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 $\beta > 2$

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

(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 $\beta = \log 5/\log 2$.

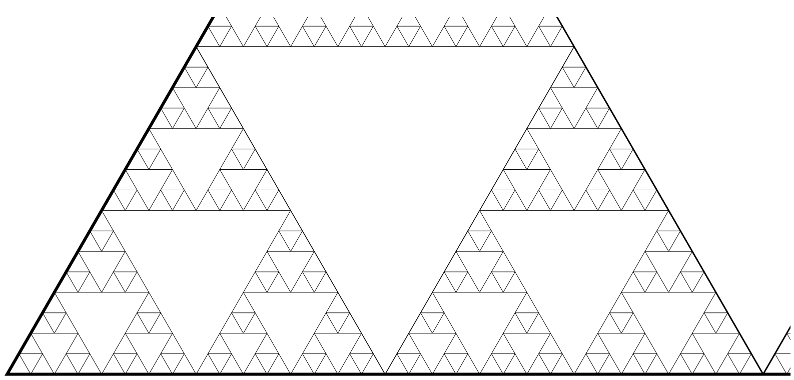

Figure 1.1: The graphical Sierpinski gasket.

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).

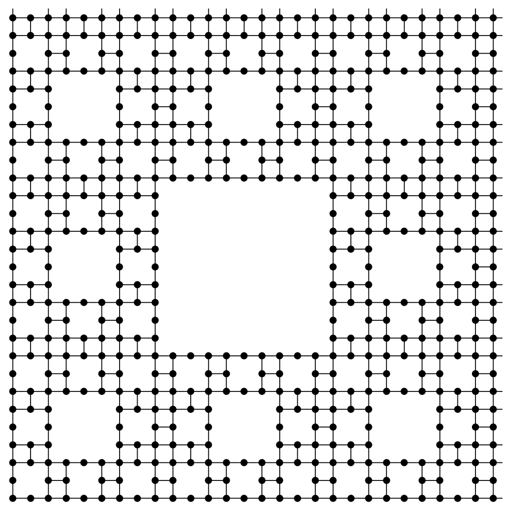

Figure 1.2: The graphical Sierpinski carpet.

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 $\{3^k, \dots, 2.3^k - 1\} \times \{3^k, \dots, 2.3^k - 1\}$. Now set

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

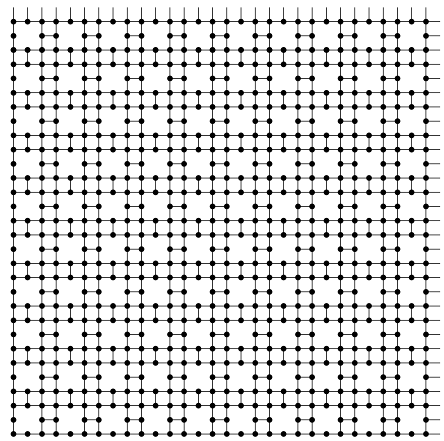

Figure 1.3: The set H_1 .

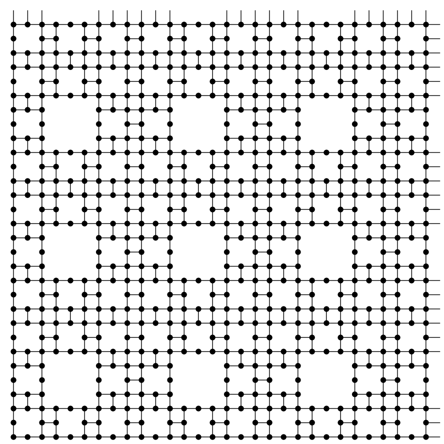

Figure 1.4: The set H_2 .

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

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

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 $n^{-1/2} X^{(1)}_{[nt]}$, $t \geq 0$ should converge weakly to a constant multiple of Brownian motion in $\mathbb{R}^2_+$. So, for large n we should have $\mathbb{E}(X^{(1)}_n)^2 \sim a_1 n$, 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 $\varphi_n(t) = \mathbb{E}^0(X^{(n)}_t)^2$, 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

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

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}.

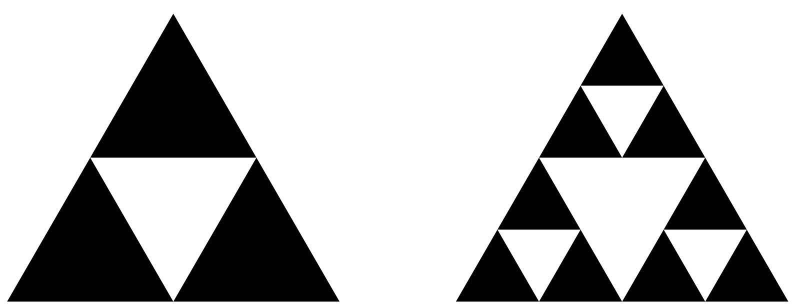

Figure 2.1: The sets H_1 and H_2.

