# PIMS 2025 – Heat kernel estimates, Harnack inequalities, and quasisymmetry

**Lecturer: Professor Mathav Murugan**

Notes by: Andrew Lin

June 2025

## 1 June 2, 2025

This course will be about heat kernels and Harnack inequalities – we'll start with a general overview.

**Definition 1**

The **heat kernel** is the fundamental solution to the heat equation $\partial_t u = \frac{1}{2}\Delta u$, which is

$$p_t(x,y) = \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{|x-y|^2}{2t}\right)$$

(the density of Brownian motion at time $t$ started at $x$).

What this actually means is that $u(t,\cdot) = p_t(x,\cdot)$ satisfies the heat equation for any fixed $x$, and as $t \downarrow 0$ this solution approximates the Dirac mass $\delta_x$. In particular, this means that for any smooth function $f \in C^\infty(\mathbb{R}^n)$,

$$u(t,x) = \int p_t(x,y) f(y) dy$$

will solve the heat equation with initial condition $f$ at time 0.

This object also has a probabilistic interpretation: we can do this procedure with any Markov process instead of the Brownian motion, and we'll be curious how information on one side (geometry) relates to information about the other (probability). The most basic relation is due to Varadhan – Varadhan's asymptotic says that for

$$\partial_t u = \frac{1}{2}\Delta u$$

for $\Delta$ the Laplace-Beltrami operator, we have that

$$\lim_{t \downarrow 0} 2t \log p_t(x,y) = -d(x,y)^2$$

for any $x, y$ on our manifold. (So the heat kernel determines, or is at least closely related to, the geometry on our space.) Manifolds are locally Euclidean, so this short-time behavior is perhaps not surprising. In contrast, the long-time behavior can look very different: if we work on hyperbolic 3-space, we instead have

$$p_t(x,y) = \frac{1}{(2\pi t)^{n/2}} \frac{d(x,y)}{\sinh d(x,y))} \exp\left(-\frac{t}{2} - \frac{d(x,y)^2}{2t}\right),$$

and in particular we can derive that the expected distance we travel grows linearly in time instead of scaling as $\sqrt{t}$. The factor of $\frac{\text{distance}^2}{\text{time}}$ reflects the "spacetime scaling," and it's natural to ask whether there are other possible scalings as well.

**Example 2**

A **symmetric $\alpha$-stable process** is a Levy process $Y_t$ with jump increments given by the Laplace transform

$$\mathbb{E}\left[e^{i\xi\cdot(Y_{t+s}-Y_t)}\right] = \exp\left(-t|\xi|^\alpha\right)$$

for some $\alpha \in (0,2)$

We don't have an exact formula for the heat kernel except for $\alpha = 1<$ but we do have asymptotics – we know that

$$p_t(x,y) \asymp \min\left(t^{-n/\alpha}, \frac{t}{|x-y|^{n+\alpha}}\right).$$

In particular, the spacetime scaling is now given by $\frac{\text{distance}^\alpha}{\text{time}}$ for some $\alpha < 2$. And it turns out to be possible to do $\alpha > 2$ as well:

**Example 3**

It is possible to define **Brownian motion on the Sierpinski gasket** (the fractal object formed by iteratively removing the median triangles of an equilateral triangle) as the limit of random walks on the approximating shapes.

The heat kernel then turns out to satisfy (due to Barlow and Perkins)

$$p_t(x,y) \asymp \frac{C}{t^{d_f/d_w}} \exp\left(-C'\left(\frac{|x-y|^{d_w}}{t}\right)^{1/(d_w-1)}\right),$$

where $d_f = \log_2 3$ (Hausdorff dimension) and $w = \log_2 5$ (walk dimension), for all $x, y$ and $0 < t < 1$. And the point in this case is that we get $\frac{\text{distance}^{d_w}}{\text{time}}$. So our goal is to understand how these different scalings can be captured by the geometry of our underlying space, and when we'll get heat kernel estimates of each of the different types.

**Remark 4.** *To convince us that the walk dimension can be greater than 2 even though we're inside a two-dimensional space, consider simple random walk on the graph*  . *Consider the expectation for how long it takes to hit another outer vertex when started from an outer vertex; this turns out to be 5. But if we ask the same question for hitting the inner triangle (scaled down by 2), the answer is just 1. This turns out to be a rather general picture: there is always a fractal with specified dimensions with $2 \le d_w \le d_f + 1$.*

We'll make use of the following **boundary trace process** as a useful tool. Suppose we have Brownian motion in the upper half-space $\mathbb{R}^{n+1} \times [0,\infty)$ reflected on the boundary, and then we erase the actual path and only look at the hitting points on the boundary. This can then be interpreted as a jump process on $\mathbb{R}^n$; what remains turns out (due to Spitzer) to exactly be like the symmetric $\alpha$-stable process with $\alpha = 1$. This can then be generalized to any other $\alpha$ as well: Molchanov and Ostrovskii showed that symmetric $\alpha$-stable processes on $\mathbb{R}^n$ can be viewed as boundary traces of some reflected diffusion processes on $\mathbb{H}^{n+1} = \mathbb{R}^n \times [0,\infty)$. (For the Brownian motion case we didn't actually have to do this reflection because of symmetry, but in general the reflection does matter.) So an interesting question to ask is the converse one of when boundary processes behave like symmetric $\alpha$-stable processes; there will be lots of domains where this behavior occurs.

A good framework to study all of this is the theory of Dirichlet forms, and we'll make use of them throughout the course. We'll mostly study symmetric processes, and there are four objects that are more or less equivalent (but it's useful for us to see how they're related): **(1)** the semigroup, **(2)** the resolvent, **(3)** the generator, and **(4)** the Dirichlet form.

For the **semigroup**, the idea is as follows. Given a Markov process $(Y_t)$ and a "nice function" $f$, define

$$P_t f = \mathbb{E}_x [f(Y_t)]$$

(where this notation means that $Y_0 = x$). We then have by the Markov property that

$$P_{t+s} f = \mathbb{E}_x[f(Y_{t+s})] = \mathbb{E}_x [\mathbb{E}_x [f(Y_{t+s})|F_t]] = \mathbb{E}_x [\mathbb{E}_{Y_t} [f(Y_s)]] ,$$

which is $\mathbb{E}_x[(P_s f)(Y_t)] = P_t P_s f$. So we get the property that $P_{t+s} = P_t P_s$:

**Definition 5**

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{R}$ with inner product denoted $\langle \cdot, \cdot \rangle$ and norm $|| \cdot ||$. A family of operators $\{T_t : \mathcal{H} \to \mathcal{H}\}_{t \geq 0}$ is a **semigroup** if the following properties hold:

1. (Symmetry) $T_t$ is symmetric for all $t$, meaning that $\langle T_t f, g \rangle = \langle f, T_f g \rangle$ for all $f, g$. (Not all Markov semigroups have this property, but many do.)
2. (Semigroup property) $T_{t+s} = T_t \circ T_s$ for all $t, s \geq 0$.
3. (Contraction) $||T_t|| \leq 1$ for all $t > 0$, meaning that $||T_t(u)|| \leq ||u||$ for all $u \in \mathcal{H}$.
4. (Strong continuity) $T_t$ approaches the identity as an operator in the sense that $\lim_{t \downarrow 0} ||T_t(u) - u|| = 0$ for all $u \in \mathcal{H}$.

Such Markov processes have right-continuous paths, so the idea is that this comes from the dominated convergence theorem. (In contrast, "weak continuity" would ask that $\langle T_t(u), v \rangle \to \langle u, v \rangle$ for all $u, v$, but this is actually equivalent here.)

**Example 6**

The **Brownian semigroup** or **heat semigroup** is a family of operators $P_t : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ defined by

$$p_t f(x) = \int \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{|x-y|^2}{2t}\right) f(y) dy.$$

This is the expected value under Brownian motion started at $x$, and we can check that this satisfies all four properties.

**Example 7**

The **Ornstein-Uhlenbeck semigroup** is defined as follows: let $\gamma_n(dx) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{||x||^2}{2}\right) dx$ be the standard Gaussian density, and defining $P_t : L^2(\mathbb{R}^n, \gamma_n) \to L^2(\mathbb{R}^n, \gamma_n)$ via the stochastic differential equation

$$dX_t = -X_t dt + \sqrt{2} dB_t \implies X_t = e^{-t} X_0 + e^{-t} B_{e^{2t}-1},$$

where $B_t$ is standard Brownian motion in $\mathbb{R}^n$.

Note that being symmetric with respect to this $L^2$ space is not the same as being symmetric with respect to the usual $L^2$. And this time we have

$$P_t f(x) = \int_{\mathbb{R}^n} f(e^{-t}x + \sqrt{1-e^{-2t}}y)\gamma_n(dy),$$

where we've used the Brownian scaling for the square root term.

**Example 8**

For a discrete state space example, consider $X = \{1, \cdots, n\}$ and define a **conductance matrix** $c : X \times X \to [0, \infty)$, which we assume to be symmetric. Letting $m(x) = \sum_{y \in X} c(x, y)$, we can then define a discrete time random walk via the transition matrix $P(x, y) = \frac{c(x,y)}{m(x)}$; this yields

$$Qf(x) = \mathbb{E}_x [f(Z_1)] = \sum_{y \in X} P(x, y) f(y) \implies Q^k f(x) = \mathbb{E}_x[f(Z_k)].$$

We can make such a chain continuous by waiting an Exp(1) amount of time between jumps and then going from $x$ to $y$ with probability $P(x, y)$. Then $Y_t = Z_{N(t)}$ for $N(t)$ an independent Poisson process of rate 1, and now we have

$$P_t f(x) = \mathbb{E}_x[f(Y_t)] = \mathbb{E}_x[f(Z_{N(t)})],$$

and we just sum over the possibilities for how many jumps we take:

$$P_t f(x) = \sum_{k=0}^{\infty} \mathbb{P}(N(t) = k)\mathbb{E}_x[f(Z_k)] = \sum_{k=0}^{\infty} \frac{e^{-t}t^k}{k!} Q^k f(x).$$

Again, we can verify that this semigroup on $L^2(X, m)$ (with $m$ the function defined above) satisfies all four properties mentioned above.

**Remark 9.** *For an example of a process which is not symmetric, suppose $X_t = X_0 + t$. This is clearly a (deterministic) Markov process, so $P_t f(x) = f(x + t)$. This does not satisfy the symmetry property, but it does the other three for $L^2(\mathbb{R})$.*

The next object is basically the Laplace transform of the semigroup:

**Definition 10**

A **resolvent** on $\mathcal{H}$ is a family of operators $\{G_\alpha : \mathcal{H} \to \mathcal{H}\}_{\alpha > 0}$ satisfying the following properties (paralleling the four above):

1. (Symmetry) $\langle G_\alpha(f), g\rangle = \langle f, G_\alpha(g)\rangle$ for all $f, g \in \mathcal{H}$.
2. (Resolvent equation) for all $\alpha, \beta > 0$, we have

   $$G_\alpha - G_\beta + (\alpha - \beta)G_\alpha G_\beta = 0.$$

   (In particular this makes $G_\alpha$ and $G_\beta$ commute.)
3. (Contraction) $||\alpha G_\alpha|| \leq 1$ for all $\alpha > 0$, meaning that $||\alpha G_\alpha(u)|| \leq ||u||$ for all $u \in \mathcal{H}$.
4. (Strong continuity) $\lim_{\alpha \to \infty} ||\alpha G_\alpha(u) - u|| = 0$ for all $U \in \mathcal{H}$.

Notice that we take $\alpha \to \infty$ instead of $\alpha \to 0$; this is typical when studying Laplace transforms.

**Proposition 11**

For any semigroup $\{T_t : t > 0\}$, define its Laplace transform

$$G_\alpha(u) = \int_0^\infty e^{-\alpha t} T_t(u) dt.$$

Then $\{G_\alpha(u)\}$ is a resolvent.

There are various ways to define this integral – one is that by strong continuity we can view $T_t$ as a continuous function and so it makes sense to take Riemann sums. But since we're in a Hilbert space, we can do something better:

**Definition 12**

Let $I \subseteq \mathbb{R}$ be any interval. We say that a function $f : I \to \mathcal{H}$ is **weakly measurable** if the following hold:

- the map $t \mapsto ||f(t)||$ is measurable,
- for any $v \in \mathcal{H}$, the function $t \mapsto \langle f(t), v \rangle$ is also measurable.

We further call a weakly measurable $f$ **integrable** if $\int_I ||f(t)|| dt < \infty$.

In such a case, $||\langle f(t), v \rangle||$ is also finite and integrable by Cauchy-Schwarz. So if $f$ is weakly measurable and integrable, we can define $\int_I f(t) dt$ via its inner product with an arbitrary vector

$$\left\langle \int_I f(t) dt, v \right\rangle = \int_I \langle f(t), v \rangle dt.$$

Indeed $T_t(u)$ satisfies both conditions, and then from there it's easy to check that we satisfy the four resolvent properties.

**Example 13**

In the simple case where $T_t = I$ for all $t > 0$ (meaning that we don't move around at all), $G_\alpha(u) = \int_0^\infty e^{-\alpha t} dt = \frac{1}{\alpha} I$. In particular, this indeed solves the resolvent equation.

It may not be clear what the motivation for this is at the moment, but the resolvent turns out to be easier to work with in this theory. And indeed Laplace transforms are injections, so the inverse Laplace transform yields the semigroup (but we'll come back to this later).

Our next object is basically the derivative of the semigroup at time $t = 0$, but we need to be precise because we're now no longer going to have a bounded operator (in fact, it isn't defined on all of $\mathcal{H}$):

**Definition 14**

Let $\{T_t : \mathcal{H} \to \mathcal{H}\}_{t>0}$ be a semigroup. The **generator** of the semigroup is

$$A(u) = \lim_{t \downarrow 0} \frac{T_t(u) - u}{t}$$

defined on the **domain** $\mathcal{D}(A)$ of $\mathcal{H}$ for which this limit exists.

It's easy to check that the domain is a subspace, and we may ask why we would want to work with generators instead of the objects above. The idea is that they are closer to the description of the Markov process and thus easier to compute:

**Example 15**

Recall the notation of the discrete Example 8, in particular the transition probabilities $P(x, y)$. Instead of waiting an Exp(1) amount of time at each state before jumping, we can specify arbitrary values $\lambda : X \to (0, \infty)$ and wait at state $x$ with parameter Exp $(\lambda(x))$ (so larger $\lambda$ means we wait a shorter amount of time) before jumping with probability $P(x, y)$.

In the case where all $\lambda$s are 1, it's easy to compute the semigroup because we have an easy expression for the number of jumps. But now the number of jumps depends on the path that we take, and it will turn out to be easier to compute the generator (which is local): we only need the semigroup up to linear terms in $t$. As $t \downarrow 0$, it's very likely that we will make only zero or one jump in total (terms with two jumps scale as $t^2$), so

$$P_t f(x) = \mathbb{E}_x[f(Y_t)] = e^{-\lambda(x)t} f(x) + (1 - e^{-\lambda(x)t}) \sum_y P(x, y) f(y) + O(t^2),$$

and subtracting off $f(x)$, dividing by $t$, and taking the limit yields the generator

$$Lf(x) = -\lambda(x) f(x) + \lambda(x) \sum_y P(x, y) f(y).$$

Furthermore, this computation tells us the measure with which we are symmetric, and we'll see that next time.

