SÉMINAIRE N. BOURBAKI

MARTIN BARLOW

Harmonic analysis on fractal spaces

Séminaire N. Bourbaki, 1991-1992, exp. n^o 755, p. 345-368.
<http://www.numdam.org/item?id=SB_1991-1992__34__345_0>

Séminaire BOURBAKI
44ème année, 1991-92, n° 755

Juin 1992

HARMONIC ANALYSIS ON FRACTAL SPACES

par Martin BARLOW

1. INTRODUCTION

The initial interest in this area came from mathematical physicists in the early 1980s, who were studying *transport properties* of disordered media. As there are good reasons for believing that fractals can provide a good model for such media, this led to their interest in such questions on the study of the wave and heat equations on fractal spaces – see for example [RT].

Mathematical work has, so far, largely centered on the more easily treated Laplace and heat equations. It began with a probabilistic treatment, by Kusuoka [K1], Goldstein [G] and Barlow–Perkins [BP], but an analytic treatment, making a substantial use of Dirichlet forms has been developed, mainly in Japan – see [Kig1, Kig2, F2, K2].

Most of the problems and difficulties arise, naturally, on the microscopic scale, from absence of any kind of Euclidean structure. However, given a regular fractal, such as the Sierpinski gasket, one can define a *pre-fractal* manifold or graph, whose large scale structure mimics that of the true fractal. These pre–fractals are entirely classical, but classical techniques fail to give, for example, the right bounds on the heat kernel on such an object.

Let me now be more specific about the kinds of problem that will be considered.

Let $F \subseteq \mathbb{R}^d$ be a connected self–similar fractal, let $d_f(F)$ be the Hausdorff dimension of F, and let μ_F be Hausdorff $x^{d_f(F)}$–measure on

F. The heat equation on F should take the form

$$\Delta_F u = \frac{\partial u}{\partial t}\ , \qquad u(x,0) = u_0(x)\ ,\ x \in F$$

where $u : F \times \mathbb{R}_+ \to \mathbb{R}$, $u_0 \in C(F)$, and Δ_F is a Laplacian operator acting on a subspace $\mathcal{D}(\Delta_F) \subset C(F)$. The following problems arise immediately:

(i) *Existence.* The construction of a suitable operator Δ_F which is F–isotropic, that is, locally invariant with respect to the local isometries of F.

(ii) *Uniqueness.* Is Δ_F characterised by the property of being F–isotropic?

(iii) *Properties.* The properties of the solutions to the Laplace and heat equation associated with Δ_F, and the form of the spectrum of Δ_F.

To these we add questions about the Markov process X_t with generator Δ_F and semigroup $P_t = \exp(t\Delta_F)$.

2. DEFINITION OF THE DIRICHLET FORM

The first work was done on what appears to be the simplest non–trivial connected fractal, the Sierpinski gasket. Later works have extended many of these results to larger classes of sets: to *nested fractals* ([L]), *p.c.f. self–similar sets* ([Kig2]), and a to still more general class which includes the Sierpinski carpets ([KZ]). However in this survey I will for the most part restrict myself to a subset of nested fractals which is large enough to capture the essential features of the subject. Indeed, it seems clear that more general sets exhibit the same kind of behaviour as the simpler ones: it is just that the results are more difficult to prove.

Let $F_0 = \{a_1, \dots, a_k\}$ be the vertices of a regular k–sided polygon in $\mathbb{R}^2$, and let H_0 be the closed convex hull of F_0. Let $\lambda \geq 1$, $M \geq k$, $a_{k+1}, \dots, a_M \in \mathbb{R}^2$ and let ψ_i, $1 \leq i \leq M$ be defined by

$$\psi_i(x) = a_i + \lambda^{-1}(x - a_i).$$

We assume:

(A1) (Symmetry) The set $\{a_i,\ 1 \le i \le M\}$ has the same symmetries as F_0,

(A2) (Nesting) $\psi_i(H_0) \subseteq H_0$,

(A3) (Connectedness) The set $H_1 = \bigcup_{i=1}^{M} \psi_i(H_0)$ is connected.

(A4) (Open set condition) $\{\mathrm{Int}(\psi_i(H_0)),\ 1 \le i \le M\}$ are disjoint.

(A5) (Finitely ramified) If $x, y \in H_0$ and $\psi_i(x) = \psi_j(y)$ for $i \ne j$ then $x, y \in F_0$.

Remark. Note that (A5) rules out fractals such as the Sierpinski carpet. This assumption is in fact very strong, and, as we will see, enables the different 'levels' of the fractal to be treated separately. Most of the work done so far has used this hypothesis in an essential way. Nevertheless, it is possible to handle the Sierpinski carpet, at least in two–dimensions, (see [BB1, BB2, BB3], [KZ]), at the cost of working rather harder.

Let $\Psi(\cdot) = \cup_{i=1}^{M} \psi_i(\cdot)$; and set $F = \cap_{n=0}^{\infty} \Psi^n(H_0)$, $F_n = \Psi^n(H_0)$. The set F is a *nested fractal* with dimension $d_f(F) = \log M / \log \lambda$, and is approximated by the finite sets F_n.

It is helpful to introduce co–ordinates via the associated abstract sequence spaces. Let $I = \{1, \dots, M\}$, and for $w = (i_1, \dots, i_n) \in I^n$ let

$$\psi_w = \psi_{i_1, i_2, \dots, i_n} = \psi_{i_1} \circ \dots \circ \psi_{i_n};$$

we will call the sets $\psi_w(F_0), w \in I^n$, *n–cells*, and the sets $\psi_w(H_0)$, $w \in I^N$, *n–complexes*. For $w \in I^{\mathbb{N}}$ write $w|n = (w_1, \cdots, w_n) \in I^n$, and note that $\psi_w(H_0) = \cap_{n=0}^{\infty} \psi_{w|n}(H_0)$ consists of a single point, $\phi(w)$ say. It is easily checked that the map $\phi : I^{\mathbb{N}} \to F$ is continuous and surjective, but not injective. In the case of the Sierpinski gasket, for example, we have $\phi(1\dot{2}) = \phi(2\dot{1})$. In fact, one can define a fractal in an entirely abstract way, by considering the set $I^{\mathbb{N}}$ under a quotient map which identifies suitable sequences of this kind – and this is done in [Kig2].

