same cell. The intersections of distinct cells consist of vertices in V_m, but not all nonboundary vertices are such junction points (see the pentagasket and hexagasket, for example). We want to construct a Dirichlet form $\mathcal{E}$ on K that is again the limit of Dirichlet forms $\mathcal{E}_m$ on Γ_m, but we can no longer rely on the simple formula (3). For $\mathcal{E}_m$, the expression

$$\mathcal{E}_m(f,f) = \sum_{x \underset{m}{\sim} y} c(x,y)(f(x) - f(y))^2 \tag{34}$$

with rather arbitrary positive coefficients $c(x,y)$ would be allowable, but this is too general. What we want is a self-similarity condition on the sequence $\{\mathcal{E}_m\}$,

$$\mathcal{E}_m(f,f) = \sum_{i=1}^{N} r_i^{-1} \mathcal{E}_{m-1}(f \circ F_i, f \circ F_i) \tag{35}$$

for certain positive coefficients $\{r_i\}$, that will translate into the self-similarity identity

$$\mathcal{E}(f,f) = \sum_{i=1}^{N} r_i^{-1} \mathcal{E}(f \circ F_i, f \circ F_i) \tag{36}$$

in the limit. We will continue to require the consistency condition (4) as well. The following theorem is proved in [Ki2].

Theorem 3. *Suppose there exist $\mathcal{E}_0$ on Γ_0 of the form (34) and coefficients $\{r_i\}$ such that if we use (35) to define $\mathcal{E}_1$, then (4) holds for $m = 1$. Then if we use (35) to define $\mathcal{E}_m$ inductively, the consistency condition (4) holds for all m. In the case that $r_i < 1$ for all i, the limit $\mathcal{E}$ defines a local Dirichlet form whose domain is contained in the continuous functions on K, and (36) holds.*

In other words, the whole construction succeeds provided it succeeds at the first step. Of course, the algebraic problem of finding the coefficients for $\mathcal{E}_0$ and the coefficients $\{r_i\}$ is nontrivial, and there is still no general existence theorem, although C. Sabot in 1997 and T. Lindstrøm in 1989 have resolved the problem in some cases. Generally speaking, one expects that there is a continuum of solutions.

For the construction of a Laplacian from the Dirichlet form via (1), it is not necessary to choose a self-similar measure for μ. It is enough to have a finite measure that gives positive values to all nonempty open sets. Many of the results discussed above for SG extend to p.c.f. fractals, with appropriate modifications and hypotheses. One new feature that does not show up in the SG example is that the matrices M_i that occur in the analog of (23) for restricting harmonic functions are not always invertible, so that the extension of harmonic functions given by (27) is not always possible. In particular, harmonic functions may be locally constant but not globally constant.

Challenges for the Future

This article has described some of the developments that have taken us quite far for a relatively narrow class of fractals, and further progress can be expected. There are some hints that these fractals have something in common with manifolds of positive curvature, although there is no obvious candidate for curvature in this context. A Liouville theorem holds for certain noncompact "blow-ups" of these fractals, and in the manifold case this requires a nonnegative curvature assumption.

However, an important challenge for the future is to extend the theory beyond the finitely ramified context. There are Brownian-motion-type processes on other fractals, notably some "Sierpinski carpets", and the infinitesimal generators give, indirectly, Laplacians. It is not clear what the natural class of fractals is for which this approach will succeed, and it is also not clear how much further information can be obtained in this mainly nonconstructive setting.

It should also be possible to go beyond the self-similar context. Of course, it is easier to work with structures obtained by iterating the same

construction, but it is the hierarchical structure on different scales that seems to be essential to the current theory. Perhaps what is needed is a concept of *fractafold*, the fractal analog of the concept of manifold.

References

[Ba] M. T. Barlow, Diffusions on fractals, *Lectures in Probability Theory and Statistics, École d'Été de Probabilités de St. Flour XXV-1996* (M. T. Barlow and D. Nualart, eds.), Lecture Notes in Mathematics, vol. 1690, Springer, Berlin, 1998.

[BST] O. Ben-Bassat, R. Strichartz, and A. Teplyaev, What is not in the domain of the Laplacian on a Sierpinski gasket type fractal, *J. Funct. Anal.*, to appear.

[DSV] K. Dalrymple, R. Strichartz, and J. Vinson, Fractal differential equations on the Sierpinski gasket, *J. Fourier Anal. Appl.* **5** (1999), 203–284.

[HK] B. M. Hambly and T. Kumagai, Transition density estimates for diffusion processes on post critically finite self-similar fractals, *Proc. London Math. Soc. (3)* **78** (1999), 431–458.

[H] J. Harrison, Continuity of the integral as a function of the domain, *J. Geom. Anal.* **8** (1999), to appear.

[JW] A. Jonsson and H. Wallin, *Function Spaces on Subsets of* $\mathbb{R}^n$, Mathematical Reports, vol. 2, no. 1, Harwood Acad. Publ., Chur, Switzerland, 1984.

[Ki1] J. Kigami, A harmonic calculus on the Sierpinski spaces, *Japan J. Appl. Math.* **8** (1989), 259–290.

[Ki2] ——, Harmonic calculus on p.c.f. self-similar sets, *Trans. Amer. Math. Soc.* **335** (1993), 721–755.

[Ki3] ——, *Analysis on Fractals*, in preparation.

[Ma] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.

[Mo] U. Mosco, Dirichlet forms and self-similarity, *New Directions in Dirichlet Forms*, AMS/IP Studies in Advanced Math., vol. 8, Amer. Math. Soc., Providence, RI, 1998, pp. 117–155.

[S] R. Strichartz, Taylor approximations on Sierpinski gasket type fractals, preprint.

[SU] R. Strichartz and M. Usher, Splines on fractals, *Math. Proc. Cambridge Philos. Soc.*, to appear.

About the Cover

The Sierpinski gasket is the connected subset of the plane obtained from an equilateral triangle by removing the open middle inscribed equilateral triangle of 1/4 the area, removing the corresponding open triangle from each of the three constituent triangles, and continuing this way. The gasket can also be obtained as the closure of the set of vertices arising in this construction. The cover shows the vertices of the constituent triangles through seven iterations of constructing midpoints. The vertices are color coded according to the stage at which they first appear, the last three stages being red, orange, and yellow. The vertices have been increased in size from points to small disks to give the illusion of connectedness for the displayed finite set of vertices.

—Peter Sykes