# 2 June 3, 2025

We saw some relations between the semigroup, generator, and resolvent last time; we mentioned that these notions end up being equivalent but that different objects will be useful in different contexts (for example, the resolvent is easier to work with, and the generator is easier to compute). In Example 15 we previously computed the generator of a continuous-time Markov chain with specified jump rates, finding that $Lf(x) = \lim_{t \to 0} \frac{P_t f(x) - f(x)}{t} = -\lambda(x) f(x) + \lambda(x) \sum_y P(x, y) f(y)$. Our goal is to calculate the symmetrizing measure, and we'll do so by rewriting

$$Lf(x) = \lambda(x)(I - Q) f(x),$$

where $Q$ is the operator defined by $Qf(x) = \sum_{y \in X} P(x, y) f(y)$ (this corresponds to the discrete-time chain). By the symmetry assumption on our conductance matrix,

$$m(x) P(x, y) = c(x, y) = c(y, x) = m(y) P(y, x),$$

and therefore we can check that

$$\langle f, Q(g) \rangle_{L^2(m)} = \langle Q(f), g \rangle_{L^2(m)}.$$

So we know that $(I - Q)$ is symmetric with respect to $m$ as well, but we need to make a correction for the $\lambda(x)$ term: if we define a new measure $\tilde{m}$ via $\tilde{m}(x) = \frac{m(x)}{\lambda(x)}$, then we will indeed find that

$$\langle f, Lg \rangle_{L^2(\tilde{m})} = \langle f, (I - Q) g \rangle_{L^2(m)} = \langle (I - Q) f, g \rangle_{L^2(m)} = \langle Lf, g \rangle_{L^2(\tilde{m})}.$$

So the measure $\tilde{m}$ makes our generator symmetric, and it will turn out to imply that the transition semigroup generates a Markov process which will also be symmetric under this same measure.

We can interpret $\tilde{m}$ in a nice probabilistic way as well. Suppose that the corresponding discrete-time chain $(Z_n)$ is irreducible; we know that after we run this chain for a long time, the discrete time chain will have occupation times

proportional to $m(x)$. But since we wait at $x$ for a total time $\text{Exp}(\lambda(x))$ at each step, we should have

$$\lim_{t\to\infty} \frac{1}{t} 1\{Y_t = y\} dt = \frac{\tilde{m}(y)}{\sum_{z\in X} \tilde{m}(z)},$$

and this is true $\mathbb{P}_x$-almost-surely (that is, no matter where we start within an irreducible component).

We're now ready to write out the relations between these general objects more carefully. The definition of the generator tells us that $A(u) = \lim_{t\downarrow 0} \frac{T_t(u)-u}{t}$, but we can also think about this as a function $f(t) = T_t(u)$ and differentiate at nonzero times as well. We can rewrite that expression in two different ways:

$$\lim_{h\downarrow 0} \frac{T_{t+h}(u) - T_t(u)}{h} = \lim_{h\downarrow 0} T_t \circ \left(\frac{T_h(u) - u}{h}\right) = \lim_{h\downarrow 0} \frac{T_h(T_t(u)) - T_t(u)}{h}.$$

So now if $u$ is in the domain of the generator, the limit in the middle expression exists because $T_t$ is a bounded operator, and thus the limit in the latter expression should also exist, and we find that

$$T_t(A(u)) = A(T_t(u))$$

for any $u \in \mathcal{D}(A)$. And we can rewrite this as a differential equation

$$\frac{df(t)}{dt} = A(f(t))$$

for some possibly unbounded operator $A$. If $A$ were bounded (like it is in our finite-dimensional example above), we would have $f(t) = e^{tA}(u)$, so in other words the semigroup is $T_t = e^{tA}$ and the resolvent "should be"

$$G_\alpha(u) = \int e^{-t\alpha} T_t(u) = \int e^{-t\alpha} e^{tA}(u) dt = (\alpha - A)^{-1}(u).$$

So we'd like to say that the generator $A$ determines the semigroup and the resolvent via those two formulas above, but to do this we need to show that the generator is a self-adjoint operator. So we'll spend some time trying to develop that theory now.

One issue we see is that this integral may not converge if $A$ has positive spectrum, and so that's the first thing we'll resolve:

**Lemma 16**

For any generator $A$, $\mathcal{D}(A)$ is a subspace of $\mathcal{H}$, and the function $A : \mathcal{D}(A) \to \mathcal{H}$ is a linear map. Furthermore, $A$ is nonpositive definite, meaning that

$$\langle A(u), u\rangle \leq 0 \text{ for all } u \in \mathcal{D}(A).$$

*Proof.* The first sentence follows directly from the definitions (limits interact nicely with linear maps and linear combinations). Thus we just need to prove the last fact, and we can look at the difference quotients. But

$$\begin{aligned}\left\langle \frac{T_t(u) - u}{t}, u \right\rangle &= \frac{1}{t}\left(\langle T_t(u), u\rangle - \langle u, u\rangle\right) \\ &= \frac{1}{t}\left(||T_t(u)|| \cdot ||u|| - ||u||^2\right) \\ &\leq 0\end{aligned}$$

by Cauchy-Schwarz and the fact that we have a contraction (so $||T_t|| \leq 1$). So by continuity of the inner product, the limiting quantity must also be nonpositive as well. □

Next, we'll see some properties of $G_\alpha$. If $A$ is nonpositive definite, then $-A$ is nonnegative definite and thus the resolvent should also be nonnegative definite. We'll prove that now (though note that we don't know yet that every resolvent comes from a semigroup):

**Lemma 17**

Let $\{G_\alpha : \alpha > 0\}$ be a resolvent. Then for all $\alpha > 0$, $G_\alpha$ is nonnegative definite, meaning that $\langle G_\alpha(u), u\rangle \geq 0$ for all $u$.

*Proof.* Fix some $u$. The function

$$f(\alpha) = \langle u, G_\alpha(u)\rangle$$

turns out to be differentiable in $\alpha$, because by the resolvent identity

$$\frac{f(\alpha) - f(\beta)}{\alpha - \beta} = \left\langle u, \frac{G_\alpha(u) - G_\beta(u)}{\alpha - \beta}\right\rangle = -\langle u, G_\alpha G_\beta(u)\rangle = -\langle G_\alpha u, G_\beta u\rangle$$

since each $G_\alpha$ is a symmetric operator. So as $\alpha \to \beta$, we have $G_\alpha \to G_\beta$, and taking the limit yields that $f'(\alpha) = -\langle G_\alpha(u), G_\alpha(u)\rangle \leq 0$. On the other hand, we know that

$$\lim_{\alpha\to\infty} \alpha f(\alpha) = \lim_{\alpha\to\infty} \langle u, \alpha G_\alpha(u)\rangle = \langle u, u\rangle \geq 0$$

by the strong continuity property. Thus $f$ is nonnegative for all $\alpha$, proving nonnegative definiteness as desired. □

If we have $G_\alpha = (\alpha - A)^{-1}$ formally, then we can solve for $A$ and find that we should get the generator $A = \alpha - G_\alpha^{-1}$ from the resolvent. For this to make sense, we must first show that the resolvent is invertible.

**Lemma 18**

For any resolvent, $G_\alpha$ is injective for all $\alpha > 0$.

*Proof.* Since $G_\alpha$ is a linear map, we just need to check that the kernel is trivial. Suppose that $G_\alpha(u) = 0$; the resolvent identity (multiplied by $\beta$) tells us that

$$\beta G_\alpha(u) = \beta G_\beta(u) + \beta(\beta - \alpha) G_\beta(G_\alpha(u)),$$

but now plugging in $G_\alpha(u) = 0$ shows that $\beta G_\beta(u) = 0$ for any arbitrary $\beta$. Thus by strong continuity this implies that $u = 0$ as well, which is what we wanted. □

We thus use that formula above to make our definition:

**Definition 19**

Let $\{G_\alpha : \alpha > 0\}$ be a resolvent on $\mathcal{H}$. The **generator of the resolvent** is defined by setting the domain to be $\mathcal{D}(A) = G_\alpha(\mathcal{H})$ (that is, the range of $G_\alpha$), and defining

$$A(u) = \alpha u - G_\alpha^{-1}(u)$$

for all $u \in \mathcal{D}(A)$.

As an exercise, we can check from the resolvent identity that $\alpha - G_\alpha^{-1}$ does not actually depend on $\alpha$, and similarly we can check that the range does not depend on $\alpha$, so this generator is in fact well-defined.

**Remark 20.** *It's clear from strong continuity that the domain of the generator of a resolvent is dense, since $\mathcal{D}(A)$ is the range of $G_\alpha$, but $\alpha G_\alpha(u)$ (which is in the domain) converges to $u$ as $\alpha \to \infty$.*

We haven't checked yet that "all of our operations commute:" if we take a Laplace transform of the semigroup (to get the resolvent) and then find the generator of that, it should be the same as the generator of the original semigroup, but it's not so clear that this is true yet. Before doing that, we'll first detour to the concept of self-adjointness:

**Definition 21**

Let $\mathcal{H}$ be a Hilbert space, and let $T$ be a densely-defined operator on $\mathcal{H}$ (that is, it is an operator $\mathcal{D}(T) \to \mathcal{H}$ where $\overline{\mathcal{D}(T)} = \mathcal{H}$). The **adjoint** of $T$, denoted $T^*$, is defined on the subspace of $\mathcal{H}$

$$\mathcal{D}(T^*) = \{x \in \mathcal{H} : \text{the map } y \mapsto \langle T(y), x\rangle \text{ is bounded on } \mathcal{D}(T)\}.$$

The idea is that Hahn-Banach allows us to extend a bounded functional to the whole space, and then Riesz representation lets us represent that result as an inner product with some other element of $\mathcal{H}$. Thus, for all $x \in \mathcal{D}(T^*)$, there is some unique vector (which we define to be $T^*(x)$) such that $\langle T(y), x\rangle = \langle y, T^*(x)\rangle$ for all $y \in \mathcal{D}(T)$.

Note that as stated right now, it's not clear that $\mathcal{D}(T^*)$ is dense in the Hilbert space, and in fact we do need additional assumptions in general for that to be case. Being self-adjoint will require an additional condition:

**Definition 22**

For any Hilbert space $\mathcal{H}$, $\mathcal{H} \times \mathcal{H}$ is also a Hilbert space with inner product $\langle (x_1, y_1), (x_2, y_2)\rangle = \langle x_1, x_2\rangle + \langle y_1, y_2\rangle$. For a linear operator $T : \mathcal{D}(T) \to \mathcal{H}$, we let $\mathcal{G}(T)$ be the **graph** of $T$ – that is, the set of points $(x, T(x)) \in \mathcal{H} \times \mathcal{H}$ for $x \in \mathcal{D}(T)$. We say that $T$ is **closed** if $\mathcal{G}(T)$ is closed in $\mathcal{H} \times \mathcal{H}$.

We may recall from ordinary functional analysis that if the domain is the whole $\mathcal{H}$, then being closed is the same as being bounded. But it turns out that for densely defined operators, being closed is the right generalization.

**Definition 23**

An operator $T : \mathcal{D}(T) \to \mathcal{H}$ is **symmetric** if $\langle T(x), y\rangle = \langle x, T(y)\rangle$ for all $x, y \in \mathcal{D}(T)$.

From the definition of $T^*$, we see that $T$ is symmetric if and only if $T^*$ is an extension of $T$. So the domain of $T^*$ might be strictly larger than the domain of $T$, but we don't want that to be the case for self-adjoint operators:

**Definition 24**

An operator $T : \mathcal{D}(T) \to \mathcal{H}$ is **self-adjoint** if $T = T^*$, meaning that $T$ is symmetric and $\mathcal{D}(T^*)$ coincides with $\mathcal{D}(T)$.

This is a property that will be satisfied by generators and resolvents of semigroups, and the reason it's so useful is that we understand exactly how they behave (up to some transformations). Notice also that symmetric operators defined on the whole space are automatically self-adjoint (because the domain of the adjoint can only be bigger), and by the closed graph theorem they are also continuous.

We'll first rephrase the relationship between an operator and its adjoint in terms of graphs:

**Definition 25**

Suppose $M$ is a subspace of $\mathcal{H}$ (or $\mathcal{H} \times \mathcal{H}$). Write $M^\perp$ for the set of vectors $u \in \mathcal{H}$ with $\langle u, m \rangle = 0$ for all $m \in M$.

Notice that $M^\perp$ is always closed even if $M$ is not (since this set is closed for any fixed $m$ and the intersection of closed sets is closed).

**Lemma 26**

Suppose $T : \mathcal{D}(T) \to \mathcal{H}$ is densely defined, and let $V : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ be the map $(x, y) \mapsto (y, -x)$ (which preserves the inner product). Then the graphs satisfy

$$\mathcal{G}(T^*) = V(\mathcal{G}(T))^\perp;$$

in particular this means the graph of $T^*$ is closed, so $T^*$ is always closed.

*Proof.* Saying that $(x, y) \in \mathcal{G}(T^*)$ is the same as saying that $\langle T(u), x \rangle = \langle u, y \rangle$ for all $u \in \mathcal{D}(T)$. But this is further equivalent to

$$\langle V(u, T(u)), (x, y) \rangle_{\mathcal{H} \times \mathcal{H}} = 0 \text{ for all } u \in \mathcal{D}(T),$$

which is exactly the definition of the space on the right-hand side. □

**Corollary 27**

Let $T : \mathcal{D}(T) \to \mathcal{H}$ be injective and self-adjoint, and let $\mathcal{R}(T)$ be the range of $T$. Then the map $T^{-1} : \mathcal{R}(T) \to \mathcal{H}$ is a densely-defined, self-adjoint operator.

*Proof.* First we prove that $\mathcal{R}(T)$ is dense in $\mathcal{H}$, or equivalently that $\mathcal{R}(T)^\perp = \{0\}$. For any $y \in \mathcal{R}(T)^\perp$, we know that $(y, 0) \in V(\mathcal{G}(T))^\perp$ (because a vector in $V(\mathcal{G}(T))$ looks like $V(x, T(x)) = (T(x), -x)$ for some $x \in \mathcal{D}(T)$, and the inner product of this with $(y, 0)$ is always zero). By our previous lemma, this means that $(y, 0) \in \mathcal{G}(T^*)$. But $\mathcal{G}(T^*) = \mathcal{G}(T)$, so because $T$ is one-to-one that must mean $y = 0$, as desired.

From here, we wish to prove self-adjointness. The graph of the inverse of a function and of a function are just related by swapping the coordinates, so let $S : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ be the map $S(x, y) = (y, x)$. We have, again by the previous lemma and then our swap map operation, that

$$\boxed{\mathcal{G}((T^{-1})^*)} = V(\mathcal{G}(T^{-1}))^\perp = V(S(\mathcal{G}(T))^\perp.$$

Now $V$ and $S$ commute up to a negative sign, but since we look at orthogonal complements that means this is the same as $S(V(\mathcal{G}(T))^\perp$. We now want to move the $\perp$ from outside to inside the $S$; to justify that step we observe that $S$ also preserves the inner product and thus orthogonal complements. Thus this space is also (using our lemma once again, and also using that $T$ is self-adjoint)

$$S(V(\mathcal{G}(T)^\perp) = S(\mathcal{G}(T^*)) = S(\mathcal{G}(T)) = \boxed{\mathcal{G}(T^{-1})}.$$

So equating the boxed expressions, the adjoint of $T^{-1}$ is indeed $T^{-1}$ itself, completing the proof. □

**Example 28**

Let $(\Omega, \mu)$ be a measure space, and let $\mathcal{H} = L^2(\Omega, \mu)$. Let $\lambda : \Omega \to \mathbb{R}$ be a measurable function, not necessarily bounded. We can then define the "multiplication operator" $M_\lambda : \mathcal{D}(M_\lambda) \to \mathcal{H}$ via

$$M_\lambda(f) = \lambda(\cdot)f(\cdot)$$

with domain exactly the set of functions $f \in \mathcal{H}$ such that $M_\lambda(f) \in \mathcal{H}$ (that is, such that $\lambda f \in L^2(\Omega, \mu)$).