We now proceed to define Dirichlet forms on the finite sets F_n.

Let $\Lambda = \{\alpha \in \mathbb{R}^{k-1} : \alpha_i = \alpha_{k-i}, 1 \le i \le k-1\}$, and define, for $x, y \in F_n$, $\alpha \in \Lambda$,

$$a_n(\alpha, x, y) = \begin{cases} \alpha_i, & \text{if } x, y \text{ belong to the same } n\text{–cell and are} \\ & i \text{ steps apart on the circumference,} \\ 0, & \text{otherwise,} \end{cases}$$

$$\mu_n(\alpha, x) = \sum_y a_n(\alpha, x, y).$$

Note that $\mu_n(\alpha, F_n) = M^n \sum \alpha_i$, and that $\mu_n(\alpha, x)$ depends on α only through $\sum \alpha_i$. Set, for $f \in B(F_n) = \mathbb{R}^{F_n}$,

$$\mathcal{E}^n_\alpha(f,f) = \sum_x \sum_y a_n(\alpha, x, y)\,(f(x) - f(y))^2. \tag{2.1}$$

We may interpret the Dirichlet form $\mathcal{E}$ in a number of ways.

1. If we regard $(F_n,\ a_n(\alpha))$ as an electric network where the wire between x and y has conductivity $a_n(\alpha, x, y)$ then $\mathcal{E}^n_\alpha(f, f)$ is the energy dissipation if a potential f is maintained on the network.

2. The discrete Laplacian Δ^n_α on F_n is defined by

$$\Delta^\alpha_n f(x) = \mu_n(\alpha, x)^{-1} \sum_y a_n(\alpha, x, y)(f(y) - f(x)),$$

and, writing (f, g) for the inner product with respect to $\mu(\alpha, .)$, Δ^n_α satisfies

$$\mathcal{E}^n_\alpha(f, g) = -(\Delta^n_\alpha f, g), \quad f, g \in C(F_n).$$

The random walk $X^n_r, r \ge 0$ with generator Δ^n_α and Dirichlet form $\mathcal{E}^n_\alpha$ is defined by

$$P(X^n_r = y | X^n_{r-1} = x) = p^n_\alpha(x, y) = a_n(\alpha, x, y)/\mu_n(\alpha, x).$$

We write P^n_α for the law of this random walk.

3. It will sometimes be helpful to write (2.1) in matrix terms. Let A ($= A^n(\alpha)$) be the matrix defined by

$$A_{xy} = \delta_{xy}\mu_n(\alpha, x, y) - a_n(\alpha, x, y);$$

then

$$\mathcal{E}^n_\alpha(f, f) = f^T A f. \tag{2.2}$$

The fundamental property of finitely ramified fractals is that the operation of 'decimation' works in a particularly straightforward fashion. We now introduce this. Given $\mathcal{E}^n_\alpha$, we define, for $g \in B(F_{n-1})$, the decimated Dirichlet form

$$\tilde{\mathcal{E}}^n_\alpha(g, g) = \inf\{\mathcal{E}^n_\alpha(f, f) : f|_{F_{n-1}} = g\}. \tag{2.3}$$

THEOREM 2.1. *$\tilde{\mathcal{E}}^n_\alpha$ is a Dirichlet form and there exists $\tilde{\alpha}(\alpha) \in \Lambda$ such that*

$$\tilde{\mathcal{E}}^n_\alpha = \mathcal{E}^{n-1}_{\tilde{\alpha}}.$$

It is straightforward to check this result. Note that the minimising function f in (2.3) satisfies

$$\Delta^\alpha_n f(x) = 0, \;\; x \in F_n - F_{n-1}, \tag{2.4}$$

and that f inside an $n-1$ cell $\Psi_w(H_0)$ depends only on the values of g on $I_w(F_0)$. Thus the proof reduces immediately to the case $n = 1$.

We now comment on the various interpretations of this operation.

1. In terms of electrical networks we can view $(F_n, a_n(\alpha))$ as a 'black box', with only the nodes in F_{n-1} being accesible. Then the network $(F_n, a_n(\alpha))$ has the same response (in terms of energy dissipation, current flows etc.) to inputs in F_{n-1} as the network $(F_{n-1}, a_{n-1}(\tilde{\alpha}))$.

2. We may 'decimate' the random walk $(X_r^n, r \geq 0)$ by defining

$$T_0 = \min\{r \geq 0 : \ X_r^n \in F_{n-1}\},$$
$$T_m = \min\{r \geq T_{m-1} : \ X_r^n \in F_{n-1} - \{X_{T_m - 1}\}\}$$
$$\widetilde{X}_r^n = X_{T_r}^n \ , r \geq 0.$$

Then the random walk $(\widetilde{X}^n, P_\alpha^n)$ is equal in law to the random walk $(X^{n-1}, P_{\widetilde{\alpha}}^{n-1})$

3. In matrix terms we write vectors $f \in \mathbb{R}^{F_n}$ with the coordinates in F_{n-1} first. So

$$A_n(\alpha) = \begin{pmatrix} B_n(\alpha) & C_n(\alpha) \\ C_n(\alpha)^T & D_n(\alpha) \end{pmatrix}$$

and the matrix corresponding to the Dirichlet form $\widetilde{\mathcal{E}}_\alpha^n$ is given by

$$\widetilde{A}_n(\alpha) = B_n(\alpha) - C_n(\alpha) D_n(\alpha)^{-1} C_n(\alpha)^T.$$

The decimation operation enables us to reduce many questions concerning the Dirichlet forms $\mathcal{E}_\alpha^n$ to the behaviour of iterates of the map $\widetilde{\alpha}$. Of particular significance are (affine) fixed points of the function $\widetilde{\alpha}$, that is $\beta \in \Lambda$ such that

$$\widetilde{\alpha}(\beta) = \beta/\rho \quad \text{for some } \rho > 0. \tag{2.5}$$

These fixed points correspond to processes and operators on F with the correct scaling properties with respect to the maps ψ_i. As $\widetilde{\alpha}$ is continuous, the fixed point theorem applied to the simplex $\Lambda' = \{\alpha \in \Lambda : \sum \alpha_i = 1\}$ shows that at least one fixed point exists. A reflection argument [L, Ch. V] shows that for nested fractals there is at least one fixed point that is non–degenerate in the sense that $\alpha_i > 0$ for all i, so that the electrical network $(F_1, a_1(a))$ is connected.

