DIFFUSIONS ON FRACTALS

Martin T. Barlow
University of British Columbia
Vancouver, B.C.
Canada V6T 1Z2

barlow@math.ubc.ca

CONTENTS

1. Introduction.

The notes are based on lectures given in St. Flour in 1995, and cover, in greater detail, most of the course given there.

The word "fractal" was coined by Mandelbrot [Man] in the 1970s, but of course sets of this type have been familiar for a long time – their early history being as a collection of pathological examples in analysis. There is no generally agreed exact definition of the word "fractal", and attempts so far to give a precise definition have been unsatisfactory, leading to classes of sets which are either too large, or too small, or both. This ambiguity is not a problem for this course: a more precise title would be "Diffusions on some classes of regular self-similar sets".

Initial interest in the properties of processes on fractals came from mathematical physicists working in the theory of disordered media. Certain media can be modelled by percolation clusters at criticality, which are expected to exhibit fractal-like properties. Following the initial papers [AO], [RT], [GAM1-GAM3] a very substantial physics literature has developed – see [HBA] for a survey and bibliography.

Let G be an infinite subgraph of $\mathbb{Z}^d$. A simple random walk (SRW) $(X_n, n \geq 0)$ on G is just the Markov chain which moves from $x \in G$ with equal probability to each of the neighbours of x. Write $p_n(x,y) = \mathbb{P}^x(X_n = y)$ for the n-step transition probabilities. If G is the whole of $\mathbb{Z}^d$ then $\mathbb{E}(X_n)^2 = n$ with many familiar consequences – the process moves roughly a distance of order $\sqrt{n}$ in time n, and the probability law $p_n(x,\cdot)$ puts most of its mass on a ball of radius $c_d n$.

If G is not the whole of $\mathbb{Z}^d$ then the movement of the process is on the average restricted by the removal of parts of the space. Probabilistically this is not obvious – but see [DS] for an elegant argument, using electrical resistance, that the removal of part of the state space can only make the process X 'more recurrent'. So it is not unreasonable to expect that for certain graphs G one may find that the process X is sufficiently restricted that for some $\beta > 2$

$$\mathbb{E}^x(X_n - x)^2 \asymp n^{2/\beta}. \qquad (1.1)$$

(Here and elsewhere I use $\asymp$ to mean 'bounded above and below by positive constants', so that (1.1) means that there exist constants c_1, c_2 such that $c_1 n^{2/\beta} \leq \mathbb{E}^x(X_n - x)^2 \leq c_2 n^{2/\beta}$). In [AO] and [RT] it was shown that if G is the Sierpinski gasket (or more precisely an infinite graph based on the Sierpinski gasket – see Fig. 1.1) then (1.1) holds with $\beta = \log 5/\log 2$.

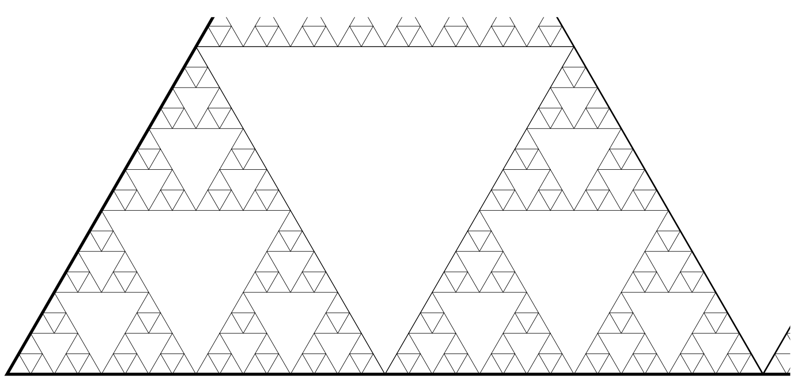

Figure 1.1: The graphical Sierpinski gasket.

Physicists call behaviour of this kind by a random walk (or a diffusion – they are not very interested in the distinction) *subdiffusive* – the process moves on average slower than a standard random walk on $\mathbb{Z}^d$. Kesten [Ke] proved that the SRW on the 'incipient infinite cluster' C (a percolation cluster at $p = p_c$ but conditioned to be infinite) is subdiffusive. The large scale structure of C is given by taking one infinite path (the 'backbone') together with a collection of 'dangling ends', some of which are very large. Kesten attributes the subdiffusive behaviour of SRW on C to the fact that the process X spends a substantial amount of time in the dangling ends.

However a graph such as the Sierpinski gasket (SG) has no dangling ends, and one is forced to search for a different explanation for the subdiffusivity. This can be found in terms of the existence of 'obstacles at all length scales'. Whilst this holds for the graphical Sierpinski gasket, the notation will be slightly simpler if we consider another example, the graphical Sierpinski carpet (GSC). (Figure 1.2).

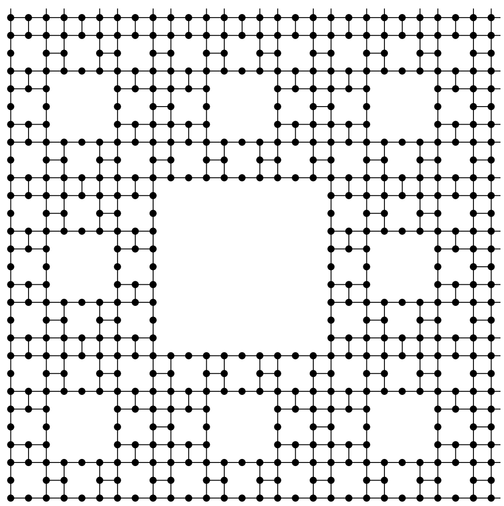

Figure 1.2: The graphical Sierpinski carpet.

This set can be defined precisely in the following fashion. Let $H_0 = \mathbb{Z}^2$. For $x = (n, m) \in H_0$ write n,m in ternary – so $n = \sum_{i=0}^{\infty} n_i 3^i$, where $n_i \in \{0, 1, 2\}$, and $n_i = 0$ for all but finitely many i. Set

$$J_k = \{(m, n) : \; n_k = 1 \text{ and } m_k = 1\},$$

so that J_k consists of a union of disjoint squares of side 3^k: the square in J_k closest to the origin is $\{3^k, \ldots, 2.3^k - 1\} \times \{3^k, \ldots, 2.3^k - 1\}$. Now set

$$H_n = H_0 - \bigcup_{k=1}^{n} J_k, \quad H = \bigcap_{n=0}^{\infty} H_n. \tag{1.2}$$

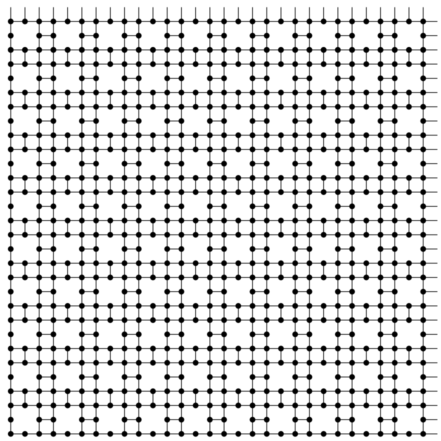

Figure 1.3: The set H_1 .

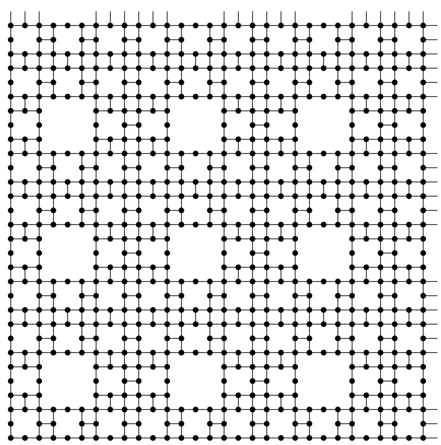

Figure 1.4: The set H_2 .

Note that $H \cap [0, 3^n]^2 = H_n \cap [0, 3^n]^2$, so that the difference between H and H_n will only be detected by a SRW after it has moved a distance of 3^n from the origin. Now let $X^{(n)}$ be a SRW on H_n, started at the origin, and let X be a SRW on H. The process $X^{(0)}$ is just SRW on $\mathbb{Z}_+^2$ and so we have

$$\mathbb{E}(X_n^{(0)})^2 \simeq n. \tag{1.3}$$

The process $X^{(1)}$ is a random walk on a the intersection of a translation invariant subset of $\mathbb{Z}^2$ with $\mathbb{Z}_+^2$. So we expect 'homogenization': the processes $n^{-1/2} X_{[nt]}^{(1)}$, $t \geq 0$ should converge weakly to a constant multiple of Brownian motion in $\mathbb{R}_+^2$. So, for large n we should have $\mathbb{E}(X_n^{(1)})^2 \sim a_1 n$, and we would expect that $a_1 < 1$, since the obstacles will on average tend to impede the motion of the process.

Similar considerations suggest that, writing $\varphi_n(t) = \mathbb{E}^0 (X_t^{(n)})^2$, we should have

$$\varphi_n(t) \sim a_n t \quad \text{ as } t \to \infty.$$

However, for small t we would expect that φ_n and φ_{n+1} should be approximately equal, since the process will not have moved far enough to detect the difference between H_n and H_{n+1}. More precisely, if t_n is such that $\varphi_n(t_n) = (3^n)^2$ then φ_n

and φ_{n+1} should be approximately equal on $[0, t_{n+1}]$. So we may guess that the behaviour of the family of functions $\varphi_n(t)$ should be roughly as follows:

$$\begin{aligned} \varphi_n(t) &= b_n + a_n(t - t_n), \quad t \geq t_n, \\ \varphi_{n+1}(s) &= \varphi_n(s), \quad 0 \leq s \leq t_{n+1}. \end{aligned} \tag{1.4}$$

If we add the guess that $a_n = 3^{-\alpha}$ for some $\alpha > 0$ then solving the equations above we deduce that

$$t_n \asymp 3^{(2+\alpha)n}, \quad b_n = 3^{2n}.$$

So if $\varphi(t) = \mathbb{E}^0(X_t)^2$ then as $\varphi(t) \simeq \lim_n \varphi_n(t)$ we deduce that φ is close to a piecewise linear function, and that

$$\varphi(t) \asymp t^{2/\beta}$$

where $\beta = 2 + \alpha$. Thus the random walk X on the graph H should satisfy (1.1) for some $\beta > 2$.

The argument given here is not of course rigorous, but (1.1) does actually hold for the set H – see [BB6, BB7]. (See also [Jo] for the case of the graphical Sierpinski gasket. The proofs however run along rather different lines than the heuristic argument sketched above).

Given behaviour of this type it is natural to ask if the random walk X on H has a scaling limit. More precisely, does there exist a sequence of constants τ_n such that the processes

$$(3^{-n} X_{[t/\tau_n]}, t \geq 0) \tag{1.5}$$

converge weakly to a non-degenerate limit as $n \to \infty$? For the graphical Sierpinski carpet the convergence is not known, though there exist τ_n such that the family (1.5) is tight. However, for the graphical Sierpinski gasket the answer is 'yes'.

Thus, for certain very regular fractal sets $F \subset \mathbb{R}^d$ we are able to define a limiting diffusion process $X = (X_t, t \geq 0, \mathbb{P}^x, x \in F)$ where $\mathbb{P}^x$ is for each $x \in F$ a probability measure on $\Omega = \{\omega \in C([0,\infty), F) : \omega(0) = x\}$. Writing $T_t f(x) = \mathbb{E}^x f(X_t)$ for the semigroup of X we can define a 'differential' operator $\mathcal{L}_F$, defined on a class of functions $\mathcal{D}(\mathcal{L}_F) \subset C(F)$. In many cases it is reasonable to call $\mathcal{L}_F$ the Laplacian on F.

From the process X one is able to obtain information about the solutions to the Laplace and heat equations associated with $\mathcal{L}_F$, the heat equation for example taking the form

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mathcal{L}_F u, \\ u(0,x) &= u_0(x), \end{aligned} \tag{1.6}$$

where $u = u(t,x)$, $x \in F$, $t \geq 0$. The wave equation is rather harder, since it is not very susceptible to probabilistic analysis. See, however [KZ2] for work on the wave equation on a some manifolds with a 'large scale fractal structure'.

The mathematical literature on diffusions on fractals and their associated infinitesimal generators can be divided into broadly three parts:

1. Diffusions on finitely ramified fractals.
2. Diffusions on generalized Sierpinski carpets, a family of infinitely ramified fractals.
3. Spectral properties of the 'Laplacian' $\mathcal{L}_F$.

These notes only deal with the first of these topics. On the whole, infinitely ramified fractals are significantly harder than finitely ramified ones, and sometimes require a very different approach. See [Bas] for a recent survey.

These notes also contain very little on spectral questions. For finitely ramified fractals a direct approach (see for example [FS1, Sh1-Sh4, KL]), is simpler, and gives more precise information than the heat kernel method based on estimating

$$\int_F p(t,x,x)dx = \sum_i e^{-\lambda_i t}.$$

In this course Section 2 introduces the simplest case, the Sierpinski gasket. In Section 3 I define a class of well-behaved diffusions on metric spaces, "Fractional Diffusions", which is wide enough to include many of the processes discussed in this course. It is possible to develop their properties in a fairly general fashion, without using much of the special structure of the state space. Section 4 contains a brief introduction to the theory of Dirichlet forms, and also its connection with electrical resistances. The remaining chapters, 5 to 8, give the construction and some properties of diffusions on a class of finitely ramified regular fractals. In this I have largely followed the analytic 'Japanese' approach, developed by Kusuoka, Kigami, Fukushima and others. Many things can now be done more simply than in the early probabilistic work – but there is loss as well as gain in added generality, and it is worth pointing out that the early papers on the Sierpinski gasket ([Kus1, Go, BP]) contain a wealth of interesting direct calculations, which are not reproduced in these notes. Any reader who is surprised by the abrupt end of these notes in Section 8 should recall that some, at least, of the properties of these processes have already been obtained in Section 3.

c_i denotes a positive real constant whose value is fixed within each Lemma, Theorem etc. Occasionally it will be necessary to use notation such as $c_{3.5.4}$ – this is simply the constant c_4 in Definition 3.5. c, c', c'' denote positive real constants whose values may change on each appearance. $B(x,r)$ denotes the open ball with centre x and radius r, and if X is a process on a metric space F then

$$\begin{aligned} T_A &= \inf\{t > 0 : X_t \in A\}, \\ T_y &= \inf\{t > 0 : X_t = y\}, \\ \tau(x,r) &= \inf\{t \geq 0 : X_t \notin B(x,r)\}. \end{aligned}$$

I have included in the references most of the mathematical papers in this area known to me, and so they contain many papers not mentioned in the text. I am grateful to Gerard Ben Arous for a number of interesting conversations on the physical conditions under which subdiffusive behaviour might arise, to Ben Hambly

for checking the final manuscript, and to Ann Artuso and Liz Rowley for their typing.

Acknowledgements. This research is supported by a NSERC (Canada) research grant, by a grant from the Killam Foundation, and by a EPSRC (UK) Visiting Fellowship.

2. The Sierpinski Gasket

This is the simplest non-trivial connected symmetric fractal. The set was first defined by Sierpinski [Sie1], as an example of a pathological curve; the name "Sierpinski gasket" is due to Mandelbrot [Man, p.142].

Let $G_0 = \{(0,0), (1,0), (1/2, \sqrt{3}/2)\} = \{a_0, a_1, a_2\}$ be the vertices of the unit triangle in $\mathbb{R}^2$, and let $\mathcal{H}u(G_0) = H_0$ be the closed convex hull of G_0. The construction of the Sierpinski gasket (SG for short) G is by the following Cantor-type subtraction procedure. Let b_0, b_1, b_2 be the midpoints of the 3 sides of G_0, and let A be the interior of the triangle with vertices $\{b_0, b_1, b_2\}$. Let $H_1 = H_0 - A$, so that H_1 consists of 3 closed upward facing triangles, each of side 2^{-1}. Now repeat the operation on each of these triangles to obtain a set H_2, consisting of 9 upward facing triangles, each of side 2^{-2}.

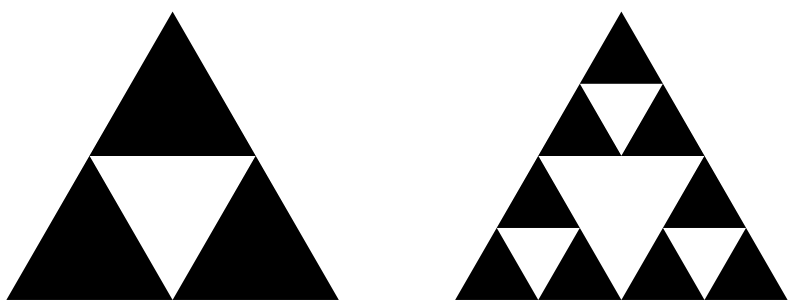

Figure 2.1: The sets H_1 and H_2.

Continuing in this fashion, we obtain a decreasing sequence of closed non-empty sets $(H_n)_{n=0}^{\infty}$, and set

$$G = \bigcap_{n=0}^{\infty} H_n. \tag{2.1}$$

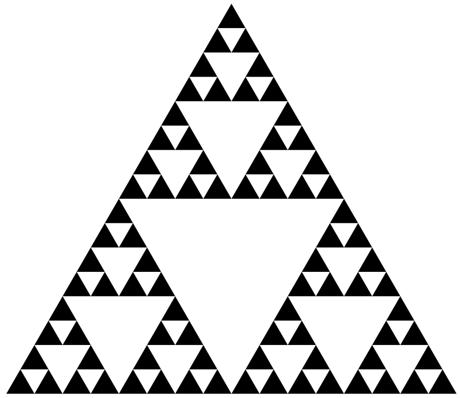

Figure 2.2: The set H_4.

It is easy to see that G is connected: just note that $\partial H_n \subset H_m$ for all $m \geq n$, so that no point on the edge of a triangle is ever removed. Since $|H_n| = (3/4)^n |H_0|$, we clearly have that $|G| = 0$.

We begin by exploring some geometrical properties of G. Call an *n-triangle* a set of the form $G \cap B$, where B is one of the 3^n triangles of side 2^{-n} which make up H_n. Let μ_n be Lebesgue measure restricted to H_n, and normalized so that $\mu_n(H_n) = 1$; that is

$$\mu_n(dx) = 2 \cdot (4/3)^n 1_{H_n}(x)\, dx.$$

Let $\mu_G = \text{wlim}\mu_n$; this is the natural "flat" measure on G. Note that μ_G is the unique measure on G which assigns mass 3^{-n} to each n-triangle. Set $d_f = \log 3/\log 2 \simeq 1.58\ldots$

Lemma 2.1. *For $x \in G$, $0 \leq r < 1$*

$$3^{-1} r^{d_f} \leq \mu_G\big(B(x,r)\big) \leq 18 r^{d_f}. \tag{2.2}$$

Proof. The result is clear if $r = 0$. If $r > 0$, choose n so that $2^{-(n+1)} < r \leq 2^{-n}$ – we have $n \geq 0$. Since $B(x,r)$ can intersect at most 6 n-triangles, it follows that

$$\begin{aligned}\mu_G\big(B(x,r)\big) &\leq 6.3^{-n} = 18.3^{-(n+1)}\\ &= 18(2^{-(n+1)})^{d_f} < 18 r^{d_f}.\end{aligned}$$

As each $(n+1)$-triangle has diameter $2^{-(n+1)}$, $B(x,r)$ must contain at least one $(n+1)$-triangle and therefore

$$\mu_G\big(B(x,r)\big) \geq 3^{-(n+1)} = 3^{-1}(2^{-n})^{d_f} \geq 3^{-1} r^{d_f}. \qquad \square$$

Of course the constants 3^{-1}, 18 in (2.2) are not important; what is significant is that the μ_G-mass of balls in G grow as r^{d_f}. Using terminology from the geometry of manifolds, we can say that G has *volume growth* given by r^{d_f}.

Detour on Dimension.

Let (F,ρ) be a metric space. There are a number of different definitions of dimension for F and subsets of F: here I just mention a few. The simplest of these is *box–counting dimension.* For $\varepsilon > 0$, $A \subset F$, let $N(A,\varepsilon)$ be the smallest number of balls $B(x,\varepsilon)$ required to cover A. Then

$$\dim_{BC}(A) = \limsup_{\varepsilon\downarrow 0} \frac{\log N(A,\varepsilon)}{\log \varepsilon^{-1}}. \tag{2.3}$$

To see how this behaves, consider some examples. We take (F,ρ) to be $\mathbb{R}^d$ with the Euclidean metric.

Examples. 1. Let $A = [0,1]^d \subset \mathbb{R}^d$. Then $N(A,\varepsilon) \asymp \varepsilon^{-d}$, and it is easy to verify that

$$\lim_{\varepsilon\downarrow 0} \frac{\log N([0,1]^d,\varepsilon)}{\log \varepsilon^{-1}} = d.$$

2. The Sierpinski gasket G. Since $G \subset H_n$, and H_n is covered by 3^n triangles of side 2^{-n}, we have, after some calculations similar to those in Lemma 2.1, that $N(G,r) \asymp (1/r)^{\log 3/\log 2}$. So,

$$\dim_{BC}(G) = \frac{\log 3}{\log 2}.$$

3. Let $A = \mathbb{Q} \cap [0,1]$. Then $N(A,\varepsilon) \asymp \varepsilon^{-1}$, so $\dim_{BC}(A) = 1$. On the other hand $\dim_{BC}(\{p\}) = 0$ for any $p \in A$.

We see that box-counting gives reasonable answers in the first two cases, but a less useful number in the third. A more delicate, but more useful, definition is obtained if we allow the sizes of the covering balls to vary. This gives us *Hausdorff dimension.* I will only sketch some properties of this here – for more detail see for example the books by Falconer [Fa1, Fa2].

Let $h : \mathbb{R}_+ \to \mathbb{R}_+$ be continuous, increasing, with $h(0) = 0$. For $U \subset F$ write $\text{diam}(U) = \sup\{\rho(x,y) : x,y \in U\}$ for the diameter of U. For $\delta > 0$ let

$$\mathcal{H}^h_\delta(A) = \inf\Big\{\sum_i h\big(d(U_i)\big) : A \subset \bigcup_i U_i, \quad \text{diam}(U_i) < \delta\Big\}.$$

Clearly $\mathcal{H}^h_\delta(A)$ is decreasing in δ. Now let

$$\mathcal{H}^h(A) = \lim_{\delta\downarrow 0} \mathcal{H}^h_\delta(A); \tag{2.4}$$

we call $\mathcal{H}^h(\cdot)$ *Hausdorff h-measure* . Let $\mathcal{B}(F)$ be the Borel σ-field of F.

Lemma 2.2. *$\mathcal{H}^h$ is a measure on $\big(F, \mathcal{B}(F)\big)$.*

For a proof see [Fa1, Chapter 1].

We will be concerned only with the case $h(x) = x^\alpha$: we then write $\mathcal{H}^\alpha$ for $\mathcal{H}^h$. Note that $\alpha \to \mathcal{H}^\alpha(A)$ is decreasing; in fact it is not hard to see that $\mathcal{H}^\alpha(A)$ is either $+\infty$ or 0 for all but at most one α.

Definition 2.3. The Hausdorff dimension of A is defined by

$$\dim_H(A) = \inf\{\alpha : \mathcal{H}^\alpha(A) = 0\} = \sup\{\alpha : \mathcal{H}^\alpha(A) = +\infty\}.$$

Lemma 2.4. $\dim_H(A) \le \dim_{BC}(A)$.

Proof. Let $\alpha > \dim_{BC}(A)$. Then as A can be covered by $N(A,\varepsilon)$ sets of diameter 2ε, we have $\mathcal{H}^\alpha_\delta(A) \le N(A,\varepsilon)(2\varepsilon)^\alpha$ whenever $2\varepsilon < \delta$. Choose θ so that $\dim_{BC}(A) < \alpha - \theta < \alpha$; then (2.3) implies that for all sufficiently small ε, $N(A,\varepsilon) \le \varepsilon^{-(\alpha-\theta)}$. So $\mathcal{H}^\alpha_\delta(A) = 0$, and thus $\mathcal{H}^\alpha(A) = 0$, which implies that $\dim_H(A) \le \alpha$. □

Consider the set $A = \mathbb{Q} \cap [0,1]$, and let $A = \{p_1, p_2, \ldots\}$ be an enumeration of A. Let $\delta > 0$, and U_i be an open internal of length $2^{-i} \wedge \delta$ containing p_i. Then (U_i) covers A, so that $\mathcal{H}^\alpha_\delta(A) \le \sum_{i=1}^\infty \left(\delta \wedge 2^{-i}\right)^\alpha$, and thus $\mathcal{H}^\alpha(A) = 0$. So $\dim_H(A) = 0$. We see therefore that $\dim_H$ can be strictly smaller than $\dim_{BC}$, and that (in this case at least) $\dim_H$ gives a more satisfactory measure of the size of A.

For the other two examples considered above Lemma 2.4 gives the upper bounds $\dim_H([0,1]^d) \le d$, $\dim_H(G) \le \log 3/\log 2$. In both cases equality holds, but a direct proof of this (which is possible) encounters the difficulty that to obtain a lower bound on $\mathcal{H}^\alpha_\delta(A)$ we need to consider all possible covers of A by sets of diameter less than δ. It is much easier to use a kind of dual approach using measures.

Theorem 2.5. *Let μ be a measure on A such that $\mu(A) > 0$ and there exist $c_1 < \infty$, $r_0 > 0$, such that*

$$\mu\big(B(x,r)\big) \le c_1 r^\alpha, \quad x \in A, \quad r \le r_0. \tag{2.5}$$

Then $\mathcal{H}^\alpha(A) \ge c_1^{-1}\mu(A)$, and $\dim_H(A) \ge \alpha$.

Proof. Let U_i be a covering of A by sets of diameter less than δ, where $2\delta < r_0$. If $x_i \in U_i$, then $U_i \subset B\big(x_i, \operatorname{diam}(U_i)\big)$, so that $\mu(U_i) \le c_1 \operatorname{diam}(U_i)^\alpha$. So

$$\sum_i \operatorname{diam}(U_i)^\alpha \ge c_1^{-1} \sum_i \mu(U_i) \ge c_1^{-1}\mu(A).$$

Therefore $\mathcal{H}^\alpha_\delta(A) \ge c_1^{-1}\mu(A)$, and it follows immediately that $\mathcal{H}^\alpha(A) > 0$, and $\dim_H(A) \ge \alpha$. □

Corollary 2.6. $\dim_H(G) = \log 3/\log 2$.

Proof. By Lemma 2.1 μ_G satisfies (2.5) with $\alpha = d_f$. So by Theorem 2.5 $\dim_H(G) \ge d_f$; the other bound has already been proved. □

Very frequently, when we wish to compute the dimension of a set, it is fairly easy to find directly a near-optimal covering, and so obtain an upper bound on $\dim_H$ directly. We can then use Theorem 2.5 to obtain a lower bound. However, we can also use measures to derive upper bounds on $\dim_H$.

Theorem 2.7. *Let μ be a finite measure on A such that $\mu\big(B(x,r)\big) \ge c_2 r^\alpha$ for all $x \in A$, $r \le r_0$. Then $\mathcal{H}^\alpha(A) < \infty$, and $\dim_H(A) \le \alpha$.*

Proof. See [Fa2, p.61].

In particular we may note:

Corollary 2.8. *If μ is a measure on A with $\mu(A) \in (0,\infty)$ and*

$$c_1 r^\alpha \le \mu\big(B(x,r)\big) \le c_2 r^\alpha, \quad x \in A, \quad r \le r_0 \tag{2.6}$$

then $\mathcal{H}^\alpha(A) \in (0,\infty)$ and $\dim_H(A) = \alpha$.

Remarks. 1. If A is a k-dimensional subspace of $\mathbb{R}^d$ then $\dim_H(A) = \dim_{BC}(A) = k$.
2. Unlike $\dim_{BC}$ $\dim_H$ is stable under countable unions: thus

$$\dim_H\Big(\bigcup_{i=1}^\infty A_i\Big) = \sup_i \dim_H(A_i).$$

3. In [Tri] Tricot defined "packing dimension" $\dim_P(\cdot)$, which is the largest reasonable definition of "dimension" for a set. One has $\dim_P(A) \ge \dim_H(A)$; strict inequality can hold. The hypotheses of Corollary 2.8 also imply that $\dim_P(A) = \alpha$. See [Fa2, p.48].
4. The sets we consider in these notes will be quite regular, and will very often satisfy (2.6): that is they will be "α-dimensional" in every reasonable sense.
5. Questions concerning Hausdorff measure are frequently much more delicate than those relating just to dimension. However, the fractals considered in this notes will all be sufficiently regular so that there is a direct construction of the Hausdorff measure. For example, the measure μ_G on the Sierpinski gasket is a constant multiple of the Hausdorff x^{d_f}-measure on G.

We note here how $\dim_H$ changes under a change of metric.

Theorem 2.9. *Let ρ_1, ρ_2 be metrics on F, and write $\mathcal{H}^{\alpha,i}$, $\dim_{H,i}$ for the Hausdorff measure and dimension with respect to ρ_i, $i = 1,2$.*
(a) If $\rho_1(x,y) \le \rho_2(x,y)$ for all $x,y \in A$ with $\rho_2(x,y) \le \delta_0$, then $\dim_{H,1}(A) \ge \dim_{H,2}(A)$.
(b) If $1 \wedge \rho_1(x,y) \asymp (1 \wedge \rho_2(x,y))^\theta$ for some $\theta > 0$, then

$$\dim_{H,2}(A) = \theta \dim_{H,1}(A).$$

Proof. Write $d_j(U)$ for the ρ_j-diameter of U. If (U_i) is a cover of A by sets with $\rho_2(U_i) < \delta < \delta_0$, then

$$\sum_i d_1(U_i)^\alpha \le \sum_i d_2(U_i)^\alpha$$

so that $\mathcal{H}_\delta^{\alpha,1}(A) \le \mathcal{H}_\delta^{\alpha,2}(A)$. Then $\mathcal{H}^{\alpha,1}(A) \le \mathcal{H}^{\alpha,2}(A)$ and $\dim_{H,1}(A) \ge \dim_{H,2}(A)$, proving (a).

(b) If U_i is any cover of A by sets of small diameter, we have

$$\sum_i d_1(U_i)^\alpha \asymp \sum_i d_2(U_i)^{\theta\alpha}.$$

Hence $\mathcal{H}^{\alpha,1}(A) = 0$ if and only if $\mathcal{H}^{\theta\alpha,2}(A) = 0$, and the conclusion follows. □

Metrics on the Sierpinski gasket.

Since we will be studying continuous processes on G, it is natural to consider the metric on G given by the shortest path in G between two points. We begin with a general definition.

Definition 2.10. Let $A \subset \mathbb{R}^d$. For $x, y \in A$ set

$$d_A(x,y) = \inf\{|\gamma| : \gamma \text{ is a path between } x \text{ and } y \text{ and } \gamma \subset A\}.$$

If $d_A(x,y) < \infty$ for all x, $y \in A$ we call d_A the *geodesic metric* on A.

Lemma 2.11. *Suppose A is closed, and that $d_A(x,y) < \infty$ for all x, $y \in A$. Then d_A is a metric on A and (A, d_A) has the geodesic property:*

For each $x, y \in A$ there exists a map $\Phi(t) : [0,1] \to A$ such that
$d_A(x, \Phi(t)) = t d_A(x,y), \quad d_A(\Phi(t), y) = (1-t) d_A(x,y).$

Proof. It is clear that d_A is a metric on A. To prove the geodesic property, let $x, y \in A$, and $D = d_A(x,y)$. Then for each $n \geq 1$ there exists a path $\gamma_n(t)$, $0 \leq t \leq 1 + D$ such that $\gamma_n \subset A$, $|d\gamma_n(t)| = dt$, $\gamma_n(0) = x$ and $\gamma_n(t_n) = y$ for some $D \leq t_n \leq D + n^{-1}$. If $p \in [0, D] \cap \mathbb{Q}$ then since $|x - \gamma_n(p)| \leq p$ the sequence $(\gamma_n(p))$ has a convergent subsequence. By a diagonalization argument there exists a subsequence n_k such that $\gamma_{n_k}(p)$ converges for each $p \in [0, D] \cap \mathbb{Q}$; we can take $\Phi = \lim \gamma_{n_k}$. □

Lemma 2.12. *For $x, y \in G$,*

$$|x - y| \leq d_G(x,y) \leq c_1 |x - y|.$$

Proof. The left hand inequality is evident.

It is clear from the structure of H_n that if A, B are n-triangles and $A \cap B = \emptyset$, then

$$|a - b| \geq (\sqrt{3}/2) 2^{-n} \qquad \text{for } a \in A,\ b \in B.$$

Let x, $y \in G$ and choose n so that

$$(\sqrt{3}/2) 2^{-(n+1)} \leq |x - y| < (\sqrt{3}/2) 2^{-n}.$$

So x, y are either in the same n-triangle, or in adjacent n-triangles. In either case choose $z \in G_n$ so it is in the same n-triangle as both x and y.

Let $z_n = z$, and for $k > n$ choose $z_k \in G_k$ such that x, z_k are in the same k-triangle. Then since z_k and z_{k+1} are in the same k-triangle, and both are contained in H_{k+1}, we have $d_G(z_k, z_{k+1}) = d_{H_{k+1}}(z_k, z_{k+1}) \le 2^{-k}$. So,

$$d_G(z, x) \le \sum_{k=n}^{\infty} d_G(z_k, z_{k+1}) \le 2^{1-n} \le 4|x - y|.$$

Hence $d_G(x, y) \le d_G(x, z) + d_G(z, y) \le 8|x - y|$. □

Construction of a diffusion on the Sierpinski gasket.

Let G_n be the set of vertices of n-triangles. We can make G_n into a graph in a natural way, by taking $\{x, y\}$ to be an edge in G_n if x, y belong to the same n-triangle. (See Fig. 2.3). Write E_n for the set of edges.

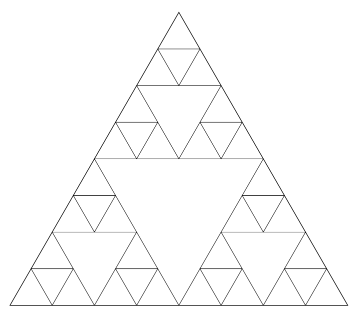

Figure 2.3: The graph G_3.

Let $Y_k^{(n)}$, $k = 0, 1, \ldots$ be a simple random walk on G_n. Thus from $x \in G_n$, the process $Y^{(n)}$ jumps to each of the neighbours of x with equal probability. (Apart from the 3 points in G_0, all the points in G_n have 4 neighbours). The obvious way to construct a diffusion process $(X_t, t \ge 0)$ on G is to use the graphs G_n, which provide a natural approximation to G, and to try to define X as a weak limit of the processes $Y^{(n)}$. More precisely, we wish to find constants $(\alpha_n, n \ge 0)$ such that

$$\left(Y^{(n)}_{[\alpha_n t]}, t \ge 0\right) \Rightarrow (X_t, t \ge 0)\,. \tag{2.7}$$

We have two problems:

(1) How do we find the right (α_n)?
(2) How do we prove convergence?

We need some more notation.

Definition 2.13. Let $\mathcal{S}_n$ be the collection of sets of the form $G \cap A$, where A is an n-triangle. We call the elements of $\mathcal{S}_n$ *n-complexes.* For $x \in G_n$ let $D_n(x) = \bigcup\{S \in \mathcal{S}_n : x \in S\}$.

The key properties of the SG which we use are, first that it is very symmetric, and secondly, that it is finitely ramified. (In general, a set A in a metric space F is finitely ramified if there exists a finite set B such that $A-B$ is not connected). For the SG, we see that each n-complex A is disconnected from the rest of the set if we remove the set of its corners, that is $A \cap G_n$.

The following is the key observation. Suppose $Y_0^{(n)} = y \in G_{n-1}$ (take $y \notin G_0$ for simplicity), and let $T = \inf\{k > 0 : Y_k^{(n)} \in G_{n-1} - \{y\}\}$. Then $Y^{(n)}$ can only escape from $D_{n-1}(y)$ at one of the 4 points, $\{x_1, \dots, x_4\}$ say, which are neighbours of y in the graph (G_{n-1}, E_{n-1}). Therefore $Y_T^{(n)} \in \{x_1, \dots, x_4\}$. Further the symmetry of the set $G_n \cap D_n(y)$ means that each of the events $\{Y_T^{(n)} = x_i\}$ is equally likely.

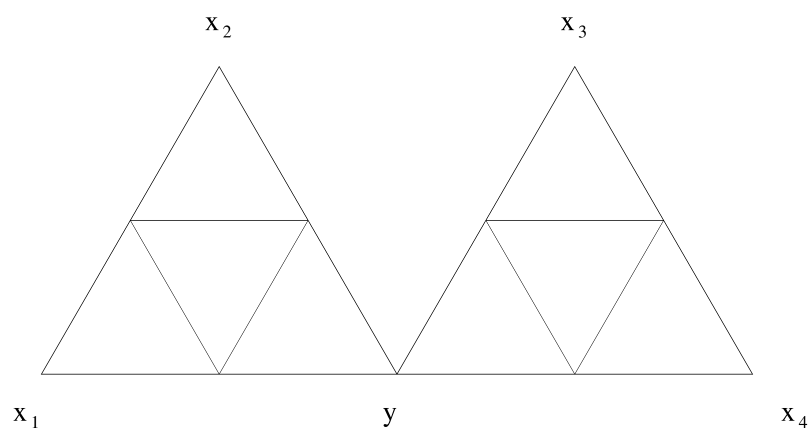

Figure 2.4: y and its neighbours.

Thus

$$\mathbb{P}\left(Y_T^{(n)} = x_i \;\middle|\; Y_0^{(n)} = y\right) = \tfrac{1}{4},$$

and this is also equal to $\mathbb{P}(Y_1^{(n-1)} = x_i | Y_0^{(n-1)} = y)$. (Exactly the same argument applies if $y \in G_0$, except that we then have only 2 neighbours instead of 4). It follows that $Y^{(n)}$ looked at at its visits to G_{n-1} behaves exactly like $Y^{(n-1)}$. To state this precisely, we first make a general definition.

Definition 2.14. Let $\mathbb{T} = \mathbb{R}_+$ or $\mathbb{Z}_+$, let $(Z_t, t \in \mathbb{T})$ be a cadlag process on a metric space F, and let $A \subset F$ be a discrete set. Then *successive disjoint hits* by Z on A are the stopping times $T_0, T_1, \dots$ defined by

$$\begin{aligned} T_0 &= \inf\{t \geq 0 : Z_t \in A\}, \\ T_{n+1} &= \inf\big\{t > T_n : Z_t \in A - \{Z_{T_n}\}\big\}, \qquad n \geq 0. \end{aligned} \tag{2.8}$$

With this notation, we can summarize the observations above.

Lemma 2.15. *Let $(T_i)_{i \geq 0}$ be successive disjoint hits by $Y^{(n)}$ on G_{n-1}. Then $(Y_{T_i}^{(n)}, i \geq 0)$ is a simple random walk on G_{n-1} and is therefore equal in law to $(Y_i^{(n-1)}, i \geq 0)$.*

Using this, it is clear that we can build a sequence of "nested" random walks on G_n. Let $N \geq 0$, and let $Y_k^{(N)}$, $k \geq 0$ be a SRW on G_N with $Y_0^{(N)} = 0$. Let $0 \leq m \leq N-1$ and $(T_i^{N,m})_{i \geq 0}$ be successive disjoint hits by $Y^{(N)}$ on G_m, and set

$$Y_i^{(m)} = Y^{(N)}(T_i^{N,m}) = Y_{T_i^{N,m}}^{(N)}, \qquad i \geq 0.$$

It follows from Lemma 2.15 that $Y^{(m)}$ is a SRW on G_m, and for each $0 \le n \le m \le N$ we have that $Y^{(m)}$, sampled at its successive disjoint hits on G_n, equals $Y^{(n)}$.

We now wish to construct a sequence of SRWs with this property holding for $0 \le n \le m < \infty$. This can be done, either by using the Kolmogorov extension theorem, or directly, by building $Y^{(N+1)}$ from $Y^{(N)}$ with a sequence of independent "excursions". The argument in either case is not hard, and I omit it.

Thus we can construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, carrying random variables $(Y_k^{(n)}, n \ge 0, k \ge 0)$ such that

(a) For each n, $(Y_k^{(n)}, k \ge 0)$ is a SRW on G_n starting at 0.
(b) Let $T_i^{n,m}$ be successive disjoint hits by $Y^{(n)}$ on G_m. (Here $m \le n$). Then

$$Y^{(n)}(T_i^{n,m}) = Y_i^{(m)}, \quad i \ge 0, \quad m \le n. \tag{2.9}$$

If we just consider the paths of the processes $Y^{(n)}$ in G, we see that we are viewing successive discrete approximations to a continuous path. However, to define a limiting process we need to rescale time, as was suggested by (2.7).

Write $\tau = T_1^{1,0} = \min\{k \ge 0 : |Y_k^{(1)}| = 1\}$, and set $f(s) = \mathbb{E}\, s^\tau$, for $s \in [0,1]$.

Lemma 2.16. *$f(s) = s^2/(4-3s)$, $\mathbb{E}\tau = f'(1) = 5$, and $\mathbb{E}\tau^k < \infty$ for all k.*

Proof. This is a simple exercise in finite state Markov chains. Let a_1, a_2 be the two non-zero elements of G_0, let $b = \frac{1}{2}(a_1 + a_2)$, and $c_i = \frac{1}{2}a_i$. Writing $f_c(s) = \mathbb{E}^{c_i} s^\tau$, and defining f_b, f_a similarly, we have $f_a(s) = 1$,

$$\begin{aligned} f(s) &= s f_c(s), \\ f_c(s) &= \tfrac{1}{4} s\big(f(s) + f_c(s) + f_b(s) + f_a(s)\big), \\ f_b(s) &= \tfrac{1}{2} s\big(f_c(s) + f_a(s)\big), \end{aligned}$$

and solving these equations we obtain $f(s)$.

The remaining assertions follow easily from this. □

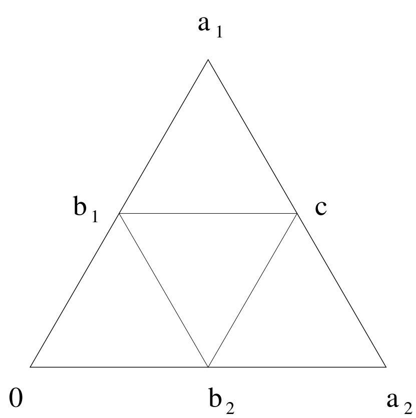

Figure 2.5: The graph G_1.

Now let $Z_n = T_1^{n,0}$, $n \geq 0$. The nesting property of the random walks $Y^{(n)}$ implies that Z_n is a simple branching process, with offspring distribution (p_n), where

$$f(s) = \sum_{k=2}^{\infty} s^k p_k. \tag{2.10}$$

To see this, note that $Y_k^{(n+1)}$, for $T_i^{n+1,n} \leq k \leq T_{i+1}^{n+1,n}$ is a SRW on $G_{n+1} \cap D_n(Y_i^{(n)})$, and that therefore $T_{i+1}^{n+1,n} - T_i^{n+1,n} \overset{(d)}{=} \tau$. Also, by the Markov property, the r.v. $\xi_i = T_{i+1}^{n+1,n} - T_i^{n+1,n}$, $i \geq 0$, are independent. Since

$$Z_{n+1} = \sum_{i=0}^{Z_n - 1} \xi_i,$$

(Z_n) is a branching process.

As $E\tau^2 < \infty$, and $\mathbb{E}\tau = 5$, the convergence theorem for simple branching processes implies that

$$5^{-n} Z_n \xrightarrow{a.s.} W$$

for some strictly positive r.v. W. (See [Har, p. 13]). The convergence is easy using a martingale argument: proving that $W > 0$ a.s. takes a little more work. (See [Har, p. 15]). In addition, if

$$\varphi(u) = Ee^{-uW}$$

then φ satisfies the functional equation

$$\varphi(5u) = f\big(\varphi(u)\big), \quad \varphi'(0) = -1. \tag{2.11}$$

We have a similar result in general.

Proposition 2.17. *Fix $m \geq 0$. The processes*

$$Z_n^{(i)} = T_i^{n,m} - T_{i-1}^{n,m}, \quad n \geq m$$

are branching processes with offspring distribution τ, and $Z^{(i)}$ are independent. Thus there exist $W_i^{(m)}$ such that for each m $(W_i^{(m)}, i \geq 0)$ are independent, $W_i^{(m)} \overset{(d)}{=} 5^{-m} W$, and

$$5^{-n} \left(T_i^{n,m} - T_{i-1}^{n,m}\right) \to W_i^{(m)} \text{ a.s.}$$

Note in particular that $\mathbb{E}(T_1^{n,0}) = 5^n$, that is that the mean time taken by $Y^{(n)}$ to cross G_n is 5^n. In terms of the graph distance on G_n we have therefore that $Y^{(n)}$ requires roughly 5^n steps to move a distance 2^n; this may be compared with the corresponding result for a simple random walk on $\mathbb{Z}^d$, which requires roughly 4^n steps to move a distance 2^n.

The slower movement of $Y^{(n)}$ is not surprising — to leave $G_n \cap B(0, 1/2)$, for example, it has to find one of the two 'gateways' $(1/2, 0)$ or $(1/4, \sqrt{3}/4)$. Thus the movement of $Y^{(n)}$ is impeded by a succession of obstacles of different sizes, which act to slow down its diffusion.

Given the space-time scaling of $Y^{(n)}$ it is no surprise that we should take $\alpha_n = 5^n$ in (2.7). Define

$$X_t^n = Y^{(n)}_{[5^n t]}, \quad t \geq 0.$$

In view of the fact that we have built the $Y^{(n)}$ with the nesting property, we can replace the weak convergence of (2.7) with a.s. convergence.

Theorem 2.18. *The processes X^n converge a.s., and uniformly on compact intervals, to a process X_t, $t \geq 0$. X is continuous, and $X_t \in G$ for all $t \geq 0$.*

Proof. For simplicity we will use the fact that W has a non-atomic distribution function. Fix for now $m > 0$. Let $t > 0$. Then, a.s., there exists $i = i(\omega)$ such that

$$\sum_{j=1}^{i} W_j^{(m)} < t < \sum_{j=1}^{i+1} W_j^{(m)}.$$

As $W_j^{(m)} = \lim_{n\to\infty} 5^{-n}\left(T_j^{n,m} - T_{j-1}^{n,m}\right)$ it follows that for $n \geq n_0(\omega)$,

$$T_i^{n,m} < 5^n t < T_{i+1}^{n,m}. \tag{2.12}$$

Now $Y^{(n)}(T_i^{n,m}) = Y_i^{(m)}$ by (2.9). Since $Y_k^{(n)} \in D_m(Y_i^{(m)})$ for $T_i^{n,m} \leq k \leq T_{i+1}^{n,m}$, we have

$$|Y^{(n)}_{[5^n t]} - Y_i^{(m)}| \leq 2^{-m} \quad \text{for all } n \geq n_0.$$

This implies that $|X_t^n - X_t^{n'}| \leq 2^{-m+1}$ for n, $n' \geq n_0$, so that X_t^n is Cauchy, and converges to a r.v. X_t. Since $X_t^n \in G_n$, we have $X_t \in G$.

With a little extra work, one can prove that the convergence is uniform in t, on compact time intervals. I give here a sketch of the argument. Let $a \in \mathbb{N}$, and let

$$\xi_m = \min_{1\leq i\leq a5^m} W_i^{(m)}.$$

Then $\xi_m > 0$ a.s. Choose n_0 such that for $n \geq n_0$

$$\left|5^{-n}T_i^{n,m} - \sum_{j=1}^{i} W_j^{(m)}\right| < \tfrac{1}{3}\xi_n, \qquad 1 \leq i \leq a5^m.$$

Then if $i = i(t,\omega)$ is such that $W_i^m \leq t < W_{i+1}^m$, and $i \leq a5^m$ we have $5^{-n}T_{i-1}^{n,m} < t < 5^{-n}T_{i+2}^{n,m}$ for all $n \geq n_0$. So, $|X_t^n - Y_i^m| \leq 2^{-m+1}$ for all $n \geq n_0$. This implies that if $T_m = \sum_{j=1}^{a5^m} W_i^{(m)}$, and $S < T_m$, then

$$\sup_{0\leq t\leq S} |X_t^n - X_t^{n'}| \leq 2^{-m+2}$$

for all n, $n' \geq n_0$. If $S < \liminf_m T_m$ then the uniform a.s. convergence on the (random) interval $[0, S]$ follows. If s, $t < T_m$ and $|t - s| < \xi_m$, then we also have $|X_t^n - X_s^n| \leq 2^{-m+2}$ for $n \geq n_0$. Thus X is uniformly continuous on $[0, S]$. Varying a we also obtain uniform a.s. convergence on fixed intervals $[0, t_0]$. □

Although the notation is a little cumbersome, the ideas underlying the construction of X given here are quite simple. The argument above is given in [BP], but Kusuoka [Kus1], and Goldstein [Go], who were the first to construct a diffusion on G, used a similar approach. It is also worth noting that Knight [Kn] uses similar methods in his construction of 1-dimensional Brownian motion.

The natural next step is to ask about properties of the process X. But unfortunately the construction given above is not quite strong enough on its own to give us much. To see this, consider the questions

(1) Is $W = \lim_{n\to\infty} 5^{-n} T_1^{n,0} = \inf\{t \geq 0 : X_t \in G - \{0\}\}$?

(2) Is X Markov or strong Markov?

For (1), we certainly have $X_W \in G - \{0\}$. However, consider the possibility that each of the random walks Y_n moves from 0 to a_2 on a path which does not include a_1, but includes an approach to a distance 2^{-n}. In this case we have $a_1 \notin \{X_t^n, 0 \leq t \leq W\}$, but $X_T = a_1$ for some $T < W$. Plainly, some estimation of hitting probabilities is needed to exclude possibilities like this.
(2). The construction above does give a Markov property for X at stopping times of the form $\sum_{j=1}^i W_j^{(m)}$. But to obtain a good Markov process $X = (X_t, t \geq 0, \mathbb{P}^x, x \in G)$ we need to construct X at arbitrary starting points $x \in G$, and to show that (in some appropriate sense) the processes started at close together points x and y are close.

This can be done using the construction given above — see [BP, Section 2]. However, the argument, although not really hard, is also not that simple.

In the remainder of this section, I will describe some basic properties of the process X, for the most part without giving detailed proofs. Most of these theorems will follow from more general results given later in these notes.

Although G is highly symmetric, the group of global isometries of G is quite small. We need to consider maps restricted to subsets.

Definition 2.19. Let (F, ρ) be a metric space. A *local isometry* of F is a triple (A, B, φ) where A, B are subsets of F and φ is an isometry (i.e. bijective and distance preserving) between A and B, and between ∂A and ∂B.

Let $(X_t, t \geq 0, \mathbb{P}^x, x \in F)$ be a Markov process on F. For $H \subset F$, set $T_H = \inf\{t \geq 0 : X_t \in H\}$. X is *invariant with respect to a local isometry* (A, B, φ) if

$$\mathbb{P}^x\left(\varphi(X_{t\wedge T_{\partial A}}) \in \cdot, t \geq 0\right) = \mathbb{P}^{\varphi(x)}\left(X_{t\wedge T_{\partial B}} \in \cdot, t \geq 0\right).$$

X is *locally isotropic* if X is invariant with respect to the local isometries of F.