We claim that $M_\lambda$ is a self-adjoint operator – this is a simple exercise using Cauchy-Schwarz. Specifically, it's clear that this is symmetric, but we need to prove that the domain of $M_\lambda^*$ is not larger than the domain of $M_\lambda$.

It turns out that every self-adjoint operator is essentially a multiplication operator:

**Theorem 29** (Spectral theorem)

Let $A$ be a densely-defined self-adjoint operator on a Hilbert space $\mathcal{H}$. Then there is a measure space $(\Omega, \mu)$, a measurable real-valued function $\lambda : (\Omega, \mu) \to \mathbb{R}$, and a unitary bijection $U : L^2(\Omega, \mu) \to \mathcal{H}$ (meaning inner products are conserved) such that $UAU^{-1} = M_\lambda$, and the domain $\mathcal{D}(A)$ is exactly $\{U(f) : f \in \mathcal{D}(M_\lambda)\}$.

We've probably seen a version of the spectral theorem in our linear algebra course (viewing symmetric matrices as diagonal matrices in some orthonormal basis); this is an infinite-dimensional version. This theorem will be useful for us because it allows us to manipulate self-adjoint operators as if they are numbers, like in the heuristic integral we did earlier.

**Remark 30.** *There's a more complicated statement of the spectral theorem using projection-valued measures which is more canonical, but we can still work with the version that we have here because we have uniqueness up to unitary transformations.*

**Definition 31** (Functional calculus)

Let $A$ be a (densely-defined) self-adjoint operator on $\mathcal{H}$, and let $A$ be nonpositive definite. For any Borel function $f : (-\infty, 0] \to \mathbb{R}$, define the operator $f(A) : \mathcal{D}(f(A)) \to \mathcal{H}$ as follows. Expressing $A$ via the equation $U^{-1}AU = M_\lambda$, the domain of $f(A)$ is

$$\mathcal{D}(f(A)) = \left\{U(g) : g \in \mathcal{D}(M_{f(\lambda(\cdot))})\right\},$$

and the operator is defined via $U^{-1}f(A)U = M_{f(\lambda(\cdot))}$.

This is basically the analog of taking powers of matrices, where we first diagonalize and then apply the powers componentwise to each diagonal element. And we can check that this definition is independent of our specific choice of $U$.

**Proposition 32**

Let $\{T_t\}_{t>0}$ be a semigroup, and let $\{G_\alpha\}_{\alpha>0}$ be the associated resolvent $G_\alpha(u) = \int_0^\infty e^{-\alpha t}T_t(u)dt$ for all $u \in \mathcal{H}$. Let $A_s$ and $A_r$ be the generators of $\{T_t\}$ and $\{G_\alpha\}$, respectively. Then $A_s = A_r$, and the generator is a nonpositive definite operator.

The non-positivity part we've already proved from taking a limit in the semigroup interpretation, and the self-adjoint part we've proved from the resolvent interpretation. But if we show that they coincide, then we get both of those properties for free!

# 3 June 5, 2025

We've now made connections between the semigroup, resolvent, and generator, and we used the spectral theorem to do some of the relevant manipulations. One important detail is that if our self-adjoint operator $A$ is also nonpositive definite, then in fact $\lambda \leq 0$ $\mu$-almost everywhere.

*Proof of Proposition 32.* To check that $A_r = A_s$, we need to also check that the domains of the two operators are the same, and it's equivalent to prove that each is a restriction of the other.

First suppose that $u \in \mathcal{D}(A_r)$; this means that $u$ is in the range of $G_\alpha$ and thus we can write $u = G_\alpha(v)$ for some $v \in \mathcal{H}$ (and in fact a unique $v$). We want to prove that $u$ also belongs to $\mathcal{D}(A_s)$, which is checking differentiability. Specifically, we'll check differentiability of $e^{-\alpha t}T_t(u)$ instead of $T_t(u)$ itself, since the two facts are equivalent. We have

$$\frac{1}{t}\left(e^{-\alpha t}T_t(u) - u\right) = \frac{1}{t}\left(e^{-\alpha t}T_t\left(\int_0^\infty e^{-\alpha s}T_s(v)dv\right) - \int_0^\infty e^{-\alpha s}T_s(v)dv\right)$$

We can now interchange integrals with $T_t$ and then shift the inner integral to $[t, \infty)$ instead of $[0, \infty)$; the result is that most of the integral cancels and we're left with

$$-\frac{1}{t}\int_0^t e^{-\alpha s}T_s(v)ds,$$

and using right-continuity this approaches $v$ as $t \downarrow 0$. So the point is that

$$\boxed{\frac{1}{t}(T_t(u) - u) = \frac{1}{t}T_t(u)(1 - e^{-\alpha t}) + \frac{1}{t}\left(e^{-\alpha t}T_t(u) - u\right)},$$

and then as $t \downarrow 0$ this converges to $\alpha u - v$. So $u$ indeed belongs to the semigroup with $A_s(u) = \alpha u - v = \alpha u - G_\alpha^{-1}(u)$, and that is exactly the definition of $A_r(u)$ from Definition 19. Therefore $A_r$ is an extension of $A_s$.

For the other direction, we basically do the same computation. Since we've already shown that

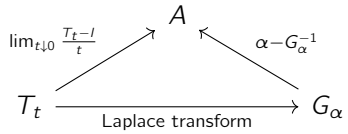
$$v = -\lim_{t\downarrow 0}\frac{1}{t}\left(e^{-\alpha t}T_t(u) - u\right)$$

for some $u \in \mathcal{D}(A_s)$, by the same argument as before this limit exists (since in the boxed expression we know the first two terms have limits) this approaches $-A_s(u) + \alpha u$. We expect $u = G_\alpha(v)$, so we define $w = u - G_\alpha(v)$ (we want to show this is zero) and compute

$$\begin{aligned}\lim_{t\downarrow 0}\frac{1}{t}\left(e^{-\alpha t}T_t(w) - w\right) &= (-v) - \lim_{t\downarrow 0}\frac{1}{t}\left(e^{-\alpha t}T_t(G_\alpha(v)) - G_\alpha(v)\right)\\ &= (-v) + v = 0,\end{aligned}$$

so indeed $G_\alpha(v) = u$ and thus $u \in G_\alpha(\mathcal{H})$, which we know to be the domain of the resolvent. And as mentioned before, we already did the work of showing the other properties of the operators (self-adjoint, nonpositive definite) through either $A_r$ or $A_s$. □

We thus have the following operators between the generator $A$, resolvent $G_\alpha$, and semigroup $T_t$, and the arrows commute:



Our goal is now to fill in some of the reverse arrows to complete the picture. Specifically, we'll try to go backwards from the generator to the other objects:

**Proposition 33**

Let $\mathcal{H}$ be a Hilbert space and $A : \mathcal{D}(A) \to \mathcal{H}$ a nonpositive definite self-adjoint operator. Then we can define $\{T_t = \exp(tA)\}_{t>0}$ via the functional calculus, and similarly we can define $\{G_\alpha = (\alpha - A)^{-1}\}_{\alpha>0}$; these are exactly the semigroup and resolvent corresponding to $A$, respectively., and $\{G_\alpha\}$ is the resolvent corresponding to $\{T_t\}$. Furthermore, there is a unique semigroup and resolvent corresponding to any $A$ (so everything here is a bijection between the set of objects of interest).

*Proof sketch.* By the spectral theorem, we can essentially assume $A$ is a multiplication operator. Specifically, let $U : L^2(\Omega, \mu) \to \mathcal{H}$ and $\lambda : \Omega \to \mathbb{R}$ come from the spectral theorem, so that $UAU^{-1} = M_\lambda$. Then by definition we must have

$$U^{-1}T_tU = M_{\exp(t\lambda(\cdot))}, \quad U^{-1}G_\alpha U = M_{(\alpha-\lambda(\cdot))^{-1}}.$$

(Notice that we never divide by zero since $\alpha > 0$.) Then $T_t$ is indeed a contraction because $\exp(t\lambda(x))$ is always nonpositive, and

$$\langle T_t(f), g\rangle_{\mathcal{H}} = \langle M_{\exp(t\lambda(\cdot))}(U^{-1}(f)), U^{-1}(g)\rangle_{L^2(\Omega,\mu)}$$

at which point we can move the multiplication operator to the other argument to get symmetry. Furthermore, the domain of each $T_t$ is always all of $\mathcal{H}$, because $M_{\exp(t\lambda(\cdot))}$ is always multiplying "component-wise" by something in $[0, 1]$ and thus sends $L^2$ functions to $L^2$ functions. The other semigroup properties are also similarly easy to check: multiplying by $e^{t\lambda(\cdot)}$ and then $e^{s\lambda(\cdot)}$ is the same as multiplying by $e^{(t+s)\lambda(\cdot)}$, so we get the semigroup property. And for strong continuity we have

$$||T_t(f) - f|| = ||e^{t\lambda(\cdot))}(U^{-1}(f)) - U^{-1}(f)||_{L^2(\Omega,\mu)};$$

now by dominated convergence theorem this goes to 0 because the inner thing is $(e^{t\lambda(\cdot)} - 1)U^{-1}(f)$, which has norm at most 2.

The arguments for checking resolvent properties are similar, and we'll omit them here. To check that the resolvent corresponds to the semigroup as well, we can just notice that

$$\int e^{-\alpha t} M_{\exp(t\lambda(\cdot))} dt = M_{(\alpha-\lambda(\cdot))^{-1}},$$

since for any constant $\mu < 0$ we have $\int_0^\infty e^{-\alpha t+\mu t} dt = (\alpha - \mu)^{-1}$, so the multiplication operator (and then applying $U$ to get back to the original operator) yields the desired result.

Finally, to get uniqueness, we just need to check that the maps we've defined are one-to-one. We'll just do this for the resolvent: assume that $\{G_\alpha\}$ and $\{G'_\alpha\}$ both correspond to the same generator $A$. We wish to show that the resolvents are the same – to do this, consider the quantity $w = G_\alpha(f) - G'_\alpha(f)$ for an arbitrary $f \in \mathcal{H}$. By the definition of the generator corresponding to a resolvent, we know that $A = \alpha - G_\alpha^{-1} = \alpha - (G'_\alpha)^{-1}$ implies that

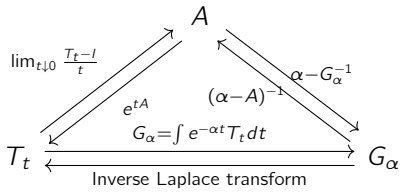
$$(\alpha - A)G_\alpha = (\alpha - A)G'_\alpha = \mathrm{Id} \implies (\alpha - A)w = f - f = 0.$$

But now $A$ is nonpositive, so

$$0 = \langle(\alpha - A)w, w\rangle \geq \alpha\langle w, w\rangle$$

and the right-hand side is nonnegative, hence must be zero and thus $w = 0$ as desired. And the injectivity of the Laplace transform also tells us that the map from generators to resolvent is also one-to-one (alternatively, we can also do this directly via some differential inequalities). □

We can thus fill in the rest of our diagram now as well:




It's left as an exercise for us to check (via the spectral theorem) that the following inverse Laplace transform is correct:

**Proposition 34**

For a resolvent $\{G_\alpha\}$, we can recover the corresponding semigroup via the formula

$$T_t(f) = \lim_{\alpha \to \infty} e^{-t\alpha} \sum_{n=0}^{\infty} \frac{(t\alpha)^n}{n!} (\alpha G_\alpha)^n(f),$$

where $G_\alpha^n$ means the repeated composition of $G_\alpha$ with itself $n$ times.

In particular, this tells us that if the resolvent is a contraction, then the semigroup is a contraction as well (by taking operator norms).

We'll next introduce a **quadratic form** which essentially encodes all of this information as well, but instead of describing "global behavior" (like the resolvent and semigroup) it will be more similar to "local behavior" (like the generator).

**Definition 35**

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. A **quadratic form** is a map $\mathcal{E} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$, where $\mathcal{F}$ is a dense subspace of $\mathcal{H}$, satisfying the following properties:

1. (Bilinearity) $\mathcal{E}(a_1 f_1 + a_2 f_2, g) = a_1 \mathcal{E}(f_1, g) + \mathcal{E}(f_2, g)$ for all $f_1, f_2, g \in \mathcal{H}$ and $a_1, a_2 \in \mathbb{R}$.
2. (Symmetry) $\mathcal{E}(f, g) = \mathcal{E}(g, f)$ for all $f, g \in \mathcal{H}$.
3. (Nonnegative definiteness) $\mathcal{E}(f, f) \geq 0$ for all $f \in \mathcal{F}$. (This is the main difference with inner products – we do not require that equality only holds for the zero vector.)

We usually write the quadratic form as $(\mathcal{E}, \mathcal{F})$. Such a quadratic form is **closed** if $\mathcal{F}$ is a Hilbert space with respect to the inner product $\mathcal{E}_1(f, g) = \mathcal{E}(f, g) + \langle f, g \rangle$.

If we're familiar with the theory of Sobolev spaces, this is similar to how we have inner products along with a gradient term. And the point is that **closed quadratic forms corresponding to nonpositive self-adjoint operators**, so we can think of them as equivalent to generators. The following is an exercise:

**Proposition 36**

We have a "Cauchy-Schwarz inequality" $|\mathcal{E}(u, v)| \leq \mathcal{E}(u, u)^{1/2} \mathcal{E}(v, v)^{1/2}$ for all $u, v \in \mathcal{F}$, and also a "triangle inequality" $\mathcal{E}(f + g, f + g)^{1/2} \leq \mathcal{E}(f, f)^{1/2} + \mathcal{E}(g, g)^{1/2}$.

**Theorem 37**

Let $A$ be a nonpositive definite, self-adjoint operator on a Hilbert space $\mathcal{H}$. Define a quadratic form on $\mathcal{F} = \mathcal{D}(\sqrt{-A})$ via

$$\mathcal{E}(f,g) = \left\langle \sqrt{-A}f, \sqrt{-A}g \right\rangle.$$

Then $(\mathcal{E}, \mathcal{F})$ is a closed quadratic form. Conversely, any closed quadratic form arises in this way from some nonpositive definite self-adjoint $A$.

*Proof.* The proof actually makes use of the resolvent construction; somehow getting from quadratic forms to resolvents is much easier than making connections in other ways. (But once we understand this basic theory, we'll typically just work with the quadratic forms directly.) From the definitions, it is clear that $(\mathcal{F}, \mathcal{E}_1)$ is an inner product space (recalling that $\mathcal{E}_1(\cdot,\cdot) = \mathcal{E}(\cdot,\cdot) + \langle \cdot, \cdot \rangle$), and all that's left is to check completeness.

This basically comes down to the fact that $\sqrt{-A}$ is a nonnegative definite self-adjoint operator, because $\sqrt{-\lambda}$ is always defined and nonnegative. Indeed, for any Cauchy sequence $(f_n)$ in $(\mathcal{F}, \mathcal{E}_1)$, we can equivalently formulate this as saying that the ordered pair $(f_n, \sqrt{-A}f_n)$ is a Cauchy sequence in $\mathcal{H} \times \mathcal{H}$ (by the definition of the inner product in a product Hilbert space). This is a sequence of points in the graph $\mathcal{G}(\sqrt{-A})$ of the operator $\sqrt{-A}$, and the graph of a self-adjoint operator (in fact any adjoint operator) is always closed. Therefore $(f_n, \sqrt{-A}f_n)$ converges to some point $(f, \sqrt{-A}f)$ also on the graph, and that's equivalent to saying that $f_n \to f$ in the $\mathcal{E}_1$-inner product, proving closedness.