The general problem of the existence and uniqueness of (non–degenerate) fixed points seems to be a hard one. The natural approach is to show that $\widetilde{\alpha}$ is a contraction on Λ' in some suitable metric d, but finding such a metric does not appear to be easy. There are some partial results in [B1], which introduces an approach which may be effective for general nested fractals, but which will not work for general p.c.f. self–similar sets. It also seems likely that any non–degenerate fixed point β will also be stable; again this would be settled if a suitable contraction could be found.

Problem. Can the map $\widetilde{\alpha}$ have more than one non–degenerate fixed point? Are non–degenerate fixed points always stable?

Remark. It is easy to see that $\widetilde{\alpha}$ can have one or more degenerate fixed points – examples are given in [L] and [B1]. It is also possible that for a p.c.f self – similar set there is no non–degenerate fixed point – see the example in [HHW].

The scalar ρ in (2.5) will play a fundamental role in the analytic properties of the fractal F. In terms of the electrical network model, replacing the network $(F_0, a_0(\beta))$ by the network $(F_1, a_1(\beta))$ gives rise, to an observer only able to access the nodes in F_0, to a network equivalent to $(F_0, a_0(\rho^{-1}\beta))$: since resistance $= 1/$ conductivity it is natural to call ρ the *resistance scale factor.*

The close connection between electric networks and symmetric random walks also enables us to give a probabilistic interpretation of ρ. Let G be a finite set, $a = (a_{xy},\ x, y \in G)$ be a 'conductivity' matrix satisfying

$$a_{xx} = 0,\ \ a_{xy} = a_{yx} \geq 0,$$

and let Y_n, $n \geq 0$ be the Markov chain with transition probabilities given by $p_{xy} = a_{xy}/\mu_x$, where $\mu_x = \sum_y a_{xy}$. Then writing $S_x = \min\{n \geq 0 : Y_n = x\}$ and $R(x,y)$ for the effective resistance of the network (G,a) between the points x and y, we have ([C])

$$E^x S_y + E^y S_x = R(x,y) \sum_x \mu_x. \tag{2.6}$$

Since the final term may be interpreted as the 'mass' of the network, we have the informal relation

$$\text{Time} = \text{Resistance} \times \text{Mass}.$$

With this motivation, it is easy to prove that, for the random walk X^n on $(F_n, a_n(\beta))$ we have

$$ET_m = M\rho.$$

We call M (the number of 1-cells in F_1) the *mass scaling factor* of the fractal F, $\tau = M\rho$ the *time scaling factor*, and λ (the contraction factor in the similitudes ψ_i) the *length scaling factor*. From these we define two new indices connected with F:

$$d_w(F) = \frac{\log \tau}{\log \lambda}$$

$$d_s(F) = -\frac{2\log M}{\log \tau};$$

following the physics literature we call these the *walk* and *spectral* dimensions of F. (The reasons for this terminology will become apparent later).

Most of the analytic properties of F can be summarised in terms of these two 'dimensions', together with the Hausdorff dimension $d_f(F)$. Of course there are really only two independent quantities, since we have

$$d_w(F) = \frac{2d_f(F)}{d_s(F)}.$$

Remarks 1. For the fractals considered here, it is easy to show that $\rho \geq 1$, or equivalently that $\tau \geq M$, so that $d_s(F) \leq 2$. We also have that $d_s(F) \leq d_f(F)$, giving $d_w(F) \geq 2$.

2. Though it does not fall into the class of nested fractals considered here, it is possible to define these dimensions for the unit cube $C \subset \mathbb{R}^d$. We then have

$$d_f(C) = d\ ,\ d_w(C) = 2\ ,\ d_s(C) = d.$$

3. LIMITING PROCESSES AND REGULARITY.

We now fix a non–degenerate fixed point α, with resistance scaling factor ρ, and drop the α from expressions like Δ^n_α. We rescale α such that $\mu_n(\alpha, F_n) = M^n$.

We wish to obtain a limit of the (suitably rescaled) Δ^n. One method is probabilistic. If X^n is the random walk on F_n with generator Δ^n, then by decimation we obtain random walks on F_{n-1}, F_{n-2}, etc. with generators $\Delta^{n-1}, \Delta^{n-2}$, etc. Taking projective limits, we obtain a sequence $X^n, n \geq 0$, tied together by this decimation property. If we rescale time by considering the processes

$$Y^n_t = X^n_{[\tau^n t]} \ , \ t \geq 0,$$

then each of the Y^n crosses F_0 in mean time 1, and it is not too hard to see that in fact $Y^n_t \to Y_t$ a.s., and that Y is a continuous F–valued process. Unfortunately, establishing the Markov property for Y is rather tiresome and technical. There are therefore some advantages in using the more analytic approach outlined in [F2].

For $f \in B(F_m)$, and $n \geq m$, let $L_{n,m}f \in B(F_n)$ be the unique function such that

$$\begin{aligned} L_{n,m}f(x) &= f(x), \quad x \in F_m \\ \Delta^n L_{n,m}f(x) &= 0 \qquad x \in F_n - F_m. \end{aligned} \tag{3.1}$$

We call a function f such that $\Delta^n f(x) = 0$ for $x \in F_n - F_m$ *m–harmonic.* It is easily verified that $L_{m,m-1}f$ is the function g which attains the minimum in (2.3); therefore

$$\mathcal{E}^m(L_{m,m-1}f, L_{m,m-1}f) = \rho^{-1}\mathcal{E}^{m-1}(f,f). \tag{3.2}$$

Thus, since $\mathcal{E}^m(f, f - L_{m,m-1}(f|_{F_{m-1}})) = 0$, we have

$$\rho^m\mathcal{E}^m(f,f)) \geq \rho^{m-1}\mathcal{E}^{m-1}(f|_{F_{m-1}}, f|_{F_{m-1}}). \tag{3.3}$$

From now on we will often avoid notation like $f|_{F_{m-1}}$ by extending $\mathcal{E}^{m-1}$ etc. to functions on F. (3.3) gives an easy definition of the limiting Dirichlet form on F: we set, for $f \in C(F)$,

$$\mathcal{E}(f,f) = \lim_n \rho^n\mathcal{E}^n(f,f) \ , \quad \mathcal{D} = \{f \in C(F) : \mathcal{E}(f,f) < \infty\}.$$

We now consider the regularity properties of $(\mathcal{E},\mathcal{D})$.