Theorem 2.20. *(a) There exists a continuous strong Markov process $X = (X_t, t \geq 0, \mathbb{P}^x, x \in G)$ on G.*
(b) The semigroup on $C(G)$ defined by

$$P_t f(x) = \mathbb{E}^x f(X_t)$$

is Feller, and is μ_G-symmetric:

$$\int_G f(x) P_t g(x) \mu_G(dx) = \int g(x) P_t f(x) \mu_G(dx).$$

(c) X is locally isotropic on the spaces $(G, |\cdot - \cdot|)$ and (G, d_G).
(d) For $n \geq 0$ let $T_{n,i}$, $i \geq 0$ be successive disjoint hits by X on G_n. Then $\widehat{Y}_i^{(n)} = X_{T_{n,i}}$, $i \geq 0$ defines a SRW on G_n, and $\widehat{Y}_{[5^n t]}^{(n)} \to X_t$ uniformly on compacts, a.s. So, in particular $(X_t, t \geq 0, \mathbb{P}^0)$ is the process constructed in Theorem 2.18.

This theorem will follow from our general results in Sections 6 and 7; a direct proof may be found in [BP, Sect. 2]. The main labour is in proving (a); given this (b), (c), (d) all follow in a relatively straightforward fashion from the corresponding properties of the approximating random walks $\widehat{Y}^{(n)}$.

The property of local isotropy on (G, d_G) characterizes X:

Theorem 2.21. *(Uniqueness). Let $(Z_t, t \geq 0, \mathbb{Q}^x, x \in \mathcal{G})$ be a non-constant locally isotropic diffusion on (G, d_G). Then there exists $a > 0$ such that*

$$\mathbb{Q}^x(Z_t \in \cdot, t \geq 0) = \mathbb{P}^x(X_{at} \in \cdot, t \geq 0).$$

(So Z is equal in law to a deterministic time change of X).

The beginning of the proof of Theorem 2.21 runs roughly along the lines one would expect: for $n \geq 0$ let $(\widetilde{Y}_i^{(n)}, i \geq 0)$ be $\widetilde{Z}$ sampled at its successive disjoint hits on G_n. The local isotropy of $\widetilde{Z}$ implies that $\widetilde{Y}^{(n)}$ is a SRW on G_n. However some work (see [BP, Sect. 8]) is required to prove that the process Y does not have traps, i.e. points x such that $\mathbb{Q}^x(Y_t = x \text{ for all } t) = 1$.

Remark 2.22. The definition of invariance with respect to local isometries needs some care. Note the following examples.
1. Let $x, y \in G_n$ be such that $D_n(x) \cap G_0 = a_0$, $D_n(y) \cap G_0 = \emptyset$. Then while there exists an isometry φ from $D_n(x) \cap G$ to $D_n(y) \cap G$, φ does not map $\partial_R D_n(x) \cap G$ to $\partial_R D_n(y) \cap G$. (∂_R denotes here the relative boundary in the set G).
2. Recall the definition of H_n, the n-th stage in the construction of G, and let $B_n = \partial H_n$. We have $G = \mathrm{cl}(\cup B_n)$. Consider the process Z_t on G, whose local motion is as follows. If $Z_t \in H_n - H_{n-1}$, then Z_t runs like a standard 1-dimensional Brownian motion on H_n, until it hits H_{n-1}. After this it repeats the same procedure on H_{n-1} (or H_{n-k} if it has also hit H_{n-k} at that time). This process is also invariant with respect to local isometries (A, B, φ) of the metric space $(G, |\cdot - \cdot|)$. See [He] for more on this and similar processes.

To discuss scale invariant properties of the process X it is useful to extend G to an unbounded set $\widetilde{G}$ with the same structure. Set

$$\widetilde{G} = \bigcup_{n=0}^{\infty} 2^n G,$$

and let $\widetilde{G}_n$ be the set of vertices of n-triangles in $\widetilde{G}_n$, for $n \geq 0$. We have

$$\widetilde{G}_n = \bigcup_{k=0}^{\infty} 2^k G_{n+k},$$

and if we define $G_m = \{0\}$ for $m < 0$, this definition also makes sense for $n < 0$. We can, almost exactly as above, define a limiting diffusion $\widetilde{X} = (\widetilde{X}_t, t \geq 0, \widetilde{\mathbb{P}}^x, x \in \widetilde{G})$ on $\widetilde{G}$:

$$\widetilde{X}_t = \lim_{n\to\infty} \widetilde{Y}^{(n)}_{[5^n t]}, \qquad t \geq 0, \text{ a.s.}$$

where $(\widetilde{Y}^{(n)}_k, n \geq 0, k \geq 0)$ are a sequence of nested simple random walks on $\widetilde{G}_n$, and the convergence is uniform on compact time intervals.

The process $\widetilde{X}$ satisfies an analogous result to Theorem 2.20, and in addition satisfies the scaling relation

$$\mathbb{P}^x(2\widetilde{X}_t \in \cdot\,, t \geq 0) = \mathbb{P}^{2x}(\widetilde{X}_{5t} \in \cdot\,, t \geq 0). \tag{2.13}$$

Note that (2.13) implies that $\widetilde{X}$ moves a distance of roughly $t^{\log 2/\log 5}$ in time t. Set

$$d_w = d_w(G) = \log 5/\log 2.$$

We now turn to the question: "What does this process look like?"

The construction of X, and Theorem 2.20(d), tells us that the 'crossing time' of a 0-triangle is equal in law to the limiting random variable W of a branching process with offspring p.g.f. given by $f(s) = s^2/(4-3s)$. From the functional equation (2.11) we can extract information about the behaviour of $\varphi(u) = E\exp(-uW)$ as $u \to \infty$, and from this (by a suitable Tauberian theorem) we obtain bounds on $\mathbb{P}(W \leq t)$ for small t. These translate into bounds on $\mathbb{P}^x(|X_t - x| > \lambda)$ for large λ. (One uses scaling and the fact that to move a distance in $\widetilde{G}$ greater than 2, X has to cross at least one 0-triangle). These bounds give us many properties of X. However, rather than following the development in [BP], it seems clearer to first present the more delicate bounds on the transition densities of $\widetilde{X}$ and X obtained there, and derive all the properties of the process from them. Write $\widetilde{\mu}_G$ for the analogue of μ_G for $\widetilde{G}$, and $\widetilde{P}_t$ for the semigroup of $\widetilde{X}$. Let $\widetilde{\mathcal{L}}$ be the infinitesimal generator of $\widetilde{P}_t$.

Theorem 2.23. *$\widetilde{P}_t$ and P_t have densities $\widetilde{p}(t,x,y)$ and $p(t,x,y)$ respectively.*
(a) $\widetilde{p}(t,x,y)$ is continuous on $(0,\infty)\times\widetilde{G}\times\widetilde{G}$.
(b) $\widetilde{p}(t,x,y) = \widetilde{p}(t,y,x)$ for all t,x,y.
(c) $t \to \widetilde{p}(t,x,y)$ is C^∞ on $(0,\infty)$ for each (x,y).
(d) For each t, y

$$|\widetilde{p}(t,x,y) - \widetilde{p}(t,x',y)| \leq c_1 t^{-1}|x-x'|^{d_w - d_f}, \quad x, x' \in \widetilde{G}.$$

(e) For $t \in (0,\infty)$, $x, y \in \widetilde{G}$

$$c_2 t^{-d_f/d_w} \exp\left(-c_3\left(\frac{|x-y|^{d_w}}{t}\right)^{1/(d_w-1)}\right) \leq \widetilde{p}(t,x,y) \leq c_4 t^{-d_f/d_w} \exp\left(-c_5\left(\frac{|x-y|^{d_w}}{t}\right)^{1/(d_w-1)}\right). \tag{2.14}$$

(f) For each $y_0 \in \widetilde{G}$, $\widetilde{p}(t,x,y_0)$ is the fundamental solution of the heat equation on $\widetilde{G}$ with pole at y_0:

$$\frac{\partial}{\partial t}\widetilde{p}(t,x,y_0) = \widetilde{\mathcal{L}}\widetilde{p}(t,x,y_0), \qquad \widetilde{p}(0,\cdot,y_0) = \delta_{y_0}(\cdot).$$

(g) $p(t,x,y)$ satisfies (a)–(f) above (with $\widetilde{G}$ replaced by G and $t \in (0,\infty]$ replaced by $t \in (0,1]$).

Remarks. 1. The proof of this in [BP] is now largely obsolete — simpler methods are now available, though these are to some extent still based on the ideas in [BP].
2. If $d_f = d$ and $d_w = 2$ we have in (2.14) the form of the transition density of Brownian motion in $\mathbb{R}^d$. Since $d_w = \log 5/\log 2 > 2$, the tail of the distribution of $|X_t - x|$ under $\mathbb{P}^x$ decays more rapidly than an exponential, but more slowly than a Gaussian.

It is fairly straightforward to integrate the bounds (2.14) to obtain information about X. At this point we just present a few simple calculations; we will give some further properties of this process in Section 3.

Definition 2.24. For $x \in \widetilde{G}$, $n \in \mathbb{Z}$, let x_n be the point in $\widetilde{G}_n$ closest to x in Euclidean distance. (Use some procedure to break ties). Let $D_n(x) = D_n(x_n)$.

Note that $\widetilde{\mu}_G\big(D_n(x_n)\big)$ is either 3^{-n} or 2.3^{-n}, that

$$|x-y| \le 2.2^{-n} \quad \text{ if } y \in D_n(x), \tag{2.15}$$

and that

$$|x-y| \ge \tfrac{\sqrt{3}}{4}2^{-(n+1)} \quad \text{ if } y \in \widetilde{G}\bigcap D_n(x)^c. \tag{2.16}$$

The sets $D_n(x)$ form a convenient collection of neighbourhoods of points in $\widetilde{G}$. Note that $\cup_{n\in\mathbb{Z}} D_n(x) = \widetilde{G}$.

Corollary 2.25. *For $x \in \widetilde{G}$,*

$$c_1 t^{2/d_w} \le \mathbb{E}^x|X_t - x|^2 \le c_2 t^{2/d_w}, \qquad t \ge 0.$$

Proof. We have

$$\mathbb{E}^x|X_t - x|^2 = \int_{\widetilde{G}} (y-x)^2 \widetilde{p}(t,x,y)\widetilde{\mu}_G(dy).$$

Set $A_m = D_m(x) - D_{m+1}(x)$. Then

$$\begin{aligned} &\int_{A_m} (y-x)^2 \widetilde{p}(t,x,y)\widetilde{\mu}_G(dy) \\ &\qquad \le c(2^{-m})^2 t^{-d_f/d_w} \exp\left(-c'\left(\left(2^{-m}\right)^{d_w}/t\right)^{1/(d_w-1)}\right)3^{-m} \\ &\qquad = c(2^{-m})^{2+d_f} t^{-d_f/d_w} \exp\left(-c'(5^{-m}/t)^{1/(d_w-1)}\right). \end{aligned} \tag{2.17}$$

Choose n such that $5^{-n} \le t < 5^{-n+1}$, and write $a_m(t)$ for the final term in (2.17). Then

$$\mathbb{E}^x(X_t - x)^2 \le \sum_{m=-\infty}^{n-1} a_m(t) + \sum_{m=n}^{\infty} a_m(t).$$

For $m < n$, $5^{-m}/t > 1$ and the exponential term in (2.17) is dominant. After a few calculations we obtain

$$\begin{aligned}\sum_{m=-\infty}^{n-1} a_m(t) &\le c(2^{-n})^{2+d_f} t^{-d_f/d_w} \\ &\le ct^{(2+d_f)/d_w - d_f/d_w} \le ct^{(2+d_f)/d_w - d_f/d_w} \le ct^{2/d_w},\end{aligned}$$

where we used the fact that $(2^{-n})^{d_w} \asymp t$. For $m \ge n$ we neglect the exponential term, and have

$$\begin{aligned}\sum_{m=n}^{\infty} a_m(t) &\le c\, t^{-d_f/d_w} \sum_{m=n}^{\infty} (2^{-m})^{2+d_f} \\ &\le ct^{-d_f/d_w}(2^{-n})^{2+d_f} \le c' t^{2/d_w}.\end{aligned}$$

Similar calculations give the lower bound. □

Remarks 2.26. 1. Since $2/d_w = \log 4/\log 5 < 1$ this implies that X is subdiffusive.
2. Since $\widetilde{\mu}_G\big(B(x,r)\big) \asymp r^{d_f}$, for $x \in \widetilde{G}$, it is tempting to try and prove Corollary 2.25 by the following calculation:

$$\begin{aligned}(2.18)\qquad \mathbb{E}^x|\widetilde{X}_t - x|^2 &= \int_0^\infty r^2 dr \int_{\partial B(x,r)} \widetilde{p}(t,x,y)\widetilde{\mu}_G(dy) \\ &\asymp \int_0^\infty dr\, r^2\, r^{d_f-1} t^{-d_f/d_w} \exp\Big(-c\big(r^{d_w}/t\big)^{1/d_w-1}\Big) \\ &= t^{2/d_w} \int_0^\infty s^{1+d_f} \exp\Big(-c\big(s^{d_w}\big)^{1/d_w-1}\Big)\, ds = ct^{2/d_w}.\end{aligned}$$

Of course this calculation, as it stands, is not valid: the estimate

$$\widetilde{\mu}_B\big(B(x,r+dr) - B(x,r)\big) \asymp r^{d_f-1}dr$$

is certainly not valid for all r. But it does hold on average over length scales of $2^n < r < 2^{n+1}$, and so splitting $\widetilde{G}$ into suitable shells, a rigorous version of this calculation may be obtained – and this is what we did in the proof of Corollary 2.25.

The λ-potential kernel density of $\widetilde{X}$ is defined by

$$u_\lambda(x,y) = \int_0^\infty e^{-\lambda t}\widetilde{p}(t,x,y)\, dt.$$

From (2.14) it follows that u_λ is continuous, that $u_\lambda(x,x) \le c\lambda^{d_f/d_w-1}$, and that $u_\lambda \to \infty$ as $\lambda \to 0$. Thus the process $\widetilde{X}$ (and also X) "hits points" – that is if

$T_y = \inf\{t > 0 : \widetilde{X}_t = y\}$ then

$$\mathbb{P}^x(T_y < \infty) > 0. \tag{2.19}$$

It is of course clear that X must be able to hit points in G_n – otherwise it could not move, but (2.19) shows that the remaining points in G have a similar status. The continuity of $u_\lambda(x, y)$ in a neighbourhood of x implies that

$$\mathbb{P}^x(T_x = 0) = 1,$$

that is that x is regular for $\{x\}$ for all $x \in \widetilde{G}$.

The following estimate on the distribution of $|\widetilde{X}_t - x|$ can be obtained easily from (2.14) by integration, but since this bound is actually one of the ingredients in the proof, such an argument would be circular.

Proposition 2.27. *For $x \in \widetilde{G}$, $\lambda > 0$, $t > 0$,*

$$\begin{aligned} c_1 \exp\left(-c_2(\lambda^{d_w}/t)^{1/d_w - 1}\right) &\leq \mathbb{P}^x(|\widetilde{X}_t - x| > \lambda) \\ &\leq c_3 \exp\left(-c_4(\lambda^{d_w}/t)^{(1/d_w - 1)}\right). \end{aligned}$$

From this, it follows that the paths of $\widetilde{X}$ are Hölder continuous of order $1/d_w - \varepsilon$ for each $\varepsilon > 0$. In fact we can (up to constants) obtain the precise modulus of continuity of $\widetilde{X}$. Set

$$h(t) = t^{1/d_w}(\log t^{-1})^{(d_w - 1)/d_w}.$$

Theorem 2.28. *(a) For $x \in G$*

$$c_1 \leq \lim_{\delta \downarrow 0} \sup_{\substack{0 \leq s \leq t \leq 1 \\ |t-s| < \delta}} \frac{|\widetilde{X}_s - \widetilde{X}_t|}{h(s-t)} \leq c_2, \qquad \mathbb{P}^x - \text{ a.s.}$$

(b) The paths of $\widetilde{X}$ are of infinite quadratic variation, a.s., and so in particular $\widetilde{X}$ is not a semimartingale.

The proof of (a) is very similar to that of the equivalent result for Brownian motion in $\mathbb{R}^d$.
For (b), Proposition 2.23 implies that $|X_{t+h} - X_t|$ is of order h^{1/d_w}; as $d_w > 2$ this suggests that X should have infinite quadratic variation. For a proof which fills in the details, see [BP, Theorem 4.5]. □

So far in this section we have looked at the Sierpinski gasket, and the construction and properties of a symmetric diffusion X on G (or $\widetilde{G}$). The following three questions, or avenues for further research, arise naturally at this point.

1. Are there other natural diffusions on the SG?
2. Can we do a similar construction on other fractals?
3. What finer properties does the process X on G have? (More precisely: what about properties which the bounds in (2.17) are not strong enough to give information on?)

The bulk of research effort in the years since [Kus1, Go, BP] has been devoted to (2). Only a few papers have looked at (1), and (apart from a number of works on spectral properties), the same holds for (3).

Before discussing (1) or (2) in greater detail, it is worth extracting one property of the SRW $Y^{(1)}$ which was used in the construction.

Let $V = (V_n, n \geq 0, \mathbb{P}^a, a \in G_0)$ be a Markov chain on G_0: clearly V is specified by the transition probabilities

$$p(a_i, a_j) = \mathbb{P}^{a_i}(V_1 = a_j), \quad 0 \leq i, j \leq 2.$$

We take $p(a, a) = 0$ for $a \in G_0$, so V is determined by the three probabilities $p(a_i, a_j)$, where $j = i + 1 \pmod 3$.

Given V we can define a Markov Chain V' on G_1 by a process we call *replication*. Let $\{b_{01}, b_{02}, b_{12}\}$ be the 3 points in $G_1 - G_0$, where $b_{ij} = \frac{1}{2}(a_i + a_j)$. We consider G_1 to consist of three 1-cells $\{a_i, b_{ij}, j \neq i\}$, $0 \leq i \leq 2$, which intersect at the points $\{b_{ij}\}$. The law of V' may be described as follows: V' moves inside each 1-cell in the way same as V does; if V'_0 lies in two 1-cells then it first chooses a 1-cell to move in, and chooses each 1-cell with equal probability. More precisely, writing $V' = (V'_n, n \geq 0, \overline{\mathbb{P}}^a, a \in G_1)$, and

$$p'(a, b) = \overline{\mathbb{P}}^a(V'_1 = b),$$

we have

$$\begin{aligned} p'(a_i, b_{ij}) &= p(a_i, a_j), \\ p'(b_{ij}, b_{ik}) &= \tfrac{1}{2} p(a_j, a_k), \quad p'(b_{ij}, a_i) = \tfrac{1}{2} p(a_j, a_i). \end{aligned} \tag{2.20}$$

Now let T_k, $k \geq 0$ be successive disjoint hits by V' on G_0, and let $U_k = V'_{T_k}$, $k \geq 0$. Then U is a Markov Chain on G_0; we say that V is *decimation invariant* if U is equal in law to V.

We saw above that the SRW $Y^{(0)}$ on G_0 was decimation invariant. A natural question is:

What other decimation invariant Markov chains are there on G_0?

Two classes have been found:

1. (See [Go]). Let $p(a_0, a_1) = p(a_1, a_0) = 1$, $p(a_2, a_0) = \frac{1}{2}$.
2. "p-stream random walks" ([Kum1]). Let $p \in (0, 1)$ and

$$p(a_0, a_1) = p(a_1, a_2) = p(a_2, a_0) = p.$$

From each of these processes we can construct a limiting diffusion in the same way as in Theorem 2.18. The first process is reasonably easy to understand: essentially its paths consist of a downward drift (when this is possible), and a behaviour like 1-dimensional Brownian motion on the portions on G which consist of line segments parallel to the x-axis.

For $p > \frac{1}{2}$ Kumagai's p-stream diffusions tend to rotate in an anti-clockwise direction, so are quite non-symmetric. Apart from the results in [Kum1] nothing is known about this process.

Two other classes of diffusions on G, which are not decimation invariant, have also been studied. The first are the "asymptotically 1-dimensional diffusions" of [HHW3], the second the diffusions, similar to that described in Remark 2.22, which are $(G, |\cdot - \cdot|)$-isotropic but not (G, d_G)- isotropic – see [He]. See also [HH1, HatK, HHK1, HHK2] for work on the self-avoiding random walk on the SG.

Diffusions on other fractal sets.

Of the three questions above, the one which has received most attention is that of making similar constructions on other fractals. To see the kind of difficulties which can arise, consider the following two fractals, both of which are constructed by a Cantor type procedure, based on squares rather than triangles. For each curve the figure gives the construction after two stages.

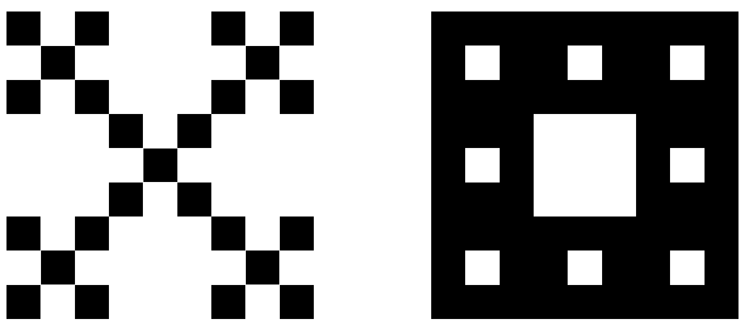

Figure 2.6: The Vicsek set and the Sierpinski carpet.

The first of these we will call the "Vicsek set" (VS for short). We use similar notation as for the SG, and write G_0, $G_1, \dots$ for the succession of sets of vertices of corners of squares. We denote the limiting set by $F = F_{VS}$. One difficulty arises immediately. Let Y_r be the SRW on G_0 which moves from any point $x \in G_0$ to each of its neighbours with equal probability. (The neighbours of x are the 2 points y in G_0 with $|x - y| = 1$). Then $Y^{(0)}$ is not decimation invariant. This is easy to see: $Y^{(0)}$ cannot move in one step from $(0,0)$ to $(1,1)$, but $Y^{(1)}$ can move from $(0,0)$ to $(1,1)$ without hitting any other point in G_0.

However it is not hard to find a decimation invariant random walk on G_0. Let $p \in [0,1]$, and consider the random walk $(Y_r, r \geq 0, \mathbb{E}_p^x, x \in G_0)$ on G_0 which moves diagonally with probability p, and horizontally or vertically with probability $\frac{1}{2}(1-p)$. Let $(Y'_r, r \geq 0, \mathbb{E}_p^x, x \in G_1)$ be the Markov chain on G_1 obtained by replication, and let T_k, $k \geq 0$ be successive disjoint hits by Y' on G_0.

Then writing $f(p) = \mathbb{P}_p^0\big(Y'_{T_1} = (1,1)\big)$ we have (after several minutes calculation)

$$f(p) = \frac{1}{4-3p}.$$

The equation $f(p) = p$ therefore has two solutions: $p = \frac{1}{3}$ and $p = 1$, each of which corresponds to a decimation invariant walk on G_0. (The number $\frac{1}{3}$ here has no general significance: if we had looked at the fractal similar to the Vicsek set, but based on a 5×5 square rather than a 3×3 square, then we would have obtained a different number).

One may now carry through, in each of these cases, the construction of a diffusion on the Vicsek set F, very much as for the Sierpinski gasket. For $p = 1$ one gets a rather uninteresting process, which, if started from $(0,0)$, is (up to a constant time change) 1-dimensional Brownian motion on the diagonal $\{(t,t), 0 \le t \le 1\}$. It is worth remarking that this process is not strong Markov: for each $x \in F$ one can take $\mathbb{P}^x$ to be the law of a Brownian motion moving on a diagonal line including x, but the strong Markov property will fail at points where two diagonals intersect, such as the point $(\frac{1}{2}, \frac{1}{2})$.

For $p = \frac{1}{3}$ one obtains a process $(X_t, t \ge 0)$ with much the same behaviour as the Brownian motion on the SG. We have for the Vicsek set (with $p = \frac{1}{3}$) $d_f(F_{VS}) = \log 5/\log 3$, $d_w(F_{VS}) = \log 15/\log 3$. This process was studied in some detail by Krebs [Kr1, Kr2]. The Vicsek set was mentioned in [Go], and is one of the "nested fractals" of Lindstrøm [L1].

This example shows that one may have to work to find a decimation invariant random walk, and also that this may not be unique. For the VS, one of the decimation invariant random walks was degenerate, in the sense that $P^x(Y \text{ hits } y) = 0$ for some $x, y \in G_0$, and we found the associated diffusion to be of little interest. But it raises the possibility that there could exist regular fractals carrying more than one "natural" diffusion.

The second example is the Sierpinski carpet (SC). For this set a more serious difficulty arises. The VS was finitely ramified, so that if Y_t is a diffusion on F_{VS}, and $(T_k, k \ge 0)$ are successive disjoint hits on G_n, for some $n \ge 0$, then $(Y_{T_k}, k \ge 0)$ is a Markov chain on G_n. However the SC is not finitely ramified: if $(Z_t, t \ge 0)$ is a diffusion on F_{SC}, then the first exit of Z from $[0, \frac{1}{3}]^2$ could occur anywhere on the line segments $\{(\frac{1}{3}, y), 0 \le y \le \frac{1}{3}\}$, $\{(x, \frac{1}{3}), 0 \le x \le \frac{1}{3}\}$. It is not even clear that a diffusion on F_{SC} will hit points in G_n. Thus to construct a diffusion on F_{SC} one will need very different methods from those outlined above. It is possible, and has been done: see [BB1-BB6], and [Bas] for a survey.

On the third question mentioned above, disappointingly little has been done: most known results on the processes on the Sierpinski gasket, or other fractals, are of roughly the same depth as the bounds in Theorem 2.23. Note however the results on the spectrum of $\mathcal{L}$ in [FS1, FS2, Sh1–Sh4], and the large deviation results in [Kum5]. Also, Kusuoka [Kus2] has very interesting results on the behaviour of harmonic functions, which imply that the measure defined formally on G by

$$\nu(dx) = |\nabla f|^2(x)\mu(dx)$$

is singular with respect to μ. There are many open problems here.

3. Fractional Diffusions.

In this section I will introduce a class of processes, defined on metric spaces, which will include many of the processes on fractals mentioned in these lectures. I have chosen an axiomatic approach, as it seems easier, and enables us to neglect (for the time being!) much of fine detail in the geometry of the space.

A metric space (F, ρ) has the *midpoint property* if for each $x, y \in F$ there exists $z \in F$ such that $\rho(x,z) = \rho(z,y) = \frac{1}{2}\rho(x,y)$. Recall that the geodesic metric d_G in Section 2 had this property. The following result is a straightforward exercise:

Lemma 3.1. *(See [Blu]). Let (F, ρ) be a complete metric space with the midpoint property. Then for each x, $y \in F$ there exists a geodesic path $(\gamma(t), 0 \le t \le 1)$ such that $\gamma(0) = x$, $\gamma(1) = y$ and $\rho(\gamma(s), \gamma(t)) = |t - s| d(x,y)$, $0 \le s \le t \le 1$.*

For this reason we will frequently refer to a metric ρ with the midpoint property as a *geodesic metric*. See [Stu1] for additional remarks and references on spaces of this type.

Definition 3.2. Let (F, ρ) be a complete metric space, and μ be a Borel measure on $\big(F, \mathcal{B}(F)\big)$. We call (F, ρ, μ) a *fractional metric space* (FMS for short) if

$$(3.1a) \qquad (F, \rho) \text{ has the midpoint property,}$$

and there exist $d_f > 0$, and constants c_1, c_2 such that if $r_0 = \sup\{\rho(x,y) : x, y \in F\} \in (0, \infty]$ is the diameter of F then

$$(3.1b) \qquad c_1 r^{d_f} \le \mu\big(B(x,r)\big) \le c_2 r^{d_f} \quad \text{for} \quad x \in F,\ 0 < r \le r_0.$$

Here $B(x,r) = \{y \in F : \rho(x,y) < r\}$.

Remarks 3.3. 1. $\mathbb{R}^d$, with Euclidean distance and Lebesgue measure, is a FMS, with $d_f = d$ and $r_0 = \infty$.
2. If G is the Sierpinski gasket, d_G is the geodesic metric on G, and $\mu = \mu_G$ is the measure constructed in Section 2, then Lemma 2.1 shows that (G, d_G, μ) is a FMS, with $d_f = d_f(G) = \log 3/\log 2$ and $r_0 = 1$. Similarly $(\widetilde{G}, d_{\widetilde{G}}, \widetilde{\mu})$ is a FMS with $r_0 = \infty$.
3. If (F_k, d_k, μ_k), $k = 1, 2$ are FMS with the same diameter r_0 and $p \in [1, \infty]$, then setting $F = F_1 \times F_2$, $d((x_1,x_2),(y_1,y_2)) = (d_1(x_1,y_1)^p + d_2(x_2,y_2)^p)^{1/p}$, $\mu = \mu_1 \times \mu_2$, it is easily verified that (F, d, μ) is also a FMS with $d_f(F) = d_f(F_1) + d_f(F_2)$.
4. For simplicity we will from now on take either $r_0 = \infty$ or $r_0 = 1$. We will write $r \in (0, r_0]$ to mean $r \in (0, r_0] \cap (0, \infty)$, and define $r_0^\alpha = \infty$ if $\alpha > 0$ and $r_0 = \infty$.

A number of properties of (F, ρ, μ) follow easily from the definition.

Lemma 3.4. (a) $\dim_H(F) = \dim_P(F) = d_f$.
(b) F is locally compact.
(c) $d_f \ge 1$.

Proof. (a) is immediate from Corollary 2.8.
(b) Let $x \in F$, $A = \overline{B}(x,1)$, and consider a maximal packing of disjoint balls $B(x_i, \varepsilon)$, $x_i \in A$, $1 \le i \le m$. As $\mu(A) \le c_2$, and $\mu\big(B(x_i,\varepsilon)\big) \ge c_1 \varepsilon^{d_f}$, we have $m \le c_2 \big(c_1 \varepsilon^{d_f}\big)^{-1} < \infty$. Also $A = \cup_{i=1}^m B(x_i, 2\varepsilon)$. Thus any bounded set in F can be

covered by a finite number of balls radius ε; this, with completeness, implies that F is locally compact.
(c) Take $x, y \in F$ with $\rho(x,y) = D > 0$. Applying the midpoint property repeatedly we obtain, for $m = 2^k$, $k \geq 1$, a sequence $x = z_0, z_1, \dots, z_m = y$ with $\rho(z_i, z_{i+1}) = D/m$. Set $r = D/2m$: the balls $B(z_i, r)$ must be disjoint, or, using the triangle inequality, we would have $\rho(x,y) < D$. But then

$$\bigcup_{i=0}^{m-1} B(z_i, r) \subset B(x, D),$$

so that

$$\begin{aligned} c_2 D^{d_f} \geq \mu\big(B(x,D)\big) &\geq \sum_{i=0}^{m-1} \mu\big(B(z_i, r)\big) \\ &\geq m c_1 D^{d_f} (2m)^{-d_f} = c m^{1-d_f}. \end{aligned}$$

If $d_f < 1$ a contradiction arises on letting $m \to \infty$. □

Definition 3.5. Let (F, ρ, μ) be a fractional metric space. A Markov process $X = (\mathbb{P}^x, x \in F, X_t, t \geq 0)$ is a *fractional diffusion* on F if
(3.2a) X is a conservative Feller diffusion with state space F.
(3.2b) X is μ-symmetric.
(3.2c) X has a symmetric transition density $p(t,x,y) = p(t,y,x)$, $t > 0$, $x, y \in F$, which satisfies, the Chapman-Kolmogorov equations and is, for each $t > 0$, jointly continuous.
(3.2d) There exist constants $\alpha, \beta, \gamma, c_1 - c_4$, $t_0 = r_0^\beta$, such that

$$\begin{aligned} c_1 t^{-\alpha} \exp\left(-c_2 \rho(x,y)^{\beta\gamma} t^{-\gamma}\right) &\leq p(t,x,y) \\ &\leq c_3 t^{-\alpha} \exp\left(-c_4 \rho(x,y)^{\beta\gamma} t^{-\gamma}\right), \ x, y \in F,\ 0 < t \leq t_0. \end{aligned} \tag{3.3}$$

Examples 3.6. 1. If F is $\mathbb{R}^d$, and $a(x) = a_{ij}(x)$, $1 \leq i, j \leq d$, $x \in \mathbb{R}^d$ is bounded, symmetric, measurable and uniformly elliptic, let $\mathcal{L}$ be the divergence form operator

$$\mathcal{L} = \sum_{ij} \frac{\partial}{\partial x_i} a_{ij}(x) \frac{\partial}{\partial x_j}.$$

Then Aronsen's bounds [Ar] imply that the diffusion with infinitesimal generator $\mathcal{L}$ is a FD, with $\alpha = d/2$, $\beta = 2$, $\gamma = 1$.
2. By Theorem 2.23, the Brownian motion on the Sierpinski gasket described in Section 2 is a FD, with $\alpha = d_f(SG)/d_w(SG)$, $\beta = d_w(SG)$ and $\gamma = 1/(\beta - 1)$.

The hypotheses in Definition 3.5 are quite strong ones, and (as the examples suggest) the assertion that a particular process is an FD will usually be a substantial theorem. One could of course consider more general bounds than those in (3.3) (with a correspondingly larger class of processes), but the form (3.3) is reasonably natural, and already contains some interesting examples.

In an interesting recent series of papers Sturm [Stu1-Stu4] has studied diffusions on general metric spaces. However, the processes considered there turn out to have an essentially Gaussian long range behaviour, and so do not include any FDs with $\beta \neq 2$.

In the rest of this section we will study the general properties of FDs. In the course of our work we will find some slightly easier sufficient conditions for a process to be a FD than the bounds (3.3), and this will be useful in Section 8 when we prove that certain diffusions on fractals are FDs. We begin by obtaining two relations between the indices d_f, α, β, γ, so reducing the parameter space of FDs to a two-dimensional one.

We will say that F is a $FMS(d_f)$ if F is a FMS and satisfies (3.1b) with parameter d_f (and constants c_1, c_2). Similarly, we say X is a $FD'(d_f, \alpha, \beta, \gamma)$ if X is a FD on a $FMS(d_f)$, and X satisfies (3.3) with constants α, β, γ. (This is temporary notation — hence the $'$).

It what follows we fix a FMS (F, ρ, μ), with parameters r_0 and d_f.

Lemma 3.7. *Let $\alpha, \gamma, x > 0$ and set*

$$I(\gamma, x) = \int_1^\infty e^{-xt^\gamma}\, dt,$$

$$S(\alpha, \gamma, x) = \sum_{n=0}^\infty \alpha^n e^{-x\alpha^{n\gamma}}.$$

Then

$$(\alpha - 1)S(\alpha, \gamma, \alpha^\gamma x) \le I(\gamma, x) \le (\alpha - 1)S(\alpha, \gamma, x), \tag{3.4}$$

and

$$I(\gamma, x) \asymp x^{-1/\gamma} \quad \text{for } x \le 1, \tag{3.5}$$

$$I(\gamma, x) \asymp x^{-1} e^{-x} \quad \text{for } x \ge 1, \tag{3.6}$$

Proof. We have

$$I(\gamma, x) = \sum_{n=0}^\infty \int_{\alpha^n}^{\alpha^{n+1}} e^{-xt^\gamma}\, dt,$$

and estimating each term in the sum (3.4) is evident.

If $0 < x \le 1$ then since

$$x^{1/\gamma} I(\gamma, x) = \int_{x^{1/\gamma}}^\infty e^{-s^\gamma}\, ds \to c(\gamma) \text{ as } x \to 0,$$

(3.5) follows.

If $x \ge 1$ then (3.6) follows from the fact that

$$xe^x I(\gamma, x) = \gamma^{-1} \int_0^\infty e^{-u}((x+u)/x)^{-1+1/\gamma} du \to \gamma^{-1} \text{ as } x \to \infty. \qquad \square$$

Lemma 3.8. *("Scaling relation"). Let X be a $FD'(d_f, \alpha, \beta, \gamma)$ on F. Then $\alpha = d_f/\beta$.*

Proof. From (3.1) we have

$$p(t, x, y) \ge c_1 t^{-\alpha} e^{-c_2} = c_3 t^{-\alpha} \qquad \text{for } \rho(x, y) \le t^{1/\beta}.$$

Set $t_0 = r_0^\beta$. So if $A = B(x, t^{1/\beta})$, and $t \le t_0$

$$1 \ge \mathbb{P}^x(\rho(x, X_t) \le t^{1/\beta}) = \int_A p(t,x,y)\mu(dy) \ge c_3 t^{-\alpha}\mu(A) \ge c t^{-\alpha + d_f/\beta}.$$

If $r_0 = \infty$ then since this holds for all $t > 0$ we must have $\alpha = d_f/\beta$. If $r_0 = 1$ then we only deduce that $\alpha \le d_f/\beta$.

Let now $r_0 = 1$, let $\lambda > 0$, $t < 1$, and $A = B(x, \lambda t^{1/\beta})$. We have $\mu(F) \le c_{3.1.2}$, and therefore

$$\begin{aligned} 1 &= \mathbb{P}^x(X_t \in A) + \mathbb{P}^x(X_t \in A^c) \\ &\le \mu(A) \sup_{y \in A} p(t,x,y) + \mu(F - A) \sup_{y \in A^c} p(t,x,y) \\ &\le c_4 t^{-\alpha + d_f/\beta} \lambda^{d_f/\beta} + c_5 t^{-\alpha} e^{-c_6 \lambda^{\beta\gamma}}. \end{aligned}$$

Let $\lambda = \big((d_f/\beta)c_6^{-1} \log(1/t)\big)^{1/\beta\gamma}$; then we have for all $t < 1$ that

$$1 \le c t^{-\alpha + d_f/\beta}(1 + (\log(1/t))^{1/\beta\gamma},$$

which gives a contradiction unless $\alpha \ge d_f/\beta$. □

The next relation is somewhat deeper: essentially it will follow from the fact that the long–range behaviour of $p(t,x,y)$ is fixed by the exponents d_f and β governing its short–range behaviour. Since γ only plays a role in (3.3) when $\rho(x,y)^\beta \gg t$, we will be able to obtain γ in terms of d_f and β (in fact, it turns out, of β only).

We begin by deriving some consequences of the bounds (3.3).

Lemma 3.9. *Let X be a $FD'(d_f, d_f/\beta, \beta, \gamma)$. Then*
(a) For $t \in (0, t_0]$, $r > 0$

$$\mathbb{P}^x\big(\rho(x, X_t) > r\big) \le c_1 \exp\big(-c_2 r^{\beta\gamma} t^{-\gamma}\big).$$

(b) There exists $c_3 > 0$ such that

$$c_4 \exp\left(-c_5 r^{\beta\gamma} t^{-\gamma}\right) \le \mathbb{P}^x\left(\rho(x, X_t) > r\right) \quad \text{for } r < c_3 r_0,\ t < r^\beta.$$

(c) For $x \in F$, $0 < r < c_3 r_0$, if $\tau(x,r) = \inf\{s > 0 : X_s \notin B(x,r)\}$ then

$$c_6 r^\beta \le \mathbb{E}^x \tau(x,r) \le c_7 r^\beta. \tag{3.7}$$

Proof. Fix $x \in F$, and set $D(a,b) = \{y \in F : a \le \rho(x,y) \le b\}$. Then by (3.1b)

$$c_{3.1.2} b^{d_f} \ge \mu\big(D(a,b)\big) \ge c_{3.1.1} b^{d_f} - c_{3.1.2} a^{d_f}.$$

Choose $\theta \ge 2$ so that $c_{3.1.1}\theta^{d_f} \ge 2c_{3.1.2}$: then we have

$$c_8 a^{d_f} \le \mu\big(D(a, \theta a)\big) \le c_9 a^{d_f}. \tag{3.8}$$

Therefore, writing $D_n = D(\theta^n r, \theta^{n+1} r)$, we have $\mu(D_n) \asymp \theta^{n d_f}$ provided $r\theta^{n+1} \le r_0$. Now

$$\text{(3.9)} \qquad \mathbb{P}^x\big(\rho(x,X_t)>r\big) = \int\limits_{B(x,r)^c} p(t,x,y)\mu(dy)$$
$$= \sum_{n=0}^{\infty} \int_{D_n} p(t,x,y)\mu(dy)$$
$$\le \sum_{n=0}^{\infty} c(r\theta^i)^{d_f} t^{-d_f/\beta} \exp\big(-c_{10} t^{-\gamma}(r\theta^n)^{\beta\gamma}\big)$$
$$= c(r^\beta/t)^{d_f/\beta} S(\theta, \beta\gamma, c_{10}(r^\beta/t)^\gamma).$$

If $c_{10}r^\beta > t$ then using (3.6) we deduce that this sum is bounded by

$$c_{11} \exp\left(-c_{12}(r^\beta/t)^\gamma\right),$$

while if $c_{10}r^\beta \le t$ then (as $\mathbb{P}^x\left(\rho(x,X_t)>r\right) \le 1$) we obtain the same bound, on adjusting the constant c_{11}.

For the lower bound (b), choose $c_3 > 0$ so that $c_3\theta < 1$. Then $\mu(D_0) \ge cr^{d_f}$, and taking only the first term in (3.9) we deduce that, since $r^\beta > t$,

$$\mathbb{P}^x\big(\rho(x,X_t)>r\big) \ge c(r^\beta/t)^{d_f/\beta} \exp(-c_{13}(r^\beta/t)^\gamma)$$
$$\ge c\exp(-c_{13}(r^\beta/t)^\gamma).$$

(c) Note first that

$$\text{(3.10)} \qquad \mathbb{P}^y(\tau(x,r)>t) \le \mathbb{P}^y(X_t \in B(x,r))$$
$$= \int_{B(x,r)} p(t,y,z)\mu(dz)$$
$$\le ct^{-d_f/\beta} r^{d_f}.$$

So, for a suitable c_{14}

$$\mathbb{P}^y(\tau(x,r) > c_{14}r^\beta) \le \tfrac{1}{2}, \quad y \in F.$$

Applying the Markov property of X we have for each $k \ge 1$

$$\mathbb{P}^y(\tau(x,r) > kc_{14}r^\beta) \le 2^{-k}, \quad y \in F,$$

which proves the upper bound in (3.7).

For the lower bound, note first that

$$\mathbb{P}^x(\tau(x,2r)<t) = \mathbb{P}^x\left(\sup_{0\le s\le t} \rho(x,X_t) \ge 2r\right)$$
$$\le \mathbb{P}^x\big(\rho(x,X_t)>r\big) + \mathbb{P}^x\big(\tau(x,2r)<t, \rho(x,X_t)<r\big)$$

Writing $S = \tau(x,2r)$, the second term above equals

$$\mathbb{E}^x 1_{(S<t)} \mathbb{P}^{X_S}\big(\rho(x,X_{t-S})<r\big) \le \sup_{y\in\partial B(x,2r)} \sup_{s\le t} \mathbb{P}^y\big(\rho(y,X_{t-s})>r\big),$$

so that, using (a),

$$\mathbb{P}^x\big(\tau(x,2r)<t\big) \le 2\sup_{s\le t}\sup_{y\in F}\mathbb{P}^y\big(\rho(y,X_s)>r\big) \le 2c_1\exp\big(-c_2(r^\beta/t)^\gamma\big). \tag{3.11}$$

So if $4c_1e^{-c_2a^\gamma}=1$ then $\mathbb{P}^x(\tau(x,2r)<ar^\beta)\le\frac{1}{2}$, which proves the left hand side of (3.7). □

Remark 3.10. Note that the bounds in (c) only used the upper bound on $p(t,x,y)$.

The following result gives sufficient conditions for a diffusion on F to be a fractional diffusion: these conditions are a little easier to verify than (3.3).

Theorem 3.11. *Let (F,ρ,μ) be a $FMS(d_f)$. Let $(Y_t, t\ge 0, \mathbb{P}^x, x\in F)$ be a μ-symmetric diffusion on F which has a transition density $q(t,x,y)$ with respect to μ which is jointly continuous in x,y for each $t>0$. Suppose that there exists a constant $\beta>0$, such that*

$$q(t,x,y)\le c_1t^{-d_f/\beta} \quad \text{for all } x,y\in F,\ t\in(0,t_0], \tag{3.12}$$

$$q(t,x,y)\ge c_2t^{-d_f/\beta} \quad \text{if } \rho(x,y)\le c_3t^{1/\beta},\ t\in(0,t_0], \tag{3.13}$$

$$c_4r^\beta\le\mathbb{E}^x\tau(x,r)\le c_5r^\beta, \quad \text{for } x\in F,\ 0<r<c_6r_0, \tag{3.14}$$

where $\tau(x,r)=\inf\{t\ge 0: Y_t\notin B(x,r)\}$. Then $\beta>1$ and Y is a FD with parameters d_f, d_f/β, β and $1/(\beta-1)$.

Corollary 3.12. *Let X be a $FD'(d_f, d_f/\beta,\beta,\gamma)$ on a FMS(d_f) F. Then $\beta>1$ and $\gamma=1/(\beta-1)$.*

Proof. By Lemma 3.8, and the bounds (3.3), the transition density $p(t,x,y)$ of X satisfies (3.12) and (3.13). By Lemma 3.9(c) X satisfies (3.14). So, by Theorem 3.11 $\beta>1$, and X is a $FD'(d_f,d_f/\beta,\beta,(\beta-1)^{-1})$. Since $p(t,x,y)$ cannot satisfy (3.3) for two distinct values of γ, we must have $\gamma=(\beta-1)^{-1}$. □

Remark 3.13. Since two of the four parameters are now seen to be redundant, we will shorten our notation and say that X is a $FD(d_f,\beta)$ if X is a $FD'(d_f,d_f/\beta,\beta,\gamma)$.

The proof of Theorem 3.11 is based on the derivation of transition density bounds for diffusions on the Sierpinski carpet in [BB4]: most of the techniques there generalize easily to fractional metric spaces. The essential idea is "chaining": in its classical form (see e.g. [FaS]) for the lower bound, and in a slightly different more probabilistic form for the upper bound. We begin with a some lemmas.

Lemma 3.14. *[BB1, Lemma 1.1] Let $\xi_1,\xi_2,\ldots,\xi_n$, V be non-negative r.v. such that $V\ge\sum_1^n\xi_i$. Suppose that for some $p\in(0,1)$, $a>0$,*

$$P\big(\xi_i\le t|\sigma(\xi_1,\ldots,\xi_{i-1})\big)\le p+at, \qquad t>0. \tag{3.15}$$

Then

$$\log P(V\le t)\le 2\Big(\frac{ant}{p}\Big)^{1/2}-n\log\frac{1}{p}. \tag{3.16}$$

Proof. If η is a r.v. with distribution function $P(\eta \le t) = (p + at) \wedge 1$, then

$$\begin{aligned} E\left(e^{-\lambda\xi_i} | \sigma(\xi_1, \ldots, \xi_{i-1})\right) &\le E e^{-\lambda\eta} \\ &= p + \int_0^{(1-p)/a} e^{-\lambda t} a dt \\ &\le p + a\lambda^{-1}. \end{aligned}$$

So

$$\begin{aligned} P(V \le t) = P\left(e^{-\lambda V} \ge e^{-\lambda t}\right) &\le e^{\lambda t} E e^{-\lambda V} \\ &\le e^{\lambda t} E \exp \lambda \sum_1^n \xi_i \le e^{\lambda t}(p + a\lambda^{-1})^n \\ &\le p^n \exp\left(\lambda t + \frac{an}{\lambda p}\right). \end{aligned}$$

The result follows on setting $\lambda = (an/pt)^{1/2}$. □

Remark 3.15. The estimate (3.16) appears slightly odd, since it tends to $+\infty$ as $p \downarrow 0$. However if $p = 0$ then from the last but one line of the proof above we obtain $\log P(V \le t) \le \lambda t + n \log \frac{a}{\lambda}$, and setting $\lambda = n/t$ we deduce that

$$\log P(V \le t) \le n \log(\frac{ate}{n}). \tag{3.17}$$

Lemma 3.16. *Let $(Y_t, t \ge 0)$ be a diffusion on a metric space (F, ρ) such that, for $x \in F$, $r > 0$,*

$$c_1 r^\beta \le \mathbb{E}^x \tau(x, r) \le c_2 r^\beta.$$

Then for $x \in F$, $t > 0$,

$$\mathbb{P}^x\big(\tau(x, r) \le t\big) \le (1 - c_1/(2^\beta c_2)) + c_3 r^{-\beta} t.$$

Proof. Let $x \in F$, and $A = B(x, r)$, $\tau = \tau(x, r)$. Since $\tau \le t + (\tau - t) 1_{(\tau > t)}$ we have

$$\begin{aligned} \mathbb{E}^x \tau &\le t + \mathbb{E}^x 1_{(\tau > t)} \mathbb{E}^{Y_t}(\tau - t) \\ &\le t + \mathbb{P}^x(\tau > t) \sup_y \mathbb{E}^y \tau. \end{aligned}$$

As $\tau \le \tau(y, 2r)$ $\mathbb{P}^y$-a.s. for any $y \in F$, we deduce

$$c_1 r^\beta \le \mathbb{E}^x \tau \le t + \mathbb{P}^x(\tau > t) c_2 (2r)^\beta,$$

so that

$$c_2 2^\beta \mathbb{P}^x(\tau \le t) \le (2^\beta c_2 - c_1) + t r^{-\beta}.$$ □

The next couple of results are needed to show that the diffusion Y in Theorem 3.11 can reach distant parts of the space F in an arbitrarily short time.

Lemma 3.17. *Let Y_t be a μ-symmetric diffusion with semigroup T_t on a complete metric space (F,ρ). If $f,g \ge 0$ and there exist $a < b$ such that*

$$\int f(x)\mathbb{E}^x g(Y_t)\mu(dx) = 0 \text{ for } t \in (a,b), \tag{3.18}$$

then $\int f(x)\mathbb{E}^x g(Y_t)\mu(dx) = 0$ for all $t > 0$.

Proof. Let $(E_\lambda, \lambda \ge 0)$ be the spectral family associated with T_t. Thus (see [FOT, p. 17]) $T_t = \int_0^\infty e^{-\lambda t} dE_\lambda$, and

$$(f, T_t g) = \int_0^\infty e^{-\lambda t} d(f, E_\lambda g) = \int_0^\infty e^{-\lambda t}\nu(d\lambda),$$

where ν is of finite variation. (3.18) and the uniqueness of the Laplace transform imply that $\nu = 0$, and so $(f, T_t g) = 0$ for all t. □

Lemma 3.18. *Let F and Y satisfy the hypotheses of Theorem 3.11. If $\rho(x,y) < c_3 r_0$ then $\mathbb{P}^x(Y_t \in B(y,r)) > 0$ for all $r > 0$ and $t > 0$.*

Remark. The restriction $\rho(x,y) < c_3 r_0$ is of course unnecessary, but it is all we need now. The conclusion of Theorem 3.11 implies that $\mathbb{P}^x(Y_t \in B(y,r)) > 0$ for all $r > 0$ and $t > 0$, for all x, $y \in F$.