For the other direction, we define a generalization of $\mathcal{E}_1$ for all $\alpha > 0$

$$\mathcal{E}_\alpha(f,g) = \mathcal{E}(f,g) + \alpha\langle f, g\rangle.$$

Each $\mathcal{E}_\alpha$ is comparable to each other one, meaning that $(\mathcal{F}, \mathcal{E}_\alpha)$ is a Hilbert space for all $\alpha > 0$. Now the map $v \mapsto \langle u, v\rangle$ is a bounded linear functional on $\mathcal{H}$; it is also a bounded linear functional on $(\mathcal{F}, \mathcal{E}_\alpha)$ for each $\alpha$, so by the Riesz representation theorem it must be of the form $\mathcal{E}_\alpha(G_\alpha(u), v)$ for some unique $G_\alpha(u) \in \mathcal{F}$. We claim that this is actually the resolvent we are looking for.

1. First, we check symmetry – we use the equation above with the roles of $u, v$ swapped to get

   $$\langle G_\alpha(u), v\rangle = \mathcal{E}_\alpha(G_\alpha(u), G_\alpha(v)),$$

   but by the symmetry of $\mathcal{E}_\alpha$ that means this is also equal to $\langle u, G_\alpha(v)\rangle$.

2. Next, we check the resolvent identity. For any $u \in \mathcal{H}$ and any $v \in \mathcal{F}$, we compute

   $$\begin{aligned}\boxed{\mathcal{E}_\alpha\left(G_\beta(u) - (\alpha-\beta)G_\alpha G_\beta(u), v\right)} &= \mathcal{E}_\alpha\left(G_\beta(u), v\right) - (\alpha-\beta)\mathcal{E}_\alpha\left(G_\alpha G_\beta(u), v\right) \\ &= \mathcal{E}_\alpha\left(G_\beta(u), v\right) - (\alpha-\beta)\langle G_\beta(u), v\rangle\end{aligned}$$

   by our definition of $G_\alpha$. And now $\mathcal{E}_\alpha$ and $\mathcal{E}_\beta$ differ exactly by the latter term, so this expression simplifies to

   $$\mathcal{E}_\beta(\langle G_\beta(u), v\rangle) = \langle u, v\rangle = \boxed{\mathcal{E}_\alpha(G_\alpha(u), v)}.$$

   Equating the boxed expressions and noting that this holds for all $v$ in a Hilbert space, we must have that $G_\beta(u) - (\alpha-\beta)G_\alpha G_\beta(u) = G_\alpha(u)$, which is what we want.

3. Next, to check contractivity, we write

   $$\boxed{||\alpha G_\alpha(u)|| \cdot ||G_\alpha(u)||} = \langle \alpha G_\alpha(u), G_\alpha(u)\rangle \leq \mathcal{E}_\alpha(G_\alpha(u), G_\alpha(u))$$

because the inner product is only one of the terms of $\mathcal{E}_\alpha$. But now by definition of $G_\alpha$ this is $\langle u, G_\alpha(u)\rangle \leq \boxed{||u|| \cdot ||G_\alpha(u)||}$, and so the boxed parts tell us that $||\alpha G_\alpha(u)|| \leq ||u||$ as desired.

4. Finally, strong continuity is a "long but easy computation." We wish to show that $\alpha G_\alpha u \to u$; we have

$$\begin{aligned}\alpha||\alpha G_\alpha(u) - u||^2 &= \langle \alpha G_\alpha(u) - u, \alpha G_\alpha(u) - u\rangle \\ &\leq \mathcal{E}_\alpha(\alpha G_\alpha(u) - u, \alpha G_\alpha(u) - u) \\ &= \alpha^2\langle u, G_\alpha u\rangle - 2\alpha\langle u, u\rangle + \mathcal{E}_\alpha(u, u)\rangle \\ &= \alpha^2\langle u, G_\alpha(u)\rangle + \mathcal{E}(u, u).\end{aligned}$$

(We used the definition of $G_\alpha$ to simplify the first two terms in the expansion of $\mathcal{E}_\alpha$.) By Cauchy-Schwarz and the fact that $\alpha G_\alpha$ is a contraction, this is bounded from above by $\alpha\langle u, u\rangle - \alpha\langle u, u\rangle + \mathcal{E}(u, u)$. But this means

$$\alpha||\alpha G_\alpha(u) - u||^2 \leq \mathcal{E}(u, u)$$

and the right-hand side does not depend on $\alpha$, so indeed as $\alpha \to \infty$ we must have $||\alpha G_\alpha(u) - u|| \to 0$ as desired.

So $\{G_\alpha\}$ is a resolvent, and we can let $A$ be its generator. Let $\mathcal{E}'$ be the closed quadratic form corresponding to $A$; this means that

$$\mathcal{E}'(u, v) = \left\langle \sqrt{-A}u, \sqrt{-A}v \right\rangle$$

for all $u, v \in \mathcal{D}(\sqrt{-A})$. We can then define $\mathcal{E}'_\alpha$ by adding an $\alpha\langle\cdot,\cdot\rangle$ term to $\mathcal{E}'$; we find that

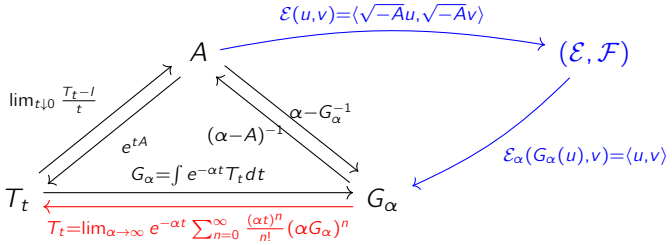
$$\mathcal{E}'_\alpha(G_\alpha(u), G_\alpha(v)) = \left\langle \sqrt{-A}(\alpha - A)^{-1}u, \sqrt{-A}(\alpha - A)^{-1}v \right\rangle + \alpha\langle u, v\rangle.$$

Moving the $\sqrt{-A}$ from one argument to the other and then using the definition of $G_\alpha$ again reduces this to $\langle G_\alpha(u), v\rangle = \mathcal{E}_\alpha(G_\alpha(u), G_\alpha(v))$. Since this was true for all $u, v \in \mathcal{H}$ and $G_\alpha$ has dense range, this means $\mathcal{E}$ and $\mathcal{E}'$ coincide on a dense set and thus agree everywhere. □

Next time, we'll see when a generator corresponds to a Markov process, and that will lead us to the theory of Dirichlet forms.

# 4 June 6, 2025

To start, let's look at the commutative diagram from last time (now complete with more arrows) and fill in a few remaining parts:




First, let's give a more probabilistic heuristic for why the inverse Laplace transform (marked in red) should be true. The expression

$$\alpha G_\alpha = \int \alpha e^{-\alpha t} T_t dt$$

involves the density of an exponential random variable, and thus we can think of this as an expected semigroup $\mathbb{E}[T_{\xi_\alpha}]$ for $\xi_\alpha$ distributed as $\text{Exp}(\alpha)$. So we can think of the expression in red as sampling $N$ from $\text{Poisson}(\alpha t)$ and finding the expectation

$$\mathbb{E}\left[T_{\xi_1+\cdots+\xi_N}\right], \text{ where } \xi_i \sim \text{Exp}(\alpha).$$

But we expect a sum of such independent exponentials $\xi_1 + \cdots + \xi_N$ to converge to $t$ for large $\alpha$, and so indeed this will converge to the semigroup $T_t$. (And with this perspective, we can in fact also use properties about exponential random variables to use this to prove the resolvent identity.)

**Proposition 38**

As an exercise, we can use the spectral theorem to show that we can go from the semigroup to the quadratic form by setting

$$\mathcal{E}(f,f) = \lim_{t \downarrow 0} \frac{1}{t} \langle (I - T_t)f, f \rangle,$$

where the expression inside the limit is nonincreasing in $t$ and thus the limit will exist. We'll thus define $\mathcal{E}$ to have domain the set of all $f \in \mathcal{F}$ where this limit is finite, and then this defines the quadratic form by the polarization identity.

**Proposition 39**

Similarly, starting from the resolvent, we can get an expression for the quadratic form by considering the map $\alpha \mapsto \alpha\langle (I - \alpha G_\alpha)f, f\rangle$. This expression is nondecreasing in $\alpha$ (by thinking of $I - \alpha G_\alpha$ as a multiplication operator with entries $1 - \frac{\alpha}{\alpha - \lambda(\cdot)}$), and thus we can define

$$\mathcal{E}(f,f) = \lim_{\alpha \to \infty} \alpha\langle (I - \alpha G_\alpha)f, f\rangle$$

where again $f$ is the set of all vectors where this limit is finite.

The notion of semigroups we've introduced so far is far too general, though, and we want to restrict ourselves to semigroups that are actually associated with Markov processes. That's what we'll do today – suppose we want to have a semigroup $P_t$ which takes the form

$$P_t f(x) = \mathbb{E}_x[f(Y_t)]$$

for some Markov process $Y_t$. Notice that this implies that nonnegative functions should be sent to nonnegative functions and that bounded functions are sent to bounded functions:

**Definition 40**

A bounded linear map $T : L^2(X,m) \to L^2(X,m)$ is **Markovian** if the following is true: for any function $f$ satisfying $0 \le f \le 1$ $m$-almost everywhere, we also have $0 \le T(f) \le 1$ $m$-almost everywhere. A semigroup $\{T_t\}$ is Markovian if each $T_t$ satisfies this property, and similarly a resolvent $\{G_\alpha\}$ is Markovian if each $\alpha G_\alpha$ satisfies this property.

It's easy to check using the Laplace transform formulas that for an associated semigroup and resolvent, one is Markovian if and only if the other one is. And so now we want to ask whether we can make a similar definition for the other objects.

**Definition 41**

Let $(X, \mathcal{M}, m)$ be a $\sigma$-finite measure space. A **Dirichlet form** on $L^2(X, m)$ is a closed quadratic form $\mathcal{E} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ satisfying the following Markov property. Let $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. Then for any function $f \in \mathcal{F}$, we also have $0 \vee (f \wedge 1) \in \mathcal{F}$, and

$$\mathcal{E}(0 \vee (f \wedge 1), 0 \vee (f \wedge 1)) \leq \mathcal{E}(f, f).$$

This definition (due to Beurling and Deny) may be a bit hard to immediately interpret, but it turns out to be equivalent to the ones we saw before under our correspondence. To prove this, we'll need a technical lemma. Since we're working in a $\sigma$-finite measure space, it'll typically suffices to consider simple functions:

**Lemma 42**

Let $P : L^2(X, m) \to L^2(X, m)$ be a bounded, linear, $m$-symmetric Markovian operator, and let $F$ be a 1-Lipschitz function $\mathbb{R} \to \mathbb{R}$ (in other words, a contraction) such that $F(0) = 0$. Then we have the following:

1. For any $n \in \mathbb{N}$ and any pairwise disjoint measurable sets $A_1, \cdots, A_n$ in $X$ such that $m(A_i)$ are all finite, and for all $a_1, \cdots, a_n \in \mathbb{R}$, define the function $f = \sum_{i=1}^n a_i 1\{A_i\}$. Then

$$\langle (I - P)f, f \rangle = \sum_{i=1}^n \mu_i a_i^2 + \frac{1}{2} \sum_{1 \leq i < j \leq n} \alpha_{ij}(a_i - a_j)^2$$

   for nonnegative coefficients $\alpha_{ij} = \langle 1\{A_i\}, P1\{A_j\} \rangle$ and $\mu_i = m(A_i) - \sum_{k=1}^n \alpha_{i,k}$.

2. For any $g \in L^2(X, m)$, $\tilde{g} = F(g)$ satisfies the contraction property, meaning that

$$\langle (I - P)\tilde{g}, \tilde{g} \rangle_{L^2} \leq \langle (I - P)f, f \rangle_{L^2}.$$

In particular, $F(t) = 0 \vee (t \wedge 1)$ satisfies this property in the statement, and the idea is that we'll prove the Markov property for a Dirichlet form by first doing so for an expression of the form $\langle (I - P)f, f \rangle$, then taking the limit in Proposition 38 or Proposition 39.

*Proof.* First note that (1) immediately implies (2). Indeed, for the case where $g$ is a simple function, $\tilde{g}$ is also a simple function with values on its intervals contracted toward the origin, so we see that every term on the right-hand side of the equation in (1) gets smaller (since $a_i^2$ and $(a_i - a_j)^2$ are both nonincreasing). Then we can use density of simple functions to get the result for all of $L^2$. (Indeed in general $\tilde{g}$ is in $L^2$ if $g$ is in $L^2$ because $|\tilde{g}(t)| \leq |g(t)|$ pointwise.)

Thus, we just need to do the computation and prove (1). We have

$$\begin{aligned}
\langle (I - P)f, f \rangle &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle (I - P)1\{A_i\}, 1\{A_j\} \rangle \\
&= \sum_{i=1}^n a_i^2 \mu(A_i) - \sum_{i=1}^n \sum_{j=1}^n a_i a_j \langle 1\{A_i\}, P1\{A_j\} \rangle \\
&= \sum_{i=1}^n a_i^2 \mu(A_i) - \sum_{i=1}^n \sum_{j=1}^n a_i a_j \alpha_{ij}
\end{aligned}$$

by separating out the $I$ and $P$ parts and noting that the $A_i$s are disjoint. We now just need to do a bit of rearrangement: this simplifies to

$$\sum_{i=1}^n a_i^2 \mu_i + \sum_{i=1}^n \sum_{i=1}^n (a_i^2 \alpha_{ij} - a_i a_j \alpha_{ij}),$$

and now using the symmetry $\alpha_{ij} = \alpha_{ji}$ we can sum over only half the terms in the latter term; we end up summing over $i < j$ of $(a_i^2 + a_j^2 - 2a_i a_j)\alpha_{ij}$, and that is exactly what we want. What remains to be shown is that each $\mu_i$ is positive – indeed,

$$\mu_i = m(A_i) - \sum_k \alpha_{i,k} = \langle 1\{A_i\}, 1\{A_i\}\rangle - \left\langle 1\{A_i\}, P\sum_{k=1}^N 1\{A_k\}\right\rangle.$$

But the function $\sum_{k=1}^N 1\{A_k\}$ is bounded between 0 and 1, so the same is true for $P\sum_{k=1}^N 1\{A_k\}$; in particular

$$\mu_i \geq \langle 1_{A_i}, 1_{A_i}\rangle - \langle 1_{A_i}, 1_X\rangle \geq 0,$$

so we have nonnegativity as desired. □

**Theorem 43**

Let $(\mathcal{E}, \mathcal{F})$ be a closed quadratic form on $L^2(X, m)$ for some $\sigma$-finite measure space $(X, m)$. Let $\{T_t\}$, $\{G_\alpha\}$, and $A : \mathcal{D}(A) \to L^2(X, m)$ be the corresponding semigroup, resolvent, and generator. Then the following are equivalent:

1. The semigroup $\{T_t\}$ is Markovian,
2. The resolvent $\{G_\alpha\}$ is Markovian (that is, $\alpha G_\alpha$ is Markovian for all $\alpha > 0$),
3. For all functions $f \in \mathcal{D}(A)$, we have $\langle A(f), (f-1)\vee 0\rangle \leq 0$,
4. The quadratic form is a Dirichlet form (that is, for all $u \in \mathcal{F}$, $\tilde{u} = 0 \vee u \wedge 1$ is also in $\mathcal{F}$ and $\mathcal{E}(\tilde{u}, \tilde{u}) = \mathcal{E}(u, u)$).