If $f \in B(F_0)$, then, as $\alpha_i > 0$ for all i, we have

$$\sup_{x,y\in F_0} |f(x)-f(y)|^2 = \mathrm{Osc}(f,F_0)^2 \le c_0\mathcal{E}^0(f,f).$$

As any pair of points in F_1 can be connected by a chain of at most M 1–cells, it follows that, for $f \in B(F_1)$,

$$\mathrm{Osc}(f,F_1)^2 \le c_0 M\mathcal{E}^1(f,f). \tag{3.4}$$

A standard argument, linking x, y by a suitable tower of r–cells, with $0 \le r \le n$, establishes the following Sobolev inequality:

$$\mathrm{Osc}(f,F_n) \le c_1\rho^n\mathcal{E}^n(f,f), \quad f \in B(F_n). \tag{3.5}$$

If x, y belong to the same m–complex, then $|x-y| < c\lambda^{-m}$, and the chain need only use r–cells with $m \le r \le n$; it follows that

$$|f(x)-f(y)|^2 \le c_2|x-y|^{d_w-d_f}\rho^n\mathcal{E}^n(f,f). \tag{3.6}$$

Let $\mathcal{H}_{n,m} = \{f \in B(F_n) : \Delta^n f(x) = 0, x \in F_n - F_m\}$ be the set of m– harmonic functions on F_n. The decimation property implies that if $f \in \mathcal{H}_{n,m}$ then $f|_{F_{n-1}} \in \mathcal{H}_{n-1,m}$ if $n-1 \ge m$. (Perhaps the quickest way to see this is in probabilistic terms – $f(X^n)$ is a martingale up to the first time X^n hits F_m, and therefore the process $f(\widetilde{X}^n)$ is also). Thus if $n \ge j \ge m$ then $L_{n,j}L_{j,m}f = L_{n,m}f$. If $g \in B(F_m)$, then for $x \in F_\infty = \cup_{n=0}^\infty F_n$, define

$$L_m g(x) = L_{n,m}g(x), \text{ where } x \in F_n.$$

Since $\rho^n\mathcal{E}^n(L_{n,m}g, L_{n,m}g) = \rho^m\mathcal{E}^m(g,g)$ it follows using (3.6) that $L_m g$ is uniformly continuous on F_∞, and so may be extended to a function (also denoted $L_m g$) in $C(F)$, satisfying $\mathcal{E}(L_m g, L_m g) = \rho^m\mathcal{E}^m(g,g)$.

This shows that

$$\bigcup_{m=0}^\infty \{L_m g, \quad g \in B(F_m)\} \subseteq \mathcal{D},$$

so that $\mathcal{D}$ is dense in $C(F)$, and this, together with the fact that $\mathcal{D} \subset C(F)$ proves that the Dirichlet form $(\mathcal{E}, \mathcal{D})$ is regular.

Let μ be the weak limit of the measures $M^{-n}\mu_n$: then $\mu(F) = 1$, and μ assigns mass M^{-n} to any n–complex. (μ is a multiple of the Hausdorff x^{d_f} – measure on F, and is also the image measure under the mapping ϕ of the uniform product measure on the abstract sequence space $I^{\mathbb{N}}$). Then since

$$\rho^n \mathcal{E}^n(f,g) = -\sum_{x\in F_n} M^{-n}\mu_n(x)\tau^n\Delta^n f(x)g(x),$$

if we define Δ_F to be the self–adjoint operator on $L^2(F,\mu)$ associated with $(\mathcal{E}, \mathcal{D})$ by $\mathcal{E}(f,g) = -(\Delta_F f, g)$, then Δ_F is approximated by the sequence $\tau^n\Delta^n$.

Let us summarise the properties of $\mathcal{E}$ and Δ_F and the associated diffusion process X in the following

THEOREM 3.1 (a) *$(\mathcal{E}, \mathcal{D})$ is a regular local Dirichlet form on $L^2(F,\mu)$.*

(b) *For $f \in \mathcal{D}$, $x, y \in F$,*

$$|f(x) - f(y)|^2 \le c|x-y|^{d_w - d_f}\mathcal{E}(f,f), \tag{3.7}$$

(c) *For $f \in \mathcal{D}$*

$$\rho^n \mathcal{E}^n(f|_{F_n}, f|_{F_n}) \uparrow \mathcal{E}(f,f).$$

(d) *Δ_F is a self–adjoint operator on $L^2(E,\mu)$ with $\mathcal{D}(\Delta_F) \subset \mathcal{D}$, satisfying*

$$\mathcal{E}(f,g) = -(\Delta_F f, g)$$
$$\Delta_F f(x) = \lim_{n\to\infty} \tau^n \Delta^n f(x)\ ,\ x \in F_\infty$$

(e) *If $(X_t, t \ge 0, P^x, x \in F)$ is the diffusion (continuous strong Markov process) with semigroup $P_t = e^{t\Delta_F}$, then X is the weak limit of the processes $Y_t^n = X^n_{[\tau^n t]}$. The process X is symmetric with respect to μ.*

The standard Laplacian on $\mathbb{R}^d$ can be characterised, up to a multiplicative constant, as the unique second order which is invariant with respect to the isometries of $\mathbb{R}^d$. It is natural to ask for a similar characterisation of Δ_F. In general the answer is not known – the problem is related to that of the uniquness of fixed points mentioned earlier. However for the Sierpinski gasket (where there is evidently only one fixed point), it is proved in [BP] that, up to a deterministic time change, X is the unique process which is locally invariant with respect to the local isometries of F: a corresponding uniqueness for Δ_F follows.