Proof. Suppose the conclusion of the Lemma fails for x, y, r, t. Choose $g \in C(F, \mathbb{R}_+)$ such that $\int_F g d\mu = 1$ and $g = 0$ outside $B(y,r)$. Let $t_1 = t/2$, $r_1 = c_3(t_1)^\beta$, and choose $f \in C(F, \mathbb{R}_+)$ so that $\int_F f d\mu = 1$, $f(x) > 0$ and $f = 0$ outside $A = B(x, r_1)$. If $0 < s < t$ then the construction of g implies that

$$0 = \mathbb{E}^x g(Y_t) = \int_F q(s,x,x') E^{x'} g(Y_{t-s})\mu(dx').$$

Since by (3.13) $q(s,x,x') > 0$ for $t/2 < s < t$, $x' \in B(x,r_1)$, we deduce that $E^{x'} g(Y_u) = 0$ for $x' \in B(x,r_1)$, $u \in (0,t/2)$. Thus as $\mathrm{supp}(f) \subset B(x,r_1)$

$$\int_F f(x') E^{x'} g(Y_u) d\mu = 0$$

for all $u \in (1, t/2)$, and hence, by Lemma 3.17, for all $u > 0$. But by (3.13) if $u = (\rho(x,y)/c_3)^\beta$ then $q(u,x,y) > 0$, and by the continuity of f, g and q it follows that $\int f \mathbb{E}^x g(Y_u) d\mu > 0$, a contradiction. □

Proof of Theorem 3.11. For simplicity we give full details of the proof only in the case $r_0 = \infty$; the argument in the case of bounded F is essentially the same. We begin by obtaining a bound on

$$\mathbb{P}^x\big(\tau(x,r) \le t\big).$$

Let $n \ge 1$, $b = r/n$, and define stopping times S_i, $i \ge 0$, by

$$S_0 = 0, \quad S_{i+1} = \inf\{t \ge S_i : \rho(Y_{S_i}, Y_t) \ge b\}.$$

Let $\xi_i = S_i - S_{i-1}$, $i \geq 1$. Let $(\mathcal{F}_t)$ be the filtration of Y_t, and let $\mathcal{G}_i = \mathcal{F}_{S_i}$. We have by Lemma 3.16

$$\mathbb{P}^x(\xi_{i+1} \leq t | \mathcal{G}_i) = \mathbb{P}^{Y_{S_i}}\big(\tau(Y_{S_i}, b) \leq t\big) \leq p + c_6 b^{-\beta} t,$$

where $p \in (0,1)$. As $\rho(Y_{S_i}, Y_{S_{i+1}}) = b$, we have $\rho(Y_0, Y_{S_n}) \leq r$, so that $S_n = \sum_1^n \xi_i \leq \tau(Y_0, r)$. So, by Lemma 3.14, with $a = c_6 (r/n)^{-\beta}$,

$$\begin{aligned} \log \mathbb{P}^x\big(\tau(x,r) \leq t\big) &\leq 2p^{-\frac{1}{2}}\big(c_6 r^{-\beta} n^{1+\beta} t\big)^{\frac{1}{2}} - n \log \frac{1}{p} \\ &= c_7 (r^{-\beta} n^{1+\beta} t)^{\frac{1}{2}} - c_8 n. \end{aligned} \tag{3.19}$$

If $\beta \leq 1$ then taking t small enough the right hand side of (3.17) is negative, and letting $n \to \infty$ we deduce $\mathbb{P}^x\big(\tau(x,r) \leq t\big) = 0$, which contradicts the fact that $\mathbb{P}^x\big(Y_t \in B(y,r)\big) > 0$ for all t. So we have $\beta > 1$. (If $r_0 = 1$ then we take r small enough so that $r < c_3$).

If we neglect for the moment the fact that $n \in \mathbb{N}$, and take $n = n_0$ in (3.19) so that

$$\tfrac{1}{2} c_8 n_0 = c_7 \Big(n_0^{1+\beta} t r^{-\beta}\Big)^{1/2},$$

then

$$n_0^{\beta-1} = (c_8^2 / 4 c_7^2) r^\beta t^{-1}, \tag{3.20}$$

and

$$\log \mathbb{P}^x\big(\tau(x,r) \leq t\big) \leq -\tfrac{1}{2} c_8 n_0.$$

So if $r^\beta t^{-1} \geq 1$, we can choose $n \in \mathbb{N}$ so that $1 \leq n \leq n_0 \vee 1$, and we obtain

$$\mathbb{P}^x\big(\tau(x,r) \leq t\big) \leq c_9 \exp\left(-c_{10}\left(\frac{r^\beta}{t}\right)^{1/(\beta-1)}\right). \tag{3.21}$$

Adjusting the constant c_9 if necessary, this bound also clearly holds if $r^\beta t^{-1} < 1$.

Now let $x, y \in F$, write $r = \rho(x,y)$, choose $\varepsilon < r/4$, and set $C_z = B(z, \varepsilon)$, $z = x, y$. Set $A_x = \{z \in F : \rho(z,x) \leq \rho(z,y)\}$, $A_y = \{z : \rho(z,x) \geq \rho(z,y)\}$. Let ν_x, ν_y be the restriction of μ to C_x, C_y respectively.

We now derive the upper bound on $q(t,x,y)$ by combining the bounds (3.12) and (3.21): the idea is to split the journey of Y from C_x to C_y into two pieces, and use one of the bounds on each piece. We have

$$\begin{aligned} \mathbb{P}^{\nu_x}(Y_t \in C_y) &= \int_{C_y}\int_{C_x} q(t, x', y') \mu(dx') \mu(dy') \\ &\leq \mathbb{P}^{\nu_x}\big(Y_t \in C_y, Y_{t/2} \in A_x\big) + \mathbb{P}^{\nu_x}(Y_t \in C_y, Y_{t/2} \in A_y). \end{aligned} \tag{3.22}$$

We begin with second term in (3.22):

$$\begin{aligned} \mathbb{P}^{\nu_x}(Y_t \in C_y, Y_{t/2} \in A_y) &= \mathbb{P}^{\nu_x}\big(\tau(Y_0, r/4) \leq t/2, Y_{t/2} \in A_y, Y_t \in C_y\big) \\ &\leq \mathbb{P}^{\nu_x}\big(\tau(Y_0, r/4) \leq t/2\big) \sup_{y' \in A_y} \mathbb{P}^{y'}\big(Y_{t/2} \in C_y\big) \end{aligned} \tag{3.23}$$

$$\le \nu_x(C_x) c_9 \exp\left(-c_{10}\left(\frac{(r/4)^\beta}{t/2}\right)^{1/(\beta-1)}\right) c_1 \nu_y(C_y) t^{-d_f/\beta}$$
$$= \mu(C_x)\mu(C_y) c_{11} t^{-d_f/\beta} \exp\left(-c_{12}(r^\beta/t)^{1/(\beta-1)}\right),$$

where we used (3.21) and (3.12) in the last but one line.

To handle the first term in (3.22) we use symmetry:

$$\mathbb{P}^{\nu_x}(Y_t \in C_y, Y_{t/2} \in A_x) = \mathbb{P}^{\nu_y}(Y_t \in C_x, Y_{t/2} \in A_x),$$

and this can now be bounded in exactly the same way. We therefore have

$$\int_{C_y}\int_{C_x} q(t,x',y')\mu(dx')\mu(dy')$$
$$\le \mu(C_x)\mu(C_y) 2c_{11} t^{-d_f/\beta} \exp\left(-c_{12}(r^\beta/t)^{1/(\beta-1)}\right),$$

so that as $q(t,\cdot,\cdot)$ is continuous

$$q(t,x,y) \le 2c_{11} t^{-d_f/\beta} \exp\left(-c_{12}(r^\beta/t)^{1/(\beta-1)}\right). \tag{3.24}$$

The proof of the lower bound on q uses the technique of "chaining" the Chapman-Kolmogorov equations. This is quite classical, except for the different scaling.

Fix x, y, t, and write $r = \rho(x,y)$. If $r \le c_3 t^{1/\beta}$ then by (3.13)

$$q(t,x,y) \ge c_2 t^{-d_f/\beta},$$

and as $\exp(-(r^\beta/t)^{1/(\beta-1)}) \ge \exp(-c_3^{1/(\beta-1)})$, we have a lower bound of the form (3.3). So now let $r > c_3 t^{1/\beta}$. Let $n \ge 1$. By the mid-point hypothesis on the metric ρ, we can find a chain $x = x_0, x_1, \ldots, x_n = y$ in F such that $\rho(x_{i-1}, x_i) = r/n$, $1 \le i \le n$. Let $B_i = B(x_i, r/2n)$; note that if $y_i \in B_i$ then $\rho(y_{i-1}, y_i) \le 2r/n$. We have by the Chapman-Kolmogorov equation, writing $y_0 = x_0$, $y_n = y$,

$$q(t,x,y) \ge \int_{B_1} \mu(dy_1) \ldots \int_{B_{n-1}} \mu(dy_{n-1}) \prod_{i=1}^n q(t/n, y_{i-1}, y_i). \tag{3.25}$$

We wish to choose n so that we can use the bound (3.13) to estimate the terms $q(t/n, y_{i-1}, y_i)$ from below. We therefore need:

$$\frac{2r}{n} \le c_3\left(\frac{t}{n}\right)^{1/\beta} \tag{3.26}$$

which holds provided

$$n^{\beta-1} \ge 2^\beta c_3^{-\beta} \frac{r^\beta}{t}. \tag{3.27}$$

As $\beta > 1$ it is certainly possible to choose n satisfying (3.27). By (3.25) we then obtain, since $\mu(B_i) \geq c(r/2n)^{d_f}$,

$$\begin{aligned} q(t,x,y) &\geq c(r/2n)^{d_f(n-1)} \Big(c_2 (t/n)^{-d_f/\beta} \Big)^n \\ &= c(r/2n)^{-d_f} \Big(c_2 (t/n)^{-1/\beta} (r/2n)^{d_f} \Big)^n \\ &= c'(r/n)^{-d_f} \Big((t/n)^{-1/\beta} (r/n) \Big)^n. \end{aligned} \tag{3.28}$$

Recall that n satisfies (3.27): as $r > c_3 t^{1/\beta}$ we can also ensure that for some $c_{13} > 0$

$$\frac{r}{n} \geq c_{13} (t/n)^{1/\beta}, \tag{3.29}$$

so that $n^{\beta-1} \leq 2^\beta c_{13}^{-\beta} r^\beta / t$. So, by (3.28)

$$\begin{aligned} q(t,x,y) &\geq c(t/n)^{-d_f/\beta} c_{14}^n \\ &\geq c_{15} t^{-d_f/\beta} \exp\left(n \log c_{14} \right) \\ &\geq c_{15} t^{-d_f/\beta} \exp\Big(-c_{16} (r^\beta/t)^{1/(\beta-1)} \Big). \end{aligned}$$

□

Remarks 3.19.
1. Note that the only point at which we used the "midpoint" property of ρ is in the derivation of the lower bound for q.
2. The essential idea of the proof of Theorem 3.11 is that we can obtain bounds on the long range behaviour of Y provided we have good enough information about the behaviour of Y over distances of order $t^{1/\beta}$. Note that in each case, if $r = \rho(x,y)$, the estimate of $q(t,x,y)$ involves splitting the journey from x to y into n steps, where $n \asymp (r^\beta/t)^{1/(\beta-1)}$.
3. Both the arguments for the upper and lower bounds appear quite crude: the fact that they yield the same bounds (except for constants) indicates that less is thrown away than might appear at first sight. The explanation, very loosely, is given by "large deviations". The off-diagonal bounds are relevant only when $r^\beta \gg t$ – otherwise the term in the exponential is of order 1. If $r^\beta \gg t$ then it is difficult for Y to move from x to y by time t and it is likely to do so along more or less the shortest path. The proof of the lower bound suggests that the process moves in a 'sausage' of radius $r/n \asymp t/r^{\beta-1}$.

The following two theorems give additional bounds and restrictions on the parameters d_f and β. Unlike the proofs above the results use the symmetry of the process very strongly. The proofs should appear in a forthcoming paper.

Theorem 3.20. *Let F be a $FMS(d_f)$, and X be a $FD(d_f, \beta)$ on F. Then*

$$2 \leq \beta \leq 1 + d_f. \tag{3.30}$$

Theorem 3.21. *Let F be a $FMS(d_f)$. Suppose X^i are $FD(d_f, \beta_i)$ on F, for $i = 1, 2$. Then $\beta_1 = \beta_2$.*

Remarks 3.22. 1. Theorem 3.21 implies that the constant β is a property of the metric space F, and not just of the FD X. In particular any FD on $\mathbb{R}^d$, with the

usual metric and Lebesgue measure, will have $\beta = 2$. It is very unlikely that every FMS F carries a FD.
2. I expect that (3.30) is the only general relation between β and d_f. More precisely, set

$$A = \{(d_f, \beta) : \text{there exists a } FD(d_f, \beta)\},$$

and $\Gamma = \{(d_f, \beta) : 2 \le \beta \le 1 + d_f\}$. Theorem 3.20 implies that $A \subset \Gamma$, and I conjecture that $\operatorname{int}\Gamma \subset A$. Since $BM(\mathbb{R}^d)$ is a $FD(d, 2)$, the points $(d, 2) \in A$ for $d \ge 1$. I also suspect that

$$\{d_f : (d_f, 2) \in A\} = \mathbb{N},$$

that is that if F is an FMS of dimension d_f, and d_f is not an integer, then any FD on F will not have Brownian scaling.

Properties of Fractional Diffusions.

In the remainder of this section I will give some basic analytic and probabilistic properties of FDs. I will not give detailed proofs, since for the most part these are essentially the same as for standard Brownian motion. In some cases a more detailed argument is given in [BP] for the Sierpinski gasket.

Let F be a $FMS(d_f)$, and X be a $FD(d_f, \beta)$ on F. Write $T_t = \mathbb{E}^x f(X_t)$ for the semigroup of X, and $\mathcal{L}$ for the infinitesimal generator of T_t.

Definition 3.23. Set

$$d_w = \beta, \qquad d_s = \frac{2d_f}{d_w}.$$

This notation follows the physics literature where (for reasons we will see below) d_w is called the "walk dimension" and d_s the "spectral dimension". Note that (3.3) implies that

$$p(t, x, x) \asymp t^{-d_s/2}, \quad 0 < t \le t_0,$$

so that the on-diagonal bounds on p can be expressed purely in terms of d_s. Since many important properties of a process relate solely to the on-diagonal behaviour of its density, d_s is the most significant single parameter of a FD.

Integrating (3.3), as in Corollary 2.25, we obtain:

Lemma 3.24. $\mathbb{E}^x \rho(X_t, x)^p \asymp t^{p/d_w}$, $x \in F$, $t \ge 0$, $p > 0$.

Since by Theorem 3.20 $d_w \ge 2$ this shows that FDs are diffusive or subdiffusive.

Lemma 3.25. *(Modulus of continuity). Let $\varphi(t) = t^{1/d_w}(\log(1/t))^{(d_w-1)/d_w}$. Then*

$$c_1 \le \lim_{\delta \downarrow 0} \sup_{\substack{0 \le s < t \le 1 \\ |t-s| < \delta}} \frac{\rho(X_s, X_t)}{\varphi(t-s)} \le c_2. \tag{3.31}$$

So, in the metric ρ, the paths of X just fail to be Hölder $(1/d_w)$. The example of divergence form diffusions in $\mathbb{R}^d$ shows that one cannot hope to have $c_1 = c_2$ in general.

Lemma 3.26. *(Law of the iterated logarithm – see [BP, Thm. 4.7]). Let* $\psi(t) = t^{1/d_w}(\log\log(1/t))^{(d_w-1)/d_w}$. *There exist* c_1, c_2 *and constants* $c(x) \in [c_1, c_2]$ *such that*

$$\limsup_{t\downarrow 0} \frac{\rho(X_t, X_0)}{\psi(t)} = c(x) \quad \mathbb{P}^x\text{-a.s.}$$

Of course, the 01 law implies that the limit above is non-random.

Lemma 3.27. *(Dimension of range).*

$$\dim_H(\{X_t : 0 \le t \le 1\}) = d_f \wedge d_w. \tag{3.32}$$

This result helps to explain the terminology "walk dimension" for d_w. Provided the space the diffusion X moves in is large enough, the dimension of range of the process (called the "dimension of the walk" by physicists) is d_w.

Potential Theory of Fractional Diffusions.

Let $\lambda \ge 0$ and set

$$u_\lambda(x,y) = \int_0^\infty e^{-\lambda s} p(s,x,y)\, ds.$$

Then if

$$U_\lambda f(x) = \mathbb{E}^x \int_0^\infty e^{-\lambda s} f(X_s)\, d_s$$

is the λ-resolvent of X, u_λ is the density of U_λ:

$$U_\lambda f(x) = \int_F u_\lambda(x,y)\mu(dy).$$

Write u for u_0.

Proposition 3.28. *Let* $\lambda_0 = 1/r_0$. *(If* $r_0 = \infty$ *take* $\lambda_0 = 0$*).*
(a) If $d_s < 2$ *then* $u_\lambda(x,y)$ *is jointly continuous on* $F \times F$ *and for* $\lambda > \lambda_0$

$$\begin{aligned} c_1 \lambda^{d_s/2-1} \exp\big(-c_2 \lambda^{1/d_w} \rho(x,y)\big) &\le u_\lambda(x,y) \\ &\le c_3 \lambda^{d_s/2-1} \exp\Big(-c_4 \lambda^{1/d_w} \rho(x,y)\Big). \end{aligned} \tag{3.33}$$

(b) If $d_s = 2$ *and* $\lambda > \lambda_0$ *then writing* $R = \rho(x,y)\lambda^{1/d_w}$

$$c_5\left(\log_+(1/R) + e^{-c_6 R}\right) \le u_\lambda(x,y) \le c_7\left(\log_+(1/R) + e^{-c_8 R}\right). \tag{3.34}$$

(c) If $d_s > 2$ *then*

$$c_9 \rho(x,y)^{d_w - d_f} \le u_{\lambda_0}(x,y) \le c_{10} \rho(x,y)^{d_w - d_f}. \tag{3.35}$$

These bounds are obtained by integrating (3.3): for (a) and (b) one uses Laplace's method. (The continuity in (b) follows from the continuity of p and the uniform bounds on p in (3.3)). Note in particular that:

(i) if $d_s < 2$ then $u_\lambda(x,x) < +\infty$ and $\lim_{\lambda\to 0} u_\lambda(x,y) = +\infty$.
(ii) if $d_s > 2$ then $u(x,x) = +\infty$, while $u(x,y) < \infty$ for $x \neq y$

Since the polarity or non-polarity of points relates to the on-diagonal behaviour of u, we deduce from Proposition 3.28

Corollary 3.29. *(a) If $d_s < 2$ then for each $x, y \in F$*

$$\mathbb{P}^x(X \text{ hits } y) = 1.$$

(b) If $d_s \geq 2$ then points are polar for X.
(c) If $d_s \leq 2$ then X is set-recurrent: for $\varepsilon > 0$

$$\mathbb{P}^y\left(\{t : X_t \in B(y,\varepsilon)\} \text{ is non-empty and unbounded}\right) = 1.$$

(d) If $d_s > 2$ and $r_0 = \infty$ then X is transient.

In short, X behaves like a Brownian motion of dimension d_s; but in this context a continuous parameter range is possible.

Lemma 3.30. *(Polar and non-polar sets). Let A be a Borel set in F.*
(a) $\mathbb{P}^x(T_A < \infty) > 0$ if $\dim_H(A) > d_f - d_w$,
(b) A is polar for X if $\dim_H(A) < d_f - d_w$.

Since X is symmetric any semipolar set is polar. As in the Brownian case, a more precise condition in terms of capacity is true, and is needed to resolve the critical case $\dim_H(A) = d_f - d_w$.

If X, X' are independent $FD(d_f, \beta)$ on F, and $Z_t = (X_t, X'_t)$, then it follows easily from the definition that Z is a FD on $F \times F$, with parameters $2d_f$ and β. If $D = \{(x,x) : x \in F\} \subset F \times F$ is the diagonal in $F \times F$, then $\dim_H(D) = d_f$, and so Z hits D (with positive probability) if

$$d_f > 2d_f - d_w,$$

that is if $d_s < 2$. So

$$\mathbb{P}^x(X_t = X'_t \text{ for some } t > 0) > 0 \quad \text{if } d_s < 2, \tag{3.36}$$

and

$$\mathbb{P}^x(X_t = X'_t \text{ for some } t > 0) = 0 \quad \text{if } d_s > 2. \tag{3.37}$$

No doubt, as in the Brownian case, X and X' do not collide if $d_s = 2$.

Lemma 3.31. *X has k-multiple points if and only if $d_s < 2k/(k-1)$.*

Proof. By [Rog] X has k-multiple points if and only if

$$\int_{B(x,1)} u_1(x,y)^k \mu(dy) < \infty;$$

the integral above converges or diverges with

$$\int_0^1 r^{kd_w-(k-1)d_f} r^{-1}\,dr,$$

by a calculation similar to that in Corollary 2.25. $\square$

The bounds on the potential kernel density $u_\lambda(x,y)$ lead immediately to the existence of local times for X – see [Sha, p. 325].

Theorem 3.32. *If $d_s < 2$ then X has jointly measurable local times $(L_t^x, x \in F, t \geq 0)$ which satisfy the density of occupation formula with respect to μ:*

$$\int_0^t f(X_s)ds = \int_F f(a)L_t^a \mu(da), \quad f \text{ bounded and measurable.} \tag{3.38}$$

In the low-dimensional case (that is when $d_s < 2$, or equivalently $d_f < d_w$) we can obtain more precise estimates on the Hölder continuity of $u_\lambda(x,y)$, and hence on the local times L_t^x. The main lines of the argument follow that of [BB4, Section 4], but on the whole the arguments here are easier, as we begin with stronger hypotheses. We work only in the case $r_0 = \infty$: the same results hold in the case $r_0 = 1$, with essentially the same prooofs.

For the next few results we fix F, a $FMS(d_f)$ with $r_0 = \infty$, and X, a $FD(d_f, d_w)$ on F. For $A \subset F$ write

$$\tau_A = T_{A^c} = \inf\{t \geq 0 : X_t \in A^c\}.$$

Let R_λ be an independent exponential time with mean λ^{-1}. Set for $\lambda \geq 0$

$$u_\lambda^A(x,y) = \mathbb{E}^x \int_0^{\tau_A} e^{-\lambda s} dL_s^y = \mathbb{E}^x L^y_{\tau_A \wedge R_\lambda},$$
$$U_\lambda^A f(x) = \int_F u_\lambda^A(x,y)\mu(dy).$$

Let

$$p_\lambda^A(x,y) = \mathbb{P}^x(T_y \leq \tau_A \wedge R_\lambda);$$

note that

$$u_\lambda^A(x,y) = p_\lambda^A(x,y)u_\lambda^A(y,y) \leq u_\lambda^A(y,y). \tag{3.39}$$

Write $u^A(x,y) = u_0^A(x,y)$, $U^A = U_0^A$, and note that $u_\lambda(x,y) = u_\lambda^F(x,y)$, $U_\lambda = U_\lambda^A$. As in the case of u we write p^A, p_λ for p_0^A, p_λ^F. As $(\mathbb{P}^x, X_t)$ is μ-symmetric we have $u_\lambda^A(x,y) = u_\lambda^A(y,x)$ for all x, $y \in F$.

The following Lemma enables us to pass between bounds on u_λ and u^A.

Lemma 3.33. *Suppose $A \subset F$, A is bounded, For x, $y \in F$ we have*

$$u^A(x,y) = u_\lambda^B(x,y) + \mathbb{E}^x\big(1_{(R_\lambda \leq \tau_A)} u^A(X_{R_\lambda}, y)\big) - \mathbb{E}^x\big(1_{(R_\lambda > \tau_A)} u_\lambda^B(X_{\tau_A}, y)\big).$$

Proof. From the definition of u^A,

$$\begin{aligned} u^A(x,y) &= \mathbb{E}^x(L^y_{\tau_A}\,; R_\lambda \le \tau_A) + \mathbb{E}^x(L^y_{\tau_A}\,; R_\lambda > \tau_A) \\ &= \mathbb{E}^x(L^y_{R_\lambda}\,; R_\lambda \le \tau_A) + \mathbb{E}^x(1_{(R_\lambda \le \tau_A)} \mathbb{E}^{X_{R_\lambda}} L^y_{\tau_A}) \\ &\quad + \mathbb{E}^x(L^y_{R_\lambda}\,; R_\lambda > \tau_A) - \mathbb{E}^x(L^y_{R_\lambda \wedge \tau_B} - L^y_{\tau_A}\,; R_\lambda > \tau_A) \\ &= u_\lambda(x,y) + \mathbb{E}^x\big(1_{(R_\lambda \le \tau_A)} u^A(X_{R_\lambda}, y)\big) - \mathbb{E}^x\big(1_{(R_\lambda > \tau_A)} u_\lambda(X_{\tau_A}, y)\big). \qquad \square \end{aligned}$$

Corollary 3.34. *Let $x \in F$, and $r > 0$. Then*

$$c_1 r^{d_w - d_f} \le u^{B(x,r)}(x,x) \le c_2 r^{d_w - d_f}.$$

Proof. Write $A = B(x,r)$, and let $\lambda = \theta r^{-d_w}$, where θ is to be chosen. We have from Lemma 3.33, writing $\tau = \tau(x,r)$,

$$u^A(x,y) \le u_\lambda(x,y) + \mathbb{E}^x 1_{(R_\lambda < \tau)} u^A(X_{R_\lambda}, y).$$

So if $v = \sup_x u^A(x,y)$ then using (3.33)

$$v \le c_3 \lambda^{d_s/2-1} + \mathbb{P}^x(R_\lambda < \tau) v. \tag{3.40}$$

Let $t_0 > 0$. Then by (3.10)

$$\begin{aligned} \mathbb{P}^x(R_\lambda < \tau) &= \mathbb{P}^x(R_\lambda < \tau, \tau \le t_0) + \mathbb{P}^x(R_\lambda < \tau, \tau > t_0) \\ &\le \mathbb{P}^x(R_\lambda < t_0) + \mathbb{P}^x(\tau > t_0) \\ &\le (1 - e^{-\lambda t_0}) + c t_0^{-d_f/d_w} r^{d_f}. \end{aligned}$$

Choose first t_0 so that the second term is less than $\frac{1}{4}$, and then λ so that the first term is also less than $\frac{1}{4}$. We have $t_0 \asymp r^{d_w} \asymp \lambda^{-1}$, and the upper bound now follows from (3.40).

The lower bound is proved in the same way, using the bounds on the lower tail of τ given in (3.11). $\square$

Lemma 3.35. *There exist constants $c_1 > 1$, c_2 such that if $x, y \in F$, $r = \rho(x,y)$, $t_0 = r^{d_w}$ then*

$$\mathbb{P}^x\big(T_y < t_0 < \tau(x, c_1 r)\big) \ge c_2.$$

Proof. Set $\lambda = (\theta/r)^{d_w}$; we have $p_\lambda(x,y) \ge c_3 \exp(-c_4\theta)$ by (3.33). So since

$$p_\lambda(x,y) = \mathbb{E}^x e^{-\lambda T_y} \le \mathbb{P}^x(T_y < t) + e^{-\lambda t},$$

we deduce that

$$\mathbb{P}^x(T_y < t) \ge c_3 \exp(-c_4\theta) - \exp(-\theta^{d_w}).$$

As $d_w > 1$ we can choose θ (depending only on c_3, c_4 and d_w) such that $\mathbb{P}^x(T_y < t) \ge \frac{1}{2} c_3 \exp(-c_4\theta) = c_5$. By (3.11) for $a > 0$

$$\mathbb{P}^x(\tau(x, aR) < R^{d_w}) \le c_6 \exp(-c_7 a^{d_w/(d_w-1)}),$$

so there exists $c_1 > 1$ such that $\mathbb{P}^x(\tau(x, c_1 r) < t_0) \le \frac{1}{2} c_5$. So

$$\mathbb{P}^x\big(T_y < t_0 < \tau(x, c_1 r)\big) \ge \mathbb{P}^x(T_y < t_0) - \mathbb{P}^x(\tau(x, c_1 r) < t_0) \ge \tfrac{1}{2} c_5. \qquad \square$$

Definition 3.36. We call a function h *harmonic* (with respect to X) in an open subset $A \subset F$ if $\mathcal{L}h = 0$ on A, or equivalently, $h(X_{t \wedge T_{A^c}})$ is a local martingale.

Proposition 3.37. *(Harnack inequality). There exist constants $c_1 > 1$, $c_2 > 0$, such that if $x_0 \in F$, and $h \geq 0$ is harmonic in $B(x_0, c_1 r)$, then*

$$h(x) \geq c_2 h(y), \quad x, y \in B(x_0, r).$$

Proof. Let $c_1 = 1 + c_{3.35.1}$, so that $B(x, c_{3.35.1}r) \subset B(x_0, c_1 r)$ if $\rho(x, x_0) \leq r$. Fix x, y, write $r = \rho(x, y)$, and set $S = T_y \wedge \tau(x, c_{3.35.1}r)$. As $h(X_{\cdot \wedge S})$ is a supermartingale, we have by Lemma 3.35,

$$h(x) \geq \mathbb{E}^x h(X_S) \geq h(y) \mathbb{P}^x(T_y < \tau(x, c_{3.35.1}r)) \geq c_{3.35.2} h(y). \qquad \square$$

Corollary 3.38. *There exists $c_1 > 0$ such that if $x_0 \in F$, and $h \geq 0$ is harmonic in $B(x_0, r)$, then*

$$h(x) \geq c_1 h(y), \quad x, y \in B(x_0, \tfrac{3}{4}r).$$

Proof. This follows by covering $B(x_0, \frac{3}{4}r)$ by balls of the form $B(y, c_2 r)$, where c_2 is small enough so that Proposition 3.37 can be applied in each ball. (Note we use the geodesic property of the metric ρ here, since we need to connect each ball to a fixed reference point by a chain of overlapping balls). $\square$

Lemma 3.39. *Let $x, y \in F$, $r = \rho(x, y)$. If $R > r$ and $B(y, R) \subset A$ then*

$$u^A(y, y) - u^A(x, y) \leq c_1 r^{d_w - d_f}.$$

Proof. We have, writing $\tau = \tau(y, r)$, $T = T_{A^c}$,

$$u^A(y, y) = \mathbb{E}^y L^y_\tau + \mathbb{E}^y \mathbb{E}^{X_\tau} L^y_T = u^B(y, y) + \mathbb{E}^y u^A(X_\tau, y),$$

so by Corollary 3.34

$$\mathbb{E}^y (u^A(y, y) - u^A(X_\tau, y)) = u^B(y, y) \leq c_2 r^{d_w - d_f}. \tag{3.41}$$

Set $\varphi(x') = u^A(y, y) - u^A(x', y)$; φ is harmonic on $A - \{y\}$. As $\rho(x, y) = r$ and ρ has the geodesic property there exists z with $\rho(y, z) = \frac{1}{4}r$, $\rho(x, z) = \frac{3}{4}r$. By Corollary 3.38, since φ is harmonic in $B(x, r)$,

$$\varphi(z) \geq c_{3.38.1} \varphi(x).$$

Now set $\psi(x') = \mathbb{E}^{x'} \varphi(X_\tau)$ for $x' \in B$. Then ψ is harmonic in B and $\varphi \leq \psi$ on B. Applying Corollary 3.38 to ψ in B we deduce

$$\psi(y) \geq c_{3.38.1} \psi(z) \geq c_{3.38.1} \varphi(z) \geq (c_{3.38.1})^2 \varphi(x).$$

Since $\psi(y) = \mathbb{E}^y (u^A(y, y) - u^A(X_\tau, y))$ the conclusion follows from (3.41). $\square$

Theorem 3.40. *(a) Let $\lambda > 0$. Then for x, x', $y \in F$, and $f \in L^1(F)$, $g \in L^\infty(F)$,*

$$|u_\lambda(x, y) - u_\lambda(x', y)| \leq c_1 \rho(x, x')^{d_w - d_f}, \tag{3.42}$$

$$|U_\lambda f(x) - U_\lambda f(x')| \leq c_1 \rho(x, x')^{d_w - d_f} ||f||_1. \tag{3.43}$$

$$|U_\lambda g(x) - U_\lambda g(x')| \leq c_2 \lambda^{-d_s/2} \rho(x, x')^{d_w - d_f} ||g||_\infty. \tag{3.44}$$

Proof. Let $x,\ x' \in F$, write $r = \rho(x,x')$ and let $R > r,\ A = B(x,R)$. Since $u_\lambda^A(y,x') \ge p_\lambda^A(y,x)u_\lambda^A(x,x')$, we have using the symmetry of X that

$$\begin{aligned} u_\lambda^A(x,y) - u_\lambda^A(x',y) &\le u_\lambda^A(y,x) - p_\lambda^A(y,x)u_\lambda^A(x,x') \\ &= p_\lambda^A(y,x)\big(u_\lambda^A(x,x) - u_\lambda^A(x,x')\big). \end{aligned} \tag{3.45}$$

Thus

$$|u_\lambda^A(x,y) - u_\lambda^A(x',y)| \le |u_\lambda^A(x,x) - u_\lambda^A(x,x')|.$$

Setting $\lambda = 0$ and using Lemma 3.39 we deduce

$$|u^A(x,y) - u^A(x',y)| \le c_3 r^{d_w - d_f}. \tag{3.46}$$

So

$$\begin{aligned} |U^A f(x) - U^A f(x')| &\le \int_A |u^A(x,y) - u^A(x',y)|\,|f(y)|\mu(dy) \\ &\le c_3 r^{d_w-d_f}||f1_A||_1. \end{aligned}$$

To obtain estimates for $\lambda > 0$ we apply the resolvent equation in the form

$$u_\lambda^A(x,y) = u^A(x,y) - \lambda U^A v(x),$$

where $v(x) = u_\lambda^A(x,y)$. (Note that $||v||_1 = \lambda^{-1}$). Thus

$$\begin{aligned} |u_\lambda^A(x,y) - u_\lambda^A(x',y)| &\le |u^A(x,y) - u^A(x',y)| + \lambda|U^A v(x) - U^A v(x')| \\ &\le c_3 r^{d_w-d_f} + \lambda c_1 r^{d_w-d_f}||v||_1 \\ &= 2c_3 r^{d_w-d_f}. \end{aligned}$$

Letting $R \to \infty$ we deduce (3.42), and (3.43) then follows, exactly as above, by integration.

To prove (3.46) note first that $p_\lambda(y,x) = u_\lambda(y,x)/u_\lambda(x,x)$. So by (3.33)

$$\begin{aligned} \int_A p_\lambda^A(y,x)|f(y)|\mu(dy) &\le ||f||_\infty u_\lambda(x,x)^{-1}\int_A u_\lambda(y,x)\mu(dy) \\ &= ||f||_\infty u_\lambda(x,x)^{-1}\lambda^{-1} \\ &\le c_4||f||_\infty \lambda^{-d_s/2}. \end{aligned} \tag{3.47}$$

From (3.45) and (3.46) we have

$$|u_\lambda^A(x,y) - u_\lambda^A(x',y)| \le c_2\big(p_\lambda^A(y,x) + p_\lambda^A(y,x')\big)r^{d_w-d_f},$$

and (3.44) then follows by intergation, using (3.47). □

The following modulus of continuity for the local times of X then follows from the results in [MR].

Theorem 3.41. *If $d_s < 2$ then X has jointly continuous local times $(L_t^x, x \in F, t \geq 0)$. Let $\varphi(u) = u^{(d_w - d_f)/2}(\log(1/u))^{1/2}$. The modulus of continuity in space of $L^{\cdot}$ is given by:*

$$\lim_{\delta \downarrow 0} \sup_{0 \leq s \leq t} \sup_{\substack{0 \leq s \leq t \\ |x-y|<\delta}} \frac{|L_s^x - L_s^y|}{\varphi(\rho(x,y))} \leq c(\sup_{x \in F} L_t^x)^{1/2}.$$

It follows that X is space-filling: for each $x, y \in F$ there exists a r.v. T such that $\mathbb{P}^x(T < \infty) = 1$ and

$$B(y,1) \subset \{X_t, 0 \leq t \leq T\}.$$

The following Proposition helps to explain why in early work mathematical physicists found that for simple examples of fractal sets one has $d_s < 2$. (See also [HHW]).

Proposition 3.42. *Let F be a FMS, and suppose F is finitely ramified. Then if X is a $FD(d_f, d_w)$ on F, $d_s(X) < 2$.*

Proof. Let F_1, F_2 be two connected components of F, such that $D = F_1 \cap F_2$ is finite. If $D = \{y_1, \ldots, y_n\}$, fix $\lambda > 0$ and set

$$M_t = e^{-\lambda t} \sum_{i=1}^{n} u_\lambda(X_t, y_i).$$

Then M is a supermartingale. Let $T_D = \inf\{t \geq 0 : X_t \in D\}$, and let $x_0 \in F_1 - D$. Since $\mathbb{P}^{x_0}(X_1 \in F_2) > 0$, we have $\mathbb{P}^{x_0}(T_D \leq 1) > 0$. So

$$\infty > \mathbb{E}^{x_0} M_0 \geq \mathbb{E}^{x_0} M_{T_D},$$

and thus $M_{T_D} < \infty$ a.s. So $u_\lambda(X_{T_D}, y_i) < \infty$ for each $y_i \in D$, and thus we must have $u_\lambda(y_i, y_i) < \infty$ for some $y_i \in D$. So, by Proposition 3.25, $d_s < 2$. $\square$

Remark 3.43. For $k = 1,2$ let (F_k, d_k, μ_k) be FMS with dimension $d_f(k)$, and common diameter r_0. Let $F = F_1 \times F_2$, let $p \geq 1$ and set $d((x_1, x_2), (y_1, y_2)) = (d_1(x_1, y_1)^p + d_2(x_2, y_2)^p)^{1/p}$, $\mu = \mu_1 \times \mu_2$. Then (F, d, μ) is a FMS with dimension $d_f = d_f(1) + d_f(2)$. Suppose that for $k = 1, 2$ X^k is a $FD(d_f(k), d_w(k))$ on F_k. Then if $X = (X^1, X^2)$ it is clear from the definition of FDs that if $d_w(1) = d_w(2) = \beta$ then X is a $FD(d_f, \beta)$ on F. However, if $d_w(1) \neq d_w(2)$ then X is not a FD on F. (Note from (3.3) that the metric ρ can, up to constants, be extracted from the transition density $p(t, x, y)$ by looking at limits as $t \downarrow 0$). So the class of FDs is not stable under products.

This suggests that it might be desirable to consider a wider class of diffusions with densities of the form:

$$p(t,x,y) \simeq t^{-\alpha} \exp\Big(- \sum_{1}^{n} \rho_i(x,y)^{\beta_i \gamma_i} t^{-\gamma_i} \Big), \tag{3.48}$$

where ρ_i are appropriate non-negative functions on $F \times F$. Such processes would have different space-time scalings in the different 'directions' in the set F given by the functions ρ_i. A recent paper of Hambly and Kumagai [HK2] suggests that diffusions on p.c.f.s.s. sets (the most general type of regular fractal which has been studied in detail) have a behaviour a little like this, though it is not likely that the transition density is precisely of the form (3.48).

Spectral properties.

Let X be a FD on a FMS F with diameter $r_0 = 1$. The bounds on the density $p(t,x,y)$ imply that $p(t,.,.)$ has an eigenvalue expansion (see [DaSi, Lemma 2.1]).

Theorem 3.44. *There exist continuous functions φ_i, and λ_i with $0 \le \lambda_0 \le \lambda_1 \le \ldots$ such that for each $t > 0$*

$$p(t,x,y) = \sum_{n=0}^{\infty} e^{-\lambda_n t} \varphi_n(x)\varphi_n(y), \tag{3.49}$$

where the sum in (3.49) is uniformly convergent on $F \times F$.

Remark 3.45. The assumption that X is conservative implies that $\lambda_0 = 0$, while the fact that $p(t,x,y) > 0$ for all $t > 0$ implies that X is irreducible, so that $\lambda_1 > 0$.

A well known argument of Kac (see [Ka, Section 10], and [HS] for the necessary Tauberian theorem) can now be employed to prove that if $N(\lambda) = \#\{\lambda_i : \lambda_i \le \lambda\}$ then there exists c_i such that

$$c_1 \lambda^{d_s/2} \le N(\lambda) \le c_2 \lambda^{d_s/2} \quad \text{for } \lambda > c_3. \tag{3.50}$$

So the number of eigenvalues of $\mathcal{L}$ grows roughly as $\lambda^{d_s/2}$. This explains the term *spectral dimension* for d_s.

4. Dirichlet Forms, Markov Processes, and Electrical Networks.

In this chapter I will give an outline of those parts of the theory of Dirichlet forms, and associated concepts, which will be needed later. For a more detailed account of these, see the book [FOT]. I begin with some general introductory remarks.

Let $X = (X_t, t \ge 0, \mathbb{P}^x, x \in F)$ be a Markov process on a metric space F. (For simplicity let us assume X is a Hunt process). Associated with X are its semigroup $(T_t, t \ge 0)$ defined by

$$T_t f(x) = \mathbb{E}^x f(X_t), \tag{4.1}$$

and its resolvent $(U_\lambda, \lambda > 0)$, given by

$$U_\lambda f(x) = \int_0^\infty T_t f(x) e^{-\lambda t}\, dt = \mathbb{E}^x \int_0^\infty e^{-\lambda s} f(X_s)\, ds. \tag{4.2}$$

While (4.1) and (4.2) make sense for all functions f on F such that the random variables $f(X_t)$, or $\int e^{-\lambda s} f(X_s)\, ds$, are integrable, to employ the semigroup or resolvent usefully we need to find a suitable Banach space $(B, \|\cdot\|_B)$ of functions on F such that $T_t : B \to B$, or $U_\lambda : B \to B$. The two examples of importance here are

$C_0(F)$ and $L^2(F,\mu)$, where μ is a Borel measure on F. Suppose this holds for one of these spaces; we then have that (T_t) satisfies the semigroup property

$$T_{t+s} = T_t T_s, \quad s,t \geq 0,$$

and (U_λ) satisfies the resolvent equation

$$U_\alpha - U_\beta = (\beta - \alpha) U_\alpha U_\beta, \quad \alpha, \beta > 0.$$

We say (T_t) is *strongly continuous* if $\|T_t f - f\|_B \to 0$ as $t \downarrow 0$. If T_t is strongly continuous then the infinitesimal generator $\big(\mathcal{L}, \mathcal{D}(L)\big)$ of (T_t) is defined by

$$\mathcal{L}f = \lim_{t\downarrow 0} t^{-1}(T_t f - f), \quad f \in \mathcal{D}(\mathcal{L}), \tag{4.3}$$

where $\mathcal{D}(L)$ is the set of $f \in B$ for which the limit in (4.3) exists (in the space B). The Hille-Yoshida theorem enables one to pass between descriptions of X through its generator $\mathcal{L}$, and its semigroup or resolvent.

Roughly speaking, if we take the analogy between X and a classical mechanical system, $\mathcal{L}$ corresponds to the equation of motion, and T_t or U_λ to the integrated solutions. For a mechanical system, however, there is another formulation, in terms of conservation of energy. The energy equation is often more convenient to handle than the equation of motion, since it involves one fewer differentiation.

For general Markov processes, an "energy" description is not very intuitive. However, for reversible, or symmetric processes, it provides a very useful and powerful collection of techniques. Let μ be a Radon measure on F: that is a Borel measure which is finite on every compact set. We will also assume μ charges every open set. We say that T_t is *μ-symmetric* if for every bounded and compactly supported f, g,

$$\int T_t f(x) g(x) \mu(dx) = \int T_t g(x) f(x) \mu(dx). \tag{4.4}$$

Suppose now (T_t) is the semigroup of a Hunt process and satisfies (4.4). Since $T_t 1 \leq 1$, we have, writing $(\cdot,\cdot)$ for the inner product on $L^2(F,\mu)$, that

$$|T_t f(x)| \leq \big(T_t f^2(x)\big)^{1/2} \big(T_t 1(x)\big)^{1/2} \leq (T_t f^2(x))^{1/2}$$

by Hölder's inequality. Therefore

$$\|T_t f\|_2^2 \leq \|T_t f^2\|_1 = (T_t f^2, 1) = (f^2, T_t 1) \leq (f^2, 1) = \|f\|_2^2,$$

so that T_t is a contraction on $L^2(F,\mu)$.

The definition of the Dirichlet (energy) form associated with (T_t) is less direct than that of the infinitesimal generator: its less intuitive description may be one reason why this approach has until recently received less attention than those based on the resolvent or infinitesimal generator. (Another reason, of course, is the more restrictive nature of the theory: many important Markov processes are not symmetric. I remark here that it is possible to define a Dirichlet form for non-symmetric Markov processes — see [MR]. However, a weaker symmetry condition, the "sector condition", is still required before this yields very much.)

Let F be a metric space, with a locally compact and countable base, and let μ be a Radon measure on F. Set $H = L^2(F, \mu)$.

Definition 4.1. Let $\mathcal{D}$ be a linear subspace of H. A *symmetric form* $(\mathcal{E}, \mathcal{D})$ is a map $\mathcal{E} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ such that
(1) $\mathcal{E}$ is bilinear
(2) $\mathcal{E}(f, f) \geq 0, \quad f \in \mathcal{D}$.

For $\alpha \geq 0$ define $\mathcal{E}_\alpha$ on $\mathcal{D}$ by $\mathcal{E}_\alpha(f, f) = \mathcal{E}(f, f) + \alpha \|f\|_2^2$, and write

$$\|f\|_{\mathcal{E}_\alpha}^2 = \|f\|_2^2 + \alpha \mathcal{E}(f, f) = \mathcal{E}_\alpha(f, f).$$

Definition 4.2. Let $(\mathcal{E}, \mathcal{D})$ be a symmetric form.
(a) $\mathcal{E}$ is *closed* if $(\mathcal{D}, \|\cdot\|_{\mathcal{E}_1})$ is complete
(b) $(\mathcal{E}, \mathcal{D})$ is *Markov* if for $f \in \mathcal{D}$, if $g = (0 \vee f) \wedge 1$ then $g \in \mathcal{D}$ and $\mathcal{E}(g, g) \leq \mathcal{E}(f, f)$.
(c) $(\mathcal{E}, \mathcal{D})$ is a *Dirichlet form* if $\mathcal{D}$ is dense in $L^2(F, \mu)$ and $(\mathcal{E}, \mathcal{D})$ is a closed, Markov symmetric form.

Some further properties of a Dirichlet form will be of importance:

Definition 4.3. $(\mathcal{E}, \mathcal{D})$ is *regular* if

$$\mathcal{D} \cap C_0(F) \text{ is dense in } \mathcal{D} \text{ in } \|\cdot\|_{\mathcal{E}_1}, \quad \text{and} \tag{4.5}$$

$$\mathcal{D} \cap C_0(F) \text{ is dense in } C_0(F) \text{ in } \|\cdot\|_\infty. \tag{4.6}$$

$\mathcal{E}$ is *local* if $\mathcal{E}(f, g) = 0$ whenever f, g have disjoint support.
$\mathcal{E}$ is *conservative* if $1 \in \mathcal{D}$ and $\mathcal{E}(1, 1) = 0$.
$\mathcal{E}$ is *irreducible* if $\mathcal{E}$ is conservative and $\mathcal{E}(f, f) = 0$ implies that f is constant.

The classical example of a Dirichlet form is that of Brownian motion on $\mathbb{R}^d$:

$$\mathcal{E}_{BM}(f, f) = \tfrac{1}{2} \int_{\mathbb{R}^d} |\nabla f|^2 \, dx, \quad f \in H^{1,2}(\mathbb{R}^d).$$

Later in this section we will look at the Dirichlet forms associated with finite state Markov chains.

Just as the Hille-Yoshida theorem gives a $1-1$ correspondence between semigroups and their generators, so we have a $1-1$ correspondence between Dirichlet forms and semigroups. Given a semigroup (T_t) the associated Dirichlet form is obtained in a fairly straightforward fashion.

Definition 4.4. (a) The semigroup (T_t) is *Markovian* if $f \in L^2(F, \mu)$, $0 \leq f \leq 1$ implies that $0 \leq T_t f \leq 1$ μ-a.e.
(b) A Markov process X on F is *reducible* if there exists a decomposition $F = A_1 \cup A_2$ with A_i disjoint and of positive measure such that $\mathbb{P}^x(X_t \in A_i \text{ for all } t) = 1$ for $x \in A_i$. X is *irreducible* if X is not reducible.

Theorem 4.5. *([FOT, p. 23]) Let $(T_t, t \geq 0)$ be a strongly continuous μ-symmetric contraction semigroup on $L^2(F,\mu)$, which is Markovian. For $f \in L^2(F,\mu)$ the function $\varphi_f(t)$ defined by*

$$\varphi_f(t) = t^{-1}(f - T_t f, f), \quad t > 0$$

is non-negative and non-increasing. Let

$$\mathcal{D} = \{f \in L^2(F,\mu) : \lim_{t \downarrow 0} \varphi_f(t) < \infty\},$$
$$\mathcal{E}(f,f) = \lim_{t \downarrow 0} \varphi_f(t), \quad f \in \mathcal{D}.$$

Then $(\mathcal{E},\mathcal{D})$ is a Dirichlet form. If $\big(\mathcal{L},\mathcal{D}(\mathcal{L})\big)$ is the infinitesimal generator of (T_t), then $\mathcal{D}(\mathcal{L}) \subset \mathcal{D}$, $\mathcal{D}(\mathcal{L})$ is dense in $L^2(F,\mu)$, and

$$\mathcal{E}(f,g) = (-\mathcal{L}f, g), \quad f \in \mathcal{D}(\mathcal{L}),\ g \in \mathcal{D}. \tag{4.7}$$

As one might expect, by analogy with the infinitesimal generator, passing from a Dirichlet form $(\mathcal{E},\mathcal{D})$ to the associated semigroup is less straightforward. Since formally we have $U_\alpha = (\alpha - \mathcal{L})^{-1}$, the relation (4.7) suggests that

$$(f,g) = \big((\alpha - \mathcal{L})U_\alpha f, g\big) = \alpha(U_\alpha f, g) + \mathcal{E}(U_\alpha f, g) = \mathcal{E}_\alpha(U_\alpha f, g). \tag{4.8}$$

Using (4.8), given the Dirichlet form $\mathcal{E}$, one can use the Riesz representation theorem to define $U_\alpha f$. One can verify that U_α satisfies the resolvent equation, and is strongly continuous, and hence by the Hille-Yoshida theorem (U_α) is the resolvent of a semigroup (T_t).

Theorem 4.6. *([FOT, p.18]) Let $(\mathcal{E},\mathcal{D})$ be a Dirichlet form on $L^2(F,\mu)$. Then there exists a strongly continuous μ-symmetric Markovian contraction semigroup (T_t) on $L^2(F,\mu)$, with infinitesimal generator $\big(\mathcal{L},\mathcal{D}(\mathcal{L})\big)$ and resolvent $(U_\alpha, \alpha > 0)$ such that $\mathcal{L}$ and $\mathcal{E}$ satisfy (4.7) and also*

$$\mathcal{E}(U_\alpha f, g) + \alpha(f,g) = (f,g), \quad f \in L^2(F,\mu),\ g \in \mathcal{D}. \tag{4.9}$$

Of course the operations in Theorem 4.5 and Theorem 4.6 are inverses of each other. Using, for a moment, the ugly but clear notation $\mathcal{E} = \text{Thm } 4.5((T_t))$ to denote the Dirichlet form given by Theorem 4.5, we have

$$\text{Thm } 4.6(\text{Thm } 4.5((T_t))) = (T_t),$$

and similarly Thm 4.5(Thm 4.6 $(\mathcal{E})) = \mathcal{E}$.