(It turns out that Markovian semigroups will correspond to Markov processes, but we'll need some additional conditions and we'll see that next time.) To prove this correspondence, we'll need another technical result (which is left as an exercise):

**Lemma 44**

Let $T : L^2(X, m) \to L^2(X, m)$ be a bounded linear operator. Then the following two conditions are equivalent:

1. If $0 \leq f \leq 1$ $m$-almost-everywhere, then $0 \leq Tf \leq 1$ $m$-almost-everywhere.
2. If $f \leq 1$ $m$-almost-everywhere, then $T(f) \leq 1$ $m$-almost-everywhere.

The main idea is that we can get a two-sided bound by considering $-f$ instead of $f$, and often it's more convenient to use the latter one-sided condition.

*Part of proof.* First of all, the equivalence of (1) and (2) follows from looking at the the Laplace and inverse Laplace transforms. To show that (1) or (2) imply (4), we're considering quadratic forms of the type

$$\mathcal{E}(f, f) = \lim_{t\downarrow 0} \frac{1}{t}\langle (I - T_t)f, f\rangle.$$

We know that $\frac{1}{t}\langle (I - T_t)f, f\rangle$ satisfy the Markov property for quadratic forms by our previous lemma; in fact, we get something stronger, because we know that for any 1-Lipschitz $F$ with $F(0) = 0$ and any $f \in \mathcal{F}$, we have $F \circ f \in \mathcal{F}$ and $\mathcal{E}(F \circ f, F \circ f) \leq \mathcal{E}(f, f)$. So we just need to plug in our particular choice of function $F(t) = 0 \vee (t \wedge 1)$ and then take the limit $t \downarrow 0$ (or in the case of resolvent use the other approximation).

For (1) or (2) implies (3), we instead apply the inequality to the function $F(t) = t \wedge 1$. This implies that $f \wedge 1 \in \mathcal{F}$ and

$$\mathcal{E}(f \wedge 1, f \wedge 1) \leq \mathcal{E}(f, f),$$

and now if we write $f$ as a sum of two functions $(f-1)_+ + (f \wedge 1)$ (indeed $(f-1)_+$ is also in the domain of the form), we can substitute in $f - (f-1)_+$ into the left-hand side to get

$$-2\mathcal{E}(f, (f-1)_+) \leq \mathcal{E}((f-1)_+, (f-1)_+) \leq 0.$$

So for any $f \in \mathcal{D}(A)$, we have

$$-\mathcal{E}(f, (f-1)_+) = -\langle \sqrt{-A}f, \sqrt{-A}(f-1)_+ \rangle = \langle Af, (f-1)_+ \rangle,$$

which we've indeed shown is nonpositive. (Here we used that it's okay to move $\sqrt{-A}$ from one side to the other, because if $f \in \mathcal{D}(A)$, then $\sqrt{-A}f \in \mathcal{D}(\sqrt{-A})$.

Out of the remaining implications, it's most important to get from (4) to (1) and (2). To get to the resolvent, recall that we already have the characterization $\mathcal{E}_\alpha(G_\alpha(u), v) = \langle u, v \rangle_{L^2}$ for all $u \in L^2(X, m)$ and all $v \in \mathcal{F}$ (recall that $\mathcal{E}_\alpha = \mathcal{E} + \alpha \langle \cdot, \cdot \rangle_{L^2}$). We'll state another characterization now: consider the function $\psi : \mathcal{F} \to \mathbb{R}$ defined by

$$\psi(w) = \mathcal{E}(w, w) + \alpha \langle w - \alpha^{-1}u, w - \alpha^{-1}u \rangle.$$

This expression turns out to have a unique minimizer $\psi_u$, and we **claim that** $\psi_u = G_\alpha(u)$. Indeed, it's enough to show that

$$\psi(w) - \psi(G_\alpha(u)) = \mathcal{E}_\alpha(G_\alpha(u) - w, G_\alpha(u) - w) \geq 0$$

by some kind of completing-the-square argument (or equivalently solving some kind of Euler-Lagrange equation to show that we end up characterizing with the same equation for $G_\alpha$ as above). Thus for (4) implies (2), we need to show that if $0 \leq u \leq 1$ $m$-almost-everywhere, then $0 \leq \alpha G_\alpha(u) \leq 1$ $m$-almost-everywhere, which is equivalent to saying that truncating at 0 and 1 does not change the function. That's equivalent to the statement that $v = 0 \vee (G_\alpha(u) \wedge \alpha^{-1})$ is exactly equal to $G_\alpha$. Since $0 \leq u \leq 1$ everywhere and $G_\alpha$ minimizes the given functional $\psi$, it's enough to show that $\psi(v) \leq \psi(G_\alpha(u))$ (because then $v$ is also the minimizer). We'll worry about one term at a time: for the first term, clearly $\mathcal{E}(v, v) \leq \mathcal{E}(G_\alpha(u), G_\alpha(u))$ by assumption (4) (since we can use bilinearity of the quadratic form to scale $[0, 1]$ to any interval $[0, \alpha^{-1}]$). And for the second term, we know that $0 \leq \alpha^{-1}u \leq \alpha^{-1}$, and therefore we also have pointwise that

$$|v - \alpha^{-1}(u)| \leq |G_\alpha(u) - \alpha^{-1}(u)| \quad m\text{-almost everywhere}$$

(because $\alpha^{-1}u$ is inside the interval of truncation, so when we truncate $G_\alpha(u)$ to $v$ that either does nothing or gets us closer). Thus the inner product with $v$ is also smaller than with $G_\alpha(u)$, as desired. (And then the inverse Laplace transform gets us to the semigroup equivalence as well.) □

Later on we'll want to show existence of the heat kernel, and it'll be useful to understand operators not just in $L^1$ but also in a general $L^p$.

- By the Markov property of a semigroup, we know that

$$f \in L^2(X, m) \cap L^\infty(X, m) \implies ||T_t(f)||_\infty \leq ||f||_\infty,$$

and we've also assumed by definition that $T_t$ is also a contraction in $L^2$. So now we'll take our $\sigma$-finite measure space and approximate it by a countable union of increasing sets $X = \bigcup A_n$, where $A_n \uparrow X$ but $m(A_n) < \infty$ for

all $n$. Then for any $f \in L^\infty$, the function $f1\{A_n\}$ is always in $L^2 \cap L^\infty$, and so it makes sense to define

$$T_t f = \lim_{n\to\infty} T_t(f1\{A_n\})$$

for $f \in L^\infty(X, m)$ nonnegative $m$-almost everywhere. (This definition turns out to be independent of the choice of $A_n$ that we make, for example by taking a common refinement.) From there, we can then split up a general $f$ into $f_+$ and $f_-$ and define $T_t f = T_t(f_+) - T_t(f_-)$; taking limits of the property above shows that $T_t$ is **also a contraction on** $L^\infty$.

- Now note that $L^1 \cap L^\infty$ is dense in $L^1$, so we claim that $T_t : L^1(X, m) \to L^1(X, m)$ will be a contraction for all $t > 0$ as well. Indeed, by density it suffices to prove control on the $L^1$ norm for any $f \in L^1 \cap L^\infty$. But by duality between $L^1$ and $L^\infty$,

$$||T_t f||_1 = \sup_{g\in L^\infty\cap L^2, ||g||_\infty=1} \int T_t(f)g\, dm,$$

where the key is that by approximating any function in $L^\infty$ by functions of our type we can now move the $T_t$ from one side to the other (because we're in $L^2$):

$$||T_t f||_1 = \sup_{g\in L^\infty\cap L^2, ||g||_\infty=1} \int fT_t(g)\, dm \le ||f||_1,$$

since $||T_t g||_\infty \le 1$. Being a contraction on a dense subspace implies that we're a contraction on the whole space.

- Finally, from here we can use Riesz-Thorin interpolation and duality to get the same contraction result for the semigroup for any general $p \in [1, \infty]$ and any $t > 0$.

# 5 June 9, 2025

We'll begin by clarifying a few points from the end of last lecture. First of all, here is a useful counterexample to keep in mind:

**Example 45**

Let $X = \{1, 2\}$ with $m$ the counting measure, and define the quadratic form

$$\mathcal{E}(f, f) = (f(1) - f(2))^2 + f(1)^2 + \lambda(f(1)^2 + f(2)^2)$$

for some parameter $\lambda$.

We can check the following facts by computing eigenvalues of a $2 \times 2$ matrix:

- $\mathcal{E}$ is a nonnegative-definite, closed quadratic form on $L^2(X, m)$ if and only if $\lambda \ge \frac{\sqrt{5}-3}{2}$.
- But $\mathcal{E}$ is a **Markovian** such form if and only if $\lambda \ge 0$. Thus there's some range of values $\lambda$ for which we are not Markovian.

We typically mostly talk about defining the semigroup on $L^2, L^1, L^\infty$ and not so much on the intermediate $p$s, but let's work through the reasoning in more detail. For any $f \in L^1 \cap L^2$, we have (all statements $m$-almost-everywhere)

$$-|f| \le f \le |F| \implies -T_t(|f|) \le T_t(f) \le T_t(|f|) \implies |T_t(f)| \le T_t(|f|).$$

Thus for any sequence of subsets $A_n \uparrow X$ with $m(A_n) < \infty$, we show contraction in $L^1$ via

$$||T_t(f)|| = \lim_{n\to\infty} \int 1\{A_n\}|T_t(f)|dm \leq \lim_{N\to\infty} 1\{A_n\}T_t|f|dm$$

(using the triangle inequality), and now interchanging the limit and the integral shows that this is equal to

$$\lim_{n\to\infty} T_t(1\{A_n\})|f|dm \leq \int |f|dm,$$

where we've used that $1\{A_n\}$ always takes on values in $[0,1]$ and thus so must $T_t(1\{A_n\})$. So we are a contraction on $L^1$, and since we also know that we are a contraction on $L^2$ we get a contraction for all $p \in (1,2)$ as well. But for our purposes going forward, what we've shown is already good enough.

We'll now talk about the **construction of Markov processes from the semigroup**. There are basically two issues that might come up:

1. Letting $\mathbf{1}$ be the constant function, we may have that $T_t\mathbf{1}(x) < 1$, meaning that the Markov process started at $x$ yields a sub-probability measure instead of a probability measure. In such a case, we need to add an extra absorbing state called the **cemetery**, such that once the Markov process reaches that state, it stays there forever, and the mass is exactly the amount needed to give us a probability measure.

2. Note however that since we're only defining the semigroup on $L^2$, the left-hand side $T_t\mathbf{1}(x)$ (and functions in general) is not even pointwise defined, so we need to be careful. We'll need to deal with a finer notion of continuity, and to define a process associated with such a semigroup (just like in the space of $L^p$ spaces) we need to throw away some "small sets" to make our construction unique.

We want to get something which is right-continuous with left limits, and we'll have to make some topological assumptions on $X$ rather than just having a measure space.

**Definition 46**

Let $(X, \mathcal{M}, m)$ be a $\sigma$-finite measure space and let $(\mathcal{E}, \mathcal{F})$ be a Dirichlet form on $L^2(X, m)$. We call $(\mathcal{E}, \mathcal{F})$ a **regular** Dirichlet form if it satisfies the following conditions:

1. $X$ is a locally compact, separable, metrizable space (that is, $X$ is **Polish**) with $\mathcal{M}$ its Borel $\sigma$-algebra and $m$ a **Radon measure** with full support (that is, a Borel measure which is finite on compact sets and nonzero on all nonempty open subsets).

2. "There are lots of continuous functions," meaning that the vector space $\mathcal{F} \cap C_c(X)$ is dense in $(\mathcal{F}, \mathcal{E}_1)$ (which we know to be a Hilbert space) and **also** dense in $(C_c(X), ||\cdot||_{\sup})$. This is for example a property of Sobolev spaces.

We're now ready to state a fundamental basic result in this area:

**Theorem 47** (Fukushima '71)

Let $(\mathcal{E}, \mathcal{F})$ be a regular Dirichlet form on $L^2(X, m)$ and let $\mathcal{B}$ be the Borel $\sigma$-algebra on $X$. Let $X_\Delta = X \cup \{\Delta\}$ be the one-point compactification of $X$ (so if $X$ is compact this adds an isolated point) and define $\mathcal{B}_\Delta = \mathcal{B} \cup \{B \cup \{\Delta\} : B \in \mathcal{B}\}$, a $\sigma$-field on $X_\Delta$. Then for all $x \in X_\Delta$, there is an $X_\Delta$-valued stochastic process $\left(\Omega, (\mathcal{F}_t)_{t\geq 0}, (Y_t)_{t\in[0,\infty]}, (\mathbb{P}_x)_{x\in X_\Delta}\right)$, which satisfies the following properties:

- The filtration $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous (to ensure that certain hitting times are stopping times), and the process $(Y_t)_{t\geq 0}$ is adapted to the filtration.
- For all $E \in \mathcal{B}_\Delta$ and all $t > 0$, the function $x \mapsto \mathbb{P}_x(Y_t \in E)$ is a measurable function of $x$ on $(X_\Delta, \mathcal{B}_\Delta)$.
- (Markov property) For all $x \in X$, $E \in \mathcal{B}_\Delta$, and $t, s \geq 0$,
$$\mathbb{P}_x\left(Y_{t+s} \in E | \mathcal{F}_t\right) = \mathbb{P}_{Y_t}(Y_s \in E).$$
- (Normality) $\mathbb{P}_x(Y_0 = x) = 1$ for all $x$, meaning that $\mathbb{P}_x$ indeed "starts our process at $x$."
- (Cemetery is absorbing) We have $Y_\infty(\omega) = \Delta$ for all $\omega \in \Omega$, and we have $\mathbb{P}_x(Y_t(\omega) = \Delta) = 1$ for all $t \geq \zeta(w)$, where $\zeta(w) = \inf\{t \geq 0 : Y_t(\omega) = \Delta\}$ is called the **lifetime** of the process (and can be infinite).
- (Càdlàg paths) The map $t \mapsto Y_t(\omega)$ is right-continuous on $[0, \infty)$ and has left limits on $(0, \infty)$.
- (Strong Markov property) For any probability measure $\mu$ on $X_\Delta$ and for any stopping time $T$, we have for all $s \geq 0$ that
$$\mathbb{P}_\mu\left(Y_{T+s} | \mathcal{F}_T\right) = \mathbb{P}_{Y_T}(Y_s \in E),$$
where $\mathbb{P}_\mu(\cdot) = \int \mathbb{P}_x(\cdot)\mu(dx)$ means that the distribution of the starting state is given by $\mu$. (Every deterministic time is a stopping time, so this is strictly stronger than the Markov property.)

Furthermore, we assume that our process has a law which is absolutely continuous with respect to the reference measure, then we can also show that (via some version of the strong Feller property) we do have uniqueness. Otherwise (see proof ideas below) we need to make some modifications.

Note that it's okay to set $Y_\infty(\omega) = \Delta$ even if we have a recurrent process which never actually gets absorbed – it's not that we're taking the limit of $Y_t(\omega)$ as $t \to \infty$. Indeed, we're not requiring left-continuity at infinity here with the càdlàg condition. And the point is that we need regularity to make this construction work.

We won't do the full proof here, but we'll give a sense of why it's true and why regularity is required. This has to do with the fact that since we can approximate every function in the domain of our form $\mathcal{F}$ with continuous functions, the semigroup maps any function in $L^2$ to $\mathcal{F}$ (this is a consequence of the spectral theorem) – the result is **quasi-continuous** rather than continuous, so we're almost pointwise defined. Then positivity allows us to represent something like Riesz representation for nonnegative functions (outside a negligible set). And the price we have to pay for applying Hilbert space methods is the loss of uniqueness.