4. ANALYTIC PROPERTIES OF Δ_F.

We begin by introducing the potential kernel densities $u_\alpha(x,y)$, $\alpha > 0$, $x, y \in F$. Formally these are the solutions to

$$(\alpha - \Delta_F)u_\alpha(\cdot, y) = \delta_y(\cdot).$$

They may be defined via the process X as the density with respect to μ of the α–resolvent

$$U_\alpha f(x) = E^x \int_0^\infty e^{-\alpha t} f(X_t)dt; \tag{4.1}$$

of course it requires some work to prove that $U_\alpha(\cdot, x) \ll \mu$. A quicker approach is via the Dirichlet form $\mathcal{E}_\alpha(f,g) = \alpha \cdot (f,g) + \mathcal{E}(f,g)$; from (3.7) it follows that

$$||f||_\infty^2 \le c_\alpha \mathcal{E}_\alpha(f,f) \ , \ f \in \mathcal{D},$$

so that the map $T_y : \mathcal{D} \to \mathbb{R}$ defined by $T_y(g) = g(y)$ is bounded. It follows, as in [F1, 3.3.3] that $\mathcal{E}_\alpha$ admits a reproducing kernel $u_\alpha(x,y)$, so that

$$\mathcal{E}_\alpha(u_\alpha(\cdot, y), g) = g(y) \ , \ g \in \mathcal{D}. \tag{4.2}$$

THEOREM 4.1 (a) $u_\alpha(x,y) = u_\alpha(y,x),\ x,y \in F$.
(b) u_α *is Hölder continuous of order* $d_w - d_f$ on $F \times F$.
(c) $U_\alpha f(x) = \int_F u_\alpha(x,y) f(y)\mu(dy),\ f \in C(F)$.

Various probabilistic properties of X follow from this result. The boundedness of the $u_\alpha(\cdot, y)$ implies that, writing $T_y = \inf\{t > 0 : X_t = y\}$ we have

$$u_\alpha(x,y) = E^x e^{-\alpha T_y} u_\alpha(y,y),$$

so that $P^y(T_y = 0) = 1$, and y is regular for $\{y\}$. The Hölder continuity of u_α implies that X has a jointly continuous local time (occupation density) $(L_t^x, x \in F,\ t \geq 0)$ and from this it follows that X is 'space–filling':

$$P^x\big(\{X_t, 0 \leq t \leq N\} = F \text{ for some } N < \infty\big) = 1.$$

Some properties of X can be expressed more simply if we consider instead the process on an unbounded version of F. Let us take $a_1 = 0$, so that the similitude ψ_1 is given by $\psi_1(x) = \lambda^{-1}x$. Then set

$$\widetilde{F} = \bigcup_{n=0}^{\infty} \lambda^n F.$$

The construction of the Dirichlet form $\mathcal{E}$, the Laplacian Δ and the diffusion X on $\widetilde{F}$ can be done by a straightforward patching argument. Then $\widetilde{X}$ satisfies the scaling relation

$$(X_t, t \geq 0, P^0) \overset{(d)}{=} (\lambda^{-1} X_{\tau t},\ t \geq 0, P^0), \tag{4.3}$$

which may be compared with the relation $\theta^{-1} B_{\theta^2 t} \overset{(d)}{=} B_t$ for standard Brownian motion on $\mathbb{R}^d$. Since $|X_1 - X_0| = 0(1)$ it follows from (4.3) that $|X_t - X_0| = 0(t^{1/d_w})$ for $t > 0$. (Recall $1/d_w = \log\lambda / \log\tau$). More precisely we have

PROPOSITION 4.2 (a) *X is Hölder continuous of order $d_w^{-1} - \varepsilon$ for any $\varepsilon > 0$, and is not Hölder contiuous of order d_w^{-1}.*
(b) *There exist constants c_1, c_2 such that for $t \geq 0$*

$$c_1 t^{2/d_w} \leq E^x |X_t - x|^2 \leq c_2 t^{2/d_w}.$$

More precise results, giving the exact modulus of continuity of the paths of X can be found in [BP, BB3, Kum]. This result gives the most intuitive explanation of the meaning of the 'dimension' d_w: it governs the space–time scaling of the diffusion X in the same way that the number 2 governs the space–time scaling of ordinary Brownian motion.

COROLLARY 4.3 (a) *X is not a semimartingale.*
(b) *If $f \in C^1(\mathbb{R}^d)$ and $f|_F \in \mathcal{D}(\Delta_F)$ then $f|_F$ is constant.*

(a) follows from the Proposition and the fact that the paths of any semimartingale have finite quadratic variation. For (b), if $f \in \mathcal{D}(\Delta)$ then

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \Delta f(X_s)ds$$

is a martingale, so $f(X_t)$ is a semimartingale, which is impossible unless $f|_F$ is constant.

We thus see that the 'smooth functions' in $\mathcal{D}(\Delta_F)$ have nothing to do with the ordinary smooth functions on $\mathbb{R}^d$. Not a lot is known at present about the structure and form of functions in $\mathcal{D}(\Delta_F)$ or $\mathcal{D}$. From Theorem 3.1 we have that they are Hölder continuous of order $d_w - d_f$ everywhere; they are also Hölder continuous of order $\frac{1}{2}d_w$ μ–a.e. on F. (Note that as $\frac{1}{2}d_w > 1$ this last result cannot hold everywhere). Some more precise results are obtained in [K2]: if ν_f is the local measure associated with $\mathcal{E}(f,f)$, so that

$$\mathcal{E}(f,f) = \int_F \nu_f(dx), \ f \in \mathcal{D},$$

then there exists a measure ν on F such that $\mu \perp \nu$, and $\nu_f \ll \nu$ for all $f \in \mathcal{D}$. Informally, we may say that for $f \in \mathcal{D}$, $|\nabla f| = 0$ μ-a.e., while $\| \, |\nabla f| \, \|_\infty = \infty$.

As the $U_\alpha(x,y)$ are continuous U_α is compact, and we can write

$$u_\alpha(x,y) = \sum_{i=1}^\infty (\alpha + \lambda_i)^{-1} \varphi_i(x) \varphi_i(y),$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq \dots$ are the eigenvalues of $-\Delta_F$, and the φ_i are the normalised eigenfunctions. Note that the λ_i and φ_i satisfy

$$\mathcal{E}(\varphi_i, f) = \lambda_i(f, g) \quad \text{for all } g \in \mathcal{D}.$$

We now examine the asymptotic behaviour of the eigenvalues. Define

$$\mathcal{D}_i = \{f \in \mathcal{D} : f = 0 \text{ on } F_i\}, \quad i = 0, 1,$$

and write Δ_i for the associated Laplacian (which corresponds to Dirichlet boundary conditions on F_i). Note that if $(\mathcal{E}_i, \mathcal{D}_i)$, $i = 1, 2$ are Dirichlet forms with $\mathcal{D}_1 \subseteq \mathcal{D}_2$ and $\mathcal{E}_2 = \mathcal{E}_1$ on $\mathcal{D}_1$, then the max–min principle (see [Ch]) shows that, writing $\lambda_n^i, n \geq 1$ for the eigenvalues of the operators $-\Delta_i$, we have $\lambda_n^1 \geq \lambda_n^2$. Hence if $N_i(x) = \#\{\lambda_n^i : \lambda_n^i \leq x\}$ we have $N_1(x) \leq N_2(x)$. We call the functions N_i the *integrated density of states* (for $(\mathcal{E}_i, \mathcal{D}_i)$).