Remark 4.7. The relation (4.7) provides a useful computational tool to identify the process corresponding to a given Dirichlet form – at least for those who find it more natural to think of generators of processes than their Dirichlet forms. For example, given the Dirichlet form $\mathcal{E}(f,f) = \int |\nabla f|^2$, we have, by the Gauss-Green formula, for $f, g \in C_0^2(\mathbb{R}^d)$, $(-\mathcal{L}f, g) = \mathcal{E}(f,g) = \int \nabla f . \nabla g = -\int g \Delta f$, so that $\mathcal{L} = \Delta$.

We see therefore that a Dirichlet form $(\mathcal{E},\mathcal{D})$ give us a semigroup (T_t) on $L^2(F,\mu)$. But does this semigroup correspond to a 'nice' Markov process? In general it need not, but if $\mathcal{E}$ is regular then one obtains a Hunt process. (Recall that a Hunt process $X = (X_t, t \geq 0, \mathbb{P}^x, x \in F)$ is a strong Markov process with cadlag sample paths, which is quasi-left-continuous.)

Theorem 4.8. *([FOT, Thm. 7.2.1.]) (a) Let $(\mathcal{E}, \mathcal{D})$ be a regular Dirichlet form on $L^2(F, \mu)$. Then there exists a μ-symmetric Hunt process $X = (X_t, t \geq 0, \mathbb{P}^x, x \in F)$ on F with Dirichlet form $\mathcal{E}$.*
(b) In addition, X is a diffusion if and only if $\mathcal{E}$ is local.

Remark 4.9. Let $X = (X_t, t \geq 0, \mathbb{P}^x, x \in \mathbb{R}^2)$ be Brownian motion on $\mathbb{R}^2$. Let $A \subset \mathbb{R}^2$ be a polar set, so that

$$\mathbb{P}^x(T_A < \infty) = 0 \text{ for each } x.$$

Then we can obtain a new Hunt process $Y = (X_t \geq 0, \mathbb{Q}^x, x \in \mathbb{R}^2)$ by "freezing" X on A. Set $\mathbb{Q}^x = \mathbb{P}^x$, $x \in A^c$, and for $x \in A$ let $\mathbb{Q}^x\big(X_t = x, \text{ all } t \in [0, \infty)\big) = 1$. Then the semigroups (T_t^X), (T_t^Y), viewed as acting on $L^2(\mathbb{R}^2)$, are identical, and so X and Y have the same Dirichlet form.

This example shows that the Hunt process obtained in Theorem 4.8 will not, in general, be unique, and also makes it clear that a semigroup on L^2 is a less precise object than a Markov process. However, the kind of difficulty indicated by this example is the only problem — see [FOT, Thm. 4.2.7.]. In addition, if, as will be the case for the processes considered in these notes, all points are non-polar, then the Hunt process is uniquely specified by the Dirichlet form $\mathcal{E}$.

We now interpret the conditions that $\mathcal{E}$ is conservative or irreducible in terms of the process X.

Lemma 4.10. *If $\mathcal{E}$ is conservative then $T_t 1 = 1$ and the associated Markov process X has infinite lifetime.*

Proof. If $f \in \mathcal{D}(\mathcal{L})$ then $0 \leq \mathcal{E}(1 + \lambda f, 1 + \lambda f)$ for any $\lambda \in \mathbb{R}$, and so $\mathcal{E}(1, f) = 0$. Thus $(-\mathcal{L}1, f) = 0$, which implies that $\mathcal{L}1 = 0$ a.e., and hence that $T_t 1 = 1$. □

Lemma 4.11. *If $\mathcal{E}$ is irreducible then X is irreducible.*

Proof. Suppose that X is reducible, and that $F = A_1 \cup A_2$ is the associated decomposition of the state space. Then $T_t 1_{A_1} = 1_{A_1}$, and hence $\mathcal{E}(1_{A_1}, 1_{A_1}) = 0$. As $1 \neq 1_{A_1}$ in $L^2(F, \mu)$ this implies that $\mathcal{E}$ is not irreducible. □

A remarkable property of the Dirichlet form $\mathcal{E}$ is that there is an equivalence between certain Sobolev type inequalities involving $\mathcal{E}$, and bounds on the transition density of the associated process X. The fundamental connections of this kind were found by Varopoulos [V1]; [CKS] provides a good account of this, and there is a very substantial subsequent literature. (See for instance [Co] and the references therein).

We say $(\mathcal{E}, \mathcal{D})$ satisfies a *Nash inequality* if

$$\|f\|_1^{4/\theta}\big(\delta\|f\|_2^2 + \mathcal{E}(f, f)\big) \geq c\|f\|_2^{2+4/\theta}, \quad f \in \mathcal{D}. \tag{4.10}$$

This inequality appears awkward at first sight, and also hard to verify. However, in classical situations, such as when the Dirichlet form $\mathcal{E}$ is the one connected with the Laplacian on $\mathbb{R}^d$ or a manifold, it can often be obtained from an isoperimetric inequality.

In what follows we fix a regular conservative Dirichlet form $(\mathcal{E}, \mathcal{D})$. Let (T_t) be the associated semigroup on $L^2(F, \mu)$, and $X = (X_t, t \geq 0, \mathbb{P}^x)$ be the Hunt process associated with $\mathcal{E}$.

Theorem 4.12. *([CKS, Theorem 2.1]) (a) Suppose $\mathcal{E}$ satisfies a Nash inequality with constants c, δ, θ. Then there exists $c' = c'(c, \theta)$ such that*

$$\|T_t\|_{1\to\infty} \leq c' e^{\delta t} t^{-\theta/2}, \quad t > 0. \tag{4.11}$$

(b) If (T_t) satisfies (4.11) with constants c', δ, θ then $\mathcal{E}$ satisfies a Nash inequality with constants $c'' = c''(c', \theta)$, δ, and θ.

Proof. I sketch here only (a). Let $f \in \mathcal{D}(\mathcal{L})$. Then writing $f_t = T_t f$, and

$$g_{th} = h^{-1}(f_{t+h} - f_t) - T_t \mathcal{L} f,$$

we have $\|g_{th}\|_2 \leq \|g_{0h}\|_2 \to 0$ as $h \to 0$. It follows that $(d/dt) f_t$ exists in $L^2(F, \mu)$ and that

$$\frac{d}{dt} f_t = T_t \mathcal{L} f = \mathcal{L} T_t f.$$

Set $\varphi(t) = (f_t, f_t)$. Then

$$h^{-1}\big(\varphi(t+h) - \varphi(t)\big) - 2(T_t \mathcal{L} f, T_t f) \quad = (g_{th}, f_t + f_{t+h}) + (T_t \mathcal{L} f, f_{t+h} - f_t),$$

and therefore φ is differentiable, and for $t > 0$

$$\varphi'(t) = 2(\mathcal{L} f_t, f_t) = -2\mathcal{E}(f_t, f_t). \tag{4.12}$$

If $f \in L^2(F, \mu)$, $T_t f \in \mathcal{D}(\mathcal{L})$ for each $t > 0$. So (4.12) extends from $f \in \mathcal{D}(\mathcal{L})$ to all $f \in L^2(F, \mu)$.

Now let $f \geq 0$, and $\|f\|_1 = 1$: we have $||f_t||_1 = 1$. Then by (4.10), for $t > 0$,

$$\varphi'(t) = -2\mathcal{E}(f_t, f_t)) \leq 2\delta ||f_t||_2^2 - c\|f_t\|_2^{2+4/\theta} = 2\delta \varphi(t)^2 - c\varphi(t)^{1+2/\theta}. \tag{4.13}$$

Thus φ satisfies a differential inequality. Set $\psi(t) = e^{-2\delta t} \varphi(t)$. Then

$$\psi'(t) \leq -2c\psi(t)^{1+2/\theta} e^{4\delta t/\theta} \leq -2c\psi(t)^{1+2/\theta}.$$

If ψ_0 is the solution of $\psi_0' = -c\psi_0^{1+2/\theta}$ then for some $a \in \mathbb{R}$ we have, for $c_\theta = c_\theta(c, \theta)$,

$$\psi_0(t) = c_\theta (t + a)^{-\theta/2}.$$

If ψ_0 is defined on $(0, \infty)$, then $a \geq 0$, so that

$$\psi_0(t) \leq c_\theta t^{-\theta/2}, \quad t > 0.$$

It is easy to verify that ψ satisfies the same bound – so we deduce that

$$\|T_t f\|_2^2 = e^{2\delta t} \psi(t) \leq c_\theta e^{2\delta t} t^{-\theta/2}, \quad f \in L^2_+, \quad \|f\|_1 = 1. \tag{4.14}$$

Now let f, $g \in L^2_+(F, \mu)$ with $\|f\|_1 = \|g\|_1 = 1$. Then

$$(T_{2t} f, g) = (T_t f, T_t g) \leq \|T_t f\|_2 \|T_t g\|_2 \leq c_\theta^2 e^{\delta 2t} t^{-\theta/2}.$$

Taking the supremum over g, it follows that $\|T_{2t}f\|_\infty \le c_\theta^2 e^{\delta 2t} t^{-\theta/2}$, that is, replacing $2t$ by t, that

$$\|T_t\|_{1\to\infty} \le c_\theta^2 e^{\delta t} t^{-\theta/2}. \qquad \square$$

Remark 4.13. In the sequel we will be concerned with only two cases: either $\delta = 0$, or $\delta = 1$ and we are only interested in bounds for $t \in (0,1]$. In the latter case we can of course absorb the constant $e^{\delta t}$ into the constant c.

This theorem gives bounds in terms of contractivity properties of the semigroup (T_t). If T_t has a 'nice' density $p(t,x,y)$, then $\|T_t\|_{1\to\infty} = \sup_{x,y} p(t,x,y)$, so that (4.11) gives global upper bounds on $p(t,\cdot,\cdot)$, of the kind we used in Chapter 3. To derive these, however, we need to know that the density of T_t has the necessary regularity properties.

So let F, $\mathcal{E}$, T_t be as above, and suppose that (T_t) satisfies (4.11). Write $P_t(x,\cdot)$ for the transition probabilities of the process X. By (4.11) we have, for $A \in \mathcal{B}(F)$, and writing $c_t = ce^{\delta t} t^{-\theta/2}$,

$$P_t(x,A) \le c_t \mu(A) \quad \text{for } \mu\text{-a.a. } x.$$

Since F has a countable base (A_n), we can employ the arguments of [FOT, p.67] to see that

$$P_t(x,A_n) \le c_t \mu(A_n), \quad x \in F - N_t, \tag{4.15}$$

where the set N_t is "properly exceptional". In particular we have $\mu(N_t) = 0$ and

$$\mathbb{P}^x(X_s \in N_t \text{ or } X_{s-} \in N_t \text{ for some } s \ge 0) = 0$$

for $x \in F - N_t$. From (4.15) we deduce that $P_t(x,\cdot) \ll \mu$ for each $x \in F - N_t$. If $s > 0$ and $\mu(B) = 0$ then $P_s(y,B) = 0$ for μ-a.a. y, and so

$$P_{t+s}(x,B) = \int P_s(x,dy) P_t(y,B) = 0, \quad x \in F - N_t.$$

So $P_{t+s}(x,.) \ll \mu$ for all $s \ge 0$, $x \in F - N_t$. So taking a sequence $t_n \downarrow 0$, we obtain a single properly exceptional set $N = \cup_n N_{t_n}$ such that $P_t(x,\cdot) \ll \mu$ for all $t \ge 0$, $x \in F - N$. Write $F' = F - N$: we can reduce the state space of X to F'.

Thus we have for each t, x a density $\tilde{p}(t,x,\cdot)$ of $P_t(x,\cdot)$ with respect to μ. These can be regularised by integration.

Proposition 4.14. *(See [Y, Thm. 2]) There exists a jointly measurable transition density $p(t,x,y)$, $t > 0$, $x,y \in F' \times F'$, such that*

$$P_t(x,A) = \int_A p(t,x,y)\mu(dy) \text{ for } x \in F', \quad t > 0,\ A \in \mathcal{B}(F),$$

$$p(t,x,y) = p(t,y,x) \text{ for all } x,y,t,$$

$$p(t+s,x,z) = \int p(s,x,y)p(t,y,z)\mu(dy) \quad \text{for all } x,z,t,s.$$

Corollary 4.15. *Suppose $(\mathcal{E}, \mathcal{D})$ satisfies a Nash inequality with constants c, δ, θ. Then, for all $x, y \in F'$, $t > 0$,*

$$p(t,x,y) \le c' e^{\delta t} t^{-\theta/2}.$$

We also obtain some regularity properties of the transition functions $p(t, x, \cdot)$. Write $q_{t,x}(y) = p(t, x, y)$.

Proposition 4.16. *Suppose $(\mathcal{E}, \mathcal{D})$ satisfies a Nash inequality with constants c, δ, θ. Then for $x \in F'$, $t > 0$, $q_{t,x} \in \mathcal{D}(\mathcal{L})$, and*

$$\|q_{t,x}\|_2^2 \le c_1 e^{2\delta t} t^{-\theta/2}, \tag{4.16}$$

$$\mathcal{E}(q_{t,x}, q_{t,x}) \le c_2 e^{\delta t} t^{-1-\theta/2}. \tag{4.17}$$

Proof. Since $q_{t,x} = T_{t/2} q_{t/2,x}$, and $q_{t/2,x} \in L^1$, we have $q_{t,x} \in \mathcal{D}(\mathcal{L})$, and the bound (4.16) follows from (4.14).

Fix x, write $f_t = q_{t,x}$, and let $\varphi(t) = \|f_t\|_2^2$. Then

$$\varphi''(t) = \frac{d}{dt}(2\mathcal{L} f_t, f_t) = 4(\mathcal{L} f_t, \mathcal{L} f_t) \ge 0.$$

So, φ' is increasing and hence

$$0 \le \varphi(t) = \varphi(t/2) + \int_{t/2}^{t} \varphi'(s)\, ds \le \varphi(t/2) + (t/2)\varphi'(t).$$

Therefore using (4.13),

$$\mathcal{E}(f_t, f_t) = -\tfrac{1}{2}\varphi'(t) \le t^{-1}\varphi(t/2) \le c e^{\delta t} t^{-1-\theta/2}. \qquad \square$$

Traces of Dirichlet forms and Markov Processes.

Let X be a μ-symmetric Hunt process on a LCCB metric space (F, μ), with semigroup (T_t) and regular Dirichlet form $(\mathcal{E}, \mathcal{D})$. To simplify things, and because this is the only case we need, we assume

$$\operatorname{Cap}(\{x\}) > 0 \text{ for all } x \in F. \tag{4.18}$$

It follows that x is regular for $\{x\}$, for each $x \in F$, that is, that

$$\mathbb{P}^x(T_x = 0) = 1, \quad x \in F.$$

Hence ([GK]) X has jointly measurable local times $(L_t^x, x \in F, t \ge 0)$ such that

$$\int_0^t f(X_s)\, ds = \int_F f(x) L_t^x \mu(dx), \quad f \in L^2(F, \mu).$$

Now let ν be a σ-finite measure on F. (In general one has to assume ν charges no set of zero capacity, but in view of (4.18) this condition is vacuous here). Let A_t be the continuous additive functional associated with ν:

$$A_t = \int L_t^a \nu(da),$$

and let $\tau_t = \inf\{s : A_s > t\}$ be the inverse of A. Let G be the closed support of ν. Let $\widetilde{X}_t = X_{\tau_t}$: then by [BG, p. 212], $\widetilde{X} = (\widetilde{X}_t, \mathbb{P}^x, x \in G)$ is also a Hunt process. We call $\widetilde{X}$ the *trace* of X on G.

Now consider the following operation on the Dirichlet form $\mathcal{E}$. For $g \in L^2(G, \nu)$ set

$$\widetilde{\mathcal{E}}(g,g) = \inf\{\mathcal{E}(f,f) : f|_G = g\}. \tag{4.19}$$

Theorem 4.17. *("Trace theorem": [FOT, Thm. 6.2.1]).*
(a) $(\widetilde{\mathcal{E}}, \widetilde{\mathcal{D}})$ is a regular Dirichlet form on $L^2(G, \nu)$.
(b) $\widetilde{X}$ is ν-symmetric, and has Dirichlet form $(\widetilde{\mathcal{E}}, \widetilde{\mathcal{D}})$.

Thus $\widetilde{\mathcal{E}}$ is the Dirichlet form associated with $\widetilde{X}$: we call $\widetilde{\mathcal{E}}$ the *trace* of $\mathcal{E}$ (on G).

Remarks 4.18. 1. The domain $\widetilde{\mathcal{D}}$ on $\widetilde{\mathcal{E}}$ is of course the set of g such that the infimum in (4.19) is finite. If $g \in \widetilde{\mathcal{D}}$ then, as $\mathcal{E}$ is closed, the infimum in (4.19) is attained, by f say. If h is any function which vanishes on G^c, then since $(f + \lambda h)|_G = g$, we have

$$\mathcal{E}(f,f) \le \mathcal{E}(f + \lambda h, f + \lambda h), \quad \lambda \in \mathbb{R}$$

which implies $\mathcal{E}(f,h) = 0$. So, if $f \in \mathcal{D}(\mathcal{L})$, and we choose $h \in \mathcal{D}$, then $(-h, \mathcal{L}f) = 0$, so that $\mathcal{L}f = 0$ a.e. on G^c.

This calculation suggests that the minimizing function f in (4.19) should be the harmonic extension of g to F; that is, the solution to the Dirichlet problem

$$\begin{aligned} f &= g \qquad \text{on } G \\ \mathcal{L}f &= 0 \qquad \text{on } G^c. \end{aligned}$$

2. We shall sometimes write

$$\widetilde{\mathcal{E}} = \mathrm{Tr}(\mathcal{E}|G)$$

to denote the trace of the Dirichlet form $\mathcal{E}$ on G.
3. Note that taking traces has the "tower property"; if $H \subseteq G \subseteq F$, then

$$\mathrm{Tr}(\mathcal{E}|H) = \mathrm{Tr}\big(\mathrm{Tr}(\mathcal{E}|G) \bigm| H\big).$$

We now look at continuous time Markov chains on a finite state space. Let F be a finite set.

Definition 4.19. A *conductance matrix* on F is a matrix $A = (a_{xy})$, $x, y \in F$, which satisfies

$$a_{xy} \geq 0, \quad x \neq y,$$
$$a_{xy} = a_{yx},$$
$$\sum_y a_{xy} = 0.$$

Set $a_x = \sum_{y \neq x} a_{xy} = -a_{xx}$. Let $E_A = \{\{x, y\} : a_{xy} > 0\}$. We say that A is *irreducible* if the graph (F, E_A) is connected.

We can interpret the pair (F, A) as an electrical network: a_{xy} is the conductance of the wire connecting the nodes x and y. The intuition from electrical circuit theory is on occasion very useful in Markov Chain theory —for more on this see [DS].

Given (F, A) as above, define the Dirichlet form $\mathcal{E} = \mathcal{E}_A$ with domain $C(F) = \{f : F \to \mathbb{R}\}$ by

$$\mathcal{E}(f, g) = \tfrac{1}{2} \textstyle\sum_{x,y} a_{xy} \big(f(x) - f(y)\big)\big(g(x) - g(y)\big). \tag{4.20}$$

Note that, writing $f_x = f(x)$ etc.,

$$\begin{aligned}
\mathcal{E}(f, g) &= \tfrac{1}{2} \sum_x \sum_{y \neq x} a_{xy}(f_x - f_y)(g_x - g_y) \\
&= \sum_x \sum_{y \neq x} a_{xy} f_x g_x - \sum_x \sum_{y \neq x} a_{xy} f_x g_y \\
&= -\sum_x a_{xx} f_x g_x - \sum_x \sum_{y \neq x} a_{xy} f_x g_y \\
&= -\sum_x \sum_y a_{xy} f_x g_y = -f^T A g.
\end{aligned}$$

In electrical terms, (4.20) gives the energy dissipation in the circuit (F, A) if the nodes are held at potential f. (A current $I_{xy} = a_{xy}\big(f(y) - f(x)\big)$ flows in the wire connecting x and y, which has energy dissipation $I_{xy}\big(f(y) - f(x)\big) = a_{xy}\big(f(y) - f(x)\big)^2$. The sum in (4.20) counts each edge twice). We can of course also use this interpretation of Dirichlet forms in more general contexts.

(4.20) gives a 1-1 correspondence between conductance matrices and conservative Dirichlet forms on $C(F)$. Let μ be any measure on F which charges every point.

Proposition 4.20. *(a) If A is a conductance matrix, then $\mathcal{E}_A$ is a regular conservative Dirichlet form.*
(b) If $\mathcal{E}$ is a conservative Dirichlet form on $L^2(F, \mu)$ then $\mathcal{E} = \mathcal{E}_A$ for a conductance matrix A.
(c) A is irreducible if and only if $\mathcal{E}$ is irreducible.

Proof. (a) It is clear from (4.20) that $\mathcal{E}$ is a bilinear form, and that $\mathcal{E}(f, f) \geq 0$. If $g = 0 \vee (1 \wedge f)$ then $|g_x - g_y| \leq |f_x - f_y|$ for all x, y, so since $a_{xy} \geq 0$ for $x \neq y$, $\mathcal{E}$ is Markov. Since $\mathcal{E}(f, f) \leq c(A, \mu)||f||_2^2$, $||.||_{\mathcal{E}_1}$ is equivalent to $||.||_2$, and so $\mathcal{E}$ is closed. It is clear from this that $\mathcal{E}$ is regular.

(b) As $\mathcal{E}$ is a symmetric bilinear form there exists a symmetric matrix A such that $\mathcal{E}(f,g) = -f^T Ag$. Let $f = f_{\alpha\beta} = \alpha 1_x + \beta 1_y$; then

$$\mathcal{E}(f,f) = -\alpha^2 a_{xx} - 2\alpha\beta a_{xy} - \beta^2 a_{yy}.$$

Taking $\alpha = 1$, $\beta = 0$ it follows that $a_{xx} \le 0$. The Markov property of $\mathcal{E}$ implies that $\mathcal{E}(f_{01}, f_{01}) \le \mathcal{E}(f_{\alpha 1}, f_{\alpha 1})$ if $\alpha < 0$. So

$$0 \le -\alpha^2 a_{xx} - 2\alpha a_{xy},$$

which implies that $a_{xy} \ge 0$ for $x \ne y$. Since $\mathcal{E}$ is conservative we have $0 = \mathcal{E}(f,1) = -f^T A1$ for all f. So $A1 = 0$, and therefore $\sum_y a_{xy} = 0$ for all x.

(c) is now evident. □

Example 4.21. Let μ be a measure on F, with $\mu(\{x\}) = \mu_x > 0$ for $x \in F$. Let us find the generator L of the Markov process associated with $\mathcal{E} = \mathcal{E}_A$ on $L^2(F,\mu)$. Let $z \in F$, $g = 1_z$, and $f \in L^2(F,\mu)$. Then

$$\mathcal{E}(f,g) = -g^T Af = -\sum_y a_{zy} f(y) = \sum_y a_{zy}(f(z) - f(y)).$$

and using (4.7) we have, writing $(\cdot,\cdot)_\mu$ for the inner product on $L^2(F,\mu)$,

$$\mathcal{E}(f,g) = (-Lf, g)_\mu = -\mu_z Lf(z).$$

So,

$$Lf(z) = \sum_{x \ne z} (a_{xz}/\mu_z)\big(f(x) - f(z)\big). \tag{4.21}$$

Note from (4.21) that (as we would expect from the trace theorem), changing the measure μ changes the jump rates of the process, but not the jump probabilities.

Electrical Equivalence.

Definition 4.22. Let (F,A) be an electrical network, and $G \subset F$. If B is a conductance matrix on G, and

$$\mathcal{E}_B = \mathrm{Tr}(\mathcal{E}_A | G)$$

we will say that the networks (F,A) and (G,B) are *(electrically) equivalent on* G.

In intuitive terms, this means that an electrician who is able only to access the nodes in G (imposing potentials, or feeding in currents etc.) would be unable to distinguish from the response of the system between the networks (F,A) and (G,B).

Definition 4.23. (Effective resistance). Let G_0, G_1 be disjoint subsets of F. The effective resistance between G_0 and G_1, $R(G_0,G_1)$ is defined by

$$R(G_0,G_1)^{-1} = \inf\{\mathcal{E}(f,f) : f|_{B_0} = 0, f|_{B_1} = 1\}. \tag{4.22}$$

This is finite if (F,A) is irreducible.

If $G = \{x, y\}$, then from these definitions we see that (F, A) is equivalent to the network (G, B), where $B = (b_{xy})$ is given by

$$b_{xy} = b_{yx} = -b_{xx} = -b_{yy} = R(x, y)^{-1}.$$

Let (F, A) be an irreducible network, and $G \subseteq F$ be a proper subset. Let $H = G^c$, and for $f \in C(F)$ write $f = (f_H, f_G)$ where f_H, f_G are the restrictions of f to H and G respectively. If $g \in C(G)$, then if $\widetilde{\mathcal{E}} = \mathrm{Tr}(\mathcal{E}_A|G)$,

$$\widetilde{\mathcal{E}}(g, g) = \inf \left\{ (f_H^T, g^T) A \begin{pmatrix} f_H \\ g \end{pmatrix}, \quad f_H \in C(H) \right\}.$$

We have, using obvious notation

$$(f_H^T, g^T) A \begin{pmatrix} f_H \\ g \end{pmatrix} = f_H^T A_{HH} f_H + 2 f_H^T A_{HG} g + g^T A_{GG} g. \tag{4.23}$$

The function f_H which minimizes (4.23) is given by $f_H = A_{HH}^{-1} A_{HG} g$. (Note that as A is irreducible, 0 cannot be an eigenvalue of A_{HH}, so A_{HH}^{-1} exists). Hence

$$\widetilde{\mathcal{E}}(g, g) = g^T (A_{GG} - A_{GH} A_{HH}^{-1} A_{HG}) g, \tag{4.24}$$

so that $\widetilde{\mathcal{E}} = \mathcal{E}_B$, where B is the conductivity matrix

$$B = A_{GG} - A_{GH} A_{HH}^{-1} A_{HG}. \tag{4.25}$$

Example 4.24. ($\Delta - Y$ transform). Let $G = \{x_0, x_1, x_2\}$ and B be the conductance matrix defined by,

$$b_{x_0 x_1} = \alpha_2, \quad b_{x_1 x_2} = \alpha_0, \quad b_{x_2 x_0} = \alpha_1.$$

Let $F = G \cup \{y\}$, and A be the conductance matrix defined by

$$\begin{aligned} a_{x_i x_j} &= 0, \quad i \neq j, \\ a_{x_i y} &= \beta_i, \quad 0 \leq i \leq 2. \end{aligned}$$

If the α_i and β_i are strictly positive, and we look just at the edges with positive conductance the network (G, B) is a triangle, while (F, A) is a Y with y at the centre. The $\Delta - Y$ transform is that (F, A) and (G, B) are equivalent if and only if

$$\begin{aligned} \alpha_0 &= \frac{\beta_1 \beta_2}{\beta_0 + \beta_1 + \beta_2}, \\ \alpha_1 &= \frac{\beta_2 \beta_0}{\beta_0 + \beta_1 + \beta_2}, \\ \alpha_2 &= \frac{\beta_0 \beta_1}{\beta_0 + \beta_1 + \beta_2}. \end{aligned} \tag{4.26}$$

Equivalently, if $S = \alpha_0 \alpha_1 + \alpha_1 \alpha_2 + \alpha_2 \alpha_0$, then

$$\beta_i = \frac{S}{\alpha_i}, \quad 0 \leq i \leq 2. \tag{4.27}$$

This can be proved by elementary, but slightly tedious, calculations. The $\Delta - Y$ transform can be of great use in reducing a complicated network to a more simple one, though there are of course networks for which it is not effective.

Proposition 4.25. *(See [Ki5]). Let (F, A) be an irreducible electric network, and $R(x, y) = R(\{x\}, \{y\})$ be the 2-point effective resistances. Then R is a metric on F.*

Proof. We define $R(x, x) = 0$. Replacing f by $1 - f$ in (4.22), it is clear that $R(x, y) = R(y, x)$, so it just remains to verify the triangle inequality. Let x_0, x_1, x_2 be distinct points in F, and $G = \{x_0, x_1, x_2\}$.

Using the tower property of traces mentioned above, it is enough to consider the network (G, B), where B is defined by (4.25). Let $\alpha_0 = b_{x_1 x_2}$, and define α_1, α_2 similarly. Let β_0, β_1, β_2 be given by (4.27); using the $\Delta - Y$ transform it is easy to see that

$$R(x_i, x_j) = \beta_i^{-1} + \beta_j^{-1}, \quad i \neq j.$$

The triangle inequality is now immediate. □

Remark 4.26. There are other ways of viewing this, and numerous connections here with linear algebra, potential theory, etc. I will not go into this, except to mention that (4.25) is an example of a Schur complement (see [Car]), and that an alternative viewpoint on the resistance metric is given in [Me6].

The following result gives a connection between resistance and crossing times.

Theorem 4.27. *Let (F, A) be an electrical network, let μ be a measure on F which charges every point, and let $(X_t, t \geq 0)$ be the continuous time Markov chain associated with $\mathcal{E}_A$ on $L^2(F, \mu)$. Write $T_x = \inf\{t > 0 : X_t = x\}$. Then if $x \neq y$,*

$$E^x T_y + E^y T_x = R(x, y)\mu(F). \tag{4.28}$$

Remark. In view of the simplicity of this result, it is rather remarkable that its first appearance (which was in a discrete time context) seems to have been in 1989, in [CRRST]. See [Tet] for a proof in a more accessible publication.

Proof. A direct proof is not hard, but here I will derive the result from the trace theorem. Fix x, y, let $G = \{x, y\}$, and let $\widetilde{\mathcal{E}} = \mathcal{E}_B = \mathrm{Tr}(\mathcal{E}|G)$. If $R = R(x, y)$, then we have, from the definitions of trace and effective resistance,

$$B = \begin{pmatrix} -R^{-1} & R^{-1} \\ R^{-1} & -R^{-1} \end{pmatrix}.$$

Let $\nu = \mu|_G$; the process $\widetilde{X}_t$ associated with $\big(\widetilde{\mathcal{E}}, L^2(G, v)\big)$ therefore has generator given by

$$\widetilde{L}f(z) = (R\mu_z)^{-1} \sum_{w \neq z} \big(f(w) - f(z)\big).$$

Writing $\widetilde{T}_x$, $\widetilde{T}_y$ for the hitting times associated with $\widetilde{X}$ we therefore have

$$E^x \widetilde{T}_y + E^y \widetilde{T}_x = R(\mu_x + \mu_y).$$

We now use the trace theorem. If $f(x) = 1_z(x)$ then the occupation density formula implies that

$$\mu_z L_t^z = \int_0^t 1_z(X_s)\, ds = |\{s \le t : X_s = z\}|.$$

So

$$A_t = \int_0^t 1_G(X_s)\, ds,$$

and thus if $S = \inf\{t \ge T_y : X_t = x\}$ and $\widetilde{S}$ is defined similarly, we have

$$\widetilde{S} = \int_0^S 1_G(X_s)\, ds.$$

However by Doeblin's theorem for the stationary measure of a Markov Chain

$$\mu(G) = (\mathbb{E}^x S)^{-1} \mathbb{E}^x \int_0^S 1_G(X_s)\, ds \mu(F). \tag{4.29}$$

Rearranging, we deduce that

$$\begin{aligned} \mathbb{E}^x S &= \mathbb{E}^x T_y + \mathbb{E}^y T_x \\ &= \big(\mu(F)/\mu(G)\big) \mathbb{E}^x \widetilde{S} \\ &= \big(\mu(F)/\mu(G)\big) \left(\mathbb{E}^x \widetilde{T}_y + \mathbb{E}^y \widetilde{T}_x\right) = R\mu(F). \end{aligned}$$ □

Corollary 4.28. *Let $H \subset F$, $x \notin H$. Then*

$$E^x T_H \le R(x, H)\mu(F).$$

Proof. If H is a singleton, this is immediate from Theorem 4.27. Otherwise, it follows by considering the network (F', H') obtained by collapsing all points in H into one point, h, say. (So $F' = (F - H) \cup \{h\}$, and $a'_{xh} = \sum_{y \in H} a_{xy}$). □

Remark. This result is actually older than Theorem 4.27 – see [Tel].

5. Geometry of Regular Finitely Ramified Fractals.

In Section 2 I introduced the Sierpinski gasket, and gave a direct "hands on" construction of a diffusion on it. Two properties of the SG played a crucial role: its symmetry and scale invariance, and the fact that it is finitely ramified. In this section we will introduce some classes of sets which preserve some of these properties, and such that a similar construction has a chance of working. (It will not always do so, as we will see).

There are two approaches to the construction of a family of well behaved regular finitely ramified fractals. The first, adopted by Lindstrøm [L1], and most of the mathematical physics literature, is to look at fractal subsets of $\mathbb{R}^d$ obtained by generalizations of the construction of the Cantor set. However when we come to study processes on F the particular embedding of F in $\mathbb{R}^d$ plays only a small role, and some quite natural sets (such as the "cut square" described below) have no simple embedding. So one may also choose to adapt an abstract approach, defining a collection of well behaved fractal metric spaces. This is the approach of Kigami [Ki2], and is followed in much of the subsequent mathematical literature on general fractal spaces. ("Abstract" fractals may also be defined as quotient spaces of product spaces – see [Kus2]).

The question of embedding has lead to confusion between mathematicians and physicists on at least one (celebrated) occasion. If G is a graph then the natural metric on G for a mathematician is the standard graph distance $d(x, y)$, which gives the length of the shortest path in G between x and y. Physicists call this the *chemical distance*. However, physicists, thinking in terms of the graph G being a model of a polymer, in which the individual strands are tangled up, are interested in the Euclidean distance between x and y in some embedding of G in $\mathbb{R}^d$. Since they regard each path in G as being a random walk path in $\mathbb{Z}^d$, they generally use the metric $d'(x, y) = d(x, y)^{1/2}$.

In this section, after some initial remarks on self-similar sets in $\mathbb{R}^d$, I will introduce the largest class of regular finitely ramified fractals which have been studied in detail. These are the *pc.f.s.s. sets* of Kigami [Ki2], and in what follows I will follow the approach of [Ki2] quite closely.

Definition 5.1. A map $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a *similitude* if there exists $\alpha \in (0, 1)$ such that $|\psi(x) - \psi(y)| = \alpha|x - y|$ for all x, $y \in \mathbb{R}^d$. We call α the *contraction factor* of ψ.

Let $M \geq 1$, and let $\psi_1, \dots, \psi_M$ be similitudes with contraction factors α_i. For $A \subset \mathbb{R}^d$ set

$$\Psi(A) = \bigcup_{i=1}^{M} \psi_i(A). \tag{5.1}$$

Let $\Psi^{(n)}$ denote the n-fold composition of Ψ.

Definition 5.2. Let $\mathcal{K}$ be the set of non-empty compact subsets of $\mathbb{R}^d$. For $A \subset \mathbb{R}^d$ set $\delta_\varepsilon(A) = \{x : |x - a| \leq \varepsilon$ for some $a \in A\}$. The *Hausdorff metric* d on $\mathcal{K}$ is defined by

$$d(A, B) = \inf \{\varepsilon > 0 : A \subset \delta_\varepsilon(B) \text{ and } B \subset \delta_\varepsilon(A)\}.$$

Lemma 5.3. *(See [Fe, 2.10.21]). (a) d is a metric on $\mathcal{K}$.*
(b) $(\mathcal{K}, d)$ is complete.
(c) If $K_N = \{K \in \mathcal{K} : K \subset \overline{B}(0, N)\}$ then K_N is compact in K.

Theorem 5.4. *Let $(\psi_1, \dots, \psi_M)$ be as above, with $\alpha_i \in (0, 1)$ for each $1 \leq i \leq M$. Then there exists a unique $F \in \mathcal{K}$ such that $F = \Psi(F)$. Further, if $G \in \mathcal{K}$ then $\Psi^n(G) \to F$ in d. If $G \in \mathcal{K}$ satisfies $\Psi(G) \subset G$ then $F = \cap_{n=0}^\infty \Psi^{(n)}(G)$.*

Proof. Note that $\Psi : \mathcal{K} \to \mathcal{K}$. Set $\alpha = \max_i \alpha_i < 1$. If A_i, $B_i \in \mathcal{K}$, $1 \leq i \leq M$ note that

$$d(\cup_{i=1}^M A_i, \cup_{i=1}^M B_i) \leq \max_i d(A_i, B_i).$$

(This is clear since if $\varepsilon > 0$ and $B_i \subset \delta_\varepsilon(A_i)$ for each i, then $\cup B_i \subset \delta_\varepsilon(\cup A_i)$). Thus

$$\begin{aligned} d\big(\Psi(A), \Psi(B)\big) &\leq \max_i d\big(\psi_i(A), \psi_i(B)\big) \\ &= \max_i \alpha_i \, d(A,B) = \alpha d(A,B). \end{aligned}$$

So Ψ is a contraction on $\mathcal{K}$, and therefore has a unique fixed point. For the final assertion, note that if $\Psi(G) \subset G$, then $\Psi^{(n)}(G)$ is decreasing. So $\cap_n \Psi^{(n)}(G)$ is non-empty, and must equal F. □

Examples 5.5. The fractal sets described in Section 2 can all be defined as the fixed point of a map Ψ of this kind.

1. *The Sierpinski gasket.* Let $\{a_1, a_2, a_3\}$ be the 3 corners of the unit triangle, and set

$$\psi_i(x) = a_i + \tfrac{1}{2}(x - a_i), \quad x \in \mathbb{R}^2, \quad 1 \leq i \leq 3. \tag{5.2}$$

2. *The Vicsek Set.* Let $\{a_1, \dots, a_4\}$ be the 4 corners of the unit square, let $M = 5$, let $a_5 = (\frac{1}{2}, \frac{1}{2})$, and let

$$\psi_i(x) = a_i + \tfrac{1}{3}(x - a_i), \quad 1 \leq i \leq 5. \tag{5.3}$$

It is possible to calculate the dimension of the limiting set F from $(\psi_1, \dots, \psi_M)$. However an "non-overlap" condition is necessary.

Definition 5.6. $(\psi_1, \dots, \psi_M)$ satisfies the *open set condition* if there exists an open set U such that $\psi_i(U)$, $1 \leq i \leq M$, are disjoint, and $\Psi(U) \subset U$. Note that, since $\Psi(\overline{U}) \subset \overline{U}$, then the fixed point F of Ψ satisfies $F = \cap \Psi^{(n)}(\overline{U})$.

For the Sierpinski gasket, if H is the convex hull of $\{a_1, a_2, a_3\}$, then one can take $U = \mathrm{int}(H)$.

Theorem 5.7. *Let $(\psi_1, \dots, \psi_M)$ satisfy the open set condition, and let F be the fixed point of Ψ. Let β be the unique real such that*

$$\sum_{i=1}^{M} \alpha_i^\beta = 1. \tag{5.4}$$

Then $\dim_H(F) = \beta$, and $0 < \mathcal{H}^\beta(F) < \infty$.

Proof. See [Fa2, p. 119].

Remark. If $\alpha_i = \alpha$, $1 \leq i \leq M$, then (5.4) simplifies to $M\alpha^\beta = 1$, so that

$$\beta = \frac{\log M}{\log \alpha^{-1}}. \tag{5.5}$$

We now wish to set up an abstract version of this, so that we can treat fractals without necessarily needing to consider their embeddings in $\mathbb{R}^d$. Let (F, d) be a compact metric space, let $I = I_M = \{1, \dots, M\}$, and let

$$\psi_i : F \to F, \quad 1 \leq i \leq M$$

be continuous injections. We wish the copies $\psi_i(F)$ to be strictly smaller than F, and we therefore assume that there exists $\delta > 0$ such that

$$d\big(\psi_i(x), \psi_i(y)\big) \le (1-\delta) d(x,y), \quad x, y \in F, \qquad i \in I_M. \tag{5.6}$$

Definition 5.8. $(F, \psi_i, 1 \le i \le M)$ is a *self-similar structure* if (F, d) is a compact metric space, ψ_i are continuous injections satisfying (5.6) and

$$F = \bigcup_{i=1}^{M} \psi_i(F). \tag{5.7}$$

Let $(F, \psi_i, 1 \le i \le M)$ be a self-similar structure. We can use iterations of the maps ψ_i to give the 'address' of a point in F. Introduce the word spaces

$$\mathbb{W}_n = I^n, \quad \mathbb{W} = I^{\mathbb{N}}.$$

We endow $\mathbb{W}$ with the usual product topology. For $w \in \mathbb{W}_n$, v in $\mathbb{W}_n$ or $\mathbb{W}$, let $w \cdot v = (w_1, \dots, w_n, v_1, \dots)$, and define the left shift σ on $\mathbb{W}$ (or $\mathbb{W}_n$) by

$$\sigma w = (w_2, \dots).$$

For $w = (w_1, ..., w_n) \in \mathbb{W}_n$ define

$$\psi_w = \psi_{w_1} \circ \psi_{w_2} \circ \ldots \circ \psi_{w_n}. \tag{5.8}$$

It is clear from (5.7) that for each $n \ge 1$,

$$F = \bigcup_{w \in \mathbb{W}_n} \psi_w(F).$$

If $a = (a_1, \dots, a_M)$ is a vector indexed by I, we write

$$a_w = \prod_{i=1}^{n} a_{w_i}, \quad w \in \mathbb{W}_n. \tag{5.9}$$

Write $A_w = \psi_w(A)$ for $w \in \cup_n \mathbb{W}_n$, $A \subset F$. If $n \ge 1$, and $w \in \mathbb{W}$ (or $\mathbb{W}_m$ with $m \ge n$) write

$$w|n = (w_1, \dots, w_n) \in \mathbb{W}_n. \tag{5.10}$$

Lemma 5.9. *For each $w \in \mathbb{W}$, there exists a $x_w \in F$ such that*

$$\bigcap_{n=1}^{\infty} \psi_{w|n}(F) = \{x_w\}. \tag{5.11}$$

Proof. Since $\psi_{w|(n+1)}(F) = \psi_{w|n}\big(\psi_{w_{n+1}}(F)\big) \subset \psi_{w|n}(F)$, the sequence of sets in (5.11) is decreasing. As ψ_i are continuous, $\psi_{w|n}(F)$ are compact, and therefore $A = \cap_n F_{w|n}$ is non-empty. But as $\text{diam}(F_{w|n}) \le (1-\delta)^n \text{diam}(F)$, we have $\text{diam}(A) = 0$, so that A consists of a single point. □

Lemma 5.10. *There exists a unique map $\pi : \mathbb{W} \to F$ such that*

$$\pi(i \cdot w) = \psi_i\big(\pi(w)\big), \quad w \in \mathbb{W}, \quad i \in I. \tag{5.12}$$

π is continuous and surjective.

Proof. Define $\pi(w) = x_w$, where x_w is defined by (5.11). Let $w \in \mathbb{W}$. Then for any n,

$$\pi(i \cdot w) \in F_{(i \cdot w)|n} = F_{i \cdot (w|n-1)} = \psi_i(F_{w|n-1}).$$

So $\pi(i \cdot w) \in \cap_m \psi_i(F_m) = \{\psi_i(x_w)\}$, proving (5.12). If π' also satisfies (5.12) then $\pi'(v \cdot w) = \psi_v\big(\pi'(w)\big)$ for $v \in \mathbb{W}_n$, $w \in \mathbb{W}$, $n \geq 1$. Then $\pi'(w) \in F_{w|n}$ for any $n \geq 1$, so $\pi' = \pi$.

To prove that π is surjective, let $x \in F$. By (5.7) there exists $w_1 \in I_M$ such that $x \in F_{w_1} = \psi_{w_1}(F) = \cup_{w_2=1}^M F_{w_1 w_2}$. So there exists w_2 such that $x \in F_{w_1 w_2}$, and continuing in this way we obtain a sequence $w = (w_1, w_2, \dots) \in \mathbb{W}$ such that $x \in F_{w|n}$ for each n. It follows that $x = \pi(w)$.

Let U be open in F, and $w \in \pi^{-1}(U)$. Then $F_{w|n} \cap U^c$ is a decreasing sequence of compact sets with empty intersection, so there exists m with $F_{w|m} \subset U$. Hence $V = \{v \in \mathbb{W} : v|m = w|m\} \subset \pi^{-1}(U)$, and since V is open in $\mathbb{W}$, $\pi^{-1}(U)$ is open. Thus π is continuous. □

Remark 5.11. It is easy to see that (5.12) implies that

$$\pi(v \cdot w) = \psi_v\big(\pi(w)\big), \qquad v \in \mathbb{W}_n, \quad w \in \mathbb{W}. \tag{5.13}$$

Lemma 5.12. *For $x \in F$, $n \geq 0$ set*

$$N_n(x) = \bigcup \{F_w : w \in \mathbb{W}_n, x \in F_w\}.$$

Then $\{N_n(x), n \geq 1\}$ form a base of neighbourhoods of x.

Proof. Fix x and n. If $v \in \mathbb{W}_n$ and $x \notin F_v$ then, since F_v is compact, $d(x, F_v) = \inf\{d(x,y) : y \in F_v\} > 0$. So, as $\mathbb{W}_n$ is finite, $d(x, N_n(x)^c) = \min\{d(x, F_v) : x \notin F_v, v \in \mathbb{W}_n\} > 0$. So $x \in \operatorname{int}(N_n(x))$. Since $\operatorname{diam} F_w \leq (1-\delta)^n \operatorname{diam}(F)$ for $w \in \mathbb{W}_n$ we have $\operatorname{diam} N_n(x) \leq 2(1-\delta)^n \operatorname{diam}(F)$. So if $U \ni x$ is open, $N_n(x) \subset U$ for all sufficiently large n. □

The definition of a self-similar structure does not contain any condition to prevent overlaps between the sets $\psi_i(F)$, $i \in I_M$. (One could even have $\psi_1 = \psi_2$ for example). For sets in $\mathbb{R}^d$ the open set condition prevents overlaps, but relies on the existence of a space in which the fractal F is embedded. A general, abstract, non-overlap condition, in terms of dimension, is given in [KZ1]. However, for finitely ramified sets the situation is somewhat simpler.

For a self-similar structure $\mathcal{S} = (F, \psi_i, i \in I_M)$ set

$$B = B(\mathcal{S}) = \bigcup_{i,j,i \neq j} F_i \cap F_j.$$

As one might expect, we will require $B(\mathcal{S})$ to be finite. However, this on its own is not sufficient: we will require a stronger condition, in terms of the word space $\mathbb{W}$. Set

$$\Gamma = \pi^{-1}\big(B(\mathcal{S})\big),$$
$$P = \bigcup_{n=1}^{\infty} \sigma^n(\Gamma).$$

Definition 5.13. A self-similar structure (F, ψ) is *post critically finite*, or p.c.f., if P is finite. A metric space (F, d) is a *p.c.f.s.s. set* if there exists a p.c.f. self-similar structure $(\psi_i, 1 \le i \le M)$ on F.

Remarks 5.14. 1. As this definition is a little impenetrable, we will give several examples below. The definition is due to Kigami [Ki2], who called Γ the *critical set* of $\mathcal{S}$, and P the *post critical set.*
2. The definition of a self-similar structure given here is slightly less general than that given in [Ki2]. Kigami did not impose the constraint (5.6) on the maps ψ_i, but made the existence and continuity of π an axiom.
3. The initial metric d on F does not play a major role. On the whole, we will work with the natural structure of neighbourhoods of points provided by the self-similar structure and the sets $F_w, w \in \mathbb{W}_n$, $n \ge 0$.

Examples 5.15. 1. *The Sierpinski gasket.* Let a_1, a_2, a_3 be the corners of the unit triangle in $\mathbb{R}^d$, and let

$$\psi_i(x) = a_i + \tfrac{1}{2}(x - a_i), \quad x \in \mathbb{R}^2, \quad 1 \le i \le 3.$$

Write G for the Sierpinski gasket; it is clear that $(G, \psi_1, \psi_2, \psi_3)$ is a self-similar structure. Writing $\dot{s} = (s, s, \ldots)$, we have

$$\pi(\dot{s}) = a_s, \quad 1 \le s \le 3.$$

So

$$B(\mathcal{S}) = \left\{ \tfrac{1}{2}(a_3 + a_1), \quad \tfrac{1}{2}(a_1 + a_2), \quad \tfrac{1}{2}(a_2 + a_3) \right\},$$
$$\Gamma = \left\{ (1\dot{3}), (3\dot{1}), (1\dot{2}), (2\dot{1}), (2\dot{3}), (3\dot{2}) \right\},$$

and

$$P = \sigma(\Gamma) = \left\{ (\dot{1}), (\dot{2}), (\dot{3}) \right\}.$$

2. *The cut square.* This is an example of a p.c.f.s.s. set which has no convenient embedding in Euclidean space. (Though of course such an embedding can certainly be found).

Start with the unit square $C_0 = [0,1]^2$. Now make 'cuts' along the line $L_1 = \{(\frac{1}{2}, y) : 0 < y < \frac{1}{2}\}$, and the 3 similar lines (L_2, L_3, L_4 say) obtained from L_1 by rotation. So the set C_1 consists of C_0, but with the points in the line segment $(\frac{1}{2}, y-)$, $(\frac{1}{2}, y+)$, viewed as distinct, for $0 < y < \frac{1}{2}$. (And similarly for the 3 similar sets obtained by rotation). Alternatively, C_1 is the closure of $A = C_0 - \cup_{i=1}^4 L_i$ in the geodesic metric d_A defined in Section 2. One now repeats this construction on each of the 4 squares of side $\frac{1}{2}$ which make up C_1 to obtain successively C_2, $C_3, \ldots$; the cut square C is the limit.

This is a p.c.f.s.s. set; one has $M = 4$, and if $a_1, \dots, a_4$ are the 4 corners of $[0,1]^2$, then the maps ψ_i agree at all points with irrational coordinates with the maps $\varphi_i(x) = a_i + \frac{1}{2}(x - a_i)$. We have

$$B = \{(0, \tfrac{1}{2}), (\tfrac{1}{2}, \tfrac{1}{2}), (1, \tfrac{1}{2}), (\tfrac{1}{2}, 0), (\tfrac{1}{2}, 1)\}$$
$$\Gamma = \{(1\dot{2}), (2\dot{1}), (2\dot{3}), (3\dot{2}), (3\dot{4}), (4\dot{3}), (4\dot{1}), (1\dot{4}), (1\dot{3}), (3\dot{1}), (2\dot{4}), (4\dot{2})\},$$

so that

$$P = \{(\dot{1}), (\dot{2}), (\dot{3}), (\dot{4})\}.$$

Note also that $\pi(1\dot{2}) = \pi(2\dot{1})$, and $\pi(1\dot{3}) = \pi(3\dot{1}) = \pi(2\dot{4}) = \pi(4\dot{2}) = z$, the centre of the square.

In both the examples above we had $P = \{(\dot{s}), s \in I_M\}$, and $P = \sigma^n P$ for all $n \geq 1$. However P can take a more complicated form if the sets $\psi_i(F)$, $\psi_j(F)$ overlap at points which are sited at different relative positions in the two sets.