To explain further, let $\left(\Omega, (\mathcal{F}_t)_{t\geq 0}, (Y_t)_{t\in[0,\infty]}, (\mathbb{P}_x)_{x\in X_\Delta}\right)$ be as above. We call a subset $\mathcal{N} \subset X$ **properly exceptional** for $\{Y_t\}_{t\in[0,\infty]}$ if $\mathcal{N}$ is a Borel set with $m(\mathcal{N}) = 0$ which is **also** negligible from the point of view of the process, meaning that
$$\mathbb{P}_x\left(\{\omega \in \Omega | \text{ for all } t, Y_t(\omega) \in X_\Delta \setminus \mathcal{N} \text{ and } Y_{t^-}(\omega) \in X_\Delta \setminus \mathcal{N}\}\right) = 1.$$
In other words, if we start outside $\mathcal{N}$, we almost surely never see $\mathcal{N}$, so it shouldn't affect the semigroup and Dirichlet form. We then have uniqueness in the following sense: for any two processes $(Y_t)$ and $(\tilde{Y}_t)$ corresponding to the

same regular Dirichlet form, there is a **common** properly exceptional set $\mathcal{N}$ (meaning that it is exceptional for both processes) such that $\mathbb{P}_x(Y_t \in \cdot)$ and $\tilde{\mathbb{P}}_x(\tilde{Y}_t \in \cdot)$ for all $t \geq 0$ and all $x \in X \setminus \mathcal{N}$. Here to be completely precise, we say that $(Y_t)$ corresponds to the Dirichlet form $(\mathcal{E}, \mathcal{F})$ if the semigroup $P_t^Y$ defined by

$$P_t^Y f(x) = \mathbb{E}_x \left[ f(Y_t) 1\{t < \zeta\} \right]$$

is exactly equal to $T_t$, the Markovian semigroup associated with the Dirichlet form in the way discussed last lecture. (The reason for the indicator is $1\{t < \zeta\}$ is that we don't know the value of $f$ on the cemetery state; this is equivalent to just extending the function $f$ via $f(\Delta) = 0$.) The following result is important for the proof, which involves modifications of quasi-continuous functions:

**Definition 48**

The 1-**capacity** of a set $A \subset X$ is defined by

$$\text{Cap}_1(A) = \inf \{\mathcal{E}_1(f, f) | f \in \mathcal{F}, f \geq 1 \ m\text{-almost-everywhere on a neighborhood of } A\}.$$

We call $\mathcal{N}$ $\mathcal{E}$-**polar** if $\text{Cap}_1(A) = 0$.

In particular, just the $L^2$ inner product term in $\mathcal{E}_1(f, f)$ ensures that $\text{Cap}_1(A) \geq m(A)$ for all Borel sets, since $\langle f, f \rangle \geq m(A) \implies \mathcal{E}_1(f, f) \geq m(A)$. We might ask whether this is equivalent to the notion above, and we'll get to that soon.

**Example 49**

An example of a set with zero measure but nonzero capacity would be a line segment from $(0, 0)$ to $(1, 0)$ in $\mathbb{R}^2$, which has zero Lebesgue measure but positive capacity with respect to the usual Dirichlet form $\mathcal{E}(f, f) = \frac{1}{2} \int |\nabla f|^2 dm$.

**Definition 50**

We say that a statement $S(x)$ (dependent on $x \in X$) holds **quasi-everywhere** (sometimes written **q.e.**) on a subset $A \subset X$ if it holds for all $A \setminus \mathcal{N}$ for some $\mathcal{E}$-polar set $\mathcal{N}$. A function $f \in L^2(X, m)$ is **quasi-continuous** if, taking any representative of the equivalence class so that we are pointwise defined, there exists $\tilde{f} : X \to \mathbb{R}$ with $\tilde{f} = f$ $m$-almost everywhere, such that for all $\varepsilon > 0$ there is some open set $\mathcal{G}$ with $\text{Cap}_1(\mathcal{G}) < \varepsilon$, and $\tilde{f}$ is continuous on the closed set $X \setminus \mathcal{G}$.

Do note however that we cannot actually get continuity outside an $\mathcal{E}$-polar set – we have to take this "nest" of approximations with smaller and smaller capacity, and that's one of the annoying aspects of this theory. This notion of almost-continuity is exactly what's used in the proof of Fukushima's theorem:

**Theorem 51**

Let $(\mathcal{E}, \mathcal{F})$ be a regular Dirichlet form on $L^2(X, m)$. Then every function $f \in \mathcal{F}$ admits a quasi-continuous modification $\tilde{f} : X \to \mathbb{R}$ (in particular, this implies $\tilde{f} = f$ $m$-almost-everywhere), and any two quasi-continuous modifications are equal quasi-almost-everywhere.

We can now discuss the relationship between our $\mathcal{E}$-polar and properly exceptional sets $\mathcal{N}$. It turns out they are more or less the same thing:

**Fact 52**

Any properly exceptional set is $\mathcal{E}$-polar, and any $\mathcal{E}$-polar set is contained in some properly exceptional set.

We'll state some more results without proofs in the remainder of this lecture and discuss how to think about them. Many properties of our processes can be read off from properties of the Dirichlet form (which is useful because it is much more infinitesimal than the semigroup), and we also have some control over the sample paths (which we know are càdlàg, but perhaps we want a **diffusion** where the paths are actually continuous):

**Definition 53**

Let $(\mathcal{E}, \mathcal{F})$ be a regular Dirichlet form on $L^2(X, m)$. We say that $(\mathcal{E}, \mathcal{F})$ is **strongly local** if for all $f, g \in \mathcal{F}$ where

- $\text{supp}_m(f), \text{supp}_m(g)$ are compact (here $\text{supp}_m(f) = \text{supp}(f \cdot m)$), and
- there exists some real number $a \in \mathbb{R}$ such that $\text{supp}_m(f - a) \cap \text{supp}_m(g) = \varnothing$,

we have $\mathcal{E}(f, g) = 0$. Similarly, we say that $\mathcal{E}(f, g)$ is **local** if whenever $\text{supp}_m(f), \text{supp}_m(g)$ are compact and $\text{supp}_m(f) \cap \text{supp}_m(g) = \varnothing$, we have $\mathcal{E}(f, g) + 0$.

**Example 54**

Let $\mathcal{E}(f, g) = \frac{1}{2} \int \nabla f \cdot \nabla g dm$ be the usual Dirichlet form. This is strongly local because whenever $f$ is constant (equal to 1) on each region where $g$ is supported, the integral evaluates to 0. On the other hand, $\mathcal{E}(f, g) = \langle f, g \rangle_{L^2(m)}$ is local but not strongly local.

**Proposition 55**

If $(\mathcal{E}, \mathcal{F})$ is strongly local, then $\{Y_t\}$ satisfies that – here consider $t \in [0, \infty)$ –

$$\mathbb{P}_x \left( t \mapsto Y_t(\omega) \in X_\Delta \text{ is continuous} \right) = 1.$$

Meanwhile if $(\mathcal{E}, \mathcal{F})$ is only local, then we instead have that statement but only considering $t \in [0, \zeta)$.

From a probabilistic point of view, the latter implication is saying that we're allowed to make jumps but only to the cemetery state. If we're familiar with Lévy processes, we might have seen a decomposition of those processes into simpler parts, and there is also a nice one here for Dirichlet forms:

**Theorem 56** (Beurling-Deny decomposition)

For any regular Dirichlet form $(\mathcal{E}, \mathcal{F})$ on $L^2(X, m)$, we have a unique triple $(\mathcal{E}^{(c)}, J, \kappa)$ (these three parts correspond to "strongly local," "jump," and "killing") where $\mathcal{E}^{(c)}$ is a strongly local nonnegative definite bilinear form on $\mathcal{F} \times \mathcal{F}$, $J$ is a symmetric Radon measure on $X \times X \setminus \text{diag}_X$ (symmetric meaning that the measure is invariant under swapping the two coordinates), and $\kappa$ is a Radon measure on $X$, such that

$$J(X \times \mathcal{N} \setminus \text{diag}_X) = 0, \quad \kappa(\mathcal{N}) = 0$$

for any $\mathcal{E}$-polar set $\mathcal{N}$, and we can write down our Dirichlet form as a sum of three parts

$$\mathcal{E}(u, v) = \mathcal{E}^{(c)}(u, v) + \iint_{X \times X \setminus \text{diag}_X} (\tilde{u}(x) - \tilde{u}(y))(\tilde{v}(x) - \tilde{v}(y)) J(dxdy) + \int_X \tilde{u}(x)\tilde{v}(x)\kappa(dx),$$

where $\tilde{u}, \tilde{v}$ are the quasi-continuous versions of $u$ and $v$ (everything is well-defined here because of our assumptions on $J$ and $\kappa$).

We call $J$ the **jumping measure** and $\kappa$ the **killing measure**; we can in fact think of the last integral here as $(\tilde{u}(x) - \tilde{u}(\Delta))(\tilde{v}(x) - \tilde{v}(\Delta))$ to match with the $J$ integral, except that we can only jump from $x$ to $\Delta$ and not vice versa. So an interpretation at the probabilistic process can be made as follows: if we assume that $1_A$ is in $\mathcal{F}$ for any Borel set $A$ (this is true for example for finite state spaces), we can let $A, B$ be two disjoint Borel sets and compute that

$$J(A \times B) = -\mathcal{E}(1_A, 1_B)$$

(because we get contributions from $x \in A, y \in B$ or vice versa), and the correspondence with semigroups yields that this expression is exactly

$$-\lim_{t \downarrow 0} \frac{\langle (I - T_t)1_B, 1_A \rangle}{t} = \lim_{t \downarrow 0} \frac{\langle T_t 1_B, 1_A \rangle}{t} = \lim_{t \downarrow 0} \frac{1}{t} \int_A \mathbb{P}_x(Y_t \in B) m(dx).$$

So this indeed captures the jumping intensity from $A$ to $B$. And similarly taking $B$ to be the whole set $X$, we can derive a similar interpretation for the killing measure as the jumping intensity into the cemetery state.

# 6 June 10, 2025

Last time, we saw the decomposition of a regular Dirichlet form $\mathcal{E}$ into three parts $(\mathcal{E}^{(c)}, J, \kappa)$, a strongly local diffusion part, a jump part, and a killing (jump to the cemetery) part. We'll saw that under the assumption that indicators $1_A$ are always quasi-continuous (strictly speaking these aren't necessary because we can use approximations), the jump part is indeed giving the measure of the jump intensity by considering disjoint sets $A, B$ and letting $u = 1_A$ and $v = 1_B$. But now if we let $u = 1_A$ and $b$ be the constant function 1, we have

$$\mathcal{E}(1_A, 1_X) = \int 1_A \kappa(dx) = \kappa(A).$$

(Indeed, the diffusion term goes away by the strongly local property, and the jump term is zero because $\tilde{v}(x) - \tilde{v}(y) = 0$.) But now by the properties of the Dirichlet form we have that

$$\kappa(A) = \lim_{t \downarrow 0} \frac{1}{t} \langle 1_A, (I - T_t)1_X \rangle = \lim_{t \downarrow 0} \frac{1}{t} \int_A \mathbb{P}_x (Y_t = \Delta) \, dm(x),$$

and thus the killing measure is indeed the rate of jumps to the cemetery state.

It turns out that $\mathcal{E}^{(c)}$ is a bit harder to describe because we can have various (sometimes fractal) constructions of diffusions on various spaces, and we'll see some examples of those later on.

We'll next see how to write probabilistic properties in terms of the associated Dirichlet forms of the Markov processes.

**Definition 57**

Let $(\mathcal{E}, \mathcal{F})$ be a regular Dirichlet form. We say that a Borel set $A \subset X$ is **$\mathcal{E}$-invariant** if for any $f \in \mathcal{F}$, $1_A f \in \mathcal{F}$ (thus $1_{A^c} f \in \mathcal{F}$ as well by linearity), and

$$\mathcal{E}(f, f) = \mathcal{E}(1_A f, 1_A f) + \mathcal{E}(1_{A^c} f, 1_{A^c} f).$$

In other words, there is no interaction between energies in $A$ and $A^c$.

This property turns out to be equivalent to the following condition on the semigroup $(P_t)$:

$$P_t(1_A f) = 1_A P_t(f) \ m - \text{almost-everywhere for all } f \in L^2(X, m).$$

That is, the evolution of the semigroup is contained in $A$, and probabilistically if our process starts in $A$ (resp. $A^c$) it never leaves $A$ (resp. $A^c$). So this definition relates to the concept of **irreducibility** of a Markov process:

**Definition 58**

We say that $(\mathcal{E}, \mathcal{F})$ is **irreducible** if for every $\mathcal{E}$-invariant set, either $m(A) = 0$ or $m(A^c) = 0$.

**Definition 59**

Let $(\mathcal{E}, \mathcal{F})$ be a Dirichlet form on $L^2(X, m)$. The **extended Dirichlet space** $\mathcal{F}_e$ is defined as

$$\mathcal{F}_e = \left\{ f : f \text{ is } m\text{-a.e. the limit } f = \lim_{n \to \infty} f_n, \text{ where } f_n \in \mathcal{F} \text{ satisfy } \mathcal{E}(f_n - f_m, f_n - f_m) \overset{m,n \to \infty}{\to} 0 \right\}.$$

We then define $\mathcal{E}(f, f) = \lim_{N \to \infty} \mathcal{E}(f_n, f_n)$ for all $f$; this turns out to be well-defined.

So we're completing the space with respect to $\mathcal{E}$ rather than an inner product. This turns out to be related to **recurrence and transience** of the underlying process.

**Definition 60**

A Markov semigroup $(P_t)$ is **transient** if for any $m$-almost-everywhere-nonnegative function $f \in L^1(X, m) \cap L^\infty(X, m)$, the Green operator

$$Gf(x) = \lim_{N \to \infty} \int_0^N P_t f(x) dt$$

is finite $m$-almost-everywhere.

For example if $f$ is the indicator of some set of finite measure, we're saying that for a transient semigroup, the amount of time spent in that set is finite in expectation. We can generally think of this as a weighted occupation time.

**Definition 61**

A Dirichlet form $(\mathcal{E}, \mathcal{F})$ is **transient** if there is some function $m$-almost-everywhere-**positive** $f \in L^1(X, m) \cap L^\infty(X, m)$ such that

$$\int_X |u| f \, dm \leq \sqrt{\mathcal{E}(u, u)}$$

for all $u \in \mathcal{F}$.

**Theorem 62**

The following are equivalent:

1. $(P_t)$ is transient,
2. $(\mathcal{E}, \mathcal{F})$ is transient,
3. the extended Dirichlet space $(\mathcal{F}_c, \mathcal{E})$ is a Hilbert space (in particular, no nonzero functions have zero energy)

(2) implies (3) is easy to see, because by the definition of transience for $(\mathcal{E}, \mathcal{F})$ we have a Cauchy sequence for $f\,dm$, and we produce a limit via some subsequence. And for the converse, we show that the Green's function is in the domain of the form and use Cauchy-Schwarz and the fact that the Laplacian and Green's function are inverses. But we won't go into the proof much beyond that.

We have a similar condition for recurrence:

**Definition 63**

A semigroup $(P_t)$ is **recurrent** if for all nonnegative functions $f \in L^1(X, m) \cap L^\infty(X, m)$, the Green operator $Gf(x) = \lim_{N \to \infty} \int_0^N P_t f \, dt$ takes on value either 0 or $\infty$ $m$-almost-everywhere. A Dirichlet form $(\mathcal{E}, \mathcal{F})$ is **recurrent** if $1_X \in \mathcal{F}$ and $\mathcal{E}(1_X, 1_X) = 0$.

**Theorem 64**

$(P_t)$ is recurrent if and only if $(\mathcal{E}, \mathcal{F})$ is recurrent.