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

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

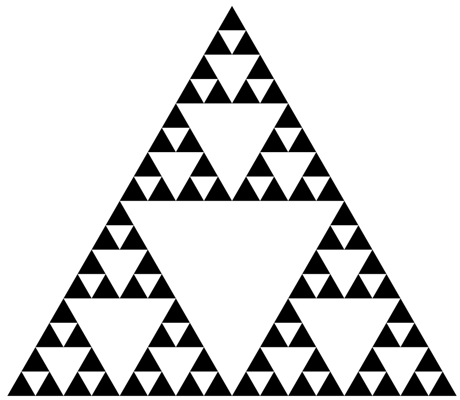

Figure 2.2: The set H_4.

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 $\mathrm{diam}(U) = \sup\{\rho(x,y) : x,y \in U\}$ for the diameter of U. For $\delta > 0$ let

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

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 $\mathcal{H}^{\alpha,1}_\delta(A) \le \mathcal{H}^{\alpha,2}_\delta(A)$. 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

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

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.

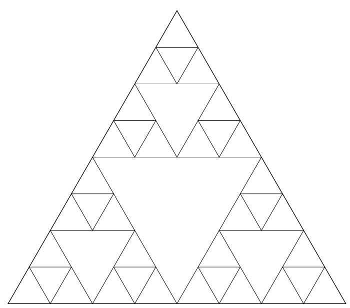

Figure 2.3: The graph G_3.

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.

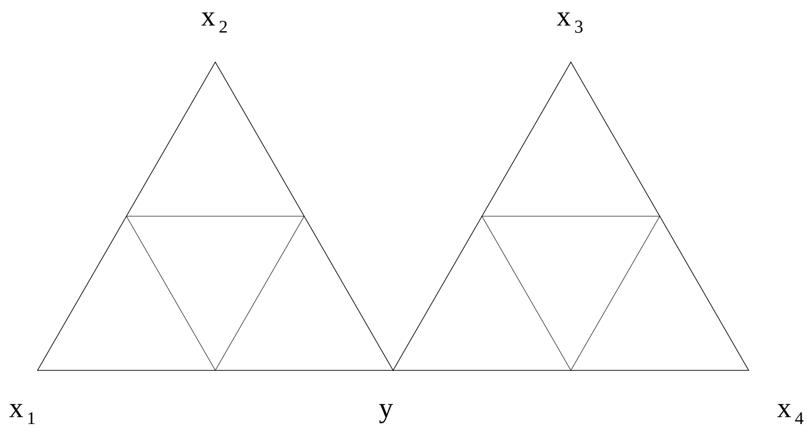

Figure 2.4: y and its neighbours.

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

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

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

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

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. □

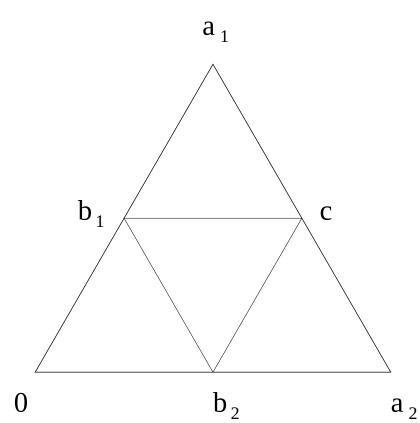

Figure 2.5: The graph G_1.

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 $\widehat{Y}^{(n)}_{[5^n t]} \to X_t$ 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 $t \in (0,\infty)$, $x, y \in \widetilde{G}$

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

$$\leq c_4 t^{-d_f/d_w} \exp\left(-c_5\left(\frac{|x-y|^{d_w}}{t}\right)^{1/(d_w-1)}\right).$$

(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

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

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

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

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

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

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:

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

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 $x \in \widetilde{G}$, $\lambda > 0$, $t > 0$,*

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

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 $x \in G$*

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

(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 [HHW4], 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, HK1, HHK] 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.