3. *Sierpinski gasket with added triangle*. (See [Kum2]). We describe this set as a subset of $\mathbb{R}^2$. Let $\{a_1, a_2, a_3\}$ be the corners of the unit triangle in $\mathbb{R}^2$, and let $\psi_i(x) = \frac{1}{2}(x - a_i) + a_i$, $1 \leq i \leq 3$. Let $a_4 = \frac{1}{3}(a_1 + a_2 + a_3)$ be the centre of the triangle, and let $\psi_4(x) = a_4 + \frac{1}{4}(x - a_4)$. Of course (ψ_1, ψ_2, ψ_3) gives the Sierpinski gasket, but $\Psi = (\psi_1, \psi_2, \psi_3, \psi_4)$ still satisfies the open set condition, and if $F = F(\Psi)$ is the fixed point of Ψ then $(F, \psi_1, \dots, \psi_4)$ is a self-similar structure. Writing b_1, b_2, b_3 for the mid-points of (a_2, a_3), (a_3, a_1), (a_1, a_2) respectively, and $c_i = \frac{1}{2}(a_i + b_i)$, $1 \leq i \leq 3$, we have

$$B = \{b_1, b_2, b_3, c_1, c_2, c_3\},$$

$\pi^{-1}(b_1) = \{(2\dot{3}), (3\dot{2})\}$, while $\pi^{-1}(c_1) = \{(12\dot{3}), (13\dot{2}), (4\dot{1})\}$, with similar expressions for $\pi^{-1}(b_j)$, $\pi^{-1}(c_j)$, $j = 2, 3$. So $\#(\Gamma) = 15$, and

$$\sigma(\Gamma) = \{(\dot{1}), (\dot{2}), (\dot{3}), (2\dot{3}), (3\dot{2}), (3\dot{1}), (1\dot{3}), (1\dot{2}), (2\dot{1})\},$$
$$\sigma^2(\Gamma) = \{(\dot{1}), (\dot{2}), (\dot{3})\}.$$

Then $P = \sigma(\Gamma)$ consists of 9 points in $\mathbb{W}$, and $\#(\pi(P)) = 6$.

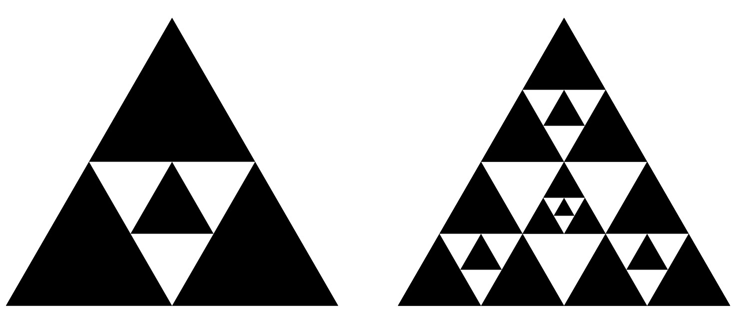

Fig. 5.1 : Sierpinski gasket with added triangle.

4. (*Rotated triangle*). Let a_i, b_i, ψ_i, $1 \leq i \leq 3$, be as above. Let $\lambda \in (0, 1)$, and let $p_1 = \lambda b_2 + (1 - \lambda) b_3$, with p_2, p_3 defined similarly. Evidently $\{p_1, p_2, p_3\}$ is an

equilateral triangle; let ψ_4 be the similitude such that $\psi_4(a_i) = p_i$. Let $F = F(\Psi)$ be the fixed point of Ψ. If H is the convex hull of $\{a_1, a_2, a_3\}$, then $\Psi(H) \subset H$, so clearly F is finitely ramified, and

$$B = \{b_1, b_2, b_3, p_1, p_2, p_3\}.$$

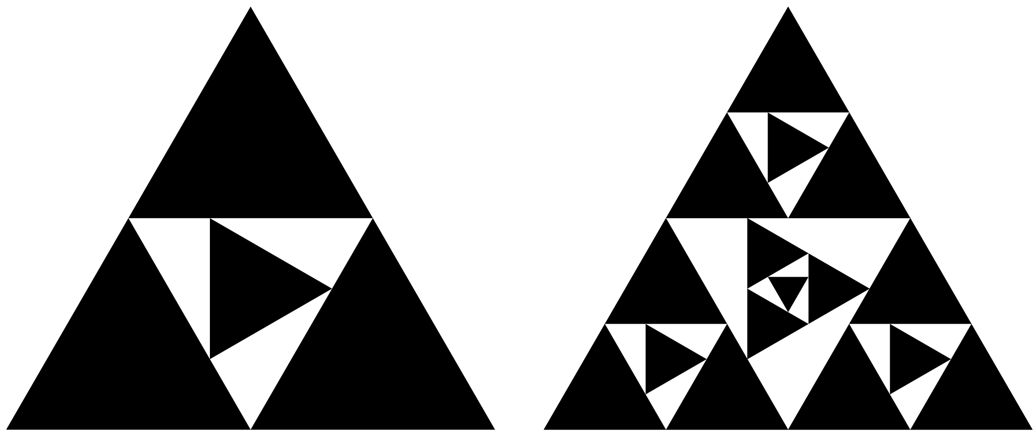

Fig. 5.2 : Rotated triangle with $\lambda = 2/3$.

As before, $\pi^{-1}(b_1) = \{(2\dot{3}), (3\dot{2})\}$. Let $y_1 = \psi_1^{-1}(p_1)$; then y_1 lies on the line segment connecting a_2 and a_3. If $A = \pi^{-1}(y_1)$ then A consists of one or two points, according to whether λ is a dyadic rational or not. Let $A = \{v, w\}$, where $v = w$ if $\lambda \notin \mathbb{D}$. Note that for each element $u \in A$, we have, writing $u = (u_1, u_2, \ldots)$, that $u_k \in \{2, 3\},\ \ k \geq 1$. Then $\pi^{-1}(p_1) = \{(4\dot{1}), (1 \cdot v), (1 \cdot w)\}$. If $\theta : \mathbb{W} \to \mathbb{W}$ is defined by $\theta(w) = w'$, where $w'_i = w_i + 1 \pmod 3$, and

$$A_n = \{(\dot{1}), \sigma^n v, \sigma^n w\},$$

then $\sigma^n(\Gamma) = A_n \cup \theta(A_n) \cup \theta^2(A_n)$.

(a) $\lambda = \frac{1}{2}$ gives Example 3 above.

(b) If λ is irrational, then $P = \cup_{n \geq 1} \sigma^n(\Gamma)$ is infinite. This example therefore shows that the "p.c.f." condition in Definition 5.13 is strictly stronger than the requirement that the set F be finitely ramified and self-similar.

(c) Let $\lambda = \frac{2}{3}$. Then $v = w = (\dot{2}\dot{3})$. Therefore B consists of p_1 and b_1, with their rotations, and $\sigma(L)$ consists of $(2\dot{3})$, $(3\dot{2})$, $(4\dot{1})$, $(123\dot{2}\dot{3})$ and their "rotations" by θ. Hence

$$P = \left\{(\dot{1}), (\dot{2}), (\dot{3}), (\dot{2}\dot{3}), (\dot{3}\dot{2}), (\dot{3}\dot{1}), (\dot{1}\dot{3}), (\dot{1}\dot{2}), (\dot{2}\dot{1})\right\}.$$

So $\lambda = \frac{2}{3}$ does give a p.c.f.s.s. set.

(d) In general, as is clear from the examples above, while F is finitely ramified for any $\lambda \in (0,1)$, F is a p.c.f.s.s. set if and only if $\lambda \in \mathbb{Q} \cap (0,1)$.

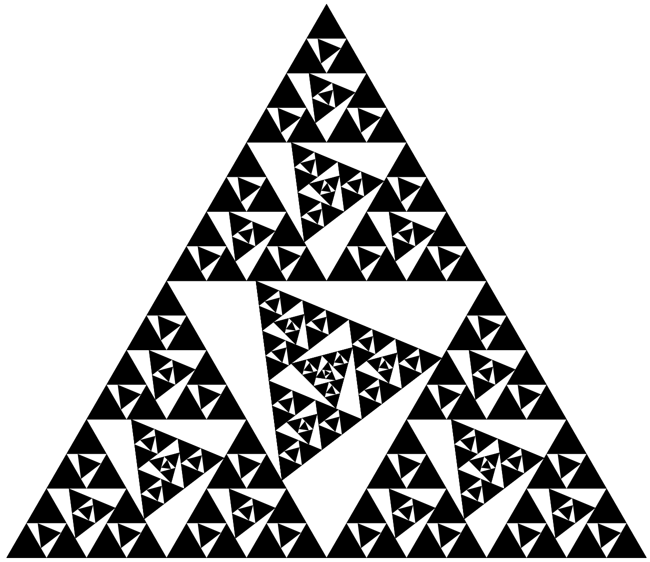

Fig. 5.3 : Rotated triangle with $\lambda = 0.721$.

We now introduce some more notation.

Definition 5.16. Let $(F, \psi_1, \dots, \psi_M)$ be a p.c.f.s.s. set. Set for $n \ge 0$,

$$P^{(n)} = \{w \in \mathbb{W} : \sigma^n w \in P\},$$
$$V^{(n)} = \pi(P^{(n)}).$$

Any set of the form F_w, $w \in \mathbb{W}_n$, we call an *n-complex*, and any set of the form $\psi_w(V^{(0)}) = V_w^{(0)}$ we call a *n-cell*.

Lemma 5.17. *(a) Let $x \in V^{(n)}$. Then $x = \psi_w(y)$, where $y \in V^{(0)}$ and $w \in \mathbb{W}_n$.*
(b) $V^{(n)} = \cup_{w \in \mathbb{W}_n} V_w^{(0)}$.

Proof. (a) From the definition, $x = \pi(w \cdot v)$, for $w \in \mathbb{W}_n$, $v \in \mathbb{W}$. Then if $y = \pi(v)$, $y \in V^{(0)}$, and by (5.13), $x = \pi(w \cdot v) = \psi_w(y)$.
(b) Let $x \in V_w^{(0)}$. Then $x = \psi_w\big(\pi(v)\big)$, where $v \in P$. Hence $x = \pi(w \cdot v)$, and since $w \cdot v \in P^{(n)}$, $x \in V^{(n)}$. The other inclusion follows from (a). □

We think of $V^{(0)}$ as being the "boundary" of the set F. The set F consists of the union of M^n n-complexes F_w (where $w \in \mathbb{W}_n$), which intersect only at their boundary points.

Lemma 5.18. *(a) If w, $v \in \mathbb{W}_n$, $w \ne v$, then $F_w \cap F_v = V_w^{(0)} \cap V_v^{(0)}$.*
(b) If $n \ge 0$, $\pi^{-1}\big(\pi(P^{(n)})\big) = \pi^{-1}(V^{(n)}) = P^{(n)}$.

Proof. (a) Let $n \ge 1$, v, $w \in \mathbb{W}_n$, and $x \in F_w \cap F_v$. So $x = \pi(w \cdot u) \ne \pi(v \cdot u')$ for $u, u' \in \mathbb{W}$. Suppose first that $w_1 \ne v_1$. Then as $F_w \subset F_{w_1}$, we have $x \in F_{w_1} \cap F_{v_1} \subset B$. So $w \cdot u$, $v \cdot u' \in \Gamma$, and thus $u = \sigma^{n-1}\sigma(w \cdot u) \in P$. Therefore $\pi(u) \in V^{(0)}$,

and $x = \psi_w\big(\pi(u)\big) \in V_w^{(0)}$. If $w_1 = v_1$ then let k be the largest integer such that $w|k = v|k$. Applying $\psi_{w|k}^{-1}$ we can then use the argument above.
(b) It is elementary that $P^{(n)} \subset \pi^{-1}\big(\pi(P^{(n)})\big)$. Let $n = 0$ and $w \in \pi^{-1}\big(\pi(P)\big)$. Then there exists $v \in P$ such that $\pi(w) = \pi(v)$. As $v \in P$, $v \in \sigma^m(\Gamma)$ for some $m \geq 1$. Hence there exists $u \in \mathbb{W}_m$ such that $u \cdot v \in \pi^{-1}(B)$. However $\pi(u \cdot w) = \psi_u\big(\pi(w)\big) = \pi(u \cdot v) \in B$, and thus $u \cdot v \in \sigma$. Hence $v \in P$.

If $n \geq 1$, and $\pi(w) \in \pi\big(P^{(n)}\big) = V^{(n)}$, then $\pi(w) \in V_v^{(0)}$ for some $v \in \mathbb{W}_n$. So $\pi(w) \in V_v^{(0)} \cap F_{w|n} = V_v^{(0)} \cap V_{w|n}^{(0)}$ by (a). Therefore $\pi(w) \in V_{w|n}^{(0)}$, and thus $\pi(w) = \psi_{w|n}\big(\pi(v)\big)$, where $v \in P$. So $\pi(w) = \pi(w|n \cdot v)$, and thus $\pi(\sigma^n w) = \pi(v)$. By the case $n = 0$ above $\sigma^n w \in P$, and hence $w \in P^{(n)}$. □

Remark 5.19. Note we used the fact that $\pi(v \cdot w) = \pi(v \cdot w')$ implies $\pi(w) = \pi(w')$, which follows from the fact that ψ_v is injective.

Lemma 5.20. *Let $s \in \{1, \ldots, M\}$. Then $\pi(\dot{s})$ is in exactly one n-complex, for each $n \geq 1$.*

Proof. Let $n = 1$, and write $x_s = \pi(\dot{s})$. Plainly $x_s \in F_s$; suppose $x_s \in F_i$ where $i \neq s$. Then $x_s = \psi_i\big(\pi(w)\big)$ for some $w \in \mathbb{W}$. Since $x_s = \psi_s^k(x_s)$ for any $k \geq 1$, $x_s = \psi_s^k\big(\pi(i \cdot w)\big) = \pi(s^k \cdot i \cdot w)$, where $s^k = (s, s, \ldots, s) \in \mathbb{W}_k$. Since $x_s \in F_i \cap F_s \subset B$, $\pi^{-1}(x_s) \in C$. But therefore $s^k \cdot i \cdot w \in C$ for each $k \geq 1$, and since $i \neq s$, C is infinite, a contradiction.

Now let $n \geq 2$, and suppose $x_s = \pi(\dot{s}) \in F_w$, where $w \in \mathbb{W}_n$ and $w \neq s^n$. Let $0 \leq k \leq n-1$ be such that $w = s^k \cdot \sigma^k w$, and $w_{k+1} \neq s$. Then applying ψ_s^{-k} to F_{s^k} we have that $x_s \in F_{\sigma^k w} \cap F_{s^{n-k}}$, which contradicts the case $n = 1$ above. □

Let $(F, \psi_1, \ldots, \psi_M, \pi)$ be a p.c.f.s.s. set. For $x \in F$, let

$$m_n(x) = \#\{w \in \mathbb{W}_n : x \in F_w\}$$

be the n-multiplicity of x, that is the number of distinct n-complexes containing x. Plainly, if $x \notin \cup_n V^{(n)}$, then $m_n(x) = 1$ for all n. Note also that $m_\cdot(x)$ is increasing.

Proposition 5.21. *For all $x \in F$, $n \geq 1$,*

$$m_n(x) \leq M\#(P).$$

Proof. Suppose $x \in F_{w^1} \cap \ldots \cap F_{w^k}$, where w^i, $1 \leq i \leq k$ are distinct elements of $\mathbb{W}_n$. Suppose first that $w_1^i \neq w_1^j$ for some $i \neq j$. Then $x \in B$, and therefore there exist $v^1, \ldots, v^k \in \mathbb{W}$ such that $\pi(w^l \cdot v^l) = x$, $1 \leq l \leq k$. Hence $w^l \cdot v^l \in \Gamma$ for each l, and so $\#(\Gamma) \geq k$. But $\#(P) \geq M^{-1}\#(\Gamma)$, and thus $k \leq M\#(P)$.

If all the w^l contain a common initial string v, then applying ψ_v^{-1} we can use the argument above. □

Nested Fractals and Affine Nested fractals.

Nested fractals were introduced by Lindstrøm [L1], and affine nested fractals (ANF) by [FHK]. These are of p.c.f.s.s. sets, but have two significant additional properties:

(1) They are embedded in Euclidean space,
(2) They have a large symmetry group.

I will first present the definition of an ANF, and then relate it to that for p.c.f.s.s. sets. Let $\psi_1, \ldots, \psi_M$ be similitudes in $\mathbb{R}^d$, and let F be the associated compact set. Writing ψ_i also for the restrictions of ψ_i to F, $(F, \psi_1, \ldots, \psi_M)$ is a self similar structure. Let $\mathbb{W}$, π, $V^{(0)}$, etc. be as above. For $x, y \in V^{(0)}$ let $g_{xy} : \mathbb{R}^d \to \mathbb{R}^d$ be reflection in the hyperplane which bisects the line segment connecting x and y. As each ψ_i is a contraction, it has a unique fixed point, z_i say. Let $\overline{V} = \{z_1, ..., z_M\}$ be the set of fixed points. Call $x \in \overline{V}$ an *essential fixed point* if there exists $y \in \overline{V}$, and $i \neq j$ such that $\psi_i(x) = \psi_j(y)$. Write $\overline{V}^{(0)}$ for the set of essential fixed points. Set also

$$\overline{V}^{(n)} = \bigcup_{w \in \mathbb{W}_n} \overline{V}^{(0)}.$$

Definition 5.22. $(F, \psi_1, \ldots, \psi_M)$ is an *affine nested fractal* if $\psi_1, \ldots, \psi_M$ satisfy the open set condition, $\#(\overline{V}^{(0)}) \geq 2$, and

(A1) (Connectivity) For any i, j there exists a sequence of 1-cells $V_{i_0}^{(0)}, \ldots, V_{i_k}^{(0)}$ such that $i_0 = i$, $i_k = j$ and $\overline{V}_{i_{r-1}}^{(0)} \cap \overline{V}_{i_r}^{(0)} \neq \emptyset$ for $1 \leq r \leq k$.
(A2) (Symmetry) For each $x, y \in \overline{V}^{(0)}$, $n \geq 0$, g_{xy} maps n cells to n cells.
(A3) (Nesting) If $w, v \in \mathbb{W}_n$ and $w \neq v$ then

$$F_w \cap F_v = \overline{V}_w^{(0)} \cap \overline{V}_v^{(0)}.$$

In addition $(F, \psi_1, \ldots, \psi_M)$ is a *nested fractal* if the ψ_i all have the same contraction factor.

If ψ_i has contraction factor α_i, then by (5.4) $\dim_H(F) = \beta$, where β solves

$$\sum_{i=1}^{M} \alpha_i^\beta = 1. \tag{5.14}$$

If $\alpha_i = \alpha$, so that F is a nested fractal, then

$$\dim_H(F) = \frac{\log M}{\log(1/\alpha)}. \tag{5.15}$$

Following Lindstrøm we will call M the *mass scale factor*, and $1/\alpha$ the *length scale factor*, of the nested fractal F.

Lemma 5.23. *Let $(F, \psi_1, \ldots, \psi_M)$ be an affine nested fractal. Write z_i for the fixed point of ψ_i. Then $z_i \notin F_j$ for any $j \neq i$.*

Proof. Suppose that $z_1 \in F_2$. Then by Definition 5.22(A3) $F_1 \cap F_2 = \overline{V}_1^{(0)} \cap \overline{V}_2^{(0)}$, so $z_1 \in \overline{V}_2^{(0)}$, and $z_1 = \psi_2(z_i)$, for some $z_i \in \overline{V}^{(0)}$. We cannot have $i = 2$, as $\psi_2(z_2) = z_2 \neq z_1$. Also, if $i = 1$ then ψ_2 would fix both z_1 and z_2, so could not be a contraction. So let $i = 3$. Therefore for any $k \geq 0$, $i \geq 0$,

$$z_1 = \psi_1^k \circ \psi_2 \circ \psi_3^i(z_3) \in F_{1^k.2.3^i}.$$

Write $D_n = \{w \in \mathbb{W}_n : z_1 \in F_w\}$: by the above $\#(D_n) \geq n$. Let U be the open set given by the open set condition. Since $F \subset \overline{U}$ we have $z_i \in \overline{U}$ for each i. So $z_1 \in \overline{U}_w$ for each $w \in D_n$, while the open set condition implies that the sets $\{U_w, w \in D_n\}$ are disjoint. So z_1 is on the boundary of at least n disjoint open sets. If (as is true for nested fractals) all these sets are congruent then a contradiction is almost immediate.

For the general case of affine nested fractals we need to work a little harder to obtain the same conclusion. Let $a > 0$ be such that

$$|B(z_i, 1) \cap U| > a \quad \text{for each } i.$$

Let α_i, $1 \leq i \leq M$ be the contraction factors of the ψ_i. Recall the notation $\alpha_w = \Pi_{i=1}^n \alpha_{w_i}$, $w \in \mathbb{W}_n$. Set $\delta = \min_{w \in D_n} \alpha_w$, and let $\beta = \min_i \alpha_i$. For each $w \in D_n$ let $w' = w \cdot 1...1$ be chosen so that $\beta\delta < \alpha_{w'} \leq \delta$. Then $z_1 \in F_{w'} \subset \overline{U}_{w'}$, for each $w \in D_n$, and the sets $\{U_{w'}, w \in D_n\}$ are still disjoint. (Since $\Psi(U) \subset U$ we have $U_{w'} \subset U_w$ for each $w \in D_n$).

Now if $w \in D_n$ then if j is such that $z_1 = \psi_{w'}(z_j)$

$$|B(z_1, \delta) \cap U_{w'}| = \alpha_{w'}^d |B(z_j, \delta/\alpha_{w'}) \cap U| \geq (\beta\delta)^d |B(z_j, 1) \cap U| \geq a(\beta\delta)^d.$$

So

$$c_d \delta^d = |B(z_1, \delta)| \geq \sum_{w \in D_n} |B(z_1, \delta) \cap U_{w'}| \geq na(\beta\delta)^d.$$

Choosing n large enough this gives a contradiction. □

Proposition 5.24. *Let $(F, \psi_1, \ldots, \psi_M)$ be an affine nested fractal. Write z_i for the fixed point of ψ_i. Then $(F, \psi_1, \ldots, \psi_M)$ is a p.c.f.s.s. set, and*

(a) $\overline{V}^{(0)} = V^{(0)}$,
(b) $P = \left\{ (\dot{s}) : z_s \in \overline{V}^{(0)} \right\}$.
(c) If $z \in V^{(0)}$ then z is in exactly one n-complex for each $n \geq 1$.
(d) Each 1-complex contains at most one element of $V^{(0)}$.

Proof. It is clear that $(F, \psi_1, \ldots, \psi_M)$ is a self-similar structure. Relabelling the ψ_i, we can assume $\overline{V}^{(0)} = \{z_1, \ldots, z_k\}$ where $2 \leq k \leq M$. We begin by calculating B, Γ and P. It is clear from (A3) that

$$B = \bigcup_{s \neq t} (\overline{V}_s^{(0)} \cap \overline{V}_t^{(0))}.$$

Let $w \in \Gamma$. Then $\pi(w) \in B$, so (as $\pi(w) \in F_{w_1}$) $\pi(w) \in \overline{V}^{(0)}_{w_1}$, and therefore $\pi(\sigma w) \in \overline{V}^{(0)}$. Say $\pi(\sigma w) = z_s$, where $s \in \{1, .., k\}$. Then since $z_s \in F_{w_2}$, by Lemma 5.23 we must have $w_2 = s$. So $\psi_s\big(\pi(\sigma^2 w)\big) = \pi(s \cdot \sigma^2 w) = \pi(\sigma w) = z_s$, and therefore $\pi(\sigma^2 w) = z_s$. So $w_3 = s$, and repeating we deduce that $\sigma w = (\dot{s})$. Therefore $\{\sigma w, w \in \Gamma\} = \{(\dot{s}), 1 \le s \le k\}$. This proves (b); as P is finite $(F, \psi_1, \dots, \psi_M)$ is a p.c.f.s.s. set. (a) is immediate, since $\pi(P) = V^{(0)} = \{\pi(\dot{s})\} = \overline{V}^{(0)}$.
(c) This is now immediate from (a), (b) and Lemma 5.23.
(d) Suppose F_i contains z_s and z_t, where $s \ne t$. Then one of s, t is distinct from i – suppose it is s. Then $z_s \in F_s \cap F_i$, which contradicts (c). □

Remarks 5.25. 1. Of the examples considered above, the SG is a nested fractal and the SG with added triangle is an ANF. The cut square is not an ANF, since if it were, the maps $\psi_i : \mathbb{R}^d \to \mathbb{R}^d$ would preserve the plane containing its 4 corners, and then the nesting axiom fails. The rotated triangle fails the symmetry axiom unless $\lambda = 1/2$. The Vicsek set defined in Section 2 is a nested fractal, but the Sierpinski carpet fails the nesting axiom.

2. The simplest examples of p.c.f.s.s. sets, and nested fractals can be a little misleading. Note the following points:
(a) Proposition 5.24(c) fails for p.c.f.s.s. sets. See for example the SG with added triangle, where $V^{(0)}$ contains the points $\{b_1, b_2, b_3\}$ as well as the corners $\{a_1, a_2, a_3\}$, and each of the points b_i lies in 2 distinct 1-cells.
(b) This example also shows that for a general p.c.f.s.s. set it is possible to have $F - V^{(0)}$ disconnected even if F is connected.
(c) Let $V_i^{(0)}$ and $V_j^{(0)}$ be two distinct 1-cells in a p.c.f.s.s. set. Then one can have $\#\big(V_i^{(0)} \cap V_j^{(0)}\big) \ge 2$. (The cut square is an example of this). For nested fractals, I do not know whether it is true that

$$\#(V_i^{(0)} \cap V_j^{(0)}) \le 1 \quad \text{if } i \ne j. \tag{5.16}$$

In [FHK, Prop. 2.2(4)] it is asserted that (5.16) holds for affine nested fractals, quoting a result of J. Murai: however, the result of Murai was proved under stronger hypotheses. While much of the work on nested fractals has assumed that (5.16) holds, this difficulty is not a serious one, since only minor modifications to the definitions and proofs in the literature are needed to handle the general case.

3. The symmetry hypothesis (A2) is very strong. We have

$$g_{xy} : V^{(0)} \to V^{(0)} \quad \text{for all} \quad x \ne y, \qquad x, y \in V^{(0)}. \tag{5.17}$$

The question of which sets $V^{(0)}$ satisfy (5.17) leads one into questions concerning reflection groups in $\mathbb{R}^d$. It is easy to see that $V^{(0)}$ satisfies (5.17) if $V^{(0)}$ is a regular planar polygon, a d-dimensional tetrahedron or a d-dimensional simplex. (That is, the set $V^{(0)} = \{e_i, -e_i, 1 \le i \le d\} \subset \mathbb{R}^d$, where $e_i = (\delta_{1i}, \dots, \delta_{di})$. I have been assured by two experts in this area that these are the only possibilities, and my web page see `(http://www.math.ubc.ca/)` contains a letter from G. Maxwell with a sketch of a proof of this fact.

Note that the cube in $\mathbb{R}^3$ fails to satisfy (5.17).

4. Note also that if F is a nested fractal in $\mathbb{R}^d$, and $V^{(0)} \subset H$ where H is a k-dimensional subspace, one does not necessarily have $F \subset H$. This is the case of the Koch curve, for example. (See [L1, p. 39]).

Example 5.26. (Lindstrøm snowflake). This nested fractal is the "classical example", used in [L1] as an illustration of the axioms. It may be defined briefly as follows. Let z_i, $1 \le i \le 6$ be the vertices of a regular hexagon in $\mathbb{R}^2$, and let $z_7 = \frac{1}{6}(z_1 + \dots z_6)$ be the centre. Set

$$\psi_i(x) = z_i + \tfrac{1}{3}(x - z_i), \qquad 1 \le i \le 7.$$

It is easy to verify that this set satisfies the axioms (A1)–(A3) above.

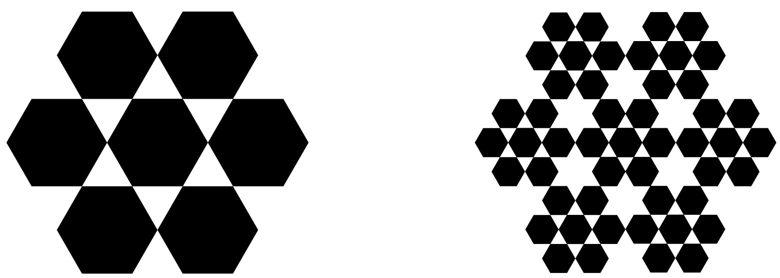

Fig. 5.4. Lindstrøm snowflake.

Measures on p.c.f.s.s. sets.

The structure of these sets makes it easy to define measures which have good properties relative to the maps ψ_i. We begin by considering measures on $\mathbb{W}$. Let $\theta = (\theta_1, \dots, \theta_M)$ satisfy

$$\sum_{i=1}^{M} \theta_i = 1, \qquad 0 < \theta_i < 1 \quad \text{for each} \quad i \in I_M.$$

Recall the notation $\theta_w = \prod_{i=1}^n \theta_{w_i}$ for $w \in \mathbb{W}_n$. We define the measure $\tilde\mu_\theta$ on $\mathbb{W}$ to be the natural product measure associated with the vector θ. More precisely, let $\xi_n : \mathbb{W} \to I_M$ be defined by $\xi_n(w) = w_n$; then $\tilde\mu_\theta$ is the measure which makes (ξ_n) i.i.d. random variables with distribution given by $\mathbb{P}(\xi_n = r) = \theta_r$. Note that for any $n \ge 1$, $w \in \mathbb{W}_n$,

$$\tilde\mu_\theta\left(\{v \in \mathbb{W} : v|n = w\}\right) = \prod_{i=1}^{n} \theta_{w_i}. \tag{5.18}$$

Definition 5.27. Let $\mathcal{B}(F)$ be the σ-field of subsets of F generated by the sets $\{F_w, w \in \mathbb{W}_n, n \ge 1\}$. (By Lemma 5.12 this is the Borel σ-field). For $A \in \mathcal{B}(F)$, set

$$\mu(A) = \tilde\mu\big(\pi^{-1}(A)\big).$$

Then for $w \in \mathbb{W}_n$

$$\mu_\theta(F_w) = \tilde{\mu}_\theta\big(\pi^{-1}(F_w)\big) = \tilde{\mu}_\theta\left(\{v : v|n = w\}\right) = \theta_w = \prod_{i=1}^{n} \theta_{w_i}. \tag{5.19}$$

In contexts when θ is fixed we will write μ for μ_θ.

Remark. If $(F, \psi_1, \ldots, \psi_M)$ is a nested fractal, then the sets $\psi_i(F)$, $1 \le i \le M$ are congruent, and it is natural to take $\theta_i = M^{-1}$. More generally, for an ANF, the 'natural' θ is given by

$$\theta_i = \alpha_i^\beta,$$

where β is defined by (5.4).

The following Lemma summarizes the self-similarity of μ in terms of the space $L^1(F, \mu)$.

Lemma 5.28. *Let $f \in L^1(F, \mu)$. Then for $n \ge 1$*

$$\int_F f\, d\mu = \sum_{w \in \mathbb{W}_n} \theta_w \int (f \circ \psi_w)\, d\mu, \qquad n \ge 1. \tag{5.20}$$

Proof. It is sufficient to prove (5.20) in the case $n = 1$: the general case then follows by iteration. Write $G = F - V^{(0)}$. Note that $G_v \cap G_w = \emptyset$ if v, $w \in \mathbb{W}_n$ and $v \ne w$. As μ is non-atomic we have $\mu(F_w) = \mu(G_w)$ for any $w \in \mathbb{W}_n$. Let $f = 1_{G_w}$ for some $w \in \mathbb{W}_n$. Then $f \circ \psi_i = 0$ if $i \ne w_1$, and $f \circ \psi_{w_1} = 1_{G_{\sigma w}}$. Thus

$$\int (f \circ \psi_i)\, d\mu = \mu(G_{\sigma w}) = \theta_{w_1}^{-1} \mu(G_w) = \theta_{w_1}^{-1} \int f d\mu,$$

proving (5.20) for this particular f. The equality then extends to L^1 by a standard argument. □

We will also need related measures on the sets $V^{(n)}$. Let $N_0 = \# V^{(0)}$. Fix θ and set

$$\mu_n(x) = N_0^{-1} \sum_{w \in \mathbb{W}_n} \theta_w 1_{V_w^{(0)}}(x), \quad x \in V^{(n)}. \tag{5.21}$$

Lemma 5.29. *μ_n is a probability measure on $V^{(n)}$ and*

$$\text{wlim}_{n\to\infty} \mu_n = \mu_\theta.$$

Proof. Since $\# V_w^{(0)} = N_0$ we have

$$\mu_n(V^{(n)}) = \sum_{x \in V^{(n)}} N_0^{-1} \sum_{w \in \mathbb{W}_n} \theta_w 1_{V_w^{(0)}}(x) = \sum_{w \in \mathbb{W}_n} \theta_w = 1,$$

proving the first assertion.

We may regard μ_n as being derived from μ by shifting the mass on each n-complex F_w to the boundary $V_w^{(0)}$, with an equal amount of mass being moved to each point. (So a point $x \in V_w^{(0)}$ obtains a contribution of θ_w from each n-complex it belongs to). So if $f : F \to \mathbb{R}$ then

$$\left| \int_F f d\mu - \int_F f d\mu_n \right| \leq \max_{w \in \mathbb{W}_n} \sup_{x,y \in F_w} |f(x) - f(y)| \tag{5.22}$$

It follows that $\mu_n \xrightarrow{w} \mu_\theta$. □

Symmetries of p.c.f.s.s. sets.

Definition 5.30. Let $\mathcal{G}$ be a group of continuous bijections from F to F. We call $\mathcal{G}$ a *symmetry group* of F if
(1) $g : V^{(0)} \to V^{(0)}$ for all $g \in \mathcal{G}$.
(2) For each $i \in I$, $g \in \mathcal{G}$ there exists $j \in I$, $g' \in \mathcal{G}$ such that

$$g \circ \psi_i = \psi_j \circ g'. \tag{5.23}$$

Note that if g, h satisfy (5.23) then

$$\begin{aligned}(g \circ h) \circ \psi_i &= g \circ (h \circ \psi_i) = g \circ (\psi_j \circ h') = (g \circ \psi_j) \circ h' \\ &= (\psi_k \circ g') \circ h' = \psi_k \circ g'',\end{aligned}$$

for some $j, k \in I$, $g', h', g'' \in \mathcal{G}$. The calculation above also shows that if $\mathcal{G}_1$ and $\mathcal{G}_2$ are symmetry groups then the group generated by $\mathcal{G}_1$ and $\mathcal{G}_2$ is also a symmetry group. Write $\mathcal{G}(F)$ for the largest symmetry group of F. If $\mathcal{G}$ is a symmetry group, and $g \in \mathcal{G}$ write $\tilde{g}(i)$ for the unique element $j \in I$ such that (5.23) holds.

Lemma 5.31. *Let $g \in \mathcal{G}$. Then for each $n \geq 0$, $w \in \mathbb{W}_n$, there exist $v \in \mathbb{W}_n$, $g' \in \mathcal{G}$ such that $g \circ \psi_w = \psi_v \circ g'$. In particular $g : V^{(n)} \to V^{(n)}$.*

Proof. The first assertion is just (5.23) if $n = 1$. If $n \geq 1$, and the assertion holds for all $v \in \mathbb{W}_n$ then if $w = i \cdot v \in \mathbb{W}_{n+1}$ then

$$g \circ \psi_w = g \circ \psi_i \circ \psi_v = \psi_j \circ g' \circ \psi_v = \psi_j \circ \psi_{v'} \circ g'',$$

for $j \in I$, $g', g'' \in \mathcal{G}$. □

Proposition 5.32. *Let $(F, \psi_1, \ldots, \psi_M)$ be an ANF. Let $\mathcal{G}_1$ be the set of isometries of $\mathbb{R}^d$ generated by reflections in the hyperplanes bisecting the line segments $[z_i, z_j]$, $i \neq j$, $z_i, z_j \in V^{(0)}$. Let $\mathcal{G}_0$ be the group generated by $\mathcal{G}_1$. Then $\mathcal{G}_R = \{g|_F : g \in \mathcal{G}_0\}$ is a symmetry group of F.*

Proof. If $g \in \mathcal{G}_1$ then $g : V^{(n)} \to V^{(n)}$ for each n and hence also $g : F \to F$. Let $i \in I$: by the symmetry axiom (A2) $g(V_i^{(0)}) = V_j^{(0)}$ for some $j \in I$. For each of the possible forms of $V^{(0)}$ given in Remark 5.25(3), the symmetry group of $V^{(0)}$ is generated by the reflections in $\mathcal{G}_1$. So, there exists $g' \in \mathcal{G}_0$ such that $g \circ \psi_i = \psi_j \circ g'$. Thus (5.23) is verified for each $g \in \mathcal{G}_1$, and hence (5.23) holds for all $g \in \mathcal{G}_0$. □

Remark 5.33. In [BK] the collection of 'p.c.f. morphisms' of a p.c.f.s.s. set was introduced. These are rather different from the symmetries defined here since the definition in [BK] involved 'analytic' as well as 'geometric' conditions.

Connectivity Properties.

Definition 5.34. Let F be a p.c.f.s.s. set. For $n \geq 0$, define a graph structure on $V^{(n)}$ by taking $\{x,y\} \in \mathbf{E}_n$ if $x \neq y$, and $x, y \in V_w^{(0)}$ for some $w \in \mathbb{W}_n$.

Proposition 5.35. *Suppose that $(V^{(1)}, \mathbf{E}_1)$ is connected. Then $(V^{(n)}, \mathbf{E}_n)$ is connected for each $n \geq 2$, and F is pathwise connected.*

Proof. Suppose that $(V^{(n)}, \mathbf{E}_n)$ is connected, where $n \geq 1$. Let $x, y \in V^{(n+1)}$. If $x, y \in V_w^{(1)}$ for some $w \in \mathbb{W}_n$, then, since $(V^{(1)}, \mathbf{E}_1)$ is connected, there exists a path $\psi_w^{-1}(x) = z_0, z_1, \dots, z_k = \psi_w^{-1}(y)$ in $\left(V^{(1)}, \mathbf{E}_1\right)$ connecting $\psi_w^{-1}(x)$ and $\psi_w^{-1}(y)$. We have $z_{i-1},\ z_i \in V_{w_i}^{(0)}$ for some $w_i \in \mathbb{W}_1$, for each $1 \leq i \leq k$. Then if $z_i' = \psi_w(z_i)$, $z_{i-1}',\ z_i' \in F_{w_i \cdot w}$ and so $\{z_{i-1}', z_i\} \in \mathbf{E}_{n+1}$. Thus x, y are connected by a path in $(V^{(n+1)}, \mathbf{E}_{n+1})$.

For general $x,\ y \in V^{(n+1)}$, as $(V^{(n)}, \mathbf{E}_n)$ is connected there exists a path $y_0, \dots, y_m$ in $(V^{(n)}, \mathbf{E}_n)$ such that $\{y_{i-1}, y_i\} \in \mathbf{E}_n$ and x, y_0, and y, y_m, lie in the same $n+1$-cell. Then, by the above, the points $x, y_0, y_1, \dots, y_m, y$ can be connected by chains of edges in $\mathbf{E}_{n+1}$.

To show that F is path-connected we actually construct a continuous path $\gamma : [0,1] \to F$ such that $F = \{\gamma(t), t \in [0,1]\}$. Let $x_0, \dots, x_N$ be a path in $(V^{(1)}, \mathbf{E}_1)$ which is "space-filling", that is such that $V^{(1)} \subset \{x_0, \dots, x_N\}$. Define $\gamma(i/N) = x_i$, $A_1 = \{i/N,\ 0 \leq i \leq N\}$. Now $x_0, x_1 \in V_w^{(0)}$, for some $w \in \mathbb{W}_1$. Let $x_0 = y_0, y_1, \dots, y_m = x_1$ be in a space-filling path in $\left(V_w^{(1)}, \mathbf{E}_2\right)$. Define $\gamma(k/Nm) = y_k$, $0 \leq k \leq m$. Continuing in this way we fill each of the sets $V_w^{(1)}$, $w \in \mathbb{W}_1$, and so can define $A_2 \subset [0,1]$ such that $A_1 \subset A_2$, and $\gamma(t)$, $t \in A_2$ is a space filling path in the graph $(V^{(2)}, \mathbf{E}_2)$. Repeating this construction we obtain an increasing sequence (A_n) of finite sets such that $\gamma(t)$, $t \in A_n$ is a space filling path in $(V^{(n)}, \mathbf{E}_n)$, and $\cup_n A_n$ is dense in $[0,1]$. If $t \in A_n$, and $t' < t < t''$ are such that $(t', t'') \cap A_n = \{t\}$, then $\gamma(s)$ is in the same n-complex as $\gamma(t)$ for $s \in (t', t'')$. So, if $t \in [0,1] - A$, and $s_n, t_n \in A_n$ are chosen so that $s_n < t < t_n$, $(s_n, t_n) \cap A_n = \emptyset$, then the points $\gamma(u)$, $u \in A \cap (s,t)$ all lie in the same n-complex. So defining $\gamma(t) = \lim_n \gamma(t_n)$, we have that the limit exists, and γ is continuous. The construction of γ also gives that γ is space filling; if $w \in \mathbb{W}$ then for any $n \geq 1$ a section of the path, $\gamma(s)$, $a_n \leq s \leq b_n$, $s \in A_n$, fills $V_{w|n}^{(0)}$.

It follows immediately from the existence of γ that F is pathwise connected. □

Remark. This proof returns to the roots of the subject – the original papers of Sierpinski [Sie1, Sie2] regarded the Sierpinski gasket and Sierpinski carpet as "curves".

Corollary 5.36. *Any ANF is pathwise connected.*

Remark 5.37. If F is a p.c.f.s.s. set, and the graph $(V^{(1)}, \mathbf{E}_1)$ is not connected, then it is easy to see that F is not connected.

For the case of ANFs, we wish to examine the structure of the graphs $(V^{(n)}, \mathbf{E}_n)$ a little more closely. Let $(F, \psi_1, \dots, \psi_M)$ be an ANF. Then let

$$a = \min \left\{ |x - y| : \ x, y \in V^{(0)},\ x \neq y \right\},$$

and set

$$\mathbf{E}_0' = \big\{\{x,y\} \in V^{(0)} : |x-y| = a\big\},$$
$$\mathbf{E}_n' = \Big\{\{x,y\} \in \mathbf{E}_n : x = \psi_w(x'), y = \psi_w(y') \quad \text{for some} \quad w \in \mathbb{W}_n,\ \{x',y'\} \in \mathbf{E}_0'\Big\},\ n \geq 1.$$

Proposition 5.38. *Let F be an ANF.*
(a) Let $x, y, z \in V^{(0)}$ be distinct points. Then there exists a path in $(V^{(0)}, \mathbf{E}_0')$ connecting x and y and not containing z.
(b) Let $x, y \in V^{(0)}$. There exists a path in $(V^{(1)}, \mathbf{E}_1')$ connecting x, y which does not contain any point in $V^{(0)} - \{x,y\}$.
(c) Let $x, y, x', y' \in V^{(0)}$ with $|x-y| = |x'-y'|$. Then there exists $g \in \mathcal{G}_R$ such that $g(x') = x$, $g(y') = y$.

Proof. If $\#\big(V^{(0)}\big) = 2$ then $\mathbf{E}_0 = \mathbf{E}_0'$, so (a) is vacuous and (b) is immediate from Corollary 5.36. So suppose $\#\big(V^{(0)}\big) \geq 3$.
(a) Since (see Remark 5.25(3)) $V^{(0)}$ is either a d-dimensional tetrahedron, or a d-dimensional simplex, or a regular polygon, this is evident. (For a proof which does not use this fact, see [L1, p. 34–35]).
(b) This now follows from (a) by the same kind of argument as that given in Proposition 5.35.
(c) Write $g[x,y]$ for the reflection in the hyperplane bisecting the line segment $[x,y]$. Let $g_1 = g[y,y']$, and $z = g_1(x')$. Then if $z = x$ we are done. Otherwise note that $|x-y| = |x'-y'| = |z-y|$, so if $g_2 = g[x,z]$ then $g_2(y) = y$. Hence $g_1 \circ g_2$ works. □

Metrics on Nested Fractals.

Nested fractals, and ANFs, are subsets of $\mathbb{R}^d$, and so of course are metric spaces with respect to the Euclidean metric. Also, p.c.f.s.s. sets have been assumed to be metric spaces. However, these metrics do not necessarily have all properties we would wish for, such as the mid-point property that was used in Section 3. We saw in Section 2 that the geodesic metric on the Sierpinski gasket was equivalent to the Euclidean metric, but for a general nested fractal there may be no path of finite length between distinct points. (It is easy to construct examples). It is however, still possible to construct a geodesic metric on a ANF.

For simplicity, we will just treat the case of nested fractals. Let $(F, (\psi_i)_{i=1}^M)$ be a nested fractal, with length scale factor L. Write $d_n(x,y)$ for the natural graph distance in the graph $(V^{(n)}, \mathbf{E}_n)$. Fix $x_0,\ y_0 \in V^{(0)}$ such that $\{x_0, y_0\} \in \mathbf{E}_0'$, and let $a_n = d_n(x_0, y_0)$, and b_0 be the maximum distance between points in $(V^{(0)}, \mathbf{E}_0')$.

Lemma 5.39. *If $x, y \in V^{(0)}$ then $a_n \leq d_n(x,y) \leq b_0 a_n$.*

Proof. Since x, y are connected by a path of length at most b_0 in $(V^{(0)}, \mathbf{E}_0')$, the upper bound is evident. Fix x, y, and let $k = d_n(x,y)$. If $\{x,y\} \in \mathbf{E}_0'$ then $d_n(x,y) = d_n(x_0,y_0) = a_n$, so suppose $\{x,y\} \notin \mathbf{E}_0'$. Choose $y' \in V^{(0)}$ such that $\{x, y'\} \in \mathbf{E}_0'$, let H be the hyperplane bisecting $[y,y']$ and let g be reflection in H. Write A, A' for the components of $\mathbb{R}^d - H$ containing y, y' respectively. As $|x-y'| < |x-y|$ we have $x \in A'$. Let $x = z_0, z_1, \dots, z_k = y$ be the shortest path in $(V^{(n)}, \mathbf{E}_n)$ connecting x and y. Let $j = \min\{i : z_i \in A\}$, and write $z_i' = z_i$ if $i < j$, $z_i' = g(z_i)$

if $i \geq j$. Then z_i', $0 \leq i \leq k$ is a path in $(V^{(n)}, \mathbf{E}_n)$ connecting x and y', and so $d_n(x,y) = k \geq d_n(x,y') = a_n$. □

Lemma 5.40. *Let $x, y \in V^{(n)}$. Then for $m \geq 0$*

$$a_m d_n(x,y) \leq d_{n+m}(x,y) \leq b_0 a_m d_n(x,y). \tag{5.24}$$

In particular

$$a_n a_m \leq a_{n+m} \leq b_0 a_n a_m, \quad n \geq 0,\ m \geq 0. \tag{5.25}$$

Proof. Let $k = d_n(x,y)$, and let $x = z_0, z_1, \dots, z_k = y$ be a shortest path connecting x and y in $(V^{(n)}, \mathbf{E}_n)$. Then since by Lemma 5.39 $d_m(z_{i-1}, z_i) \leq b_0 a_m$, the upper bound in (5.24) is clear.

For the lower bound, let $r = d_{n+m}(x,y)$, and let $(z_i)_{i=0}^r$ be a shortest path in $(V^{(n+m)}, \mathbf{E}_{n+m})$ connecting x, y. Let $0 = i_0, i_1, \dots, i_s = r$ be successive disjoint hits by this path on $V^{(n)}$. (Recall the definition from Section 2: of course it makes sense for a deterministic path as well as a process). We have $s = d_n(x,y) \geq a_n$. Then since $z_{i_{j-1}}, z_{i_j}$ lie in the same n-cell, $i_j - i_{j-1} = d_m(z_{i_{j-1}}, z_{i_j}) \geq a_m$, by Lemma 5.39. So $r = \sum_{j=1}^s (i_j - i_{j-1}) \geq a_n a_m$. □

Corollary 5.41. *There exists $\gamma \in [L, b_0 a_1]$ such that*

$$b_0^{-1} \gamma^n \leq a_n \leq \gamma^n. \tag{5.26}$$

Proof. Note that $\log(b_0 a_n)$ is a subadditive sequence, and that $\log a_n$ is superadditive. So by the general theory of these sequences there exist θ_0, θ_1 such that

$$\theta_0 = \lim_{n\to\infty} n^{-1} \log(b_0 a_n) = \inf_{n\geq 0} n^{-1} \log(b_0 a_n),$$

$$\theta_1 = \lim_{n\to\infty} n^{-1} \log(a_n) = \sup_{n\geq 0} n^{-1} \log(a_n).$$

So $\theta_0 = \theta_1$, and setting $\gamma = e^{\theta_0}$, (5.26) follows.

To obtain bounds on γ note first that as $a_n \leq b_0 a_1 a_{n-1}$ we have $\gamma \leq b_0 a_1$. Also,

$$|x_0 - y_0| \leq a_n L^{-n} |x_0 - y_0|,$$

so $\gamma \geq L$. □

Definition 5.42. We call $d_c = \log \gamma / \log L$ the *chemical exponent* of the fractal F, and γ the *shortest path scaling factor.*

Theorem 5.43. *There exists a metric d_F on F with the following properties.*
(a) There exists $c_1 < \infty$ such that for each $n \geq 0$, $w \in \mathbb{W}_n$,

$$d_F(x,y) \leq c_1 \gamma^{-n} \text{ for } x, y \in F_w, \tag{5.27}$$

and

$$d_F(x,y) \geq c_2 \gamma^{-n} \text{ for } x \in V^{(n)},\ y \in N_n(x)^c. \tag{5.28}$$

(b) d_F induces the same topology on F as the Euclidean metric.

(c) d_F has the midpoint property.
(d) The Hausdorff dimension of F with respect to the metric d_F is

$$d_f(F) = \frac{\log M}{\log \gamma}. \tag{5.29}$$

Proof. Write $V = \cup_n V^{(n)}$. By Lemma 5.41 for $x, y \in V$ we have

$$b_0^{-1} \gamma^m d_n(x,y) \le d_{n+m}(x,y) \le b_0 \gamma^m d_n(x,y). \tag{5.30}$$

So $(\gamma^{-m} d_{n+m}(x,y), m \ge 0)$ is bounded above and below. By a diagonalization argument we can therefore find a subsequence $n_k \to \infty$ such that

$$d_F(x,y) = \lim_{k\to\infty} \gamma^{-n_k} d_{n_k}(x,y) \text{ exists for each } x, y \in V.$$

So, if $x, y \in V_w^{(0)}$ where $w \in \mathbb{W}_n$ then

$$c_0^{-1} \gamma^{-n} \le d_F(x,y) \le c_0 \gamma^{-n}. \tag{5.31}$$

It is clear that d_F is a metric on V.

Let $n \ge 0$ and $y \in V^{(n)}$. For $m = n-1, n-2, \dots, 0$ choose inductively $y_m \in V^{(m)}$ such that y_m is in the same m-cell as $y_{m+1}, ..., y_n$. Then

$$d_{m+1}(y_m, y_{m+1}) \le \max\{d_1(x', y') : x', y' \in V^{(1)}\} = c < \infty.$$

So by (5.30) $d_n(y_m, y_{m+1}) \le b_0 \gamma^{n-(m+1)} c$, and therefore

$$d(y_k, y) \le c \sum_{i=k}^{\infty} \gamma^{-i-1} = c' \gamma^{-k}.$$

So if $x, y \in V$ are in the same k-cell, choosing x_k in the same way we have

$$d_F(x,y) \le d_F(x, x_k) + d_F(x_k, y_k) + d_F(y_k, y) \le c_1 \gamma^k, \tag{5.32}$$

since $d_k(x_k, y_k) \le b_0$. Thus d_F is uniformly continuous on $V \times V$, and so extends by continuity to a metric d_F on F. (a) is immediate from (5.31).