In fact, an irreducible Markov process is always recurrent or transient.

**Example 65**

Consider the measure $m(dx) = x^{n-1} dx$ on $(0, \infty)$ (the symmetric measure for the $n$-dimensional Bessel process). The Dirichlet form for that process then takes the form $\mathcal{E}(f, g) = \int_X f'(x) g'(x) m dx$ on $L^2(X, m)$. The Bessel process is recurrent for $n = 1, 2$ and transient otherwise; we can indeed show (exercise) that $(\mathcal{E}, \mathcal{F})$ is irreducible, and it is recurrent if $n \in (0, 2]$ and transient if $n \in (2, \infty)$.

We can also obtain a description in terms of the associated process $Y_t$.

**Theorem 66**

Let $Y_t$ be the associated process to $(\mathcal{E}, \mathcal{F})$. For any Borel set $B \subset X$, let $T_B = \inf\{t > 0 : Y_t \in B\}$. Then if $(\mathcal{E}, \mathcal{F})$ is irreducible and $B$ is a non-polar Borel set, then "the process does see the set" in the sense that $\mathbb{P}_x(T_B < \infty) > 0$ for $q$-almost-every starting point $x \in X$.

Furthermore if $(\mathcal{E}, \mathcal{F})$ is recurrent, then for any non-polar Borel set $B$ we have

$$\mathbb{P}_x\left(T_B \circ \theta_n < \infty \text{ for all } n \in \mathbb{N}\right) = 1,$$

where $\theta_n$ is the "time-shift by $n$" operator ($Y_t \circ \theta_s = Y_{t+s}$), for $q$-almost-every $x \in X$. On the other hand if $(\mathcal{E}, \mathcal{F})$ is transient, then the process is either killed or escapes to infinity in the sense that

$$\mathbb{P}_x\left(\zeta = \infty, \lim_{t\to\infty} Y_t = \Delta\right) = \mathbb{P}_x(\zeta = \infty)$$

($\lim Y_t = \Delta$ meaning that we escape every compact set) for $q$-almost-every $x \in X$.

The main takeaway is thus that we can deduce global properties (long-time behavior) from an infinitesimal description at time $t = 0$. Everything here is qualitative, but we'll see now that we can do the same with quantitative properties such as the **heat kernel**.

**Definition 67**

Let $(\mathcal{E}, \mathcal{F})$ be a Dirichlet form on $L^2(X, m)$ and $(P_t)$ its corresponding semigroup. Consider a family $(p_t)_{t>0}$ of Borel measurable functions on $X \times X$ such that for any $f \in L^2(X, m)$, we have

$$P_t f(x) = \int_X p_t(x, y) f(y) m(dy).$$

If such a family exists, it is called a **heat kernel**.

We can think of $p_t(x, y)m(dy)$ as the law of $Y_t$ given $Y_0 = x$ (at least up to quasi-almost-everywhere considerations) – indeed if $f$ is an indicator, this tells us probabilities of being in a particular set at time $t$. Heat kernels do not always exist, for example if the Dirichlet form is just zero and thus we would need $p_t$ to be the Dirac mass.

The point is that if we know something about $p_t$, then we have a way of measuring how the process evolves. The following is a measure-theoretic result:

**Proposition 68**

Suppose $(X, \mathcal{M}, m)$ is a $\sigma$-finite separable measure space, and $T : L^1(X, m) \to L^\infty(X, m)$ is a bounded operator. Then there exists a jointly measurable function $K : X \times X \to \mathbb{R}$ which is bounded (that is, $K \in L^\infty(X \times X, m \otimes m)$) with $||T||_{1\to\infty} = ||K||_\infty$, and such that

$$Tf(x) = \int K(x, y) f(y) m(dy).$$

If we already know $T$ is of this integral kernel form, then the bound is easy because we can approximate $f$ by a Dirac mass at the maximum of $K$. So if we can show that the heat operator is a bounded operator from $L^1$ to $L^\infty$, it will show the existence of $p_t$ and in fact give us a bound on how large it is. We will call such a bounded operator from $L^1$ to $L^\infty$ **ultracontractive** – the idea is that this also smooths out the operator, since $L^1$ functions can in fact be very spiky. So the question now is how we would prove such a bound for $P_t$. (Note that it's possible for the heat

kernel to only exist after some cutoff time, or for the heat kernel to exist even without ultracontractivity. But we're starting with the simplest case.)

**Remark 69.** *The theory of regular Dirichlet forms assumes that we have a locally compact space. This precludes infinite-dimensional systems of interest, but there is also a theory of quasi-regular Dirichlet forms where we can construct processes similar to Fukushima's theorem.*

**Proposition 70**

Let $(\mathcal{E}, \mathcal{F})$ be a Dirichlet form on $L^2(X, m)$, and let $(P_t)$ be the associated semigroup. Assume that we have constants $C_1 \geq 0, C_2 > 0, n \geq 1$ (all real numbers) such that the following **Nash inequality** holds for all $f \in L^1(X, m) \cap \mathcal{F}$:

$$||f||_2^{1+n/2} \leq ||f||_1 \left(C_1||f||_2^2 + C_2\mathcal{E}(f, f)\right)^{n/4}.$$

Then $(P_t)$ admits a heat kernel $(p_t)$ satisfying the bound

$$\text{ess sup}_{x,y\in X} p_t(x, y) \leq \max\left(2C_1, \frac{nC_2}{t}\right)^{n/2}.$$

In particular if $C_1 = 0$, then this tells us a bound of the form $t^{-n/2}$ as $t \to 0$.

(It's always good to check that multiplying $f$ by a constant in functional inequalities like the Nash inequality above doesn't change the statement.) This inequality was proven in the 1950s for the Dirichlet form of Brownian motion with $C_1 = 0$ and $n$ the dimension of the Brownian motion; we'll see Nash's proof later on. And in fact this fact applies to all three examples we saw in the first lecture (Brownian motion, Brownian motion on the Sierpinski gasket, and $\alpha$-stable processes) and is sharp up to constants.

*Proof.* We wish to show that $||P_t||_{1\to\infty} \leq \max\left(2C_1, \frac{nC_2}{t}\right)^{n/2}$, so that we can apply Proposition 68. Instead of showing this bound, we'll **first show a bound from $L^1$ to $L^2$ and then use duality** (since we then also get a bound from $L^2$ to $L^\infty$ for the adjoint and then compose them together).

Our first goal is to show

$$\boxed{||P_t(f)||_2 \leq \max\left(2C_1, \frac{nC_2}{2t}\right)^{n/4} ||f||_1}$$

for all $t > 0$ and $f \in L^1(X, m)$. Indeed, without loss of generality (by scaling by constants) we can assume that $||f||_1 = 1$ and we want to understand how the $L^2$ norm evolves over time. One way to do this is to find a differential inequality: we have

$$\frac{d}{dt}||P_t f||_2^2 = \int 2\,(P_t f)\,\frac{d}{dt}\,(P_t f)\,dm,$$

and now we can justify this next step via the spectral theorem, since $P_t = e^{tA} \implies \frac{d}{dt}P_t = Ae^{tA}$: the above expression becomes

$$-2\int (P_t f)(-A)(P_t f)dm = -2\mathcal{E}(P_t f, P_t f).$$

In particular, this tells us that the Dirichlet energy captures the rate of decrease of the $L^2$ norm of the contracted $P_t f$. Defining $\psi(t) = ||P_t f||_2^2$, we've shown that

$$\psi'(t) = -2\mathcal{E}(P_t f, P_t f).$$

Now applying the Nash inequality to $P_t f$, we have

$$\begin{aligned} ||P_t f||_2^{1+2/n} &\leq ||P_t(f)||_1 \left(C_1||P_t f||_2^2 + C_2\mathcal{E}(P_t f, P_t f)\right)^{n/4} \\ &\leq \left(C_1||P_t f||_2^2 + C_2\mathcal{E}(P_t f, P_t f)\right)^{n/4} \\ \implies \psi(t)^{(1+n/2)/2} &\leq \left(C_1\psi(t) - C_2\frac{\psi'(t)}{2}\right)^{n/4} \end{aligned}$$

by plugging in our expression for $\psi'$ and using that $P_t$ is also a contraction in $L^1$. Raising both sides to the power of $4/n$ yields

$$\psi(t)^{1+2/n} \leq C_1\psi(t) - \frac{C_2}{2}\psi'(t).$$

Now if $\psi(t) \leq (2C_1)^{n/2}$ the boxed inequality is already true; otherwise if $\psi(t) \geq (2C_1)^{n/2}$ we have

$$\begin{aligned} -\frac{C_2}{2}\psi'(t) &\geq \psi(t)^{1+2/n} - C_1\psi(t) \\ &\geq \psi(t)\left(\psi(t)^{2/n} - C_1\right) \\ &\geq \psi(t)\left(\psi(t)^{2/n} - \frac{1}{2}\psi(t)^{2/n}\right) \\ &= \frac{1}{2}\psi(t)^{1+2/n}, \end{aligned}$$

and therefore we have $\psi(t)^{1+2/n} \leq -C_1\psi'(t)$ as long as $\psi(t) \geq (2C_1)^{n/2}$. Solving this differential inequality yields

$$\frac{d}{dt}(\psi(t)^{-2/n}) \geq \frac{2}{nC_2},$$

which, after integrating and rearranging, is the boxed inequality that we wanted. And finally to get the theorem from the boxed inequality, we have

$$||P_t||_{1\to\infty} = ||P_{t/2} \circ P_{t/2}||_{1\to\infty} \leq ||P_{t/2}||_{1\to 2}||P_{t/2}||_{2\to\infty},$$

and now since everything is self-adjoint $P_t : L^2 \to \infty$ is the adjoint of the operator $P_t : L^1 \to L^2$. Thus both terms on the right-hand side are $\max\left(2C_1, \frac{nC_2}{2(t/2)}\right)^{n/4}$, which is what we wanted. □

So the overall message is that because the Dirichlet energy tells us about the norm of the semigroup, we can encode that via differential inequalities to get ultracontractive bounds and thus existence of the heat kernel. It may seem weird why Nash did all of this when we already know the exact formulas for the heat kernel of Brownian motion, but the point is that this method is **robust to perturbations** as well: if we have

$$\mathcal{E}(f, f) = \int \nabla f(x)^T \mathcal{A}(x)\nabla f(x)dx$$

for $\mathcal{A} : \mathbb{R}^n \to M^{n\times n}$ a positive definite, measurable, symmetric, matrix-valued function, and we have some $\Lambda$ such that

$$\Lambda^{-1}||\xi||^2 \leq \xi^T\mathcal{A}(x)\xi \leq \Lambda||\xi||^2$$

(so $\mathcal{A}$ is almost constant but we can have shears and various speeds in different places), we can think of having a perturbation of Brownian motion, but still getting comparable Nash inequalities and thus comparable bounds on the heat kernel.

# 7 June 12, 2025

Last time, we discussed Fukushima's theorem (which lets us construct processes) and found ways to connect properties of the Dirichlet form to properties of the process. We then also saw a quantitative result in the form of getting bounds on the heat kernel from the (functional) Nash inequality and commented on the stability of the result under perturbations.

Today, we'll see a proof of the Nash inequality for the setting of Brownian motion:

*Proof of Nash inequality on $\mathbb{R}^n$.* The proof uses Fourier analysis; we'll use the convention that $\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-ix\cdot\xi}dx$. By Parseval's identity, we can write down $||f||_2^2 = (2\pi)^{-n}||\hat{f}||_2^2$ and break up the resulting integral into two terms, within and outside a ball of radius $R$:

$$\begin{aligned}||f||_2^2 &= (2\pi)^{-n}\left(\int_{B_R(0)}|\hat{f}(\xi)|^2d\xi + \int_{B_R(0)^c}|\hat{f}(\xi)|^2d\xi\right)\\ &\leq (2\pi)^{-n}\left(\int_{B_R(0)}|\hat{f}(\xi)|^2d\xi + \int_{B_R(0)^c}\frac{|\xi|^2}{cR^2}|\hat{f}(\xi)|^2d\xi\right)\\ &\leq (2\pi)^{-n}||F||_1^2 w_n R^n + \frac{1}{R^2}\int|\nabla f(x)|^2dx\end{aligned}$$

where in the last inequality we bound the integrand by $||f||_1^2$ pointwise for the first term, and we use $\widehat{\nabla f} = -2\xi\hat{f}(\xi)$ and Parseval's theorem again for the second term. But this is true for any choice of $R > 0$, so we can minimize it as a function of $R$ – doing the appropriate calculus optimization exactly yields the desired result

$$||f||_2^{1+n/2} \leq ||f||_1\left(\int|\nabla f|^2dx\right)^{n/4}$$

for all $f \in W^{1,2}(\mathbb{R}^n)\cap L^1(\mathbb{R}^n)$ (in particular we have $C_1 = 0$, so our heat kernel bound is $\left(\frac{nC}{t}\right)^{n/2}$, which is tight when $x = y$). □

It's worth mentioning at this point that the existence of the heat kernel cannot be taken for granted:

**Example 71**

Consider Brownian motion on the circle $\mathbb{T} = \frac{\mathbb{R}}{2\pi\mathbb{Z}}$ (that is, Brownian motion projected down by the quotient map); this is symmetric with respect to the Haar probability measure on $\mathbb{T}$. We can write down the heat kernel in this case:

$$p_t(x,y) = g_t(x-y), g_t(x) = \sqrt{\frac{2\pi}{t}}\sum_{k\in\mathbb{Z}}\exp\left(-\frac{(x+2\pi k)^2}{2t}\right)$$

(the $2\pi$ is in the numerator of the square root because our torus is of length $2\pi$).

In this one-dimensional example, the heat kernel indeed exists for all time. But now consider the infinite-dimensional torus $\mathbb{T}^\infty = \prod_{k=1}^\infty \mathbb{T}$ and define a Brownian motion on this space via

$$Y_t = (B^1_{a_1t}, B^2_{a_2t}, \cdots)$$

where $(a_1, a_2, \cdots)$ is some sequence of positive real numbers and $B^1, B^2, \cdots$ are independent Brownian motions on $\mathbb{T}$. Since everything is independent, we can write down the law as a product of one-dimensional measures, and we can assume that we start the process at the origin because the process is translation-invariant. So the corresponding law $\mu_t$ now takes the form

$$\mu_t(dx) = \prod_{k=1}^\infty g_{a_kt}(x)m(dx);$$

the infinite-dimensional Brownian motion is indeed symmetric with respect to the Haar measure on $\mathbb{T}^\infty$, which is $m_\infty = \prod_{k=1}^\infty m$. We are then curious when this law has a density with respect to the Haar measure – it turns out to always be absolutely continuous or singular (nothing in between):

**Theorem 72** (Kakutani's dichotomy theorem)
With the notation above, $\mu_t \ll m_\infty$ if and only if

$$\prod_{n\in\mathbb{N}} \int_{\mathbb{T}} \sqrt{g_{a_n t}(x)} m(dx) > 0,$$

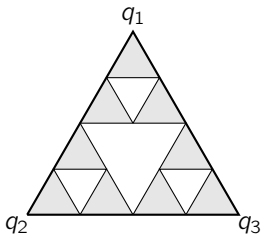
and $\mu_t \perp m_\infty$ if and only if the product is zero. It turns out that the product is positive if and only if $\sum_{n=1}^\infty e^{-ta_n} < \infty$.

(By Cauchy-Schwarz each of these integrals is at most 1, so the product definitely exists and we want the limit to be strictly positive to get a density.) And if we choose the coefficients to be $a_n = \frac{1}{\alpha}\log(n+1)$ for all $n$, then it turns out $\mu_t$ is singular with respect to $m_\infty$ if and only if $t \le \alpha$.