Write N, N_i for the integrated density of states for $(\mathcal{E}, \mathcal{D}), (\mathcal{E}, \mathcal{D}_i)$ respectively. If $f, g \in B(F_n)$ then

$$\mathcal{E}^n(f, g) = \sum_{i=1}^{M} \mathcal{E}^{n-1}(f \circ \psi_i, g \circ \psi_i),$$

so for $f, g \in \mathcal{D}$ it follows that

$$\mathcal{E}(f, g) = \sum_{i=1}^{M} \rho \mathcal{E}(f \circ \psi_i, g \circ \psi_i). \tag{4.4}$$

As $\mathcal{D}_1 \subseteq \mathcal{D}_0$, $N_1(x) \leq N_0(x)$. However, if g is an eigenfunction of $-\Delta_0$, with $\Delta_0 g + \lambda g = 0$, then for $1 \leq i \leq M$ define $h_i \in C(F)$ by

$$h_i(x) = \begin{cases} g \circ \psi_i^{-1}(x) & \text{for } x \in \psi_i(F). \\ 0 & \text{for } x \notin \psi_i(F). \end{cases}$$

Then for $v \in \mathcal{D}_1$, by (4.4), and as $h_i \circ \psi_i = g$,

$$\mathcal{E}(h_i, v) = \rho \mathcal{E}(g, v \circ \psi_i) = \rho \lambda(g, v \circ \psi_i) = M \rho \lambda(h_i, v).$$

So $\tau\lambda$ is an eigenvalue of Δ_1 with multiplicity M and it follows that $N_1(x) = MN_0(x/\tau)$.

We perform a similar kind of surgery for the Neumann eigenvalues on F; this time it involves cuts at the points in F_1, so replacing F by a disjoint union of the spaces $\psi_i(F), 1 \leq i \leq M$. Writing $\mathcal{D}'$ for the associated Dirichlet space, and N' for the integrated density of states, we have (taking a suitable embedding) $\mathcal{D} \subseteq \mathcal{D}'$. The same argument as for the Dirichlet case gives (writing $y = x/\tau$)

$$\begin{aligned} N_0(y) = M^{-1}N_1(\tau y) \leq M^{-1}N_0(\tau y) &\leq M^{-1}N(\tau y) \\ &\leq M^{-1}N'(Ty) = N(y). \end{aligned}$$

Thus there exist c_1, c_2 such that

$$c_1 x^{\frac{1}{2}d_s} \leq N_0(x) \leq N(x) \leq c_2 x^{\frac{1}{2}d_s} \quad \text{for } x \geq x_0. \tag{4.5}$$

It is now natural to ask about finer details of the asymptotics of $N(\cdot)$. However, Fukushima and Shima [FS] have proved that, for the Sierpinski gasket,

$$\liminf_{x\to\infty} x^{-\frac{1}{2}d_s} N(x) < \limsup_{x\to\infty} x^{-\frac{1}{2}d_s} N(x) \ ;$$

thus one cannot hope for an asymptotic expansion of N of the kind that occurs for domains in $\mathbb{R}^d$.

Remarks 1. [S] and [FS] give a very detailed description of the spectrum of Δ for the Sierpinski gasket.

2. Equation (4.5) explains the term *spectral dimension* for the number d_s.

5. HEAT KERNEL BOUNDS

So far we have had to pay little attention to the detailed geometry of F. However, when we consider the form of the heat kernel the metric $|x - y|$

is inadequate: it is better to consider an intrinsic metric $d(x,y)$, related to the shortest path between x and y in F. Let

$$H_n = \Psi^n(H_0)$$

be the closed connected set obtained after n iterations of the set map Ψ defined by (1.2), and $d_n(x,y)$ be the length of the shortest path in H_n connecting x and y.

For fractals such on the Sierpinski gasket and carpet one has $|x-y| \le d_n(x,y) \le c|x-y|$ for all n, but for more general fractals, where the limiting set does not contain straight line segments, this fails. In general there exists $b \ge \lambda$ such that points in F_0 may be connected by a chain of $O(b^n)$ n–cells, so that, writing

$$d_c = \log b / \log \lambda \ ,$$

one has

$$d_n(x,y) \approx |x-y|^{d_c}(b/\lambda)^n.$$

Taking limits along a subsequence one obtains a metric d on F satisfying

$$d(x,y) \approx |x-y|^{d_c}.$$

This metric is called the *chemical distance* in the physics literature, and d_c is known as the *chemical exponent.* (Contrary to what a mathematician might expect, physicists do not usually measure the length of a polymer with this metric).

It is helpful to redefine the fractal and walk dimensions in terms of this new metric, setting

$$d_f^l = \log M / \log b = d_f/d_c,$$
$$d_w^l = \log \tau / \log b = d_w/d_c.$$

Note that the spectral dimension satisfies $\frac{1}{2}d_s = d_f/d_w = d_f^l/d_w^l$ and is unaffected by the change of metric.