If $x, y \in V^{(n)}$ and $x \ne y$ then $d_F(x,y) \ge b_0^{-1} \gamma^{-n}$. This, together with (5.30), implies (b).

If $x, y \in V^{(n)}$ then there exists $z \in V^{(n)}$ such that

$$|d_n(u,z) - \tfrac{1}{2} d_n(x,y)| \le 1, \quad u = x, y.$$

So the metrics d_n have an approximate midpoint property: (c) follows by an easy limiting argument.

Let μ be the measure on F associated with the vector $\theta = (M^{-1}, ..., M^{-1})$. Thus $\mu(F_w) = M^{-|w|}$ for each $w \in \mathbb{W}_n$. Since we have $\text{diam}_{d_F}(F_w) \asymp \gamma^{-|w|}$, it follows that, writing $d_f = \log M / \log \gamma$,

$$c_5 r^{d_f} \le \mu(B_{d_F}(x,r)) \le c_6 r^{d_f}, \quad x \in F$$

and the conclusion then follows from Corollary 2.8. □

Remark 5.44. The results here on the metric d_F are not the best possible. The construction here used a subsequence, and did not give a procedure for finding the scale factor γ. See [BS], [Kum2], [FHK], [Ki6] for more precise results.

6. Renormalization on Finitely Ramified Fractals.

Let $(F, \psi_1, \dots, \psi_M)$ be a p.c.f.s.s. set. We wish to construct a sequence $Y^{(n)}$ of random walks on the sets $V^{(n)}$, nested in a similar fashion to the random walks on the Sierpinski gasket considered in Section 2. The example of the Vicsek set shows that, in general, some calculation is necessary to find such a sequence of walks. As the random walks we treat will be symmetric, we will find it convenient to use the theory of Dirichlet forms, and ideas from electrical networks, in our proofs.

Fix a p.c.f.s.s. set $\big(F, (\psi_i)_{i=1}^M\big)$, and a Bernouilli measure $\mu = \mu_\theta$ on F, where each $\theta_i > 0$. We also choose a vector $r = (r_1, \dots, r_M)$ of positive "weights": loosely speaking r_i is the size of the set $\psi_i(F) = F_i$, for $1 \le i \le M$. We call r a *resistance vector*.

Definition 6.1. Let $\mathbb{D}$ be the set of Dirichlet forms $\mathcal{E}$ defined on $C\big(V^{(0)}\big)$. From Section 4 we have that each element $\mathcal{E} \in \mathbb{D}$ is of the form $\mathcal{E}_A$, where A is a conductance matrix. Let also $\mathbb{D}_1$ be the set of Dirichlet forms on $C\big(V^{(1)}\big)$.

We consider two operations on $\mathbb{D}$:

(1) Replication – i.e. extension of $\mathcal{E} \in \mathbb{D}$ to a Dirichlet form $\mathcal{E}^R \in \mathbb{D}_1$.
(2) Decimation/Restriction/Trace. Reduction of a form $\mathcal{E} \in \mathbb{D}_1$ to a form $\widetilde{\mathcal{E}} \in \mathbb{D}$.

Note. In Section 4, we defined a Dirichlet form $(\mathcal{E}, \mathcal{D})$ with domain $\mathcal{D} \subset L^2(F, \mu)$. But for a finite set F, as long as μ charges every point in the set it plays no role in the definition of $\mathcal{E}$. We therefore will find it more convenient to define $\mathcal{E}$ on $C(F) = \{f : F \to \mathbb{R}\}$.

Definition 6.2. Given $\mathcal{E} \in \mathbb{D}$, define for $f, g \in C\big(V^{(1)}\big)$,

$$\mathcal{E}^R(f, g) = \sum_{i=1}^M r_i^{-1} \mathcal{E}(f \circ \psi_i, g \circ \psi_i). \tag{6.2}$$

(Note that as $\psi_i : V^{(0)} \to V^{(1)}$, $f \circ \psi_i \in C\big(V^{(0)}\big)$.) Define $R : \mathbb{D} \to \mathbb{D}_1$ by $R(\mathcal{E}) = \mathcal{E}^R$.

Lemma 6.3. *Let $\mathcal{E} = \mathcal{E}_A$, and let*

$$a^R_{xy} = \sum_{i=1}^M 1_{\left(x \in V_i^{(0)}\right)} 1_{\left(y \in V_i^{(0)}\right)} r_i^{-1} a_{\psi_i^{-1}(x), \psi_i^{-1}(y)}. \tag{6.3}$$

Then

$$\mathcal{E}^R(f, g) = \tfrac{1}{2} \sum_{x,y} a^R_{xy} \big(f(x) - f(y)\big)\big(g(x) - g(y)\big). \tag{6.4}$$

$A^R = (a^R_{xy})$ is a conductance matrix, and $\mathcal{E}^R$ is the associated Dirichlet form.

Proof. As the maps ψ_i are injective, it is clear that $a^R_{xy} \geq 0$ if $x \neq y$, and $a^R_{xx} \leq 0$. Also $a^R_{xy} = a^R_{yx}$ is immediate from the symmetry of A. Writing $x_i = \psi_i^{-1}(x)$ we have

$$\sum_{y \in V^{(1)}} a^R_{xy} = \sum_i r_i^{-1} 1_{V_i^{(0)}}(x) \sum_{y \in V^{(1)}} 1_{V_i^{(0)}}(y) a_{\psi_i^{-1}(x), \psi_i^{-1}(y)}$$
$$= \sum_i r_i^{-1} 1_{V_i^{(0)}}(x) \sum_{y \in V_i^{(0)}} a_{x,y} = 0,$$

so A^R is a conductance matrix.

To verify (6.4), it is sufficient by linearity to consider the case $f = g = \delta_z$, $z \in V^{(1)}$. Let $B = \{i \in \mathbb{W}_1 : z \in V_i^{(0)}\}$. If $i \notin B$, then $f \circ \psi_i(x) = 0$, since $\psi_i(x)$ cannot equal z. If $i \in B$, then $f \circ \psi_i(x) = \delta_{z_i}(x)$, where $z_i = \psi_i^{-1}(z)$. So,

$$\mathcal{E}(f \circ \psi_i, f \circ \psi_i) = \mathcal{E}(\delta_{z_i}, \delta_{z_i}) = -a_{z_i z_i}.$$

Thus

$$\mathcal{E}^R(f, f) = -\sum_{i \in B} r_i^{-1} a_{z_i z_i} = -\sum_{i=1}^{M} r_i^{-1} 1_{V_i^{(0)}}(z) a_{\psi_i^{-1}(z), \psi_i^{-1}(z)} = -a^R_{zz},$$

while

$$\tfrac{1}{2} \sum_{x,y} a^R_{xy} \big(f(x) - f(y)\big)^2 = -f^T A^R f = -a^R_{zz}.$$

So (6.4) is verified. □

The most intuitive explanation of the replication operation is in terms of electrical networks. Think of $V^{(0)}$ as an electric network. Take M copies of $V^{(0)}$, and rescale the ith one by multiplying the conductance of each wire by r_i^{-1}. (This explains why we called r a resistance vector). Now assemble these to form a network with nodes $V^{(1)}$, using the ith network to connect the nodes in $V_i^{(0)}$. Then $\mathcal{E}^R$ is the Dirichlet form corresponding to the network $V^{(1)}$.

As we saw in the previous section, for $x, y \in V^{(1)}$ there may in general be more than one 1-cell which contains both x and y: this is why the sum in (6.3) is necessary. If x and y are connected by k wires, with conductivities $c_1, \ldots, c_k$ then this is equivalent to connection by one wire of conductance $c_1 + \ldots + c_k$.

Remark 6.4. The replication of conductivities defined here is not the same as the replication of transition probabilities discussed in Section 2. To see the difference, consider again the Sierpinski gasket. Let $V^{(0)} = \{z_1, z_2, z_3\}$, and y_3 be the midpoint of $[z_1, z_2]$, and define y_1, y_2 similarly. Let A be a conductance matrix on $V^{(0)}$, and write $a_{ij} = a_{z_i z_j}$. Take $r_1 = r_2 = r_3 = 1$. While the continuous time Markov Chains $X^{(0)}$, $X^{(1)}$ associated with $\mathcal{E}_A$ and $\mathcal{E}^R_A$ will depend on the choice of a measure on $V^{(0)}$ and $V^{(1)}$, their discrete time skeletons that is, the processes $X^{(i)}$

sampled at their successive jump times do not – see Example 4.21. Write $Y^{(i)}$ for these processes. We have

$$\mathbb{P}^{y_3}\left(Y_1^{(1)} \in \{z_2, y_1\}\right) = \frac{a_{12}+a_{31}}{2a_{12}+a_{31}+a_{23}}.$$

On the other hand, if we replicate probabilities as in Section 2,

$$\mathbb{P}^{y_3}\left(Y_1^{(1)} \in \{z_2, y_1\}\right) = \mathbb{P}^{y_3}\left(Y_1^{(1)} \in \{z_1, y_2\}\right) = \tfrac{1}{2};$$

in general these expressions are different. So, even when we confine ourselves to symmetric Markov Chains, replication of conductivities and transition probabilities give rise to different processes.

Since the two replication operations are distinct, it is not surprising that the dynamical systems associated with the two operations should have different behaviours. In fact, the simple symmetric random walk on $V^{(0)}$ is stable fixed point if we replicate conductivities, but an unstable one if we replicate transition probabilities.

The second operation on Dirichlet forms, that of restriction or trace, has already been discussed in Section 4.

Definition 6.5. For $\mathcal{E} \in \mathbb{D}_1$ let

$$T(\mathcal{E}) = Tr(\mathcal{E}|V^{(0)}). \tag{6.5}$$

Define $\Lambda : \mathbb{D} \to \mathbb{D}$ by $\Lambda(\mathcal{E}) = T\big(R(\mathcal{E})\big)$. Note that Λ is homogeneous in the sense that if $\theta > 0$,

$$\Lambda(\theta\mathcal{E}) = \theta\Lambda(\mathcal{E}).$$

Example 6.6. (The Sierpinski gasket). Let A be the conductance matrix corresponding to the simple random walk on $V^{(0)}$, so that

$$a_{xy} = 1, \quad x \neq y, \quad a_{xx} = -2.$$

Then A^R is the network obtained by joining together 3 symmetric triangular networks. If $\Lambda(\mathcal{E}_A) = \mathcal{E}_B$, then B is the conductance matrix such that the networks $(V^{(1)}, A^R)$ and $(V^{(0)}, B)$ are electrically equivalent on $V^{(0)}$. The simplest way to calculate B is by the $\Delta - Y$ transform. Replacing each of the triangles by an (upside down) Y, we see from Example 4.24 that the branches in the Y each have conductance 3. Thus $(V^{(1)}, A^R)$ is equivalent to a network consisting of a central triangle of wires of conductance 3/2, and branches of conductance 3. Applying the transform again, the central triangle is equivalent to a Y with branches of conductance 9/2. Thus the whole network is equivalent to a Y with branches of conductance 9/5, or a triangle with sides of conductance 3/5.

Thus we deduce

$$\Lambda(\mathcal{E}_A) = \mathcal{E}_B, \qquad \text{where} \quad B = \tfrac{3}{5}A.$$

The example above suggests that to find a decimation invariant random walk we need to find a Dirichlet form $\mathcal{E} \in \mathbb{D}$ such that for some $\lambda > 0$

$$\Lambda(\mathcal{E}) = \lambda \mathcal{E}. \tag{6.6}$$

Thus we wish to find an eigenvector for the map Λ on $\mathbb{D}$. Since however (as we will see shortly) Λ is non-linear, this final formulation is not particularly useful. Two questions immediately arise: does there always exist a non-zero $(\mathcal{E}, \lambda)$ satisfying (6.6) and if so, is this solution (up to constant multiples) unique? We will abuse terminology slightly, and refer to an $\mathcal{E} \in \mathbb{D}$ such that (6.6) holds as a *fixed point* of Λ. (In fact it is a fixed point of Λ defined on a quotient space of $\mathcal{D}$.)

Example 6.7. ("*abc* gaskets" – see [HHW1]).

Let m_1, m_2, m_3 be integers with $m_i \geq 1$. Let z_1, z_2, z_3 be the corners of the unit triangle in $\mathbb{R}^2$, H be the closed convex hull of $\{z_1, z_2, z_3\}$. Let $M = m_1 + m_2 + m_3$, and let ψ_i, $1 \leq i \leq M$ be similitudes such that (writing for convenience $\psi_{M+j} = \psi_j$, $1 \leq j \leq M$) $H_i = \psi_i(H) \subset H$, and the M triangles H_i are arranged round the edge of H, such that each triangle H_i touches only H_{i-1} and H_{i+1}. (H_1 touches H_M and H_2 only). In addition, let $z_1 \in H_1$, $z_2 \in H_{m_3+1}$, $z_3 \in H_{m_3+m_1+1}$. So there are $m_3 + 1$ triangles along the edge $[z_1, z_2]$, and $m_1 + 1$, $m_2 + 1$ respectively along $[z_2, z_3]$, $[z_3, z_1]$. We assume that ψ_i are rotation-free. Note that the triangles H_2 and H_M do not touch, unless $m_1 = m_2 = m_3 = 1$.

Let F be the fractal obtained by Theorem 5.4 from $(\psi_1, \ldots, \psi_M)$. To avoid unnecessarily complicated notation we write ψ_i for both ψ_i and $\psi_i|_F$.

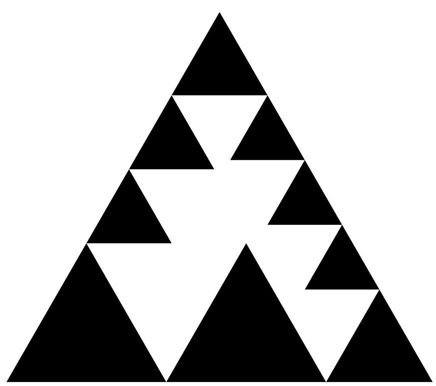

Figure 6.1: *abc* gasket with $m_1 = 4$, $m_2 = 3$, $m_3 = 2$.

It is easy to check that $(F, \psi_1, \ldots, \psi_M)$ is a p.c.f.s.s. set. Write $r = 1$, $s = m_3+1$, $t = m_3 + m_1 + 1$. We have $\pi(i\dot{s}) = \pi\big((i+1)\dot{r}\big)$ for $1 \leq i \leq m_3$, $\pi(i\dot{t}) = \pi\big((i+1)\dot{s}\big)$ for $m_3 + 1 \leq i \leq m_3 + m_1$, $\pi(i\dot{r}) = \pi\big((i+1)\dot{t}\big)$ for $m_3 + m_1 + 1 \leq i \leq M - 1$, and $\pi(M\dot{r}) = \pi(1\dot{t})$. The set $B = \cup(H_i \cap H_j)$ consists of these points. Hence

$$P = \{(\dot{r}), (\dot{s}), (\dot{t})\}, \quad V^{(0)} = \{z_1, z_2, z_3\}.$$

While it is easier to define F in $\mathbb{R}^2$, rather than abstractly, doing so has the misleading consequence that it forces the triangles H_i to be of different sizes. However, we will view F as an abstract metric space in which all the triangles H_i are of equal size, and so we will take $r_i = 1$ for $1 \le i \le M$.

We now study the renormalization map Λ. If $\mathcal{E} = \mathcal{E}_A \in \mathbb{D}$, then A is specified by the conductivities

$$\alpha_1 = a_{z_2,z_3}, \quad \alpha_2 = a_{z_3,z_1}, \quad \alpha_3 = a_{z_1,z_2}.$$

Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be the renormalization map acting on $(\alpha_1, \alpha_2, \alpha_3)$. (So if $A = A(\alpha)$ then $\Lambda(\mathcal{E}) = \mathcal{E}_{A\left(f(\alpha)\right)}$).

It is easier to compute the action of the renormalization map on the variables β_i given by the $\triangle - Y$, transform. So let $\varphi : (0,\infty)^3 \to (0,\infty)^3$ be the $\triangle - Y$ map given in Example 4.24. Note that φ is bijective. Let $\beta = \varphi(\alpha)$ be the $Y-$conductivities, and write $\widetilde{\beta} = (\widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3)$ for the renormalized $Y-$conductivities: then $\widetilde{\beta} = \varphi(f(\alpha))$.

Applying the $\triangle - Y$ transform on each of the small triangles, we obtain a network with nodes $z_1, z_2, z_3, y_1, y_2, y_3$, where $\{z_i, y_i\}$ has conductivity β_i, and if $i \neq j$ $\{y_i, y_j\}$ has conductivity β_i, and if $i \neq j$, $\{y_i, y_i\}$ has conductivity

$$\gamma_k = \frac{\beta_i \beta_j}{(\beta_i + \beta_j) m_k},$$

where $k = k(i,j)$ is such that $k \in \{1,2,3\} - \{i,j\}$.

Apply the $\triangle - Y$ transform again to $\{y_1, y_2, y_3\}$, to obtain a Y, with conductivities δ_1, δ_2, δ_3, in the branches where

$$\delta_i \gamma_i = S = \gamma_1 \gamma_2 + \gamma_2 \gamma_3 + \gamma_3 \gamma_1, \qquad 1 \le i \le 3.$$

Then

$$(6.7) \qquad \widetilde{\beta}_1^{-1} = \beta_1^{-1} + \delta_1^{-1} = \beta_1^{-1} + \tfrac{\beta_2 \beta_3}{(\beta_2 + \beta_3) m_1 S}.$$

Suppose that $\alpha \in (0,\infty)^3$ is such that $\varphi(\alpha) = \lambda\alpha$ for some $\lambda > 0$. Then since $\varphi(\theta\alpha) = \theta\,\varphi(\alpha)$ for any $\theta > 0$, we deduce that $\widetilde{\beta} = \varphi(f(\alpha)) = \lambda\beta$. So, from (6.7),

$$\lambda^{-1}\beta_1^{-1} = \beta_1^{-1} + \tfrac{\beta_2 \beta_2}{(\beta_2 + \beta_3) m_1 S},$$

which implies that $\lambda^{-1} > 1$. Writing $T = \beta_1\beta_2\beta_3/S$, and $\theta = T\lambda(1-\lambda)^{-1}$, we therefore have

$$m_1(\beta_2 + \beta_3) = \theta,$$

and (as S, T are symmetric in the β_i) we also obtain two similar equations. Hence

$$(6.8) \qquad \beta_2 + \beta_3 = \theta/m_1, \quad \beta_3 + \beta_1 = \theta/m_2, \quad \beta_1 + \beta_2 = \theta/m_3,$$

which has solution

$$(6.9) \qquad 2\beta_1 = \theta(m_2^{-1} + m_3^{-1} - m_1^{-1}), \qquad \text{etc.}$$

Since, however we need the $\beta_i > 0$, we deduce that a solution to the conductivity renormalization problem exists only if m_i^{-1} satisfy the triangle condition, that is that

$$(6.10) \qquad m_2^{-1} + m_3^{-1} > m_1^{-1}, \quad m_3^{-1} + m_1^{-1} > m_2^{-1}, \quad m_1^{-1} + m_2^{-1} > m_3^{-1}.$$

If (6.10) is satisfied, then (6.9) gives β_i such that the associated $\alpha = \varphi^{-1}(\beta)$ does satisfy the eigenvalue problem.

In the discussion above we looked for strictly positive α such that $\varphi(\alpha) = \lambda\alpha$. Now suppose that just one of the α_i, α_3 say, equals 0. Then while z, and z_2 are only connected via z_3 in the network $V^{(0)}$ they are connected via an additional path in the network $V^{(1)}$. So, $\varphi(\alpha)_3 > 0$, and α cannot be a fixed point. If now $\alpha_1 > 0$, and $\alpha_2 = \alpha_3 = 0$ then we obtain $\varphi(\alpha)_2 = \varphi(\alpha)_3 = 0$. So $\alpha = (1, 0, 0)$ satisfies $\varphi(\alpha) = \lambda\alpha$ for some $\lambda > 0$. Similarly $(0, 1, 0)$ and $(0, 0, 1)$ are also fixed points. Note that in these cases the network $(V^{(0)}, A(\alpha))$ is not connected.

The example of the *abc* gaskets shows that, even if fixed points exist, they may correspond to a reducible (ie non-irreducible) $\mathcal{E} \in \mathbb{D}$. The random walks (and limiting diffusion) corresponding to such a fixed point will be restricted to part of the fractal F. We therefore wish to find a *non-degenerate fixed point* of (6.6), that is an $\mathcal{E}_A \in \mathbb{D}$ such that the network $(V^{(0)}, A)$ is connected.

Definition 6.8. Let $\mathbb{D}^i$ be the set of $\mathcal{E} \in \mathbb{D}_0$ such that $\mathcal{E}$ is irreducible – that is the network $(V^{(0)}, A)$ is connected. Call $\mathcal{E} \in \mathbb{D}$ *strongly irreducible* if $\mathcal{E} = \mathcal{E}_A$ and $a_{xy} > 0$ for all $x \neq y$. Write $\mathbb{D}^{si}$ for the set of strongly irreducible Dirichlet forms on $V^{(0)}$.

The existence problem therefore takes the form:

Problem 6.9. (Existence). Let $(F, \psi_1, ..., \psi_M)$ be a p.c.f.s.s. set and let $r_i > 0$. Does there exist $\mathcal{E} \in \mathbb{D}^i$, $\lambda > 0$, such that

$$(6.12 \qquad \Lambda(\mathcal{E}) = \lambda\mathcal{E}?$$

Before we pose the uniqueness question, we need to consider the role of symmetry. Let $(F, (\psi_i))$ be a p.c.f.s.s. set, and let $\mathcal{H}$ be a symmetry group of F.

Definition 6.10. $\mathcal{E} \in \mathbb{D}$ is $\mathcal{H}$*-invariant* if for each $h \in \mathcal{H}$

$$\mathcal{E}(f \circ h, g \circ h) = \mathcal{E}(f, g), \quad f, g \in C(V^{(0)}).$$

r is $\mathcal{H}$*-invariant* if $r_{\tilde{h}(i)} = r_i$ for all $h \in \mathcal{H}$. (Here $\tilde{h}$ is the bijection on I associated with h).

Lemma 6.11. *(a) Let $\mathcal{E} = \mathcal{E}_A$. Then $\mathcal{E}$ is $\mathcal{H}$-invariant if and only if:*

$$(6.13) \qquad a_{h(x)\,h(y)} = a_{xy} \text{ for all } x, y \in V^{(0)}, \ h \in \mathcal{H}.$$

(b) Let $\mathcal{E}$ and r be $\mathcal{H}$-invariant. Then $\Lambda\mathcal{E}$ is $\mathcal{H}$-invariant.

Proof. (a) This is evident from the equation $\mathcal{E}(1_x, 1_y) = -a_{xy}$.

(b) Let $f \in C(V^{(1)})$. Then if $h \in \mathcal{H}$,

$$\begin{aligned}\mathcal{E}^R(f \circ h, f \circ h) &= \sum_i r_i^{-1}\, \mathcal{E}(f \circ h \circ \psi_i, f \circ h \circ \psi_i) \\ &= \sum r_i^{-1}\, \mathcal{E}(f \circ \psi_{\tilde{h}(i)} \circ h, f \circ \psi_{\tilde{h}(i)} \circ h,) \\ &= \sum r_{\tilde{h}(i)}^{-1}\, \mathcal{E}(f \circ \psi_{\tilde{h}(i)}, f \circ \psi_{\tilde{h}(i)}) = \mathcal{E}^R(f, f).\end{aligned}$$

If $g \in C(V^{(0)})$ then writing $\widetilde{\mathcal{E}} = \Lambda(\mathcal{E})$, if $f|_{V^{(0)}} = g$ then as $f \circ h|_{V^{(0)}} = g \circ h$,

$$\widetilde{\mathcal{E}}(g \circ h, g \circ h) \le \mathcal{E}^R(f \circ h, f \circ h) = \mathcal{E}^R(f, f),$$

and taking the infimum over f, we deduce that for any $h \in \mathcal{H}$, $\widetilde{\mathcal{E}}(g \circ h, g \circ h) \le \widetilde{\mathcal{E}}(g, g)$. Replacing g by $g \circ h$ and h by h^{-1} we see that equality must hold. $\square$

If the fractal F has a non-trivial symmetry group $\mathcal{G}(F)$ then it is natural to restrict our attention to $\mathcal{G}(F)$-symmetric diffusions. We can now pose the uniqueness problem.

Problem 6.12. (Uniqueness). Let $(F, (\psi_i))$ be a p.c.f.s.s. set, let $\mathcal{H}$ be a symmetry group of F, and let r be $\mathcal{H}$-invariant. Is there at most one $\mathcal{H}$-invariant $\mathcal{E} \in \mathbb{D}^i$ such that $\Lambda(\mathcal{E}) = \lambda \mathcal{E}$?

(Unless otherwise indicated, when I refer to fixed points for nested fractals, I will assume they are invariant under the symmetry group $\mathcal{G}_R$ generated by the reflections in hyperplanes bisecting the lines $[x, y]$, $x, y \in V^{(0)}$).

The following example shows that uniqueness does not hold in general.

Example 6.13. (Vicsek sets – see [Me3].) Let $(F, \psi_i, 1 \le i \le 5)$ be the Vicsek set – see Section 2. Write $\{z_1, z_2, z_3, z_{4,}\}$ for the 4 corners of the unit square in $\mathbb{R}^2$. For $\alpha, \beta, \gamma > 0$ let $A(\alpha, \beta, \gamma)$ be the conductance matrix given by

$$a_{12} = a_{23} = a_{34} = a_{41} = \alpha, \quad \alpha_{13} = \beta, \quad a_{24} = \gamma,$$

where $a_{ij} = a_{z_i\, z_j}$. If $\mathcal{H}$ is the group on F generated by reflections in the lines $[z_1, z_3]$ and $[z_2, z_4]$ then A is clearly $\mathcal{H}$-invariant. Define $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\gamma}$ by

$$\Lambda(\mathcal{E}_A) = \mathcal{E}_{A(\widetilde{\alpha},\, \widetilde{\beta},\, \widetilde{\gamma})}.$$

Then several minutes calculation with equivalent networks shows that

$$\begin{aligned}\widetilde{\alpha} &= \frac{\alpha(\alpha+\beta)(\alpha+\gamma)}{5\alpha^2 + 3\alpha\beta + 3\alpha\gamma + \beta\gamma}, \\ \widetilde{\beta} &= \tfrac{1}{3}(\alpha+\beta) - \widetilde{\alpha}, \\ \widetilde{\gamma} &= \tfrac{1}{3}(\alpha+\gamma) - \widetilde{\alpha}.\end{aligned} \tag{6.14}$$

If $(1, \beta, \gamma)$ is a fixed point then $(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}) = (\theta, \theta\beta, \theta\gamma)$ for some $\theta \ge 0$, so that $\widetilde{\beta} = \widetilde{\alpha}\beta$, $\widetilde{\gamma} = \widetilde{\alpha}\gamma$. So $\widetilde{\alpha} = \frac{1}{3}$, and this implies that $\beta\gamma = 1$. We therefore have that $(1, \beta, \beta^{-1})$ is a fixed point (with $\lambda = \frac{1}{3}$) for any $\beta \in (0, \infty)$ Thus for the group $\mathcal{H}$ uniqueness does not hold.

However if we replace $\mathcal{H}$ by the group $\mathcal{G}_R = \mathcal{G}(F)$, generated by all the symmetries of the square then for $\mathcal{E}_A$ to be $\mathcal{G}_R$-invariant we have to have $\beta = \gamma$. So in this case we obtain

$$\begin{aligned} \widetilde{\alpha}(\alpha,\beta) &= \frac{\alpha(\alpha+\beta)^2}{5\alpha^2+6\alpha\beta+\beta^2}\,, \\ \widetilde{\beta}(\alpha,\beta) &= \tfrac{1}{3}(\alpha+\beta) - \widetilde{\alpha}. \end{aligned} \tag{6.15}$$

This has fixed points $(0,\beta)$, $\beta > 0$, and (α,α), $\alpha > 0$. The first are degenerate, the second not, so in this case, as we already saw in Section 2, uniqueness does hold for Problem 6.12.

This example also shows that Λ is in general non-linear.

As these examples suggest, the general problem of existence and uniqueness is quite hard. For all but the simplest fractals, explicit calculation of the renormalization map Λ is too lengthy to be possible without computer assistance – at least for 20th century mathematicians. Lindstrøm [L1] proved the existence of a fixed point $\mathcal{E} \in \mathbb{D}^{si}$ for nested fractals, but did not treat the question of uniqueness. After the appearance of [L1], the uniqueness of a fixed point for Lindstrøm's canonical example, the snowflake (Example 5.26) remained open for a few years, until Green [Gre] and Yokai [Yo] proved uniqueness by computer calculations.

The following analytic approach to the uniqueness problem, using the theory of quadratic forms, has been developed by Metz and Sabot – see [Me2-Me5, Sab1, Sab2]. Let $\mathbb{M}_+$ be set of symmetric bilinear forms $Q(f,g)$ on $C(V^{(0)})$ which satisfy

$$\begin{aligned} &Q(1,1) = 0, \\ &Q(f,f) \geq 0 \quad \text{for all } f \in C(V^{(0)}). \end{aligned}$$

For Q_1, $Q_2 \in \mathbb{M}_+$ we write $Q_1 \geq Q_2$, if $Q_2 - Q_1 \in \mathbb{M}_+$ or equivalently if $Q_2(f,f) \geq Q_1(f,f)$ for all $f \in C(V^{(0)})$. Then $\mathbb{D} \subset \mathbb{M}_+$; it turns out that we need to consider the action of Λ on $\mathbb{M}_+$, and not just on $\mathbb{D}$. For $Q \in \mathbb{M}_+$, the replication operation is defined exactly as in (6.2)

$$Q^R(f,g) = \sum_{i=1}^{M} r_i^{-1}\, Q(f\circ\psi_i, g\circ\psi_i), \quad f,g \in C(V^{(1)}). \tag{6.16}$$

The decimation operation is also easy to extend to $\mathbb{M}_+$:

$$T(Q^R)(g,g) = \inf\{Q^R(f,f) : f \in C(V^{(0)}), f|_{V^{(0)}} = g\};$$

we can write $T(Q^R)$ in matrix terms as in (4.24). We set $\Lambda(Q) = T(Q^R)$.

Lemma 6.14. *The map Λ on $\mathbb{M}_+$ satisfies:*
(a) $\Lambda : \mathbb{M}_+ \to \mathbb{M}_+$, and is continuous on $\mathrm{int}(\mathbb{M}_+)$.
(b) $\Lambda(Q_1 + Q_2) \geq \Lambda(Q_1) + \Lambda(Q_2)$.
(c) $\Lambda(\theta Q) = \theta\Lambda(Q)$

Proof. (a) is clear from the formulation of the trace operation in matrix terms.

Since the replication operation is linear, we clearly have $Q^R = Q_1^R + Q_2^R$, $(\theta Q)^R = \theta Q^R$. (c) is therefore evident, while for (b), if $g \in C(V^{(0)})$,

$$\begin{aligned} T(Q^R)(g,g) &= \inf\{Q_1^R\,(f,f) + Q_2^R\,(f,f) :\ f|_{V^{(0)}} = g\} \\ &\geq \inf\{Q_1^R\,(f,f) :\ f|_{V^{(0)}} = g\} + \inf\{Q_2^R\,(f,f) :\ f|_{V^{(0)}} = g\} \\ &= T(Q_1^R)(g,g) + T(Q_2^R)(g,g). \end{aligned}$$

□

Note that for $\mathcal{E} \in \mathbb{D}^i$, we have $\mathcal{E}(f,f) = 0$ if only if f is constant.

Definition 6.15. For $\mathcal{E}_1, \mathcal{E}_2 \in \mathbb{D}^i$ set

$$\begin{aligned} m(\mathcal{E}_1/\mathcal{E}_2) &= \sup\{\alpha \geq 0;\ \mathcal{E}_1 - \alpha\mathcal{E}_2 \in \mathbb{M}_+\}. \\ &= \inf\{\frac{\mathcal{E}_1(f,f)}{\mathcal{E}_2(f,f)}\ : f \text{ non constant}\}. \end{aligned}$$

Similarly let

$$M(\mathcal{E}_1/\mathcal{E}_2) = \sup\{\frac{\mathcal{E}_1(f,f)}{\mathcal{E}_2(f,f)}\ : f \text{ non constant}\}.$$

Note that

$$M(\mathcal{E}_1/\mathcal{E}_2) = m(\mathcal{E}_2/\mathcal{E}_1)^{-1}. \tag{6.18}$$

Lemma 6.16. *(a) For $\mathcal{E}_1, \mathcal{E}_2 \in \mathbb{D}^i$, $0 < m(\mathcal{E}_1, \mathcal{E}_2) < \infty$.*
(b) If $\mathcal{E}_1$, $\mathcal{E}_2 \in \mathbb{D}^i i$ then $m(\mathcal{E}_1/\mathcal{E}_2) = M(\mathcal{E}_1/\mathcal{E}_2)$ if and only if $\mathcal{E}_2 = \lambda\mathcal{E}_1$ for some $\lambda > 0$.
(c) If $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3 \in \mathbb{D}^i$ then

$$m(\mathcal{E}_1/\mathcal{E}_3) \geq m(\mathcal{E}_1/\mathcal{E}_2)\, m(\mathcal{E}_2/\mathcal{E}_3),$$

$$M(\mathcal{E}_1/\mathcal{E}_3) \leq M(\mathcal{E}_1/\mathcal{E}_2)\, M(\mathcal{E}_2/\mathcal{E}_3).$$

Proof. (a) This follows from the fact that $\mathcal{E}_i$ are irreducible, and so vanish only on the subspace of constant functions.
(b) is immediate from the definition of m and M.
(c) We have

$$m(\mathcal{E}_1/\mathcal{E}_3) = \inf_f \frac{\mathcal{E}_1(f,f)}{\mathcal{E}_2(f,f)} \frac{\mathcal{E}_2(f,f)}{\mathcal{E}_3(f,f)} \geq m(\mathcal{E}_1/\mathcal{E}_2) m(\mathcal{E}_2/\mathcal{E}_3);$$

while the second assertion is immediate from (6.18). □

Definition 6.17. Define

$$d_H(\mathcal{E}_1, \mathcal{E}_2) = \log \frac{M(\mathcal{E}_1 \mathcal{E}_2)}{m(\mathcal{E}_1\, \mathcal{E}_2)}, \quad \mathcal{E}_1, \mathcal{E}_2 \in \mathbb{D}^i.$$

Let $p\,\mathbb{D}^i$ be the projective space $\mathbb{D}^i/\sim$, where $\mathcal{E}_1 \sim \mathcal{E}_2$ if $\mathcal{E}_1 = \lambda\mathcal{E}_2$. d_H is called *Hilbert's projective metric* – see [Nus], [Me4].

Proposition 6.18. *(a) $d_H(\mathcal{E}_1,\mathcal{E}_2)=0$ if and only if $\mathcal{E}_1=\lambda\mathcal{E}_2$ for some $\lambda>0$.*
(b) d_H is a pseudo-metric on $\mathbb{D}^i$, and a metric on $p\,\mathbb{D}^i$.
(c) If $\mathcal{E},\mathcal{E}_0,\mathcal{E}_1\in\mathbb{D}^i$ then for $\alpha_0,\alpha_1>0$,

$$d_H(\mathcal{E},\alpha_0\,\mathcal{E}_0+\alpha_1\mathcal{E}_1)\le\max(d_H(\mathcal{E},\mathcal{E}_0),d_H(\mathcal{E},\mathcal{E}_1)).$$

In particular open balls in d_H are convex.
(d) $(p\,\mathbb{D}^i,d_H)$ is complete.

Proof. (a) is evident from Lemma 6.17(b). To prove (b) note that $d_H(\mathcal{E}_1,\mathcal{E}_2)\ge0$, and that $d_H(\mathcal{E}_1,\mathcal{E}_2)=d_H(\mathcal{E}_2,\mathcal{E}_1)$ from (6.18). The triangle inequality is immediate from Lemma 6.17(c). So d_H is a pseudo metric on $\mathbb{D}^i$.

To see that d_H is a metric on $p\,\mathbb{D}^i$, note that

$$m(\lambda\mathcal{E}_1/\mathcal{E}_2)=\lambda m(\mathcal{E}_1/\mathcal{E}_2),\quad\lambda>0,$$

from which it follows that $d_H(\lambda\mathcal{E}_1,\mathcal{E}_2)=d_H(\mathcal{E}_1,\mathcal{E}_2)$ and thus d_H is well defined on $p\,\mathbb{D}^i$. The remaining properties are now immediate from those of d_H on $\mathbb{D}^i$.
(c) Replacing $\mathcal{E}_1$ by $\big(m(\mathcal{E}_1/\mathcal{E}_0)/m(\mathcal{E}/\mathcal{E}_1)\big)\mathcal{E}_1$ we can suppose that

$$m(\mathcal{E}/\mathcal{E}_0)=m(\mathcal{E}\,/\mathcal{E}_1)=m.$$

Write $M_i=M(\mathcal{E}/\mathcal{E}_i)$. Then if $\mathcal{F}=\alpha_0\,\mathcal{E}_0+\alpha_1\mathcal{E}_1$,

$$\begin{aligned}M(\mathcal{E}/\mathcal{F})&=\inf_f\frac{\alpha_0\mathcal{E}_0(f,f)+\alpha_1\mathcal{E}_1(f,f)}{\mathcal{E}(f,f)}\\&\ge\alpha_0 m(\mathcal{E}/\mathcal{E}_0)+\alpha_1 m(\mathcal{E}/\mathcal{E}_1)=\alpha_0+\alpha_1\,.\end{aligned}$$

Similarly $M(\mathcal{E}/\mathcal{F})\le\alpha_0M_0+\alpha_1M_1$. Therefore

$$\begin{aligned}\exp d_H(\mathcal{E},\mathcal{F})&\le(\alpha_0/(\alpha_0+\alpha_1))(M_0/m)+(\alpha_1/(\alpha_0+\alpha_1))(M_1/m)\\&\le\max(M_0/m,M_1/m).\end{aligned}$$

It is immediate that if $\mathcal{E}_i\in B(\mathcal{E},r)$ then $d_H(\mathcal{E},\lambda\mathcal{E}_0+(1-\lambda)\mathcal{E}_1)<r$, so that $B(\mathcal{E},r)$ is convex. For (d) see [Nus, Thm. 1.2]. □

Theorem 6.19. *Let $\mathcal{E}_1,\mathcal{E}_2\in\mathbb{D}^i$. Then*

$$m(\Lambda(\mathcal{E}_1),\Lambda(\mathcal{E}_2))\ge m(\mathcal{E}_1,\mathcal{E}_2),\tag{6.19}$$

$$M(\Lambda(\mathcal{E}_1),\Lambda(\mathcal{E}_2))\le M(\mathcal{E}_1,\mathcal{E}_2).\tag{6.20}$$

In particular Λ is non-expansive in d_H :

$$d_H(\Lambda(\mathcal{E}_1),\Lambda(\mathcal{E}_2))\le d_H(\mathcal{E}_1,\mathcal{E}_2).\tag{6.21}$$

Proof. Suppose $\alpha<m(\mathcal{E}_1,\mathcal{E}_2)$. Then $Q=\mathcal{E}_1-\alpha\,\mathcal{E}_2\in\mathbb{M}_+$, and $Q(f,f)>0$, for all non-constant $f\in C(V^{(0)})$. So by Lemma 6.14

$$\Lambda(\mathcal{E}_1)=\Lambda(Q+\alpha\mathcal{E}_2)\ge\Lambda(Q)+\alpha\Lambda(\mathcal{E}_2),$$

and since $\Lambda(Q) \geq 0$, this implies that $\Lambda(\mathcal{E}_1) - \alpha\Lambda(\mathcal{E}_2) \geq 0$. So $\alpha < m(\Lambda(\mathcal{E}_1), \Lambda(\mathcal{E}_2))$, and thus $m(\mathcal{E}_1, \mathcal{E}_2) \leq m(\Lambda(\mathcal{E}_1), \Lambda(\mathcal{E}_2))$, proving (6.19). (6.20) and (6.21) then follow immediately from (6.19), and the definition of d_H. □

A strict inequality in (6.21) would imply the uniqueness of fixed points. Thus the example of the Vicsek set above shows that strict inequality cannot hold in general. So this Theorem gives us much less than we might hope. Nevertheless, we can obtain some useful information.

Corollary 6.20. *(See [HHW1, Cor. 3.7]) Suppose $\mathcal{E}_1, \mathcal{E}_2$ are fixed points satisfying $\Lambda(\mathcal{E}_i) = \lambda_i \mathcal{E}_i$, $i = 1, 2$. Then $\lambda_1 = \lambda_2$.*

Proof. From (6.19)

$$m(\mathcal{E}_1/\mathcal{E}_2) \leq m\big(\Lambda(\mathcal{E}_1)/\Lambda(\mathcal{E}_2)\big) = (\lambda_1/\lambda_2) m(\mathcal{E}_1/\mathcal{E}_2),$$

so that $\lambda_1 \geq \lambda_2$. Interchanging $\mathcal{E}_1$ and $\mathcal{E}_2$ we obtain $\lambda_1 = \lambda_2$. □

We can also deduce the existence of $\mathcal{H}$-invariant fixed points.

Proposition 6.21. *Let $\mathcal{H}$ be a symmetry group of F. If Λ has a fixed point $\mathcal{E}_1$ in $\mathbb{D}^i$ then Λ has an $\mathcal{H}$-invariant fixed point in $\mathbb{D}^i$.*

Proof. Let $A = \{\mathcal{E} \in \mathbb{D}^i : \mathcal{E}$ is $\mathcal{H}$-invariant.$\}$. (It is clear from Lemma 6.11 that A is non-empty). Then by Lemma 6.11(b) $\Lambda : A \to A$. Let $\mathcal{E}_0 \in A$, and write $r = d_H(\mathcal{E}_1, \mathcal{E}_0)$, $B = B_{d_H}(\mathcal{E}_1, 2r)$. By Theorem 6.20 $\Lambda : B \to B$. So $\Lambda : A \cap B \to A \cap B$. Each of A, B is convex (A is convex as the sum of two $\mathcal{H}$-invariant forms is $\mathcal{H}$-invariant, B by Proposition 6.18(c)), and so $A \cap B$ is convex. Since Λ is a continuous function on a convex space, by the Brouwer fixed point theorem Λ has a fixed point $\mathcal{E}' \in A \cap B$, and $\mathcal{E}'$ is $\mathcal{H}$-invariant. □

We will not make use of the following result, but is useful for understanding the general situation.

Corollary 6.22. *Suppose Λ has two distinct fixed points $\mathcal{E}_1$ and $\mathcal{E}_2$ (with $\mathcal{E}_1 \neq \lambda\mathcal{E}_2$ for any λ). Then Λ has uncountably many fixed points.*

Proof. (Note that the example of the Vicsek set shows that $\frac{1}{2}(\mathcal{E}_1 + \mathcal{E}_2)$ is not necessarily a fixed point). Let $\mathbb{F} \subset \mathbb{D}^i$ be the set of fixed points. Let $\mathcal{E}_0, \mathcal{E}_1 \in \mathbb{F}$; multiplying $\mathcal{E}_1$ by a scalar we can take $m(\mathcal{E}_0, \mathcal{E}_1) = 1$. Write $R = d_H(\mathcal{E}_0, \mathcal{E}_1)$. If $\mathcal{E}_\lambda = \lambda\mathcal{E}_1 + (1-\lambda)\mathcal{E}_0$ then as in Proposition 6.19(c)

$$\exp d_{\mathcal{H}}(\mathcal{E}_\lambda, \mathcal{E}_0) \leq (1-\lambda) + \lambda M(\mathcal{E}_1, \mathcal{E}_0)$$

and so

$$d_{\mathcal{H}}(\mathcal{E}_{1/2}, \mathcal{E}_0) \leq \log((1+e^R)/2).$$

Thus there exists δ, depending only on R, such that

$$A = \{\mathcal{E} \in \mathbb{D}^i : \mathcal{E} \in B(\mathcal{E}_0, (1-\delta)R) \bigcap B(\mathcal{E}_1, (1-\delta)R)\}$$

is non-empty. Since Λ preserves A, Λ has a fixed point in A. $\mathbb{F}$ thus has the property:

if $\mathcal{E}_1, \mathcal{E}_2$ are distinct elements of $\mathbb{F}$ then there exists $\mathcal{E}_3 \in \mathbb{F}$
such that $0 < d_{\mathcal{H}}(\mathcal{E}_3, \mathcal{E}_1) < d_{\mathcal{H}}(\mathcal{E}_2, \mathcal{E}_1)$.

As $\mathbb{F}$ is closed (since Λ is continuous) we deduce that $\mathbb{F}$ is perfect, and therefore uncountable. $\square$

This if as far as we will go in general. For nested fractals the added structure – symmetry and the embedding in $\mathbb{R}^d$, enables us to obtain stronger results. If $(F, (\psi_i))$ is a nested fractal, or an ANF, we only consider the set $\mathbb{D}^i \cap \{\mathcal{E} : \mathcal{E}$ is $\mathcal{G}_R$-invariant$\}$, so that in discussing the existence and uniqueness of fixed points we will be considering only $\mathcal{G}_R$-invariant ones.

Let $(F, (\psi_i))$ be a nested fractal, write $\mathcal{G} = \mathcal{G}_R$ and let $\mathcal{E}_A$ be a ($\mathcal{G}$-invariant) Dirichlet form on $C(V^{(0)})$. $\mathcal{E}_A$ is determined by the conductances on the equivalence classes of edges in $(V^{(0)}, \mathcal{E}_0)$ under the action of $\mathcal{G}$. By Proposition 5.38(c) if $|x-y| = |x'-y'|$ then the edges $\{x, y\}$ and $\{x', y'\}$ are equivalent, so that $A_{xy} = A_{x'y'}$.

List the equivalence classes in order of increasing Euclidean distance, and write $\alpha_1, \alpha_2 ..., \alpha_k$ for the common conductances of the edges. Since $\widetilde{A} = \Lambda(A)$ is also $\mathcal{G}$-invariant, Λ induces a map $\Lambda' : \mathbb{R}^k_+ \to \mathbb{R}^k_+$ such that, using obvious notation, $\Lambda(A(\alpha)) = A(\Lambda'(\alpha))$.

Set $\mathbb{D}^* = \{\alpha : \alpha_1 \geq \alpha_2 \geq ... \geq \alpha_k > 0\}$. Clearly we have $\mathbb{D}^* \subset \mathbb{D}^{si}$. We have the following existence theorem for nested fractals.

Theorem 6.23. *(See [L1, p. 48]). Let $(F, (\psi_i))$ be a nested fractal (or an ANF). Then Λ has a fixed point in $\mathbb{D}^*$.*

Proof. Let $\mathcal{E}_A \in \mathbb{D}^*$, and let $\alpha_1, ...\alpha_k$ be the associated conductivities. Let $(Y_t, t \geq 0, \mathbb{Q}^x, x \in V^{(0)})$ be the continuous time Markov chain associated with $\mathcal{E}_A$, and let $(\widehat{Y}_n, n \geq 0,\ \mathbb{Q}^x, x \in V^{(0)})$ be the discrete time skeleton of Y.

Let $E_0^{(1)}, ..., E_0^{(k)}$ be the equivalence classes of edges in $(V^{(0)}, E_0)$, so that $A_{xy} = \alpha_j$ if $\{x, y\} \in E_0^{(j)}$. Then if $\{x, y\} \in E_0^{(j)}$,

$$\mathbb{Q}^x(\widehat{Y}_1 = y) = \frac{\alpha_j}{\sum_{y \neq x} A_{xy}}.$$

As $c_1 = \sum_{y \neq x} A_{xy}$ does not depend on x (by the symmetry of $V^{(0)}$) the transition probabilities of $\widehat{Y}$ are proportional to the α_j.

Now let $R(A)$ be the conductivity matrix on $V^{(1)}$ attained by replication of A. Let $(X_t, t \geq 0,\ \mathbb{P}^x,\ x \in V^{(1)})$ and $(\widehat{X}_n, n \geq 0,\ \mathbb{P}^x, x \in V^{(1)})$ be the associated Markov Chains. Let $T_0, T_1, ...$ be successive disjoint hits (see Definition 2.14) on $V^{(0)}$ by $\widehat{X}_n$.

Write $\widetilde{A} = \Lambda(A)$, and $\widetilde{\alpha}$ for the edge conductivities given by A. Using the trace theorem,

$$\mathbb{P}^x(\widehat{X}_{T_1} = y) = \widetilde{\alpha}_j / c_1 \quad \text{ if } \{x, y\} \in E_0^{(j)}.$$

Now let $x_1, y_1, y_2 \in V^{(0)}$, with $|x_1 - y_1| < |x_1 - y_2|$. We will prove that

$$\mathbb{P}^{x_1}(\widehat{X}_{T_1} = y_2) < \mathbb{P}^{x_1}(\widehat{X}_{T_1} = y_1). \tag{6.23}$$

Let H be the hyperplane bisecting $[y_1, y_2]$, let g be reflection in H, and $x_2 = g(x_1)$. Let $B = V^{(0)} - \{x_1\}$, and

$$T = \min\{n \geq 0 :\ \widehat{X}_n \in B,$$

so that $T_1 = T$ $\mathbb{P}^{x_1}$-almost surely. Set

$$f_n(x) = \mathbb{E}^x \, 1_{(T \le n)}(1_{y_1}(\widehat{X}_T) - 1_{y_2}(\widehat{X}_T)).$$

Let $p(x,y), \ x,y \in V^{(1)}$ be the transition probabilities of $\widehat{X}$. Then

(6.24) $$f_{n+1}(x) = 1_B(x)\, f_0(x) + 1_{B^c}(x) \sum_y p(x,y) f_n(y).$$

Let $J_{12} = \{x \in V^{(1)} : |x - y_1| \le |x - y_2|\}$, and define J_{21} analogously. We prove by induction that f_n satisfies

(6.25a) $$f_n(x) \ge 0, \qquad x \in J_{12},$$

(6.25b) $$f_n(x) + f_n(g(x)) \ge 0, \qquad x \in J_{12}.$$

Since $f_0 = 1_{y_1} - 1_{y_2}$, and $y_1 \in J_{12}$, f_0 satisfies (6.25). Let $x \in B^c \cap J_{12}$ and suppose f_n satisfies (6.25). If $p(x,y) > 0$, and $y \in J_{12}^c$, then x, y are in the same 1-cell so if $y' = g(y)$, y' is also in the same 1-cell as x_1 and $|x - y'| \le |x - y|$. So (since $\mathcal{E}_A \in \mathbb{D}^*$), $p(x,y') \ge p(x,y)$ and using (6.25b), as $f_n(y') \ge 0$,

$$p(x,y)f_n(y) + p(x,y')f_n(y') \ge p(x,y)(f_n(y) + f_n(g(y)) \ge 0.$$

Then by (6.24), $f_{n+1}(x) \ge 0$. A similar argument implies that f_{n+1} satisfies (6.25b).