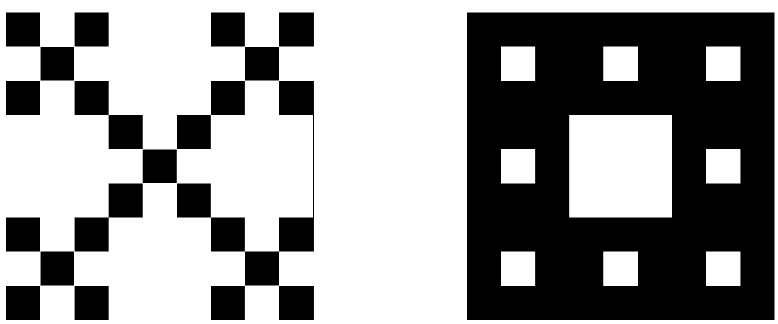

Figure 2.6: The Vicsek set and the Sierpinski carpet.

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 $x = z_0, z_1, \ldots, z_m = y$ 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

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

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

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

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

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

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 $\mathbb{P}^x(\rho(x,X_t) > r) \le 1$) 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 $\tau(x,r)=\inf\{t\ge0: Y_t\notin B(x,r)\}$. 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 $\xi_1,\xi_2,\dots,\xi_n$, V be non-negative r.v. such that $V\ge\sum_1^n\xi_i$. Suppose that for some $p\in(0,1)$, $a>0$,*

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

Then

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

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

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

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

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

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 $f,g \geq 0$ and there exist $a < b$ such that*

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

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

Proof. Let $(E_\lambda, \lambda \geq 0)$ 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 $\text{supp}(f) \subset B(x,r_1)$

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

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

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

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

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

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

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

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 $a = c_6(r/n)^{-\beta}$,

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

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 $\mathbb{P}^x\left(\tau(x,r) \leq t\right) = 0$, which contradicts the fact that $\mathbb{P}^x\left(Y_t \in B(y,r)\right) > 0$ 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

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

then

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

and

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

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

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

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

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

We begin with second term in (3.22):

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

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

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

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

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

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

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 $r \leq c_3 t^{1/\beta}$ then by (3.13)

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

and as $\exp(-(r^\beta/t)^{1/(\beta-1)}) \geq \exp(-c_3^{1/(\beta-1)})$, we have a lower bound of the form (3.3). So now let $r > c_3 t^{1/\beta}$. Let $n \geq 1$. 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 $\rho(x_{i-1}, x_i) = r/n$, $1 \leq i \leq n$. Let $B_i = B(x_i, r/2n)$; note that if $y_i \in B_i$ then $\rho(y_{i-1}, y_i) \leq 2r/n$. We have by the Chapman-Kolmogorov equation, writing $y_0 = x_0$, $y_n = y$,

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

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:

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

which holds provided

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

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

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

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)

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

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* $\lambda > \lambda_0$

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

(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 $x, y \in F$*

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

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

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

(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. □

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*

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

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 $u^A_\lambda(y,x') \geq p^A_\lambda(y,x)u^A_\lambda(x,x')$, we have using the symmetry of X that

$$\begin{aligned} u^A_\lambda(x,y) - u^A_\lambda(x',y) &\leq u^A_\lambda(y,x) - p^A_\lambda(y,x)u^A_\lambda(x,x') \\ &= p^A_\lambda(y,x)\big(u^A_\lambda(x,x) - u^A_\lambda(x,x')\big). \end{aligned} \tag{3.45}$$

Thus

$$|u^A_\lambda(x,y) - u^A_\lambda(x',y)| \leq |u^A_\lambda(x,x) - u^A_\lambda(x,x')|.$$

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

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

So

$$\begin{aligned} |U^A f(x) - U^A f(x')| &\leq \int_A |u^A(x,y) - u^A(x',y)|\,|f(y)|\mu(dy) \\ &\leq c_3 r^{d_w - d_f}||f 1_A||_1. \end{aligned}$$

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

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

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

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

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)

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

From (3.45) and (3.46) we have

$$|u^A_\lambda(x,y) - u^A_\lambda(x',y)| \leq c_2\big(p^A_\lambda(y,x) + p^A_\lambda(y,x')\big)r^{d_w - d_f},$$

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, $d_s < 2$. □

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 $0 \le \lambda_0 \le \lambda_1 \le \dots$ 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 $X = (X_t, t \ge 0, \mathbb{P}^x)$ 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*

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

(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 $\|g_{th}\|_2 \le \|g_{0h}\|_2 \to 0$ 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 $f \ge 0$, and $\|f\|_1 = 1$: we have $||f_t||_1 = 1$. Then by (4.10), for $t > 0$,

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

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

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

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 $a \ge 0$, so that

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

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

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

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

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

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

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

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

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

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

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

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 $\mathcal{E}(f,g) = -f^T A g$. 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 $0 = \mathcal{E}(f,1) = -f^T A 1$ 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

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

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].