Now consider X on the unbounded fractal $\widetilde{F}$, and let $p(t,x,y)$ be the associated transition density. (The existence of such a function follows from the μ –symmetry of X and the existence of a resolvent density U_α). p is the solution to the heat equation on $\widetilde{F}$:

$$\frac{\partial}{\partial t}p(t,x,y) = \Delta_x p(t,x,y)$$
$$p(0,x,y) = \delta_y(x)$$

Upper bounds on $p(t,x,x)$ can be obtained from the scaling of $\mathcal{E}(\cdot,\cdot)$, and the general theory of [CKS]. The argument here is due to Fitzsimmons and Hambly. Since (see [F, p21])

$$\mathcal{E}(f,f) = \sup_\alpha(\alpha f, f - \alpha U_\alpha f) \geq (f, f - U_1 f),$$

we have

$$\|f\|_2^2 \leq \|u_1\|_\infty \|f\|_1^2 + \mathcal{E}(f,f).$$

Replacing f by $\widetilde{f}(x) = f(\lambda^n x)$, using scaling, and optimising over n, one obtains the *Nash inequality*

(5.1) $$\|f\|_2^{2+4/d_s} \leq c\mathcal{E}(f,f).\|f\|_1^{4/d_s}.$$

This implies ([CKS]) that

(5.2) $$p(t,x,x) \leq c_1 t^{\frac{1}{2}d_s}, \quad 0 < t < \infty.$$

On-diagonal lower bounds are also quite easy, and follow from the estimate

$$\lim_{\theta\uparrow\infty} \sup_{x,t}\ P^x(|X_t - x| > \theta t^{1/d_w}) = 0.$$

This implies that, for a suitable $r > 0$, writing $z = rt^{1/d_w}$,

$$\tfrac{1}{2} < \int_{B(x,z)} p(t,x,y)\mu(dy) \leq \mu(B(x,z))p(t,x,x)^{\frac{1}{2}} \sup_y p(t,y,y)^{\frac{1}{2}}.$$

Using $\mu(B(x,z)) \geq cz^{d_f}$ gives the lower bound

$$p(t,x,x) \geq ct^{\frac{1}{2}d_s}, \quad 0 < t < \infty. \tag{5.3}$$

Remarks. 1. The link between the asymptotics of the heat kernel and the number of eigenvalues means that one would expect bounds of this form for small t from (4.5). The scaling relation (4.2) then implies that these bounds hold for all t.

2. Note that the exponent in (5.2) and (5.3) is not related to the isoperimetric dimension, which is 0 for finitely ramified fractals. The situation here is therefore rather different from the groups studied in [V], where an isoperimetric inequality yields a Nash inequality with the correct exponent. Compare also the bounds on the heat kernel given in [O] and [BB3].

The off–diagonal bounds are more tricky. There is a general 'machine' for obtaining off–diagonal upper bounds – see the account of 'Davies' method' in [CKS]. This essentially involves looking at the semigroup $e^{-\psi}P_t e^{\psi}$, using the global upper bound given by the on–diagonal upper bound, and varying ψ suitably. Unfortunately this does not seem to work in the fractal case. The essential reason is that, on one hand, one requires $\psi \in \mathcal{D}$, while the upper bound obtained involves, in a more or less explicit form, expressions involving $||\nabla\psi||_\infty$, which as is proved in [Kus2], is $+\infty$ for nested fractals.

The technique one adopts, for both upper and lower bounds, is that of *chaining*. For the lower bound this is classical. Suppose we have that

$$p(t,x,y) \geq ct^{-d_s/2} \text{ for } t > 0, \quad |x-y| < c_2 t^{1/d_w}. \tag{5.4}$$

Writing $D = |x-y|$, we have that x and y may be connected by a chain of $n = c(D/\varepsilon)^{d_c}$ balls of Euclidean radius ε; denote these $B_i = B(x_i,\varepsilon)$,

$1 \le i \le n$, where $x_0 = x, x_n = y$. Choose ε so that $\varepsilon = c_2(t/n)^{1/d_w}$; then

$$\begin{aligned}
&p(t,x,y)\\
&\quad \ge \int_{B_1} \cdots \int_{B_{n-1}} p(t/n, x_0, x_1) \dots p(t/n, x_{n-1}, y)\mu(dx_1) \dots \mu(dx_{n-1})\\
&\quad \ge \prod_1^{n-1} \mu(B_i) c_1^n (t/n)^{-nd_s/2}\\
&\quad \ge (c\varepsilon^{d_f})^{n-1} c_1^n (t/n)^{-d_s/2} .(\varepsilon/c_2)^{-d_f(n-1)}\\
&\quad \ge t^{-d_s/2} c_3^n\\
&\quad = t^{-d_s/2} \exp(-c_4 n).
\end{aligned}$$

Substituting for n, one obtains

$$p(t,x,y) \ge t^{-d_s/2} \exp\big(-c(|x-y|^{d_w}/t)^{d_c/(d_w-d_c)}\big); \tag{5.5}$$

the exponent in (5.5) may be rewritten as

$$c\big(d(x,y)^{d_w^l}/t\big)^{1/(d_w^l-1)}.$$

Thus the chaining argument transforms the crude lower bound of (5.4) into a global lower bound which is, up to constants, of the best possible form.

It remains to prove (5.4). However, it holds for $y = x$, and the Hölder continuity of functions in $\mathcal{D}$ allows this to be extended to a ball.

A similar technique works for the upper bound, but the argument here is more probabilistic. For $x \in \widetilde{F}$, $\varepsilon > 0$ write

$$S(x,\varepsilon) = \inf\{t \ge 0 : |X_t - x| < \varepsilon\}.$$

The space–time scaling of X suggests that $\varepsilon^{-1/d_w} S(x,\varepsilon)$ should be $0(1)$, and using this and the elementary inequality

$$P(\xi \le t) \le (\mathrm{Var}(\xi) + 2t\mathrm{E}\xi)/\mathrm{E}\xi^2$$

which holds for any non–negative random variable ξ, one obtains

$$P(S(x,\varepsilon) < t) \le p + a\varepsilon^{-1/d_w} t, \quad t \ge 0. \tag{5.6}$$

The following lemma ([BB1, Lemma 1.1]) is the key to the chaining argument.

LEMMA 5.1 *Let $T, S_1, \dots, S_n$ be non–negative random variables satisfying, for some $p < 1, a > 0$,*

(a) $T \ge \sum_1^n S_i$,

(b) $P(S_i < t \,|\, \sigma(S_1, \dots S_{i-1})) \le p + at.$

Then $P(T \le t) \le \exp(2(ant/p)^{\frac{1}{2}} - n \log p^{-1})$.