So (f_n) satisfies (6.25) for all n, and hence its limit f_∞ does. Thus $f_\infty(x_1) = \mathbb{P}^{x_1}(\widehat{X}_T = y_1) - \widehat{\mathbb{P}}(\widehat{X}_T = y_2) \ge 0$, proving (6.23).

From (6.23) we deduce that $\widetilde{\alpha}_1 \ge \widetilde{\alpha}_2 \ge ... \ge \widetilde{\alpha}_k$, so that $\Lambda : \mathbb{D}^* \to \mathbb{D}^*$. As $\Lambda'(\theta\alpha) = \theta\Lambda'(\alpha)$, we can restrict the action of Λ' to the set

$$\{\alpha \in \mathbb{R}_+^k : \alpha_1 \ge ... \ge \alpha_k \ge 0, \sum \alpha_i = 1\}.$$

This is a closed convex set, so by the Brouwer fixed point theorem, Λ' has a fixed point in $\mathbb{D}^*$. □

Remark 6.24. The proof here is essentially the same as that in Lindstrøm [L1]. The essential idea is a kind of reflection argument, to show that transitions along shorter edges are more probable. This probabilistic argument yields (so far) a stronger existence theorem for nested fractals than the analytic arguments used by Sabot [Sab1] and Metz [Me7]. However, the latter methods are more widely applicable.

It does not seem easy to relax any of the conditions on ANFs without losing some link in the proof of Theorem 6.23. This proof used in an essential fashion not only the fact that $V^{(0)}$ has a very large symmetry group, but also the Euclidean embedding of $V^{(0)}$ and $V^{(1)}$.

The following uniqueness theorem for nested fractals was proved by Sabot [Sab1]. It is a corollary of a more general theorem which gives, for p.c.f.s.s. sets, sufficient conditions for existence and uniqueness of fixed points. A simpler proof of this result has also recently been obtained by Peirone [Pe].

Theorem 6.25. *Let $(F,(\psi_i))$ be a nested fractal. Then Λ has a unique $\mathcal{G}_R$-invariant non-degenerate fixed point.*

Definition 6.26. Let $\mathcal{E}$ be a fixed point of Λ. The *resistance scaling factor* of $\mathcal{E}$ is the unique $\rho > 0$ such that

$$\Lambda(\mathcal{E}) = \rho^{-1}\,\mathcal{E}.$$

Very often we will also call ρ the resistance scaling factor of F: in view of Corollary 6.21, ρ will have the same value for any two non-degenerate fixed points.

Proposition 6.27. *Let $(F,(\psi_i))$ be a p.c.f.s.s. set, let (r_i) be a resistance vector, and let $\mathcal{E}_A$ be a non-degenerate fixed point of Λ. Then for each $s \in \{1,...M\}$ such that $\pi(\dot{s}) \in V^{(0)}$,*

$$r_s\,\rho^{-1} < 1. \tag{6.27}$$

Proof. Fix $1 \le s \le M$, let $x = \pi(\dot{s})$, and let $f = 1_x \in C(V^{(0)})$. Then

$$\mathcal{E}_A(f,f) = \sum_{y\in V^{(0)},\, y\neq x} A_{xy} = |A_{xx}|.$$

Let $g = 1_x \in C(V^{(1)})$. As $\Lambda(\mathcal{E}_A) = \rho^{-1}\mathcal{E}_A$,

$$\rho^{-1}|A_{xx}| = \Lambda(\mathcal{E}_A)(f,f) < \mathcal{E}_A^R(g,g): \tag{6.28}$$

since g is not harmonic with respect to $\mathcal{E}_A^R$, strict inequality holds in (6.28). By Proposition 5.24(c), x is in exactly one 1-complex. So

$$\mathcal{E}_A^R(g,g) = \sum_i r_i^{-1}\mathcal{E}_A(g\circ\psi_i, g\circ\psi_i) = r_s^{-1}|A_{xx}|,$$

and combining this with (6.28) gives (6.27). □

Since $r_s = 1$ for nested fractals, we deduce

Corollary 6.28. *Let $(F,(\psi_i))$ be a nested fractal. Then $\rho > 1$.*

For nested fractals, many properties of the process can be summarized in terms of certain scaling factors.

Definition 6.29. Let $(F,(\psi_i))$ be a nested fractal, and $\mathcal{E}$ be the (unique) non-degenerate fixed point. See Definition 5.22 for the length and mass scale factors L and M. The *resistance scale factor* ρ of F is the resistance scaling factor of $\mathcal{E}$. Let

(6.29 $$\tau = M\rho;$$

we call τ the *time scaling factor*. (In view of the connection between resistances and crossing times given in Theorem 4.27, it is not surprising that τ should have a connection with the space-time scaling of processes on F.)

It may be helpful at this point to draw a rough distinction between two kinds of structure associated with the nested fractal (F,ψ). The quantities introduced in Section 5, such as L, M, the geodesic metric d_F, the chemical exponent γ and the dimension $d_w(F)$ are all *geometric* – that is, they can be determined entirely by a geometric inspection of F. On the other hand, the resistance and time scaling factors ρ and τ are *analytic* or *physical* – they appear in some sense to lie deeper than the geometric quantities, and arise from the solution to some kind of equation on the space. On the Sierpinski gasket, for example, while one obtains $L = \gamma = 2$, and $M = 3$ almost immediately, a brief calculation (Lemma 2.16) is needed to obtain ρ. For more complicated sets, such as some of the examples given in Section 5, the calculation of ρ would be very lengthy.

Unfortunately, while the distinction between these two kinds of constant arises clearly in practice, it does not seem easy to make it precise. Indeed, Corollary 6.20 shows that the geometry does in fact determine ρ: it is not possible to have one nested fractal (a geometric object) with two distinct analytic structures which both satisfy the symmetry and scale invariance conditions.

We have the following general inequalities for the scaling factors.

Proposition 6.30. *Let $(F, (\psi_i))$, be a nested fractal with scaling factors L, M, ρ, τ. Then*

$$L > 1, \qquad M \geq 2, \qquad M \geq L, \qquad \tau = M\rho \geq L^2. \tag{6.30}$$

Proof. $L > 1$, $M \geq 2$ follow from the definition of nested fractals. If $\theta = \operatorname{diam}(V^{(0)})$, then, as $V^{(1)}$ consists of M copies of $V^{(0)}$ each of diameter $L^{-1}\theta$, by the connectivity axiom we deduce $ML^{-1}\theta \geq \theta$. Thus $M \geq L$.

To prove the final inequality in (6.30) we use the same strategy as in Proposition 6.27, but with a better choice of minimizing function.

Let $\mathcal{H}$ be the set of functions f of the form $f(x) = Ox + a$, where $x \in \mathbb{R}^d$ and O is an orthogonal matrix. Set $\mathcal{H}_n = \{f|_{V^{(n)}},\ f \in \mathcal{H}\}$. Let $\theta = \sup\{\mathcal{E}(f,f) : f \in \mathcal{H}_0\}$: clearly $\theta < \infty$. Choose f to attain the supremum, and let $g \in \mathcal{H}$ be such that $f = g|_{V^{(0)}}$. Then if $f_1 = g|_{V^{(1)}}$

$$\rho^{-1}\theta = \rho^{-1}\mathcal{E}(f,f) = \Lambda(\mathcal{E})(f,f) \leq \mathcal{E}^R(g_1, g_1) = \sum_{i=1}^{M} \mathcal{E}(g_1 \circ \psi_i, g_1 \circ \psi_i).$$

However, $g_1 \circ \psi_i$ is the restriction to $V^{(0)}$ of a function of the form $L^{-1}Ox + a_i$, and so $\mathcal{E}(g \circ \psi_i, g \circ \psi_i) \leq L^{-2}\theta$. Hence $\rho^{-1}\theta \leq ML^{-2}\theta$, proving (6.30). □

The following comparison theorem provides a technique for bounding ρ in certain situations.

Proposition 6.31. *Let $(F_1, \{\psi_i, 1 \leq i \leq M_1\})$ be a p.c.f.s.s. set. Let $F_0 \subset F_1$, $M_0 \leq M_1$, and suppose that $(F_0, \{\psi_i, 1 \leq i \leq M_0\})$ is also a p.c.f.s.s. set, and that $V^{(0)}_{F_1} = V^{(0)}_{F_0}$. Let $(r^{(k)}_i, 1 \leq i \leq M_k)$ be resistance vectors for $k = 0, 1$, and suppose that $r^{(0)}_i \geq r^{(1)}_i$ for $1 \leq i \leq M_0$. Let Λ_k be the renormalization map for $(F_k, (\psi_i)^{M_k}_{i=1}, (r^{(k)}_i)^{M_k}_{i=1})$. If $\mathcal{E}_k$ are non-degenerate Dirichlet forms satisfying $\Lambda_k(\mathcal{E}_k) = \rho_k^{-1}\mathcal{E}_k$, $k = 0, 1$, then $\rho_1 \leq \rho_0$.*

Proof. Since $V^{(0)}_{F_1} \subset V^{(1)}_{F_1}$, we have, writing R_i for the replication maps associated with F_i,

$$R_1\mathcal{E}(f,f) \geq R_0\mathcal{E}(f,f), \quad f \in C(V^{(1)}_{F_1}).$$

So $\Lambda_1(\mathcal{E}) \geq \Lambda_0(\mathcal{E})$ for any $\mathcal{E} \in \mathbb{D}$. If $m = m(\mathcal{E}_1/\mathcal{E}_0)$, then

$$\rho_1^{-1}\mathcal{E}_1 = \Lambda_1(\mathcal{E}_1) \geq \Lambda_1(m\,\mathcal{E}_0) \geq \Lambda_0(m\,\mathcal{E}_0) = m\rho_0^{-1}\mathcal{E}_0 \geq \rho_0^{-1}\mathcal{E}_1\,,$$

which implies that $\rho_0 \geq \rho_1$. □

7. Diffusions on p.c.f.s.s. sets.

Let $\big(F,(\psi_i)\big)$ be a p.c.f.s.s. set, and r_i be a resistance vector. We assume that the graph $(V^{(1)},\mathbf{E}_1)$ is connected. Suppose that the renormalization map Λ has a non-degenerate fixed point $\mathcal{E}^{(0)} = \mathcal{E}_A$, so that $\Lambda(\mathcal{E}^{(0)}) = \rho^{-1}\mathcal{E}^{(0)}$. Fixing F, r, and $\mathcal{E}_A$, in this section we will construct a diffusion X on F, as a limit of processes on the graphical approximations $V^{(n)}$. In Section 2 this was done probabilistically for the Sierpinski gasket, but here we will use Dirichlet form methods, following [Kus2, Fu1, Ki2].

Definition 7.1. For $f \in C(V^{(n)})$, set

$$\mathcal{E}^{(n)}(f,f) = \rho^n \sum_{w\in\mathbb{W}_n} r_w^{-1}\mathcal{E}^{(0)}(f\circ\psi_w, f\circ\psi_w). \tag{7.1}$$

This is the Dirichlet form on $V^{(n)}$ obtained by replication of scaled copies of $\mathcal{E}^{(0)}$, where the scaling associated with the map ψ_w is $\rho^n r_w^{-1}$.

These Dirichlet forms have the following nesting property.

Proposition 7.2. *(a) For $n \geq 1$, $Tr(\mathcal{E}^{(n)}|V^{(n-1)} = \mathcal{E}^{(n-1)}$.*
(b) If $f \in C(V^{(n)})$, and $g = f|_{V^{(n-1)}}$ then $\mathcal{E}^{(n)}(f,f) \geq \mathcal{E}^{(n-1)}(g,g)$.
(c) $\mathcal{E}^{(n)}$ is non-degenerate.

Proof. (a) Let $f \in C(V^{(n)})$. Then decomposing $w \in \mathbb{W}_n$ into $v\cdot i$, $v \in \mathbb{W}_{n-1}$,

$$\begin{aligned}\mathcal{E}^{(n)}(f,f) &= \rho^n \sum_{v\in\mathbb{W}_{n-1}} r_v^{-1}\sum_i r_i^{-1}\mathcal{E}^{(0)}(f\circ\psi_v\circ\psi_i, f\circ\psi_v\circ\psi_i)\\ &= \rho^{n-1}\sum_{v\in\mathbb{W}_{n-1}} r_v^{-1}\mathcal{E}^{(1)}(f_v,f_v),\end{aligned} \tag{7.2}$$

where $f_v = f\circ\psi_v \in C(V^{(1)})$. Now let $g \in C(V^{(n-1)})$. If $f|_{V^{(n-1)}} = g$ then $f_v|_{V^{(0)}} = g\circ\psi_v = g_v$. As $\mathcal{E}^{(0)}$ is a fixed point of Λ,

$$\begin{aligned}\inf\Big\{\mathcal{E}^{(1)}(h,h) : h|_{V^{(0)}} = g_v\Big\} &= \rho\inf\Big\{R\mathcal{E}^{(0)}(h,h) : h|_{V^{(0)}} = g_v\Big\}\\ &= \rho\Lambda(\mathcal{E}^{(0)})(g_v,g_v) = \mathcal{E}^{(0)}(g_v,g_v).\end{aligned} \tag{7.3}$$

Summing over $v \in \mathbb{W}_{n-1}$ we deduce therefore

$$\inf\Big\{\mathcal{E}^{(n)}(f,f) : f|_{V^{(n-1)}} = g\Big\} \leq \rho^{n-1}\sum_v r_v^{-1}\mathcal{E}^{(0)}(g,g) = \mathcal{E}^{(n-1)}(g,g).$$

For each $v \in \mathbb{W}_{n-1}$, let $h_v \in C(V^{(1)})$ be chosen to attain the infimum in (7.3). We wish to define $f \in C(V^{(n)})$ such that

$$f \circ \psi_v = h_v, \qquad v \in \mathbb{W}_{n-1}. \tag{7.4}$$

Let $v \in \mathbb{W}_{n-1}$. We define

$$f\big(\psi_v(y)\big) = h_v(y), \qquad y \in V^{(1)}.$$

We need to check f is well-defined; but if v, u are distinct elements of $\mathbb{W}_{n-1}$ and $x = \psi_v(y) = \psi_u(z)$, then $x \in V^{(n-1)}$ by Lemma 5.18, and so y, $z \in V^{(0)}$. Therefore

$$f\big(\psi_v(y)\big) = h_v(y) = g_v(y) = g(x) = f\big(\psi_u(z)\big),$$

so the definitions of f at x agree. (This is where we use the fact that F is finitely ramified: it allows us to minimize separately over each set of the form $V_v^{(1)}$).

So

$$\mathcal{E}^{(n)}(f,f) = \mathcal{E}^{(n-1)}(g,g),$$

and therefore $Tr\left(\mathcal{E}^{(n)}|V^{(n-1)}\right) = \mathcal{E}^{(n-1)}$.
(b) is evident from (a).
(c) We prove this by induction. $\mathcal{E}^{(0)}$ is non-degenerate by hypothesis. Suppose $\mathcal{E}^{(n-1)}$ is non-degenerate, and that $\mathcal{E}^{(n)}(f,f) = 0$. From (7.2) we have

$$\mathcal{E}^{(n)}(f,f) = \rho \sum_{v \in \mathbb{W}_1} r_v^{-1} \mathcal{E}^{(n-1)}(f \circ \psi_v, f \circ \psi_v),$$

and so $f \circ \psi_v$ is constant for each $v \in \mathbb{W}_1$. Thus f is constant on each 1-complex, and as $(V^{(1)}, \mathbf{E}_1)$ is connected this implies that f is constant. $\square$

To avoid clumsy notation we will identify functions with their restrictions, so, for example, if $f \in C(V^{(n)})$, and $m < n$, we will write $\mathcal{E}^{(m)}(f,f)$ instead of $\mathcal{E}^{(m)}\left(f|_{V^{(m)}}, f|_{V^{(m)}}\right)$.

Definition 7.3. Set $V^{(\infty)} = \cup_{n=0}^{\infty} V^{(n)}$. Let $U = \{f : V^{(\infty)} \to \mathbb{R}\}$. Note that the sequence $\left(\mathcal{E}^{(n)}(f,f)\right)_{n=1}^{\infty}$ is non-decreasing. Define

$$\mathcal{D}' = \{f \in U : \sup_n \mathcal{E}^{(n)}(f,f) < \infty\},$$

$$\mathcal{E}'(f,g) = \sup_n \mathcal{E}^{(n)}(f,g); \quad f,g \in \mathcal{D}'.$$

$\mathcal{E}'$ is the initial version of the Dirichlet form we are constructing.

Lemma 7.4. *$\mathcal{E}'$ is a symmetric Markov form on $\mathcal{D}'$.*

Proof. $\mathcal{E}'$ clearly inherits the properties of symmetry, bilinearity, and positivity from the $\mathcal{E}^{(n)}$. If $f \in \mathcal{D}'$, and $g = (0 \vee f) \wedge 1$ then $\mathcal{E}^{(n)}(g,g) \le \mathcal{E}^{(n)}(f,f)$, as the $\mathcal{E}^{(n)}$ are Markov. So $\mathcal{E}'(g,g) \le \mathcal{E}'(f,f)$. $\square$

What we have done here seems very easy. However, more work is needed to obtain a 'good' Dirichlet form $\mathcal{E}$ which can be associated with a diffusion on F. Note the following scaling result for $\mathcal{E}'$.

Lemma 7.5. *For $n \geq 1$, $f \in \mathcal{D}'$,*

$$\mathcal{E}'(f,f) = \sum_{w \in \mathbb{W}_n} \rho^n r_w^{-1} \mathcal{E}'(f \circ \psi_w, f \circ \psi_w). \tag{7.5}$$

Proof. We have, for $m \geq n$, $f \in \mathcal{D}'$,

$$\mathcal{E}^{(m)}(f,f) = \sum_{w \in \mathbb{W}_n} \rho^n r_w^{-1} \mathcal{E}^{(m-n)}(f \circ \psi_w, f \circ \psi_w).$$

Letting $m \to \infty$ it follows, first that $f \circ \psi_w \in \mathcal{D}'$, and then that (7.5) holds. $\square$

If H is a set, and $f : H \to \mathbb{R}$, we write

$$\operatorname{Osc}(f,B) = \sup_{x,y \in B} |f(x) - f(y)|, \quad B \subset H. \tag{7.6}$$

Lemma 7.6. There exists a constant c_0, depending only on $\mathcal{E}$, such that

$$\operatorname{Osc}(f, V^{(0)}) \leq c_0 \mathcal{E}^{(0)}(f,f), \qquad f \in C(V^{(0)}).$$

Proof. Let $\widetilde{E}_0 = \big\{\{x,y\} : A_{xy} > 0\big\}$. As $\mathcal{E}^{(0)}$ is non-degenerate, $(V^{(0)}, \widetilde{E}_0)$ is connected; let N be the maximum distance between points in this graph. Set $\alpha = \min\{A_{xy}, \{x,y\} \in \widetilde{E}_0\}$. If x, $y \in V^{(0)}$, there exists a chain $x = x_0, x_1, \ldots, x_n = y$ connecting x, y with $n \leq N$, and therefore,

$$\begin{aligned} |f(x) - f(y)|^2 &\leq \left(\sum_{i=1}^n |f(x_i) - f(x_{i-1})| \right)^2 \\ &\leq n \sum_{i=1}^n |f(x_i) - f(x_{i-1})|^2 \\ &\leq n\alpha^{-1} \sum_{i=1}^n A_{x_{i-1},x_i} |f(x_i) - f(x_{i-1})|^2 \\ &\leq N\alpha^{-1} \mathcal{E}^{(0)}(f,f). \end{aligned}$$

$\square$

Since $V^{(1)}$ consists of M copies of $V^{(0)}$ we deduce a similar result for $V^{(1)}$.

Corollary 7.7. *There exists a constant $c_1 = c_1(F,r,A)$ such that*

$$\operatorname{Osc}(f, V^{(1)}) \leq c_1 \mathcal{E}^{(1)}(f,f), \qquad f \in \mathcal{D}'. \tag{7.7}$$

Proof. For $i \in \mathbb{W}_1$, $f \in C(V^{(1)})$,

$$\operatorname{Osc}(f, V_i^{(0)}) = \operatorname{Osc}(f \circ \psi_i, V^{(0)}) \leq c_0 \mathcal{E}^{(0)}(f \circ \psi_i, f \circ \psi_i).$$

So, as $V^{(1)}$ is connected,

$$\begin{aligned} \operatorname{Osc}(f, V^{(1)}) &\leq \sum_i \operatorname{Osc}(f, V_i^{(0)}) \\ &\leq \sum_i c_0 \mathcal{E}^{(0)}\left(f \circ \psi_i, f \circ \psi_i\right) \leq c_1 \mathcal{E}^{(1)}(f,f), \end{aligned}$$

where c_1 is chosen so that $c_0 \le c_1 \rho r_i^{-1}$ for each $i \in \mathbb{W}_1$. □

Corollary 7.8. *Let $w \in \mathbb{W}_n$, and x, $y \in V_w^{(1)}$. Then*

$$\mathrm{Osc}(f, V_w^{(1)}) \le c_1 r_w \rho^{-n} \mathcal{E}'(f,f), \qquad f \in \mathcal{D}'.$$

Proof. We have $\mathrm{Osc}(f, V_w^{(1)}) = \mathrm{Osc}(f \circ \psi_w, V^{(1)}) \le c_1 \mathcal{E}^{(1)}(f \circ \psi_w, f \circ \psi_w)$. Since $\mathcal{E}^{(1)} \le \mathcal{E}'$, and by (7.5)

$$\mathcal{E}'(f \circ \psi_w, f \circ \psi_w) \le r_w \rho^{-n} \mathcal{E}'(f,f),$$

the result is immediate. □

Definition 7.9. We will call the fixed point $\mathcal{E}^{(0)}$ a *regular fixed point* if

$$r_i < \rho \qquad \text{for} \quad 1 \le i \le M. \tag{7.8}$$

Proposition 6.27 implies that (7.8) holds for any $s \in \{1, \dots, M\}$ such that $\pi(\dot{s}) \in V^{(0)}$. In particular therefore, for nested fractals, where every point in $V^{(0)}$ is of this form and r is constant, any fixed point is regular.

It is not hard to produce examples of non-regular fixed points. Consider the Lindstrøm snowflake, but with $r_i = 1$, $1 \le i \le 6$, $r_7 = r > 1$. Writing $\rho(r)$ for the resistance scale factor, we have (by Proposition 6.31) that $\rho(r)$ is increasing in r. However, also by Proposition 6.31, $\rho(r) \le \rho_0$, where ρ_0 is the resistance scale factor of the nested fractal obtained just from ψ_i, $1 \le i \le 6$. So if we choose $r_7 > \rho_0$, then as $r_7 > \rho_0 \ge \rho(r_7)$, we have an example of an affine nested fractal with a non-regular fixed point.

From now on we take $\mathcal{E}^{(0)}$ to be a regular fixed point. (See [Kum3] for the general situation). Write $\gamma = \max_i r_i/\rho < 1$. For x, $y \in F$, set $w(x,y)$ to be the longest word w such that x, $y \in F_w$.

Proposition 7.10. *(Sobolev inequality). Let $f \in \mathcal{D}'$. Then if $\mathcal{E}^{(0)}$ is a regular fixed point*

$$|f(x) - f(y)|^2 \le c_2 r_{w(x,y)} \rho^{-|w(x,y)|} \mathcal{E}'(f,f), \qquad x, y \in V^{(\infty)}. \tag{7.8}$$

Proof. Let x, $y \in V^{(n)}$, let $w = w(x,y)$ and let $|w| = m$. We prove (7.8) by a standard kind of chaining argument, similar to those used in continuity results such as Kolmogorov's lemma. (But this argument is deterministic and easier). We may assume $n \ge m$.

Let $u \in \mathbb{W}_n$ be an extension of w, such that $x \in V_u^{(0)}$: such a u certainly exists, as $x \in V_n^{(0)} \cap F_w$. Write $u_k = u|k$ for $m \le k \le n$. Now choose a sequence z_k, $m \le k \le n$ such that $z_n = x$, and $z_k \in V_{u_k}^{(0)}$ for $k \le m \le n-1$. For each $k \in \{m, \dots, n-1\}$ we have z_k, $z_{k+1} \in V_{u_k}^{(1)}$. So

$$\begin{aligned} |f(z_n) - f(z_m)| &\le \sum_{k=m}^{n-1} |f(z_{k+1}) - f(z_k)| \\ &\le \sum_{k=m}^{n-1} \left(c_1 r_{u_k} \rho^{-k} \mathcal{E}(f,f)\right)^{1/2} \end{aligned} \tag{7.9}$$

$$= \big(c_1 r_w \rho^{-m} \mathcal{E}(f,f)\big)^{1/2} \Big(\sum_{k=m}^{n-1} \frac{r_{u_k}}{r_w} \rho^{-k+m}\Big)^{1/2}.$$

As $\mathcal{E}$ is a regular fixed point, $\gamma = \max_i r_i/\rho < 1$, so the final sum in (7.9) is bounded by $(\sum_{k=m}^{\infty} \gamma^{k-m})^{1/2} = c_3 < \infty$. Thus we have

$$|f(x) - f(z_m)|^2 \le c_1 c_3 r_w \rho^{-n} \mathcal{E}'(f,f),$$

and as a similar bound holds for $|f(y) - f(z_m)|^2$, this proves (7.8). □

We have not so far needed a measure on F. However, to define a Dirichlet form we need some L^2 space in which the domain of $\mathcal{E}$ is closed. Let μ be a probability measure on $(F, \mathcal{B}(F))$ which charges every set of the form F_w, $w \in \mathbb{W}_n$. Later we will take μ to be the Bernouilli measure μ_θ associated with a vector of weights $\theta \in (0,\infty)^M$, but for now any measure satisfying the condition above will suffice.

As $\mu(F) = 1$, $C(F) \subset L^2(F,\mu)$. Set

$$\begin{aligned} &\mathcal{D} = \{f \in C(F) : f|_{V^{(\infty)}} \in \mathcal{D}'\} \\ &\mathcal{E}(f,f) = \mathcal{E}'\left(f|_{V^{(\infty)}}, f|_{V^{(\infty)}}\right), \quad f \in \mathcal{D}. \end{aligned}$$

Proposition 7.11. *$(\mathcal{E}, \mathcal{D})$ is a closed symmetric form on $L^2(F,\mu)$.*

Proof. Note first that the condition on μ implies that if $f, g \in \mathcal{D}$ then $||f-g||_2 = 0$ implies that $f = g$. We need to prove that $\mathcal{D}$ is complete in the norm $\|f\|^2_{\mathcal{E}_1} = \mathcal{E}(f,f) + \|f\|_2^2$. So suppose (f_n) is Cauchy in $\|\cdot\|_{\mathcal{E}_1}$. Since (f_n) is Cauchy in $\|\cdot\|_2$, passing to a subsequence there exists $\widetilde{f} \in L^2(F,\mu)$ such that $f_n \to \widetilde{f}$ μ–a.e. Fix $x_0 \in F$ such that $f_n(x_0) \to \widetilde{f}(x)$. Then since $f_n - f_m$ is continuous, (7.8) extends to an estimate on the whole of F and so

$$\begin{aligned} |f_n(x) - f_m(x)| &\le |(f_n - f_m)(x) - (f_n - f_m)(x_0)| + |(f_n - f_m)(x_0)| \\ &\le c_2^{1/2} \mathcal{E}(f_n - f_m, f_n - f_m)^{1/2} + |f_n(x_0) - f_m(x_0)|. \end{aligned}$$

So (f_n) is Cauchy in the uniform norm, and thus there exists $f \in C(F)$ such that $f_n \to f$ uniformly.

Let $n \ge 1$. Then as $\mathcal{E}^{(n)}(g,g)$ is a finite sum,

$$\begin{aligned} \mathcal{E}^{(n)}(f,f) = \lim_{m\to\infty} \mathcal{E}^{(n)}(f_m, f_m) &\le \limsup_{m\to\infty} \mathcal{E}(f_m, f_m) \\ &\le \sup_m \|f_m\|_{\mathcal{E}_1} < \infty. \end{aligned}$$

Hence $\mathcal{E}^{(n)}(f,f)$ is bounded, so $f \in \mathcal{D}$. Finally, by a similar calculation, for any $N \ge 1$,

$$\mathcal{E}^{(N)}(f_n - f, f_n - f) \le \lim_{m\to\infty} \mathcal{E}(f_n - f_m, f_n - f_m).$$

So $\mathcal{E}(f_n - f, f_n - f) \to 0$ as $n \to \infty$, and thus $\|f - f_n\|^2_{\mathcal{E}_1} \to 0$. □

To show that $(\mathcal{E}, \mathcal{D})$ is a Dirichlet form, it remains to show that $\mathcal{D}$ is dense in $L^2(F,\mu)$. We do this by studying the harmonic extension of a function.

Definition 7.12. Let $f \in C(V^{(n)})$. Recall that $\mathcal{E}^{(n)}(f,f) = \inf\{\mathcal{E}^{(n+1)}(g,g) : g|_{V^{(n)}} = f\}$. Let $\widetilde{H}_{n+1}f \in C(V^{(n+1)})$ be the (unique, as $\mathcal{E}^{(n+1)})$ is non-degenerate) function which attains the infimum.

For $x \in V^{(\infty)}$ set

$$\widehat{H}_n f(x) = \lim_{m\to\infty} \widetilde{H}_m \widetilde{H}_{m-1} \dots \widetilde{H}_{n+1} f(x);$$

note that (as $\widetilde{H}_{n+1}f = f$ on $V^{(n)}$) this limit is ultimately constant.

Proposition 7.13. *Let $\mathcal{E}$ be a regular fixed point.*
(a) $\widehat{H}_n f$ has a continuous extension to a function $H_n f \in \mathcal{D} \cap C(F)$, which satisfies

$$\mathcal{E}(H_n f, H_n f) = \mathcal{E}^{(n)}(f,f).$$

(b) If f, $g \in C(F)$

$$\mathcal{E}(H_n f, g) = \mathcal{E}^{(n)}(f,g). \tag{7.10}$$

Proof. From the definition of $\widetilde{H}_{n+1}$, $\mathcal{E}^{(n+1)}(\widetilde{H}_{n+1}f, \widetilde{H}_{n+1}f) = \mathcal{E}^{(n)}(f,f)$. Thus $\mathcal{E}^{(m)}(\widehat{H}_n f, \widehat{H}_n f) = \mathcal{E}^{(n)}(f,f)$ for any m, so that $\widehat{H}_n f \in \mathcal{D}'$ and

$$\mathcal{E}(\widehat{H}_n f, \widehat{H}_n f) = \mathcal{E}^{(n)}(f,f), \qquad f \in C(V^{(n)}).$$

If $w \in \mathbb{W}_m$, and x, $y \in V^{(\infty)} \cap F_w$ then by Proposition 7.10

$$|\widehat{H}_n f(x) - \widehat{H}_n f(y)|^2 \le c_2 r_w \rho^{-m} \mathcal{E}^{(n)}(f,f). \tag{7.11}$$

Since $r_w \rho^{-m} \le \gamma^m$, (7.11) implies that $\mathrm{Osc}(\widehat{H}_n f, V^{(\infty)} \cap F_w)$ converges to 0 as $|w| = m \to \infty$. Thus $\widehat{H}_n f$ has a continuous extension $H_n f$, and $H_n f \in \mathcal{D}$ since $\widehat{H}_n f \in \mathcal{D}'$.
(b) Note that, by polarization, we have

$$\mathcal{E}^{(n+1)}(\widetilde{H}_{n+1}f, \widetilde{H}_{n+1}g) = \mathcal{E}^{(n)}(f,g).$$

Since $\mathcal{E}^{(n+1)}(\widetilde{H}_{n+1}f, h) = 0$ for any h such that $h|_{V^{(n)}} = 0$, it follows that

$$\mathcal{E}^{(n+1)}(\widetilde{H}_{n+1}f, g) = \mathcal{E}^{(n)}(f,g).$$

Iterating, we obtain (7.10). □

Theorem 7.14. *$(\mathcal{E}, \mathcal{D})$ is an irreducible, regular, local Dirichlet form on $L^2(F,\mu)$.*

Proof. Let $f \in C(F)$. Since for any $n \ge 1$, $w \in \mathbb{W}_n$ we have

$$\inf_{F_w} f \le H_n f(x) \le \sup_{F_w} f, \qquad x \in F_w$$

it follows that $H_n f \to f$ uniformly. As $H_n f \in \mathcal{D}$, we deduce that $\mathcal{D}$ is dense in $C(F)$ in the uniform norm. Hence also $\mathcal{D}$ is dense in $L^2(F,\mu)$. As (4.5) is immediate, we deduce that $\mathcal{D}$ is a regular Dirichlet form. If $\mathcal{E}(f,f) = 0$ then $\mathcal{E}^{(n)}(f,f) = 0$ for each n. Since $\mathcal{E}^{(n)}$ is irreducible, $f|_{V^{(n)}}$ is constant for each n. As f is continuous, f is therefore constant. Thus $\mathcal{E}$ is irreducible.

To prove that $\mathcal{E}$ is local, let f, g be functions in $\mathcal{D}$ with disjoint closed supports, S_f, S_g say. If $\mathcal{E}^{(n)}(f,g) \neq 0$ then one of the terms in the sum (7.1) must be non-zero, so there exists $w_n \in \mathbb{W}_n$, and points $x_n \in S_f \cap V^{(0)}_{w_n}$, $y_n \in S_g \cap V^{(0)}_{w_n}$. Passing to a subsequence, there exists z such that $x_n \to z$, $y_n \to z$, and as therefore $z \in S_f \cap S_g$, this is a contradiction. □

By Theorem 4.8 there exists a continuous μ-symmetric Hunt process $(X_t, t \geq 0, \mathbb{P}^x, x \in F)$ associated with $(\mathcal{E}, \mathcal{D})$ and $L^2(F,\mu)$.

Remark 7.15. Note that we have constructed a process $X = X^{(\mu)}$ for each Radon measure μ on F. So, at first sight, the construction given here has built much more than the probabilistic construction outlined in Section 2. But this added generality is to a large extent an illusion: Theorem 4.17 implies that these processes can all be obtained from each other by time-change.

On the other hand the regularity of $(\mathcal{E}, \mathcal{D})$ was established without much pain, and here the advantage of the Dirichlet form approach can be seen: all the probabilistic approaches to the Markov property are quite cumbersome.

The general probabilistic construction, such as given in [L1] for example, encounters another obstacle which the Dirichlet form construction avoids. As well as finding a decimation invariant set of transition probabilities, it also appears necessary (see e.g. [L1, Chapter VI]) to find associated transition times. It is not clear to me why these estimates appear essential in probabilistic approaches, while they do not seem to be needed at all in the construction above.

We collect together a number of properties of $(\mathcal{E}, \mathcal{D})$.

Proposition 7.16. *(a) For each $n \geq 0$*

$$\mathcal{E}(f,g) = \sum_{w \in \mathbb{W}_n} \rho^n r_w^{-1} \mathcal{E}(f \circ \psi_w, g \circ \psi_w). \tag{7.12}$$

(b) For $f \in \mathcal{D}$,

$$|f(x) - f(y)|^2 \leq c_1 r_w \rho^{-n} \mathcal{E}(f,f) \quad \textit{if} \quad x, y \in F_w,\ w \in \mathbb{W}_n \tag{7.13}$$

$$\int f^2 d\mu \leq c_2 \mathcal{E}(f,f) + \Big(\int f d\mu\Big)^2, \tag{7.14}$$

$$f(x)^2 \leq 2\int f^2 d\mu + 2c_1 \mathcal{E}(f,f), \qquad x \in F. \tag{7.15}$$

Proof. (a) is immediate from Lemma 7.5, while (b) follows from Proposition 7.10 and the continuity of f. Taking $n = 0$ in (7.13) we deduce that

$$\big(f(x) - f(y)\big)^2 \leq c_1 \mathcal{E}(f,f), \qquad f \in \mathcal{D}.$$

So as $\mu(F) = 1$,

$$\int\int c_1 \mathcal{E}(f,f)\mu(dx)\mu(dy) = c_1\mathcal{E}(f,f)$$
$$\leq \int\int \big(f(x) - f(y)\big)^2 \mu(dx)\mu(dy)$$
$$= 2\int f^2\, d\mu - 2\Big(\int f\, d\mu\Big)^2,$$

proving (7.14).

Since $f(x)^2 \leq 2f(y)^2 + 2|f(x) - f(y)|^2$ we have from (7.13) that

$$f(x)^2 = \int f(x)^2 \mu(dy)$$
$$\leq 2\int f(y)^2\mu(dy) + 2c_1 \int \mathcal{E}(f,f)\mu(dy),$$

which proves (7.15). □

We need to examine further the resistance metric introduced in Section 4.

Definition 7.17. Let $R(x,x) = 0$, and for $x \neq y$ set

$$R(x,y)^{-1} = \inf\left\{\mathcal{E}(f,f) : f(x) = 0, f(y) = 1, f \in \mathcal{D}\right\}.$$

Note that

(7.16) $$R(x,y) = \sup\Big\{\frac{|f(x) - f(y)|^2}{\mathcal{E}(f,f)} : f \in \mathcal{D}, \quad f \text{ non constant}\Big\}.$$

Proposition 7.18. *(a) If $x \neq y$ then $0 < R(x,y) \leq c_1 < \infty$.*
(b) If $w \in \mathbb{W}_n$ then

(7.17) $$R(x,y) \leq c_1 r_w \rho^{-n}, \qquad x, y \in F_w.$$

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

(7.18) $$|f(x) - f(y)|^2 \leq R(x,y)\mathcal{E}(f,f).$$

(d) R is a metric on F, and the topology induced by R is equal to the original topology on F.

Proof. Let x, y be distinct points in F. As $\mathcal{D}$ is dense in $C(F)$, there exists $f \in \mathcal{D}$ with $f(x) \geq 1$, $f(y) \leq 0$. Since $\mathcal{E}$ is irreducible, $\mathcal{E}(f,f) > 0$, and so by (7.16) $R(x,y) > 0$. (7.17) is immediate from Proposition 7.16, proving (b). Taking $n = 0$, and w to be the empty word in (7.17) we deduce $R(x,y) \leq c_1$ for any x, $y \in F$, completing the proof of (a).
(c) is immediate from (7.16).
(d) R is clearly symmetric. The triangle inequality for R is proved exactly as in Proposition 4.25, by considering the trace of $\mathcal{E}$ on the set $\{x, y, z\}$.

It remains to show that the topologies induced by R and d (the original metric on F) are the same. Let $R(x_n, x) \to 0$. If $\varepsilon > 0$, there exists $f \in \mathcal{D}$ with $f(x) = 1$ and $\mathrm{supp}(f) \subset B_d(x, \varepsilon)$. By (7.16) $R(x, y) \geq \mathcal{E}(f, f)^{-1} > 0$ for any $y \in B_d(x, \varepsilon)^c$. So $x_n \in B_d(x, \varepsilon)$ for all sufficiently large n, and hence $d(x_n, x) \to 0$.

If $d(x_n, x) \to 0$ then writing

$$N_m(x) = \bigcup \{F_w : w \in \mathbb{W}_m, \quad x \in F_w\}$$

we have by Lemma 5.12 that $x_n \in N_m(x)$ for all sufficiently large n. However if $\gamma = \max_i r_i/\rho < 1$ we have by, (7.17), $R(x, y) \leq c_1 \gamma^m$ for $y \in N_m(x)$. Thus $R(x_n, x) \to 0$. □

Remark 7.19. The resistance metric R on F is quite well adapted to the study of the diffusion X on F. Note however that $R(x, y)$ is obtained by summing (in a certain sense) the resistance of all paths from x to y. So it is not surprising that R is not a geodesic metric. (Unless F is a tree).

Also, R is not a geometrically natural metric on F. For example, on the Sierpinski gasket, since $r_i = 1$, and $\rho = 5/3$, we have that if x, y are neighbours in $(V^{(n)}, \mathbf{E}_n)$ then

$$R(x, y) \asymp (3/5)^n.$$

However, for general p.c.f.s.s. sets it is not easy to define a metric which is well-adapted to the self-similar structure. (And, if one imposes strict conditions of exact self-similarity, it is not possible in general – see the examples in [Ki6]). So, for these general sets the resistance metric plays an extremely useful role. The next section contains some additional results on R.

It is also worth remarking that the balls $B_R(x, r) = \{y : R(x, y) < r\}$ need not in general be connected. For example, consider the wire network corresponding to the graph consisting of two points x, y, connected by n wires each of conductivity 1. Let z be the midpoint of one of the wires. Then $R(x, y) = 1/n$, while the conductivities in the network $\{x, y, z\}$ are given by $C(x, z) = C(z, y) = 2$, $C(x, y) = n - 1$. So, after some easy calculations,

$$R(x, z) = \frac{n+1}{4n-1} > \tfrac{1}{4}.$$

So if $n = 4$, $R(x, y) = \frac{1}{4}$ while $R(x, z) = \frac{1}{3}$. Hence if $\frac{1}{4} < r < \frac{1}{3}$ the ball $B_R(x, r)$ is not connected. (In fact, y is an isolated point of $\overline{B}_R(x, \frac{1}{4}) = \{x' : d(x, x') \leq \frac{1}{4}\}$). (Are the balls $B_R(x, r)$ in the Sierpinski gasket connected? I do not know).

Recall the notation $\mathcal{E}_\alpha(f, g) = \mathcal{E}(f, g) + \alpha(f, g)$. Let $(U_\alpha, \alpha > 0)$ be the resolvent of X. Since by (4.8) we have

$$\mathcal{E}_\alpha(U_\alpha f, g) = (f, g),$$

if U_α has a density $u_\alpha(x, y)$ with respect to μ, then a formal calculation suggests that

$$\mathcal{E}_\alpha\big(u_\alpha(x, \cdot), g\big) = \mathcal{E}_\alpha(U_\alpha \delta_x, g) = (\delta_x, g) = g(x).$$

We can use this to obtain the existence and continuity of the resolvent density u_α. (See [FOT, p. 73]).

Theorem 7.20. *(a) For each $x \in F$ there exists $u_\alpha^x \in \mathcal{D}$ such that*

$$\mathcal{E}_\alpha(u_\alpha^x, f) = f(x) \qquad \text{for all} \quad f \in \mathcal{D}. \tag{7.19}$$

(b) Writing $u_\alpha(x,y) = u_\alpha^x(y)$, we have

$$u_\alpha(x,y) = u_\alpha(y,x) \qquad \text{for all} \quad x, y \in F.$$

(c) $u_\alpha(\cdot,\cdot)$ is continuous on $F \times F$ and in particular

$$|u_\alpha(x,y) - u_\alpha(x,y')|^2 \le R(y,y')u_\alpha(x,x). \tag{7.20}$$

(d) $u_\alpha(x,y)$ is the resolvent density for X: for $f \in C(F)$,

$$E^x \int_0^\infty e^{-\alpha t} f(X_t)dt = U_\alpha f(x) = \int u_\alpha(x,y)f(y)\mu(dy).$$

(e) There exists $c_2(\alpha)$ such that

$$u_\alpha(x,y) \le c_2(\alpha), \qquad x, y \in F. \tag{7.21}$$

Proof. (a) The existence of u_α^x is given by a standard argument with reproducing kernel Hilbert spaces. Let $x \in F$, and for $f \in \mathcal{D}$ let $\phi(f) = f(x)$. Then by (7.15)

$$|\phi(f)|^2 = |f(x)|^2 \le 2\|f\|_2^2 + 2c_1\mathcal{E}(f,f) \le c_\alpha \mathcal{E}_\alpha(f,f),$$

where $c_\alpha = 2\max(c_1, \alpha^{-1})$. Thus ϕ is a bounded linear functional on the Hilbert space $(\mathcal{D}, \| \ \|_{\mathcal{E}_\alpha})$, and so there exists a $u_\alpha^x \in \mathcal{D}$ such that

$$\phi(f) = \mathcal{E}_\alpha(u_\alpha^x, f) = f(x), \qquad f \in \mathcal{D}.$$

(b) This is immediate from (a) and the symmetry of $\mathcal{E}$:

$$u_\alpha^y(x) = \mathcal{E}_\alpha(u_\alpha^x, u_\alpha^y) = \mathcal{E}_\alpha(u_\alpha^y, u_\alpha^x) = u_\alpha^x(y).$$

(c) As $u_\alpha^x \in \mathcal{D}$, $u_\alpha(x,x) < \infty$. Since $\mathcal{E}(u_\alpha^x, u_\alpha^x) = u_\alpha(x,x) < \infty$, the estimate (7.20) follows from (7.18). It follows immediately that u is jointly continuous on $F \times F$.
(d) This follows from (7.19) and linearity. For a measure ν on F set

$$V_\nu f(x) = \int u_\alpha(x,y)f(y)\nu(dy), \qquad f \in C(F).$$

As u_α is uniformly continuous on $F \times F$, we can choose $\nu_n \xrightarrow{w} \mu$ so that $V_{\nu_n} f \to Vf$ uniformly, and ν_n are atomic with a finite number of atoms. Write $V_n = V_{\nu_n}$, $V = V_\mu$. Since by (7.19)

$$\begin{aligned}\mathcal{E}_\alpha(V_n f, g) &= \sum_x \nu_n(\{x\}) f(x) \mathcal{E}_\alpha(u_a^x, g) \\ &= \sum_x f(x)g(x)\nu_n(\{x\}) = \int fg \, d\nu_n,\end{aligned}$$

we have

$$\mathcal{E}_\alpha(V_nf - V_mf, V_nf - V_mf) = \\ \int f(V_nf - V_mf)\, d\nu_n - \int f(V_nf - V_mf)\, d\nu_m.$$

Thus $\mathcal{E}_\alpha(V_nf - V_mf, V_nf - V_mf) \to 0$ as m, $n \to \infty$, and so, as $\mathcal{E}$ is closed, we deduce that $Vf \in \mathcal{D}$ and $\mathcal{E}_\alpha(Vf, g) = \lim_n \mathcal{E}_\alpha(V_nf, g) = \lim_n \int fg\, d\nu_n = \int fg\, d\mu$. So $\mathcal{E}_\alpha(Vf, g) = \mathcal{E}_\alpha(U_\alpha f, g)$ for all g, and hence $Vf = U_\alpha f$.
(e) As $R(y, y') \le c_1$ for y, $y' \in F$, we have from (7.20) that

$$u_\alpha(x, y) \ge u_\alpha(x, x) - \big(c_1 u_\alpha(x, x)\big)^{1/2}. \tag{7.22}$$

Since $\int u_\alpha(x, y)\mu(dy) = \alpha^{-1}$, integrating (7.22) we obtain

$$u_\alpha(x, x) \le \big(c_1 u_\alpha(x, x)\big)^{1/2} + \alpha^{-1},$$

and this implies that $u_\alpha(x, x) \le c_2(\alpha)$, where $c(\alpha)$ depends only on α and c_1. Using (7.20) again we obtain (7.21). □

Theorem 7.21. *(a) For each $x \in F$, x is regular for $\{x\}$.*
(b) X has a jointly continuous local time $(L^x_t, x \in F, t \ge 0)$ such that for all bounded measurable f

$$\int_0^t f(X_s)\, ds = \int f(a) L^a_t \mu(da), \quad \text{a.s.}$$

Proof. These follow from the estimates on the resolvent density u_α. As u_α is bounded and continuous, we have that x is regular for $\{x\}$. Thus X has jointly measurable local times $(L^x_t, x \in F, t \ge 0)$.

Since X is a symmetric Markov process, by Theorem 8.6 of [MR], L^x_t is jointly continuous in (x, t) if and only if the Gaussian process Y_x, $x \in F$ with covariance function given by

$$EY_aY_b = u_1(a, b), \qquad a, b \in F$$

is continuous. Necessary and sufficient conditions for continuity of Gaussian processes are known (see [Tal]), but here a simple sufficient condition in terms of metric entropy is enough. We have

$$E(Y_a - Y_b)^2 = u_1(a, a) - 2u_1(a, b) + u_1(b, b) \le c_1 R(a, b)^{1/2}.$$

Set $r(a, b) = R(a, b)^{1/2} : r$ is a metric on F. Write $N_r(\varepsilon)$ for the smallest number of sets of r-diameter ε needed to cover F. By (7.17) we have $R(a, b) \le c\gamma^n$ if a, $b \in F_w$ and $w \in \mathbb{W}_n$. So $N_r(c'\gamma^{n/2}) \le \#\mathbb{W}_n = M^n$, and it follows that

$$N_r(\varepsilon) \le c_2 \varepsilon^{-\beta},$$

where $\beta = 2 \log M / \log \theta^{-1}$. So

$$\int_{0+} \big(\log N_r(\varepsilon)\big)^{1/2} d\varepsilon < \infty,$$

and thus by [Du, Thm. 2.1] Y is continuous. $\square$

We can use the continuity of the local time of X to give a simple proof that X is the limit of a natural sequence of approximating continuous time Markov chains. For simplicity we take μ to be a Bernouilli measure of the form $\mu = \mu_\theta$, where $\theta_i > 0$. Let μ_n be the measure on $V^{(n)}$ given in (5.21). Set

$$\begin{aligned} A_t^n &= \int_F L_t^x \mu_n(dx), \\ \tau_t^n &= \inf\{s : A_s^n > t\}, \\ X_t^n &= X_{\tau_t^n}. \end{aligned}$$

Theorem 7.22. *(a) $\big(X_t^n, t \geq 0, \mathbb{P}^x, x \in V^{(n)}\big)$ is the symmetric Markov process associated with $\mathcal{E}^{(n)}$ and $L^2(V^{(n)}, \mu_n)$.*
(b) $X_t^n \to X_t$ a.s. and uniformly on compacts.

Proof. (a) By Theorem 7.21(a) points are non-polar for X. So by the trace theorem (Theorem 4.17) X^n is the Markov process associated with the trace of $\mathcal{E}$ on $L^2(V^{(n)}, \mu_n)$. But for $f \in \mathcal{D}$, by the definition of $\mathcal{E}$,

$$Tr\big(\mathcal{E}|V^{(n)}\big)(f,f) = \mathcal{E}^{(n)}\big(f|_{V^{(n)}}, f|_{V^{(n)}}\big).$$

(b) As F is compact, for each $T > 0$, $(L_t^x, 0 \leq t \leq T, x \in F)$ is uniformly continuous. So, using (5.22), if $T_2 < T_1 < T$ then $A_t^n \to t$ uniformly in $[0, T_1]$, and so $\tau_t^n \to t$ uniformly on $[0, T_2]$. As X is continuous, $X_t^n \to X$ uniformly in $[0, T_2]$. $\square$

Remark 7.23. As in Example 4.21, it is easy to describe the generator L_n of X^n. Let $a^{(n)}(x,y)$, x, $y \in V^{(n)}$ be the conductivity matrix such that

$$\mathcal{E}^{(n)}(f,f) = \tfrac{1}{2} \sum_{x,y} a^{(n)}(x,y)\big(f(x) - f(y)\big)^2.$$

Then by (7.1) we have

$$a^{(n)}(x,y) = \sum_{w \in \mathbb{W}_n} 1_{(x,y \in V_w^{(0)})} \rho^n r_w^{-1} A\big(\psi_w^{-1}(x), \psi_w^{-1}(y)\big), \tag{7.23}$$

where A is such that $\mathcal{E}^{(0)} = \mathcal{E}_A$, and $A(x,y) = A_{xy}$. Then for $f \in L^2(V^{(n)}, \mu_n)$,

$$L_n f(x) = \mu_n(\{x\})^{-1} \sum_{y \in V^{(n)}} a^{(n)}(x,y)\big(f(y) - f(x)\big). \tag{7.24}$$

Of course Theorem 7.22 implies that if (Y^n) is a sequence of continuous time Markov chains, with generators given by (7.24), then $Y^n \xrightarrow{w} X$ in $D([0,\infty), F)$.

8. Transition Density Estimates.

In this section we fix a connected p.c.f.s.s. set $\big(F,(\psi_i)\big)$, a resistance vector r_i, and a non-degenerate regular fixed point $\mathcal{E}_A$ of the renormalization map Λ. Let $\mu = \mu_\theta$ be a measure on F, and let $X = (X_t, t \geq 0, \mathbb{P}^x, x \in F)$ be the diffusion process constructed in Section 7. We investigate the transition densities of the process X: initially in fairly great generality, but as the section proceeds, I will restrict the class of fractals.

We begin by fixing the vector θ which assigns mass to the 1-complexes $\psi_i(F)$, in a fashion which relates $\mu_\theta\big(\psi_i(F)\big)$ with r_i. Let $\beta_i = r_i \rho^{-1}$: by (7.8) we have

$$\beta_i < 1, \quad 1 \leq i \leq M. \tag{8.1}$$

Let $\alpha > 0$ be the unique positive real such that

$$\sum_{i=1}^{M} \beta_i^\alpha = 1. \tag{8.2}$$

Set

$$\theta_i = \beta_i^\alpha, \qquad 1 \leq i \leq M, \tag{8.3}$$

and let $\mu = \mu_\theta$ be the associated Bernouilli type measure on F. Write $\beta_+ = \max_i \beta_i$, $\beta_- = \min_i \beta_i$: we have $0 < \beta_- \leq \beta_i \leq \beta_+ < 1$.

We wish to split the set F up into regions which are, "from the point of view of the process X", all roughly the same size. The approximation Theorem 7.22 suggests that if $w \in \mathbb{W}_n$ then the 'crossing time' of the region F_w is of the order of $\rho^{-n} r_w \theta_w^{-1} = \beta_w \theta_w^{-1} = \beta_w^{1-\alpha}$. (See Proposition 8.10 below for a more precise statement of this fact). So if $r.$ is non-constant the decomposition $F = \cup\{F_w,\, w \in \mathbb{W}_n\}$ of F into n complexes is unsuitable; instead we need to use words w of different lengths. (This idea is due to Hambly – see [Ham2]).

Let $\mathbb{W}_\infty = \cup_{n=0}^\infty \mathbb{W}_n$ be the space of all words of finite length. $\mathbb{W}_\infty$ has a natural tree structure: if $w \in \mathbb{W}_n$ then the parent of w is $w|n-1$, while the offspring of w are the words $w \cdot i$, $1 \leq i \leq M$. (We define the truncation operator τ on $\mathbb{W}_\infty$ by $\tau w = w|\big(|w|-1\big)$.) Write also for $w \in \mathbb{W}_\infty$

$$w \cdot \mathbb{W} = \{w \cdot v, v \in \mathbb{W}\} = \{v \in \mathbb{W} : v_i = w_i,\ 1 \leq i \leq |w|\}.$$

Lemma 8.1. *(a) For $\lambda > 0$ let*

$$\mathbb{W}_\lambda = \{w \in \mathbb{W}_\infty : \beta_w \leq \lambda,\ \beta_{\tau w} > \lambda\}.$$

Then the sets $\{w \cdot \mathbb{W}, w \in \mathbb{W}_\lambda\}$ are disjoint, and

$$\bigcup_{w \in \mathbb{W}_\lambda} w \cdot \mathbb{W} = \mathbb{W}.$$

(b) For $f \in L^1(F,\mu)$,

$$\int f\,d\mu = \sum_{w\in\mathbb{W}_\lambda} \theta_w \int f_w\,d\mu$$
$$\mathcal{E}(f,f) = \sum_{w\in\mathbb{W}_\lambda} \beta_w^{-1}\mathcal{E}(f_w,f_w).$$

Proof. (a) Suppose w, $w' \in \mathbb{W}_\lambda$ and $v \in (w\cdot\mathbb{W})\cap(w'\cdot\mathbb{W})$. Then there exist u, $u' \in \mathbb{W}$ such that $v = w\cdot u = w'\cdot u'$. So one of w, w' (say w) is an ancestor of the other. But if $\beta_w \le \lambda$, $\beta_{\tau w} > \lambda$ then as $\beta_i < 1$ we can only have $\beta_{\tau w'} > \lambda$ if $w' = w$. So if $w \neq w'$, $w\cdot\mathbb{W}$ and $w'\cdot\mathbb{W}$ are disjoint.

Let $v \in \mathbb{W}$. Then $\beta_{v|n} = \prod_{i=1}^n \beta_{v_i} \to 0$ as $n \to \infty$. So there exists m such that $v|m \in \mathbb{W}_\lambda$, and then $v \in (v|m)\cdot\mathbb{W}$, completing the proof of (a).
(b) This follows in a straightforward fashion from the decompositions given in (7.12) and Lemma 5.28. □

Note that $\beta_- > 0$ and that

$$\beta\lambda \le \beta_w \le \lambda, \qquad (\beta_-)^\alpha\lambda^\alpha \le \theta_w \le \lambda^\alpha, \quad w \in \mathbb{W}_\lambda. \tag{8.4}$$

Definition 8.2. The *spectral dimension* of F is defined by

$$d_s = d_s(F,\mathcal{E}_A) = 2\alpha/(1+\alpha).$$

Theorem 8.3. *For* $f \in \mathcal{D}$,

$$\|f\|_2^{2+4/d_s} \le c_1\left(\mathcal{E}(f,f) + \|f\|_2^2\right)\|f\|_1^{4/d_s}. \tag{8.5}$$

Proof. It is sufficient to consider the case f non-negative, so let $f \in \mathcal{D}$ with $f \ge 0$. Let $0 < \lambda < 1$: by Lemma 8.1, (7.14) and (8.4) we have

$$\begin{aligned}
\|f\|_2^2 &= \sum_{w\in\mathbb{W}_\lambda} \theta_w \int f_w^2\,d\mu \\
&\le \sum_w \theta_w\left(c_1\mathcal{E}(f_w,f_w) + \left(\int f_w\,d\mu\right)^2\right) \\
&\le c_2\sum_w \lambda^\alpha\mathcal{E}(f_w,f_w) + c_2\sum_w \lambda^\alpha\left(\int f_w\,d\mu\right)^2 \\
&\le c_3\lambda^{\alpha+1}\sum_w \beta_w^{-1}\mathcal{E}(f_w,f_w) + c_2\lambda^\alpha\left(\sum_w\int f_w\,d\mu\right)^2 \\
&\le c_3\lambda^{\alpha+1}\mathcal{E}(f,f) + c_4\lambda^{-\alpha}\left(\sum_w \theta_w\int f_w\,d\mu\right)^2 \\
&= c_3\lambda^{\alpha+1}\mathcal{E}(f,f) + c_4\lambda^{-\alpha}\|f\|_1^2.
\end{aligned} \tag{8.6}$$

The final line of (8.6) is minimized if we take $\lambda^{2\alpha+1} = c_5\|f\|_1^2/\mathcal{E}(f,f)$. If $\mathcal{E}(f,f) \geq c_5\|f\|_1^2$ then $\lambda < 1$ and so we obtain from (8.6) that

$$\|f\|_2^2 \leq c\mathcal{E}(f,f)^{\alpha/(2\alpha+1)}\big(\|f\|_1^2\big)^{(\alpha+1)/(2\alpha+1)}, \tag{8.7}$$

which implies that that

$$\|f\|_2^{2+4/d_s} \leq c\mathcal{E}(f,f)\|f\|_1^{4/d_s} \qquad \text{if} \quad \mathcal{E}(f,f) \geq c_5\|f\|_1^2. \tag{8.8}$$

If $\mathcal{E}(f,f) \leq c_5\|f\|_1^2$ then by (7.14)

$$\|f\|_2^2 \leq c_1\left(\mathcal{E}(f,f) + \|f\|_1^2\right) \leq c\|f\|_1^2,$$

and so

$$\|f\|_2^{2+4/d_s} \leq c\|f\|_2^2\|f\|_1^{4/d_s} \qquad \text{if} \quad \mathcal{E}(f,f) \leq c_5\|f\|_1^2. \tag{8.9}$$

Combining (8.8) and (8.9) we obtain (8.5). □

From the results in Section 4 we then deduce

Theorem 8.4. *X has a transition density $p(t,x,y)$ which satisfies*

$$p(t,x,y) \leq c_1 t^{-d_s/2}, \qquad 0 < t \leq 1, \quad x,y \in F, \tag{8.10}$$

$$\big|p(t,x,y) - p(t,x,y')\big|^2 \leq c_2 t^{-1-d_s/2}R(y,y'), \quad 0 \leq t \leq 1, \quad x,y,y' \in F. \tag{8.11}$$

Proof. By Proposition 4.14 X has a jointly measurable transition density, and by Corollary 4.15 we have for x, $y \in F$, $0 < t \leq 1$,

$$p(t,x,y) \leq ct^{-d_s/2}e^{ct} \leq c't^{-d_s/2}.$$

By (4.17) the function $q_{t,x} = p(t,x,\cdot)$ satisfies $\mathcal{E}(q_{t,x},q_{t,x}) \leq ct^{-1-d_s/2}$, and so $q_{t,x} \in \mathcal{D}$ and is continuous. Further, by Proposition 7.18

$$|p(t,x,y) - p(t,x,y')|^2 \leq cR(y,y')t^{-d_s/2-1}, \qquad x,y,y' \in F.$$

Thus $p(t,\cdot,\cdot)$ is jointly Hölder continuous in the metric R on F. □

Remarks 8.5. 1. As $\alpha > 0$, we have $0 < d_s = 2\alpha(1+\alpha)^{-1} < 2$.
2. The estimate (8.10) is good if $t \in (0,1]$ and x close to y. It is poor if t is small compared with $R(x,y)$, and in this case we can obtain a better estimate by chaining, as was done for fractional diffusions in Section 3. For this we need some additional properties of the resistance metric.

Lemma 8.6. *If v, $w \in \mathbb{W}_\lambda$ and $v \neq w$ then $F_v \cap F_w = V_v^{(0)} \cap V_w^{(0)}$.*

Proof. This follows easily from the corresponding property for $\mathbb{W}_n$. Let v, $w \in \mathbb{W}_\lambda$, with $|v| = m \leq |w| = n$, $v \neq w$. Let $x \in F_v \cap F_w$. Set $w' = w|m$; then as $F_w \subset F_{w'}$, $x \in F_v \cap F_w$, and so by Lemma 5.17(a) $x \in V_v^{(0)} \cap V_{w'}^{(0)}$. Further, as $x \in F_v$ there exists $v' \in \mathbb{W}_n$ such that $v'|m = v$, and $x \in F_{v'}$. Then $x \in F_{v'} \cap F_w = V_{v'}^{(0)} \cap V_w^{(0)}$. So $x \in V_v^{(0)} \cap V_w^{(0)}$. □

Definition 8.7. Set

$$V_\lambda^{(0)} = \bigcup_{w \in \mathbb{W}_\lambda} V_w^{(0)}.$$

Let $G_\lambda = \left(V_\lambda^{(0)}, \mathbf{E}_\lambda\right)$ be the graph with vertex set $V_\lambda^{(0)}$, and edge set $\mathbf{E}_\lambda$ such that $\{x, y\}$ is an edge if and only if $x,\, y \in V_w^{(0)}$ for some $w \in \mathbb{W}_\lambda$. For $A \subset F$ set

$$\begin{aligned} N_\lambda(A) &= \bigcup \{F_w : w \in \mathbb{W}_\lambda,\, F_w \cap A \neq \emptyset\}, \\ \widetilde{N}_\lambda(x) &= N_\lambda\left(N_\lambda(\{x\})\right). \end{aligned}$$

As we will see, $\widetilde{N}_\lambda(x)$ is a neighbourhood of x with a structure which is well adapted to the geometry of F in the metric R. We write $N_\lambda(y) = N_\lambda\left(\{y\}\right)$.

Lemma 8.8. *(a) If $x,\, y \in V_\lambda^{(0)}$ and $x \neq y$ then*

$$R(x, y) \geq c_1 \lambda.$$

(b) If $\{x, y\} \in \mathbf{E}_\lambda$ then $R(x, y) \leq c_2 \lambda$.

Proof. (b) is immediate from the definition of $\mathbb{W}_\lambda$ and Proposition 7.18(b). For (a), note first that if $x \in F$ then by Proposition 5.21 x can belong to at most $M_1 = M\#(P)$ n–complexes, for any n. So there are at most M_1 distinct elements $w \in \mathbb{W}_\lambda$ such that $x \in F_w$.

As $V^{(0)}$ is a finite set, and $\mathcal{E}_A^{(0)}$ is non-degenerate, there exists $c_3,\, c_4 > 0$ such that,

$$c_4 \geq R(x, V^{(0)} - \{x\}) \geq c_3, \qquad x \in V^{(0)}. \tag{8.12}$$

(Recall that this resistance is, by the construction of $\mathcal{E}$, the same in $(F, \mathcal{E})$ as in $(V^{(0)}, \mathcal{E}_A^{(0)})$). Now fix $x \in V_\lambda^{(0)}$. If $w \in \mathbb{W}_\lambda$, and $x \in V_w^{(0)}$, let $x' = \psi_w^{-1}(x)$, and g_w be the function on F such that $g_w(x') = 1$, $g_w(y) = 0$, $g \in V^{(0)} - \{x'\}$, and

$$\mathcal{E}(g_w, g_w)^{-1} = R\left(x', V^{(0)} - \{x'\}\right) \geq c_3.$$

Define g'_w on F_w by $g'_w = g_w \circ \psi_w^{-1}$, and extend g'_w to F by setting $g'_w = 0$ on $F - F_w$.

Now let $g'_v = 0$ if $x \notin V_v^{(0)}$, $V \in \mathbb{W}_\lambda$, and set

$$g = \sum_{v \in \mathbb{W}_\lambda} g'_v.$$

Then $g(x) = 1$, $g(y) = 0$ if $y \in V_\lambda^{(0)}$, $y \neq x$, and

$$\begin{aligned} \mathcal{E}(g, g) &= \sum_{w \in \mathbb{W}_\lambda} \beta_w^{-1} \mathcal{E}(g \circ \psi_w, g \circ \psi_w) \\ &= \sum_w \beta_w^{-1} 1_{(x \in F_w)} \mathcal{E}(g_w, g_w) \leq c_5 \lambda^{-1} M_1. \end{aligned}$$

Hence if $y \neq x$, $y \in V_\lambda^{(0)}$, we have

$$R(x,y)^{-1} \leq \mathcal{E}(g,g) \leq \lambda^{-1} M_1 c_5^{-1},$$

so that $R(x,y) \geq c_6 \lambda$. □

Remark. For $x \in V_\lambda^{(0)}$ the function g constructed above is zero outside $N_\lambda(\{x\})$. So we also have

$$R(x,y) \geq c_6 \lambda, \qquad x \in V_\lambda^{(0)}, \quad y \in N_\lambda(\{x\})^c. \tag{8.13}$$

Proposition 8.9. *There exist constants c_i such that for $x \in F$, $\lambda > 0$,*

$$B_R(x, c_1\lambda) \subset \widetilde{N}_\lambda(x) \subset B_R(x, c_2\lambda), \tag{8.14}$$

$$c_3 \lambda^\alpha \leq \mu\big(B_R(x,\lambda)\big) \leq c_4 \lambda^\alpha \tag{8.15}$$

$$c_5 \lambda \leq R\big(x, \widetilde{N}_\lambda(x)^c\big) \leq c_6 \lambda, \tag{8.16}$$

$$c_7 \lambda \leq R\big(x, B_R(x,\lambda)^c\big) \leq c_8 \lambda. \tag{8.17}$$

Proof. Let $x \in F$. If $y \in N_\lambda(\{x\})$ then by (7.17), $R(x,y) \leq c\lambda$. So if $z \in \widetilde{N}_\lambda(x)$, since there exists $y \in N_\lambda(\{x\})$ with $z \in N_\lambda(\{y\})$, $R(x,z) \leq c'\lambda$, proving the right hand inclusion in (8.14).

If $x \in V_\lambda^{(0)}$ then by (8.13), if $c_9 = c_{8\cdot 7\cdot 6}$,

$$B_R(x, c_9\lambda) \subset N_\lambda(x).$$

Now let $x \notin V_\lambda^{(0)}$, so that there exists a unique $w \in \mathbb{W}_\lambda$ with $x \in F_w$. For each $y \in V_w^{(0)}$ let $f_y(\cdot)$ be the function constructed in Lemma 8.8, which satisfies $f_y(y) = 1$, $f_y = 0$ outside $N_\lambda(y)$, $f_y(z) = 0$ for each $z \in V_\lambda^{(0)} - \{y\}$, and $\mathcal{E}(f_y, f_y) \leq c_{10}\lambda^{-1}$. Let $f = \sum_y f_y$: then $f(y) = 1$ for each $y \in V_w^{(0)}$. So if

$$g = 1_{F_w} + 1_{F_w^c} f,$$

$\mathcal{E}(g,g) \leq \mathcal{E}(f,f) \leq \#(V_w^{(0)}) c_{10} \lambda^{-1} \leq c_{11}\lambda^{-1}$. As $g(x) = 1$, and $g(z) = 0$ for $z \notin \widetilde{N}_\lambda(x)$, we have for $z \notin \widetilde{N}_\lambda(x)$ that $R(x,z)^{-1} \leq \mathcal{E}(g,g) \leq c_{11}\lambda^{-1}$. So $B_R(x, c_{11}\lambda) \subset N_\lambda(x)$. This proves (8.14), and also that $R\big(x, \widetilde{N}_\lambda(x)^c\big) \geq c_{11}^{-1}\lambda$.

The remaining assertions now follow fairly easily. For $w \in \mathbb{W}_\lambda$ we have $c_{12}\lambda^\alpha \leq \mu(F_w) \leq c_{13}\lambda^\alpha$. As $\widetilde{N}_\lambda(x)$ contains at least one λ–complex, and at most $M^2\#(P)^2$ λ–complexes, we have

$$\mu\big(\widetilde{N}_\lambda(x)\big) \asymp \lambda^\alpha,$$

and using (8.14) this implies (8.15).

If $A \subset B$ then it is clear that $R(x,A) \geq R(x,B)$. So (provided λ is small enough) if $x \in F$ we can find a chain x, y_1, y_2, y_3 where $y_i \in V_\lambda^{(0)}$, $\{y_i, y_{i+1}\}$ is an edge in $\mathbf{E}_\lambda$, $y_3 \notin \widetilde{N}_\lambda(x)$, and x and y are in the same λ–complex. Then $R(x,y_3) \leq c\lambda$ by (7.17), and so, using Lemma 8.8(b) we have $R(x,y_3) \leq c'\lambda$. Thus $R\big(x, \widetilde{N}_\lambda(x)^c\big) \leq R(x,y_3) \leq c'\lambda$ proving the right hand side of (8.16): the left hand side was proved above.

(8.17) follows easily from (8.14) and (8.16). □

Corollary 8.10. *In the metric R, the Hausdorff dimension of F is α, and further*

$$0 < \mathcal{H}_R^\alpha(F) < \infty.$$

Proof. This is immediate from Corollary 2.8 and (8.15). □

Proposition 8.11. *For $x \in F$, $r > 0$ set $\tau(x,r) = T_{B_R(x,r)^c}$. Then*

$$c_1 r^{\alpha+1} \le \mathbb{E}^x \tau(x,r) \le c_2 r^{\alpha+1}, \qquad x \in F, \quad r > 0. \tag{8.18}$$

Proof. Let $B = B_R(x,r)$. Then by Theorem 4.25 and the estimates (8.15) and (8.17)

$$E^x \tau(x,r) \le \mu(B) R(x, B^c) \le c_3 r^{\alpha+1},$$

which proves the upper bound in (8.18).

Let $(X_t^B, t \ge 0)$ be the process X killed at $\tau = T_{B^c}$, and let $g(x,y)$ be the Greens' function for X^B. In view of Theorem 7.19, we can write

$$g(x,y) = \mathbb{E}^x L_\tau^y, \qquad x, y \in F.$$

Then if $f(y) = g(x,y)/g(x,x)$, $f \in \mathcal{D}$ and by the reproducing kernel property of g we have

$$\mathcal{E}(f,f) = g(x,x)^{-2} \mathcal{E}\big(g(x,\cdot), g(x,\cdot)\big) = g(x,x)^{-1},$$

and as in Theorem 4.25 $g(x,x) = R(x, B^c) \ge c_4 r$. By (7.18)

$$|f(x) - f(y)|^2 \le R(x,y)\mathcal{E}(f,f) \le R(x,y)(c_4 r)^{-1} \le \tfrac{1}{4}$$

if $R(x,y) \le \frac{1}{4} c_4 r$. Thus $f(y) \ge \frac{1}{2}$ on $B_R(x, \frac{1}{4} c_4 r)$, and hence

$$\begin{aligned} \mathbb{E}^x \tau &= \int_B g(x,y)\mu(dy) \\ &\ge \tfrac{1}{2} g(x,x) \mu\left(B_R\left(x, \tfrac{1}{4} c_4 r\right)\right) \ge c_5 r^{1+\alpha}, \end{aligned}$$

proving (8.18). □

We have a spectral decomposition of $p(t,x,y)$. Write $(f,g) = \int_F f g \, d\mu$.

Theorem 8.12. *There exist functions $\varphi_i \in \mathcal{D}$, $\lambda_i \ge 0$, $i \ge 0$, such that $(\varphi_i, \varphi_i) = 1$, $0 = \lambda_0 < \lambda_1 \le \cdots$, and*

$$\mathcal{E}(\varphi_i, f) = \lambda_i (\varphi_i, f), \qquad f \in \mathcal{D}.$$

The transition density $p(t,x,y)$ of X satisfies

$$p(t,x,y) = \sum_{i=0}^\infty e^{-\lambda_i t} \varphi_i(x) \varphi_i(y), \tag{8.19}$$

where the sum in (8.19) converges uniformly and absolutely. So p is jointly continuous in (t,x,y).

Proof. This follows from Mercer's Theorem, as in [DaSi]. Note that $\varphi_0 = 1$ as $\mathcal{E}$ is irreducible and $\mu(F) = 1$. □

The following is an immediate consequence of (8.19)

Corollary 8.13. *(a) For $x,\, y \in F$, $t > 0$,*

$$p(t,x,y)^2 \le p(t,x,x)p(t,y,y).$$

(b) For each $x, y \in F$

$$\lim_{t\to\infty} p(t,x,y) = 1.$$

Lemma 8.14.

$$p(t,x,y) \ge c_0 t^{-d_s/2}, \qquad 0 \le t \le 1, \quad R(x,y) \le c_1 t^{1/(1+\alpha)}. \tag{8.20}$$

Proof. We begin with the case $x = y$. From Proposition 8.11 and Lemma 3.16 we deduce that there exists $c_2 > 0$ such that

$$\mathbb{P}^x\big(\tau(x,r) \le t\big) \le (1-2c_2) + c_3 t r^{-\alpha-1}.$$

Choose $c_4 > 0$ such that $c_3 t r_0^{-\alpha-1} = c_2$ if $r_0 = c_4 t^{1/(1+\alpha)}$. Then

$$\mathbb{P}^x\big(X_t \in B_R(x,r_0)\big) \ge \mathbb{P}^x\big(\tau(x,r_0) \le t\big) \ge c_2.$$

So using Cauchy-Schwarz and the symmetry of p, and writing $B = B_R(x,r_0)$,

$$\begin{aligned} 0 < c_2^2 &\le \Big(\int_B p(t,x,y)\mu(dy)^2\Big) \\ &\le \int_{B(x,r_0)} \mu(dy) \int_B p(t,x,y)p(t,y,x)\mu(dy) \\ &\le \mu\big(B)\big)\ p(2t,x,x) \\ &\le c_5 t^{\alpha/(1+\alpha)} p(2t,x,x). \end{aligned}$$

Replacing t by $t/2$ we have

$$p(t,x,x) \ge c_0 t^{-d_s/2}.$$

Fix t, x, and write $q(y) = p(t,x,y)$. By (4.16) and (8.5) $\mathcal{E}(q,q) \le c_6 t^{-1-d_s/2}$ for $t \le 1$, so using (7.18), if $R(x,y) \le c_7 t^{1/(1+\alpha)}$ then, as $1 + d_s/2 = (1+2\alpha)/(1+\alpha)$,

$$\begin{aligned} q(y) &\ge q(x) - |q(x) - q(y)| \\ &\ge c_0 t^{-\alpha/(1+\alpha)} - \big(R(x,y)\mathcal{E}(q,q)\big)^{1/2} \\ &\ge c_0 t^{-\alpha/(1+\alpha)} - \Big(c_7 c_6 t^{-2\alpha/(1+\alpha)}\Big)^{1/2} \\ &= t^{-\alpha/(1+\alpha)}(c_0 - (c_7c_6)^{1/2}). \end{aligned}$$

Choosing c_7 suitably gives (8.20). ☐

We can at this point employ the chaining arguments used in Theorem 3.11 to extend these bounds to give upper and lower bounds on $p(t,x,y)$. However, as R is not in general a geodesic metric, the bounds will not be of the form given in Theorem 3.11. The general case is given in a paper of Hambly and Kumagai [HK2], but since the proof of Theorem 3.11 does not use the geodesic property for the upper bound we do obtain:

Theorem 8.15. *The transition density $p(t,x,y)$ satisfies*

$$p(t,x,y) \le c_1 t^{-\alpha/(1+\alpha)} \exp\left(-c_2\big(R(x,y)^{1+\alpha}/t\big)^{1/\alpha}\right). \tag{8.21}$$

Note. The power $1/\alpha$ in the exponent is not in general best possible.

Theorem 8.16. *Suppose that there exists a metric ρ on F with the midpoint property such that for some $\theta > 0$*

$$c_1\rho(x,y)^\theta \le R(x,y) \le c_2\rho(x,y)^\theta \; x,y \in F. \tag{8.22}$$

Then if $d_w = \theta(1+\alpha)$, $d_f = \alpha\theta$, (F,ρ,μ) is a fractional metric space of dimension d_f, and X is a fractional diffusion with indices d_f, d_w.

Proof. Since $B_\rho(x,(r/c_2)^\theta) \subset B_R(x,r) \subset B_\rho(x,(r/c_1)^\theta)$, it is immediate from (8.15) that (F,ρ) is a $FMS(d_f)$. Write $\tau_\rho(x,r) = \inf\{t : X_t \notin B_\rho(x,r)\}$. Then from (8.18) and (8.22)

$$cr^{\theta(1+\alpha)} \le \mathbb{E}^x\tau_\rho(x,r) \le c_2 r^{\theta(1+\alpha)}.$$

So, by (8.10) and (8.20), X satisfies the hypotheses of Theorem 3.11, and so X is a $FD(d_f,d_w)$. □

Remark. Note that in this case the estimate (7.20) on the Hölder continuity of $u_\lambda(x,y)$ implies that

$$|u_\lambda(x,y) - u_\lambda(x',y)| \le cR(x,x')^{\frac{1}{2}} \le c'\rho(x,x')^{\theta/2}, \tag{8.23}$$

while by Theorem 3.40 we have

$$|u_\lambda(x,y) - u_\lambda(x',y)| \le c\rho(x,x')^\theta. \tag{8.24}$$

The difference is that (8.23) used only the fact that $u_\lambda(.,y) \in \mathcal{D}$, while the proof of (8.24) used the fact that it is the λ-potential density.

Diffusions on nested fractals.

We conclude by treating briefly the case of nested fractals. Most of the necessary work has already been done. Let $(F,(\psi_i))$ be a nested fractal, with length, mass, resistance and shortest path scaling factors L, M, ρ, γ. Recall that in this context we take $r_i = 1$, $\theta_i = 1/M$, $1 \le i \le M$, and $\mu = \mu_\theta$ for the measure associated with θ. Write $d = d_F$ for the geodesic metric on F defined in Section 5.

Lemma 8.17. *Set $\theta = \log\rho/\log\gamma$. Then*

$$c_1 d(x,y)^\theta \le R(x,y) \le c_2 d(x,y)^\theta, \quad x,y \in F. \tag{8.24}$$

Proof. Let $\lambda \in (0,1)$. Since all the r_i are equal, $\widetilde{N}_\lambda(x)$ is a union of n-complexes, where $\rho^{-n} \le \lambda \le \rho^{-n+1}$. So by Theorem 5.43 and Proposition 8.8, since $\gamma^{-n} = (\rho^{-n})^\theta$,

$$y \in \widetilde{N}_\lambda(x) \text{ implies that } R(x,y) \le c_1\lambda, \text{ and } d(x,y) \le c_2\lambda^\theta, \tag{8.26}$$

(8.27) $\qquad y \notin \widetilde{N}_\lambda(x)$ implies that $R(x,y) \geq c_3\lambda$, and $d(x,y) \geq c_4\lambda^\theta$.

The result is immediate from (8.26) and (8.27). □

Applying Lemma 8.17 and Theorem 8.15 we deduce:

Theorem 8.18. *Let F be a nested fractal, with scaling factors L, M, ρ, γ. Set*

$$d_f = \log M/\log\gamma, \quad d_w = \log M\rho/\log\gamma.$$

Then (F, d_F, μ) is a fractional metric space of dimension d_f, and X is a $FD(d_f, d_w)$. In particular, the transition density $p(t,x,y)$ of X is jointly continuous in (t,x,y) and satisfies

(8.28)
$$c_1 t^{-d_f/d_w} \exp\left(-c_2\big(d(x,y)^{d_w}/t\big)^{1/(d_w-1)}\right) \leq p(t,x,y) \leq c_3 t^{-d_f/d_w} \exp\left(-c_4\big(d(x,y)^{d_w}/t\big)^{1/(d_w-1)}\right).$$

References.

[AO] S. Alexander and R. Orbach: Density of states on fractals: "fractons ". *J. Physique (Paris) Lett.* **43**, L625–L631 (1982).

[AT] W.N. Anderson and G.E. Trapp: Shorted operators, II, *SIAM J. Appl. Math.* **28**, 60–71 (1975).

[Ar] D.G. Aronson, Bounds on the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* **73**, 890-896 (1967).

[BS] C. Bandt and J. Stahnke: Self-similar sets 6. Interior distance on deterministic fractals. Preprint 1990.

[Bar1] M.T. Barlow: Random walks, electrical resistance and nested fractals. In *Asymptotic Problems in Probability Theory*, ed. K.D. Elworthy, N. Ikeda, 131–157, Longman Scientific, Harlow UK, 1990.

[Bar2] M.T. Barlow: Random walks and diffusions on fractals. *Proc. Int. Congress Math. Kyoto 1990*, 1025–1035. Springer, Tokyo 1991.

[Bar3] M.T. Barlow: Harmonic analysis on fractal spaces. *Séminaire Bourbaki Volume 1991/1992, Astérisque* **206** (1992).

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

[BB2] M. T. Barlow and R.F. Bass: Local times for Brownian motion on the Sierpinski carpet. *Probab. Th. Rel. Fields,* **85**, 91–104 (1990).

[BB3] M. T. Barlow and R.F. Bass: On the resistance of the Sierpinski carpet. *Proc. R. Soc. London A.* **431**, 345–360 (1990).

[BB4] M.T. Barlow and R.F. Bass: Transition densities for Brownian motion on the Sierpinski carpet. *Probab. Th. Rel. Fields* **91**, 307–330 (1992).

[BB5] M.T. Barlow and R.F. Bass: Coupling and Harnack inequalities for Sierpinski carpets. *Bull. A.M.S.* **29**, 208–212 (1993).

[BB6] M.T. Barlow and R.F. Bass: Brownian motion and harmonic analysis on Sierpinski carpets. Preprint 1997.

[BB7] M.T. Barlow and R.F. Bass: Random walks on graphical Sierpinski carpets. Preprint 1997.

[BBS] M.T. Barlow, R.F. Bass, and J.D. Sherwood: Resistance and spectral dimension of Sierpinski carpets. *J. Phys. A*, **23**, L253–L258 (1990).

[BH] M.T. Barlow and B.M. Hambly: Transition density estimates for Brownian motion on scale irregular Sierpinski gaskets. *Ann. IHP.* **33**, 531-557 (1997).

[BHHW] M.T. Barlow, K. Hattori, T. Hattori and H. Watanabe: Weak homogenization of anisotropic diffusion on pre-Sierpinski carpets. To appear *Comm Math. Phys.*

[BK] M.T. Barlow and J. Kigami: Localized eigenfunctions of the Laplacian on p.c.f. self-similar sets. To appear *J. Lond. Math. Soc.*

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

[Bas] R.F. Bass: Diffusions on the Sierpinski carpet. *Trends in Probability and related Analysis: Proceedings of SAP '96.* World Scientific, Singapore, 1997. To appear.

[BAH] D. Ben-Avraham and S. Havlin: Exact fractals with adjustable fractal and fracton dimensionalities. *J. Phys. A.* **16**, L559–563 (1983).

[Blu] L.M. Blumenthal: *Theory and applications of distance geometry.* Oxford, 1953.

[BG] R.M. Blumenthal and R.K. Getoor: *Markov processes and potential theory.* Academic Press, New York, 1968.

[CKS] E.A. Carlen, S. Kusuoka and D.W. Stroock: Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincare Sup. no. 2*, 245–287 (1987).

[Ca1] D. Carlson: What are Schur complements, anyway? *Linear Alg. Appl.* **74**, 257–275 (1986).

[CRRST] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky and 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.

[Co] T.C. Coulhon: Ultracontractivity and Nash type inequalities. *J. Funct. Anal.* **141**, 510-539 (1996).

[DSV] K. Dalrymple, R.S. Strichartz and J.P. Vinson: Fractal differential equations on the Sierpinski gasket. Preprint 1997.

[Da] E.B. Davies: *Heat kernels and spectral theory.* Cambridge University Press 1989.

[DaSi] E.B. Davies and B. Simon: Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Funct. Anal.* **59**, 335–395, (1984).

[DS] P.G. Doyle and J.L. Snell: *Random walks and electrical networks.* Washington, Math. Assoc. of America, 1984.

[Dud] R.M. Dudley: Sample functions of the Gaussian process. *Ann. Prob.* **1**, 66–103 (1971).

[FaS] E.B. Fabes and D.W. Stroock, A new proof of Moser's parabolic Harnack inequality via the old ideas of Nash. *Arch. Mech. Rat. Anal.* **96**, 327–338 (1986).

[Fa1] K.J. Falconer: *Geometry of fractal sets.* Cambridge Univ. Press, 1985.

[Fa2] K.J. Falconer: *Fractal Geometry.* Wiley, 1990

[Fe] H. Federer: *Geometric measure theory.* Springer, New York, 1969.

[FHK] P.J. Fitzsimmons, B.M. Hambly, T. Kumagai: Transition density estimates for Brownian motion on affine nested fractals. *Comm. Math. Phys.* **165**, 595–620 (1995).

[FOT] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet Forms and Symmetric Markov Processes.* de Gruyter, Berlin, 1994.

[Fu1] M. Fukushima, Dirichlet forms, diffusion processes, and spectral dimensions for nested fractals. *Ideas and methods in stochastic analysis, stochastics and applications*, 151–161. Cambridge Univ. Press., Cambridge, 1992.

[FS1] M. Fukushima and T. Shima: On a spectral analysis for the Sierpinski gasket. *J. of Potential Analysis.* **1**, 1–35 (1992).

[FS2] M. Fukushima and T. Shima, On discontinuity and tail behaviours of the integrated density of states for nested pre-fractals. *Comm. Math. Phys.* **163**, 461–471 (1994).

[FST] M. Fukushima, T. Shima and M. Takeda: Large deviations and related LILs for Brownian motions on nested fractals. Preprint 1997.

[GAM1] Y. Gefen, A. Aharony and B.B. Mandelbrot: Phase transitions on fractals.I. Quasilinear lattices. *J. Phys. A* **16**, 1267–1278 (1983).

[GAM2] Y. Gefen, A. Aharony, Y. Shapir and B.B. Mandelbrot: Phase transitions on fractals. II. Sierpinski gaskets. *J. Phys. A* **17**, 435–444 (1984).

[GAM3] Y. Gefen, A. Aharony and B. Mandelbrot: Phase transitions on fractals. III. Infinitely ramified lattices. *J. Phys. A* **17**, 1277–1289 (1984).

[GK] R.K. Getoor and H. Kesten: Continuity of local times of Markov processes. *Comp. Math.* **24**, 277-303 (1972).

[Go] 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.

[Gra] P.J. Grabner: Functional equations and stopping times for the Brownian motion on the Sierpinski gasket. To appear in *Mathematica.*

[GrT] P.J. Grabner and R.F. Tichy: Equidistribution and Brownian motion on the Sierpinski gasket. Preprint 1997.

[GrW] P.J. Grabner and W Woess: Functional iterations and periodic oscillations for simple random walk on the Sierpinski gasket. To appear in *Stoch. Proc. Appl.*

[Gre] R.F. Green: A simple model for the construction of Brownian motion on Sierpinski's hexagonal gasket. Technical report, Univ. of Minnesota, Duluth. 1989.

[HS] L. deHaan and U. Stadtmuller: Dominated variation and related concepts and Tauberian theorems for Laplace transforms. *J. Math. Anal. Appl.* **108**, 344-365 (1985).

[Ham1] B.M. Hambly: Brownian motion on a homogeneous random fractal. *Probab. Th. Rel. Fields* **94**, 1–38, (1992).

[Ham2] B.M. Hambly: Brownian motion on a random recursive Sierpinski gasket. *Ann. Prob.* **25**, 1059-1102 (1997).

[HK1] B.M. Hambly, T. Kumagai: Heat kernel estimates and homogenization for asymptotically lower dimensional processes on some nested fractals. To appear *Potential Analysis.*

[HK2] B.M. Hambly, T. Kumagai: Transition density estimates for diffusion processes on p.c.f. self-similar fractals. To appear *Proc. London Math. Soc.*

[Har] T.E. Harris: *The theory of branching processes.* Springer 1963.

[Hat1] T. Hattori: Asympotically one-dimensional diffusions on scale-irregular gaskets. Preprint 1994.

[HH] K. Hattori and T. Hattori: Self-avoiding process on the Sierpinski gasket. *Probab. Th. Rel. Fields* **88**, 405–428 (1991).

[HatK] T. Hattori, S. Kusuoka: The exponent for mean square displacement of self-avoiding random walk on Sierpinski gasket. *Probab. Th. Rel. Fields* **93**, 273–284 (1992).

[HHK1] K. Hattori, T. Hattori, and S. Kusuoka, Self-avoiding paths on the three-dimensional Sierpinski gasket. *Publ. Res. Inst. Math. Sci.* **29**, 455–509 (1993).

[HHK2] K. Hattori, T. Hattori and S. Kusuoka: Self-avoiding paths on the pre-Sierpinski gasket. *Probab. Th. Rel. Fields* **84**, 1–26 (1990).

[HHW] K. Hattori, T. Hattori and H. Watanabe: Reasoning out the empirical rule $\tilde{d} < 2$, *Physics Letters A* **115**, 207–212 (1986).

[HHW1] K. Hattori, T. Hattori and H. Watanabe: Gaussian field theories and the spectral dimensions. *Prog. Th. Phys. Supp. No.* **92**, 108–143, (1987).

[HHW2] K. Hattori, T. Hattori and H. Watanabe: New approximate renormalisation method on fractals. *Phys Rev. A* **32**, 3730–3733 (1985).

[HHW3] K. Hattori, T. Hattori and H. Watanabe: Asymptotically one-dimensional diffusions on the Sierpinski gasket and the *abc*-gaskets. *Prob. Th. Rel. Fields* **100**, 85–116 (1994).

[HBA] S. Havlin and D. Ben-Avraham: Diffusion in disordered media, *Adv. Phys.* **36**, 695–798 (1987).

[He] M.K. Heck: Homogeneous diffusions on the Sierpinski gasket. Preprint 1996.

[Hu] J.E. Hutchinson: Fractals and self-similarity. *Indiana J. Math.* **30**, 713–747 (1981).

[Jo] O.D. Jones: Transition probabilities for the simple random walk on the Sierpinski graph. *Stoch. Proc. Appl.* **61**, 45-69 (1996).

[J] A. Jonsson: Brownian motion on fractals and function spaces. *Math. Zeit.* **122**, 495-504 (1986).

[Ka] M. Kac: On some connections between probability theory and differential and integral equations. *Proc. 2nd Berkelely Symposium, 1950*, 189-215, (1951).

[Ke] H. Kesten: Subdiffusive behaviour of random walk on random cluster. *Ann. Inst. H. Poincare B* **22**, 425–487 (1986).

[Ki1] J. Kigami: A harmonic calculus on the Sierpinski space. *Japan J. Appl. Math.* **6**, 259–290 (1989).

[Ki2] J. Kigami: A harmonic calculus for p.c.f. self-similar sets. *Trans. A.M.S.* **335**, 721–755 (1993).

[Ki3] J. Kigami: Harmonic metric and Dirichlet form on the Sierpinski gasket. In *Asymptotic Problems in Probability Theory*, ed. K.D. Elworthy, N. Ikeda, 201–218, Longman Scientific, Harlow UK, 1990.

[Ki4] J. Kigami: A calculus on some self-similar sets. To appear in Proceedings on the 1st IFIP Conference "Fractal 90", Elsevier.

[Ki5] J. Kigami: Effective resistance for harmonic structures on P.C.F. self-similar sets. *Math. Proc. Camb. Phil. Soc.* **115**, 291–303 (1994).

[Ki6] J. Kigami: Hausdorff dimension of self-similar sets and shortest path metric. *J. Math. Soc. Japan* **47**, 381–404 (1995).

[Ki7] J. Kigami: Harmonic calculus on limits of networks and its application to dendrites. *J. Functional Anal.* **128**, 48–86 (1995).

[KL] J. Kigami, M. Lapidus: Weyl's spectral problem for the spectral distribution of Laplacians on P.C.F. self-similar fractals. *Comm. Math. Phys.* **158**, 93–125 (1993).

[Kn] F.B. Knight: *Essentials of Brownian motion and diffusion.* AMS Math. Surveys, Vol. 18. AMS 1981.

[Koz] S.M. Kozlov: Harmonization and homogenization on fractals. *Comm. Math. Phys.* **153**, 339-357 (1993).

[Kr1] W.B. Krebs: A diffusion defined on a fractal state space. *Stoch. Proc. Appl.* **37**, 199–212 (1991).

[Kr2] W.B. Krebs: Hitting time bounds for Brownian motion on a fractal. *Proc. A.M.S.* **118**, 223–232 (1993).

[Kr3] W.B. Krebs: Brownian motion on the continuum tree. *Prob. Th. Rel. Fields* **101**, 421-433 (1995). Preprint 1993.

[Kum1] T. Kumagai: Construction and some properties of a class of non- symmetric diffusion processes on the Sierpinski gasket. In *Asymptotic Problems in Probability Theory*, ed. K.D. Elworthy, N. Ikeda, 219–247, Longman Scientific, Harlow UK, 1990.

[Kum2] T. Kumagai: Estimates of the transition densities for Brownian motion on nested fractals. *Prob. Th. Rel. Fields* **96**, 205–224 (1993).

[Kum3] T. Kumagai: Regularity, closedness, and spectral dimension of the Dirichlet forms on p.c.f. self-similar sets. *J. Math. Kyoto Univ.* **33**, 765–786 (1993).

[Kum4] T. Kumagai: Rotation invariance and characterization of a class of self-similar diffusion processes on the Sierpinski gasket. *Algorithms, fractals and dynamics*, 131–142. Plenum, New York, 1995.

[Kum5] T. Kumagai: Short time asymptotic behaviour and large deviation for Brownian motion on some affine nested fractals. *Publ. R.I.M.S. Kyoto Univ.* **33**, 223-240 (1997).

[KK1] T. Kumagai and S. Kusuoka, Homogenization on nested fractals. *Probab. Th. Rel. Fields* **104**, 375–398 (1996).

[Kus1] 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

[Kus2] S. Kusuoka: Dirichlet forms on fractals and products of random matrices. *Publ. RIMS Kyoto Univ.,* **25**, 659–680 (1989).

[Kus3] S. Kusuoka: Diffusion processes on nested fractals. In: *Statistical mechanics and fractals*, Lect. Notes in Math. **1567**, Springer, 1993.

[KZ1] S. Kusuoka and X.Y. Zhou: Dirichlet form on fractals: Poincaré constant and resistance. *Prob. Th. Rel. Fields* **93**, 169–196 (1992).

[KZ2] S. Kusuoka and X.Y. Zhou: Waves on fractal-like manifolds and effective energy propagation. Preprint 1993.

[L1] T. Lindstrøm: Brownian motion on nested fractals. *Mem. A.M.S.* **420**, 1990.

[L2] T. Lindstrøm: Brownian motion penetrating the Sierpinski gasket. In *Asymptotic Problems in Probability Theory*, ed. K.D. Elworthy, N. Ikeda, 248–278, Longman Scientific, Harlow UK.

[Mae] F-Y Maeda: *Dirichlet Integrals on Harmonic Spaces*. Springer L.N.M. **803** 1980.

[Man] B.B. Mandelbrot: *The fractal geometry of nature.* W.H. Freeman, San Fransisco, 1982.

[MR] M.B. Marcus, J. Rosen: Sample path functions of the local times of strongly symmetric Markov processes via Gaussian processes. *Ann. Prob.* **20**, 1603–1684 (1992).

[MW] R.D. Mauldin, S.C. Williams: Hausdorff dimension in graph directed constructions. *Trans. A.M.S.* **309**, 811–829 (1988).

[Me1] V. Metz: Potentialtheorie auf dem Sierpinski gasket. *Math. Ann.* **289**, 207–237 (1991).

[Me2] V. Metz, Renormalization of finitely ramified fractals. *Proc. Roy. Soc. Edinburgh Ser A* **125**, 1085–1104 (1995).

[Me3] V. Metz: How many diffusions exist on the Vicsek snowflake? *Acta Appl. Math.* **32**, 224–241 (1993).

[Me4] V. Metz: Hilbert's projective metric on cones of Dirichlet forms. *J. Funct. Anal.* **127**, 438–455, (1995).

[Me5] V. Metz, Renormalization on fractals. *Proc. International Conf. Potential Theory 94*, 413–422. de Gruyter, Berlin, 1996.

[Me6] V. Metz: Maeda's energy norm explains effective resistance. Preprint.

[Me7] V. Metz, Renormalization contracts on nested fractals. *C.R. Acad. Sci. Paris* **332**, 1037–1042 (1996).

[MS] V. Metz and K.-T. Sturm, Gaussian and non-Gaussian estimates for heat kernels on the Sierpinski gasket. In: *Dirichlet forms and stochastic processes*, 283–289, de Gruyter, Berlin, 1995.

[Mor] P.A.P. Moran: Additive functions of intervals and Hausdorff measure. *Proc. Camb. Phil. Soc.* **42**, 15–23 (1946).

[Mos] U. Mosco: Composite media and asymptotic Dirichlet forms. *J. Funct. Anal.* **123**, 368–421 (1994).

[Nus] R.D. Nussbaum: Hilbert's projective metric and iterated non-linear maps. *Mem. A.M.S.* **75**, 1988.

[OSS] M. Okada, T. Sekiguchi and Y. Shiota: Heat kernels on infinte graph networks and deformed Sierpinski gaskets. *Japan J. App. Math.* **7**, 527–554 (1990).

[O1] H. Osada: Isoperimetric dimension and estimates of heat kernels of pre–Sierpinski carpets. *Probab. Th. Rel. Fields* **86**, 469–490 (1990).

[O2] H. Osada: Cell fractals and equations of hitting probabilities. Preprint 1992.

[OS] B. O'Shaughnessy, I. Procaccia: Analytical solutions for diffusion on fractal objects. *Phys. Rev. Lett.* **54**, 455–458, (1985).

[Pe] R. Peirone: Homogenization of functionals on fractals. Preprint, 1996.

[PP] K. Pietruska-Paluba: The Lifchitz singularity for the density of states on the Sierpinski gasket. *Probab. Th. Rel. Fields* **89**, 1–34 (1991).

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

[R] R. Rammal: Spectrum of harmonic excitations on fractals. *J. de Physique* **45**, 191–206 (1984).

[Rog] L.C.G. Rogers, Multiple points of Markov processes in a complete metric space. *Sém. de Probabilités XXIII*, 186–197. Springer, Berlin, 1989.

[Sab1] C.Sabot: Existence and uniqueness of diffusions on finitely ramified self-similar fractals. Preprint 1996.

[Sab2] C.Sabot: Espaces de Dirichlet reliés par un nombre fini de points et application aux diffusions sur les fractales. Preprint 1996.

[Sha] M. Sharpe: *General Theory of Markov Processes.* Academic Press. New York, 1988.

[Sh1] T. Shima: On eigenvalue problems for the random walk on the Sierpinski pre-gaskets. *Japan J. Appl. Ind. Math.,* **8**, 127–141 (1991).

[Sh2] T. Shima: The eigenvalue problem for the Laplacian on the Sierpinski gasket. In *Asymptotic Problems in Probability Theory* , ed. K.D. Elworthy, N. Ikeda, 279–288, Longman Scientific, Harlow UK.

[Sh3] T. Shima: On Lifschitz tails for the density of states on nested fractals. *Osaka J. Math.* **29**, 749–770 (1992).

[Sh4] T. Shima: On eigenvalue problems for Laplacians on P.C.F. self- similar sets. *Japan J. Indust. Appl. Math.* **13**, 1–23 (1996).

[Sie1] W. Sierpinski: Sur une courbe dont tout point est un point de ramification. *C.R. Acad. Sci. Paris* **160**, 302–305 (1915).

[Sie2] W. Sierpinski: Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée. *C.R. Acad. Sci. Paris* **162**, 629–632 (1916).

[Stu1] K.T. Sturm: Diffusion processes and heat kernels on metric spaces. To appear *Ann. Probab.*

[Stu2] K.T. Sturm: Analysis on local Dirichlet spaces I. Recurrence, conservativeness and L^p-Liouville properties. *J. reine angew. Math.* **456**, 173–196 (1994).

[Stu3] K.T. Sturm: Analysis on local Dirichlet spaces II. Upper Gaussian estimates for the fundamental solutions of parabolic equations. *Osaka J. Math.* **32**, 275–312 (1995).

[Stu4] K.T. Sturm: Analysis on local Dirichlet spaces III. The parabolic Harnack inequality. *J. Math. Pures Appl.* **75**, 273–297 (1996).

[Tal] M. Talagrand: Regularity of Gaussian processes *Acta Math.* **159**, 99-149 (1987).

[Tel] A. Telcs: Random walks on graphs, electric networks and fractals. *Prob. Th. Rel. Fields* **82**, 435-451 (1989).

[Tet] P. Tetali: Random walks and the effective resistance of networks. *J. Theor. Prob.* **4**, 101–109 (1991).

[Tri] C. Tricot: Two definitions of fractional dimension. *Math. Proc. Camb. Phil. Soc.* **91**, 54–74, (1982).

[V1] N. Th. Varopoulos: Isoperimetric inequalities and Markov chains. *J. Funct. Anal.* **63**, 215–239 (1985).

[Wat1] H. Watanabe: Spectral dimension of a wire network. *J. Phys. A.* **18**, 2807–2823 (1985).

[Y1] J.-A. Yan: A formula for densities of transition functions. *Sem. Prob. XXII*, 92-100. Lect Notes Math. **1321**.