Now fix x, y, t, and set $A = \{Z \in F : |z-y| < |z-x|\}$. If we set $S_0 = 0$, $S_{i+1} = S(X_{S_i}, \varepsilon)$, then since the shortest path from x to A in F crosses $n = c(|x-y|/\varepsilon)^{d_c}$ balls of radius ε, we have, writing $T_A = \inf\{t \ge 0 : X_t \in A\}$,

$$\begin{aligned}\log P^x(T_A \le t) &\le (2a\varepsilon^{-1/d_w} nt/p)^{\frac{1}{2}} - n\log p^{-1} \\ &\le c(|x-y|^{d_w}/t)^{d_c/(d_w-d_c)}.\end{aligned}$$

Combining this with the global upper bound (5.1), one obtains:

THEOREM 5.2 *Let $\phi(z,t) = (z^{d_w^l}/t)^{1/(d_w^l-1)}$. The heat kernel density $p(t,x,y)$ on $\widetilde{F}$ satisfies, for $0 < t < \infty$, $x, y \in \widetilde{F}$,*

$$c_1 t^{-d_s/2} \exp(-c_2\phi(d(x,y),t)) \le p(t,x,y) \le c_3 t^{-d_s/2} \exp(-c_4\phi(d(x,y),t)).$$

Remark. Bounds of this type are quite unfamiliar in classical situations, but they do nevertheless arise in setups less unfamiliar than the fractals treated here. For example, let $\widetilde{G}$ be the unbounded two–dimensional Sierpinski gasket (for which $d_c = 1, d_w = \log 5/\log 2$ and $d_f = \log 3/\log 2$). Embed $\widetilde{G}$ in $\mathbb{R}^3$, set

$$M' = \bigcup_{x \in \widetilde{G}} B(x,1),$$

and let $M = \partial M'$. Then M is a manifold with a large scale structure which mimics that of the Sierpinski gasket. If $p(t,x,y)$ is the heat kernel on M, then for small t, and $|x-y| = 0(1)$ its behaviour is similar to that of the heat kernel on $\mathbb{R}^2$. However one expects the estimates of Theorem 5.2 to hold in the region $1 \leq t \leq |x-y|$. (For $|x-y| > t$ one is back in the classical large deviation situation).

REFERENCES

[B1] M.T. BARLOW: Random walks, electrical resistance and nested fractals. To appear in *Asymptotic Problems in Probability Theory,* ed. K.D. Elworthy, N. Ikeda, Pitman.

[BB1] M. T. BARLOW and R. F. BASS: The construction of Brownian motion on the Sierpinski carpet. *Ann. Inst. H. Poincaré* **25** (1989) 225–257.

[BB2] M. T. BARLOW and R. F. BASS: On the resistance of the Sierpinski carpet. *Proc. R. Soc. London A.* **431** (1990) 345-360.

[BB3] M.T. BARLOW and R.F. BASS: Transition densities for Brownian motion on the Sierpinski carpet. *Probab. Th. Rel. Fields* **91** (1992) 307-330.

[BP] M. T. BARLOW and E. A. PERKINS: Brownian motion on the Sierpinski gasket. *Probab. Th. Rel. Fields* **79** (1988) 543–623.

[CKS] E.A. CARLEN, S. KUSUOKA and D.W. STROOCK: Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincaré Sup no. 2* (1987) 245-287.

[C] A.K. CHANDRA, P. RAGHAVEN, W.L. RUZZO, R. SMOLENSKY, P. TIWARI: The electrical resistance of a graph captures its commute and cover times. Proceedings of the 21st ACM Symposium on theory of computing, 1989.

[Ch] I. CHAVEL: *Eigenvalues in Riemannian Geometry.* Academic Press, 1984.

[F1] M. FUKUSHIMA: *Dirichlet forms and Markov processes*, North Holland 1980.

[F2] M. FUKUSHIMA: Dirichlet forms, diffusion processes, and spectral dimensions for nested fractals. To appear in "Ideas and methods in stochastic analysis, stochastics and applications", ed. S Albeverio et. al., Cambridge Univ. Press., Cambridge.

[FS] M. FUKUSHIMA and T. SHIMA: On a spectral analysis for the Sierpinski gasket. To appear *J. of Potential Analysis.*

[G] S. GOLDSTEIN: Random walks and diffusion on fractals. In: Kesten, H. (ed.) *Percolation theory and ergodic theory of infinite particle systems* (IMA Math. Appl., vol.8.) Springer, New York, 1987, pp.121–129.

[HHW] K. HATTORI, T. HATTORI and H. WATANABE: Gaussian field theories and the spectral dimensions. *Prog. Th. Phys. Supp. No.* **92** (1987) 108-143.

[Kig1] J. KIGAMI: A harmonic calculus on the Sierpinski space. Japan J. Appl. Math. **6** (1989) 259–290.

[Kig2] J. KIGAMI: A harmonic calculus for p.c.f. self-similar sets. To appear *Trans. A.M.S.*

[Kum] T. KUMAGAI: Estimates of the transition densities for Brownian motion on nested fractals. Preprint 1991.

[K1] S. KUSUOKA: A diffusion process on a fractal. In: Ito, K., N. Ikeda, N. (ed.) Symposium on Probabilistic Methods in Mathematical Physics, Taniguchi, Katata. Academic Press, Amsterdam, 1987, pp.251–274

[K2] S. KUSUOKA: Dirichlet forms on fractals and products of random matrices. *Publ. RIMS Kyoto Univ.,* **25** (1989) 659-680.

[KZ] S. KUSUOKA and X.Y.ZHOU: Dirichlet form on fractals: Poincaré constant and resistance. *Probab. Th. Rel. Fields* **93**, (1992) 169-186.

[L] T. LINDSTRØM: Brownian motion on nested fractals. *Mem. A.M.S.* **420**, 1990.

[O] H. OSADA: Isoperimetric dimension and estimates of heat kernels of pre–Sierpinski carpets. *Probab. Th. Rel. Fields* **86** (1990) 469–490.

[RT] R. RAMMAL and G. TOULOUSE: Random walks on fractal structures and percolation clusters, *J. Physique Lettres* **44** (1983) L13–L22.

[S] T. SHIMA: On eigenvalue problems for the random walk on the Sierpinski pre-gaskets. *Japan J. Appl. Ind. Math.*, **8** (1991) 127-142.

[V] N. Th. VAROPOULOS: Isoperimetric inequalities and Markov chains. *J. Funct. Anal.* **63** (1985) 215-239.

M.T. BARLOW
Department of Mathematics
University of British Columbia
Vancouver
British Columbia
Canada V6T 1Z2