**Remark 73.** *There's a conjecture that even if the components of the Brownian motion are not independent but still correspond to a translation-invariant diffusion, there will still be such a cutoff time (possibly infinite) where we're singular before and absolutely continuous afterward. But that conjecture is still open, since we can't just compute a dichotomy of this form and we can't simply diagonalize the covariance matrix because we won't be working with the standard torus anymore.*

On the other hand, do note that it's not possible for the heat kernel to first exist and then fail to exist later because of the Markov property: at time $t+s$ we can view things as a shifted version of the law at time $t$, so it will always exist.

**Example 74**
Next, we'll construct the Brownian motion on the Sierpinski gasket. Recall that the definition is as follows: start with an equilateral triangle and remove the median triangle in the middle, and then continue this process on each of the three remaining half-side-length triangles.




Let $S = \{1,2,3\}$. We'll define the maps $f_1, f_2, f_3 : \mathbb{R}^2 \to \mathbb{R}^2$ which are scale-1/2 contractions centered at the three vertices $q_1, q_2, q_3$ respectively; that is, $f_j(x) = \frac{1}{2}(x+q_j)$ for all $j \in S$ and $x \in \mathbb{R}^2$. Formally, the **Sierpinski gasket** is then the unique non-empty compact set satisfying the identity

$$K = \bigcup_{j\in S} f_j(K).$$

We'll now define an inductive set of points

$$V_0 = \{q_1, q_2, q_3\}, \quad V_m = \bigcup_{j\in S} f_j(V_{m-1})$$

(so $V_1$ will also include the midpoints of the big triangle, then $V_2$ also all of the intersections between line segments in the figure shown above, and so on). This is an increasing family of sets, and we can think of these as a sequence of graphs which approximate the Sierpinski gasket.

Define $F_j$ to be the map $f_j$ restricted to the Sierpinski gasket $K$, and let

$$W_m = S^m = \{w_1 w_2 \cdots w_m : w_i \in S\}$$

be the set of words of length $m$ in the alphabet $\{1, 2, 3\}$. For any $x, y \in V_m$, we will define the adjacency relation

$$x \overset{m}{\sim} y \quad \text{if } x \neq y \text{ and } x, y \in F_w(V_1) \text{ for some } w \in W_m,$$

where $F_w = F_{w_1} \circ F_{w_2} \circ \cdots \circ F_{w_m}$. In other words, we're saying that we started off with two vertices of the big triangle and then performed some sequence of $m$ contractions, so that $x$ and $y$ are two of the three resulting vertices.

We want to construct Dirichlet forms on these discrete graphs and take a limit to get a Dirichlet form on $K$. From a probabilistic point of view, we have a random walk on the graphs $V_m$ and want to take scaling limits. To do this, let

$$E^{(m)}(u, v) = \frac{1}{2} \sum_{\substack{x,y \in V_m \\ x \overset{m}{\sim} y}} (u(x) - u(y))(v(x) - v(y)) \text{ for all } u, v \in \mathbb{R}^{V_m}.$$

be the unscaled energy corresponding to simple random walk. We now want to rescale this (that is, speed up the walk), and it turns out we should set

$$\mathcal{E}^{(m)}(u, v) = \left(\frac{5}{3}\right)^m E^{(m)}(u, v).$$

The $3^m$ factor comes from there being roughly $3^m$ vertices at scale $m$, and the $5^m$ came from the explanation of the first lecture (to travel twice as far, we need to spend 5 times as much time). But we can actually do a computation of the energies that shows this scaling in another way:

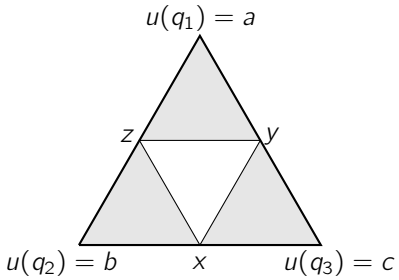
**Proposition 75**

For any $m$, suppose we have any function $u \in \mathbb{R}^{V_m}$ at level $m$ and want to extend it to level $m+1$ while minimizing the energy. Then we have

$$\min_{\substack{v \in \mathbb{R}^{V_{m+1}} \\ v|_{V_m} = u}} E^{(m+1)}(v, v) = \frac{3}{5} E^{(m)}(u, u).$$

Thus the rescaled energies satisfy a monotonicity property without any constants.

*Proof.* First, we claim we can reduce to the case $m = 0$. Indeed, the energy at level $(m+1)$ is the sum of the energies in the three smaller triangles, each of which is the energy at level $m$, and thus we can just keep going down until the smallest level. Suppose we have some boundary conditions for $u$, and we have to optimize over the three remaining values:



We then have

$$E^{(1)}(v,v) = \frac{1}{2}\left((a-x)^2 + (a-y)^2 + (z-y)^2 + (b-z)^2 + (b-x)^2 + (x-z)^2 + (y-x)^2 + (y-c)^2 + (x-c)^2\right)$$

and for fixed $a, b, c$ we want to minimize the result as a function of $x, y, z$. By completing the squares it turns out the minimizer is

$$x = \frac{a+2b+2c}{5}, \quad y = \frac{2a+b+2c}{5}, \quad z = \frac{2a+2b+c}{5}.$$

(Indeed, the minimizer can be viewed in terms of harmonic measures – if we start at $x$ and do simple random walk until we hit any of $q_1, q_2, q_3$, the probabilities turn out to be $\frac{1}{5}, \frac{2}{5}, \frac{2}{5}$.) And if we plug these values in we indeed get that $E^{(1)}(v,v) = \frac{3}{5}E^{(0)}(u,u)$, as desired. □

**Remark 76.** *We'll see a different probabilistic interpretation in terms of a time-change later on. This is the simplest example of diffusion on fractals – the Sierpinski carpet is actually much more difficult for many reasons, since we can't actually compute the rescaling factors very easily.*

With this, we'll now make the following natural definition of the limiting Dirichlet form:

**Definition 77**

Let $V_* = \bigcup_{n=1}^{\infty} V_n$ be the limit of the increasing sets. It turns out $\overline{V_*} = K$, and we define

$$\mathcal{F}_* = \left\{u \in \mathbb{R}^{V_*} : \lim_{m\to\infty} \mathcal{E}^{(m)}\left(u|_{V_m}, u|_{V_m}\right) < \infty\right\}.$$

(The limit $\lim_{m\to\infty} \mathcal{E}^{(m)}\left(u|_{V_m}, u|_{V_m}\right)$ always exists because we have a nondecreasing nonnegative sequence, and we're just asking for it to be finite.) We then define the bilinear form $\mathcal{E}_*$ via

$$\mathcal{E}^{(*)}(u,v) = \lim_{m\to\infty} \mathcal{E}^{(m)}\left(u|_{V_m}, v|_{V_m}\right)$$

for all $u, v \in \mathcal{F}_*$ (via polarization).

It's easy to check that $\mathcal{E}^*$ is a bilinear, nonnegative definite form on $\mathcal{F}_*$. Notice however that functions in $\mathcal{F}_*$ are only defined on a dense subset of $K$, not all of $K$ itself, and we would like to extend this to the whole space $K$. It turns out we can do so in a unique way as a continuous function, but for that we must show that functions in $\mathcal{F}_*$ are already uniformly continuous by proving some estimates.

**Proposition 78**

For all $x, y \in V_*$ and all $u \in \mathcal{F}_*$, we have

$$|u(x) - u(y)|^2 \leq 400|x-y|^{\alpha}\mathcal{E}^{(*)}(u,u)$$

for $\alpha = \log(5/3)/\log(2)$. In other words, this tells us that the function $u$ is always $\frac{\alpha}{2}$-Hölder continuous, hence uniformly continuous, and so it admits a unique continuous extension to $K = \overline{V_*}$.

*Proof sketch.* For neighboring vertices, the claim is easier to prove, and then we'll use a chaining argument after that. First consider any two vertices $y, z \in V_2$; the graph distance $d(y,z)$ is at most 2, so for all $y, z \in V_2$ there exists some

vertex $q \in V_1$ such that either $y = z = q$, $y \overset{2}{\sim} q = z$, or $y \overset{2}{\sim} q \overset{2}{\sim} z$. So then in all cases we have the inequality

$$\begin{aligned} |v(y) - v(z)| &\le \sqrt{2}\left(|v(y) - v(q)|^2 + |v(q) - v(z)|^2\right)^{1/2} \\ &\le \sqrt{2}E^{(2)}(v,v)^{1/2} \\ &\le \sqrt{\frac{6}{5}}\mathcal{E}^{(2)}(v,v)^{1/2}. \end{aligned}$$

If we now consider $u \in \mathcal{F}_*$ and an arbitrary $x \in V_*$, meaning that $x \in V_m$ for some fixed $m$ and there is some $j \in S$ and some word $w_1 w_2 \cdots w_m \in W_m$ such that $x = F_w(q_j)$, we can define

$$x_k = F_{w_1} \circ \cdots \circ F_{w_{k-1}}(q_j)$$

for all $1 \le k \le m$. Here we take the convention that $w_0$ is the empty word so $F_0$ is the identity map, and we can think of these as the approximations of $x$ by smaller values, since we start from $q_j$ and then apply a sequence of contraction maps. We can then estimate, using the same argument but at a smaller scale as before, that

$$\begin{aligned} |u(x_{k-1}) - u(x_k)| &= \left|u \circ F_{w_1\cdots w_{k-1}}(F^{-1}_{w_1\cdots w_{k-1}}(x_{k-1})) - u \circ F_{w_1\cdots w_{k-1}}(F^{-1}_{w_1\cdots w_{k-1}}(x_k))\right| \\ &\le \sqrt{\frac{6}{5}}\mathcal{E}^{(2)}\left(u \circ F_{w_1\cdots w_{k-1}}|_{V_2}, \quad u \circ F_{w_1\cdots w_{k-1}}|_{V_2}\right) \end{aligned}$$

because $x_{k-1}$ and $x_k$ are adjacent to each other at a smaller level. Applying the rescaling factors by summing over all cells at this scale, we can bound this as

$$\begin{aligned} |u(x_{k-1}) - u(x_k)| &\le \sqrt{\frac{6}{5}}\left(\frac{3}{5}\right)^{(k-1)/2}\left(\left(\frac{5}{3}\right)^{(k-1)/2}\sum_{z \in W_{k-1}} \mathcal{E}^{(*)}(u \circ F_z, u \circ F_z)\right) \\ &\le \sqrt{\frac{6}{5}}\left(\frac{3}{5}\right)^{k-1}\mathcal{E}^*(u,u)^{1/2} \end{aligned}$$

using self-similarity and taking limits. Here in the last step we've used the fact that $\mathcal{E}^{(m)}(u|_{V_m}, u|_{V_m}) = \left(\frac{5}{3}\right)^m \sum_{x \overset{m}{\sim} y}(u(x) - u(y))^2$ and therefore $(u(x) - u(y))^2 \le \left(\frac{3}{5}\right)^m \mathcal{E}^{(m)}(u|_{V_m}, u|_{V_m}) \le \left(\frac{3}{5}\right)^m \mathcal{E}^{(*)}(u,u)$. Since $|x - y| = 2^{-m+1}$ for any $x, y$ neighbors at level $m$, this therefore means that for any $m$ we have

$$|u(x) - u(y)|^2 \lesssim |x - y|^\alpha \mathcal{E}^{(*)}(u,u), \quad \alpha = \frac{\log(5/3)}{\log 2}.$$

All of this has only been used for the case where $x, y$ are neighbors, but in general we can always zoom in in the gasket until $x$ and $y$ are in different sub-triangles but we're maximally zoomed in. We can then let $q$ be the common vertex of those two sub-triangles and use that $|u(x) - u(y)| \lesssim |u(x) - u(q)| + |u(q) - u(y)|$; now we can approximate $x$ by a series of contractions of $q$ and similarly do the same for $q$, which gives us convergent geometric series. This yields the desired result. □

So the bound

$$|u(x) - u(y)|^2 \le 400|x - y|^\alpha \mathcal{E}^{(*)}(u,u)$$

also holds for all $x, y \in K$ (not just $V_*$). In particular, this bound tells us that the only functions in $\mathcal{F}_*$ of zero energy $\mathcal{E}^{(*)}(u,u)$ are the constant functions (because it forces $u(x) = u(y)$ for all $x, y$), and so we can define a quadratic form on functions on the whole gasket

$$\mathcal{F} = \{u \in C(K) : u|_{V_*} \in \mathcal{F}_*\}.$$

(To show that we have nonconstant functions in this space, we can always start with any values on the outer vertices

and do the "harmonic measure" idea to extend it to all other points.) Then we define the quadratic form

$$\mathcal{E}(u,v) = \mathcal{E}^{(*)}(u|_{V_*}, v|_{V_*}).$$

$(\mathcal{E}, \mathcal{F})$ is then a nonnegative definite quadratic form.

**Proposition 79**

$(\mathcal{E}, \mathcal{F})$ satisfies the self-similarity property

$$\mathcal{F} = \{u \in C(K) : u \circ F_j \in \mathcal{F} \text{ for all } j \in S\}.$$

In other words, restricting to only one of the sub-triangles yields another function, and that is also continuous. Furthermore, the energy also satisfies

$$\mathcal{E}(u,u) = \frac{5}{3}\sum_{j\in S} \mathcal{E}(u \circ F_j, u \circ F_j)$$

(this is easy to check for the discrete approximations, and then we take limits).

To ensure this is actually a Dirichlet form, we must first check the Markov property. For any $u \in \mathcal{F}$, it's clear that $\tilde{u} = (0 \vee u) \wedge 1 \in \mathcal{F}$ and that

$$\mathcal{E}^{(m)}(\tilde{u}|_{V_m}, \tilde{u}|_{V_m}) \le \mathcal{E}^{(m)}(u|_{V_m}, u|_{V_m})$$

because it's even true in a pointwise sense: $(\tilde{u}(x) - \tilde{u}(y))^2 \le (u(x) - u(y))^2$ for all $x, y \in K$, and we're looking at sums of expressions of this type. So taking $m \to \infty$ yields the desired property.

Now we check regularity, and we do so by claiming that $\mathcal{F}$ forms an **algebra under multiplication**; in fact,

$$u, v \in \mathcal{F} \implies uv \in \mathcal{F}, \quad \mathcal{E}(uv, uv) \le 2||u||_\infty^2 \mathcal{E}(v,v) + 2||v||_\infty^2 \mathcal{E}(u,u).$$

Indeed, this again comes down to a deterministic pointwise identity

$$(u(x)v(x) - u(y)v(y))^2 \le 2u(x)^2((v(x) - v(y))^2 + 2v(y)^2(u(x) - u(y))^2$$

and then taking limits. Furthermore, we claim that $\mathcal{F}$ **separates points**, meaning that for any finite subset $V \subset K$, $\{u|_V : u \in \mathcal{F}\} = \mathbb{R}^V$ (that is, we can always construct a function which restricts to anything on $V$). Indeed, if $|V| = 1$ then this is clear because the constant functions are all in $\mathcal{F}$, and otherwise we can go to a small enough scale $m$ so that all of our finite points are in sufficiently different cells of the Sierpinski gasket. Precisely, we can choose $m$ so that

$$2^{1-m} < \min_{\substack{x \ne y \\ x,y \in V}} |x - y|,$$

which means that for any $v, w \in W_m$ and any $x, y \in V$, such that $x \in F_v(K)$ and $y \in F_w(K)$, the cells $F_v(K)$ and $F_w(K)$ are disjoint. Next time, we'll show how to use this to prescribe values of the function and then make use of the **Stone-Weierstrass theorem**.

# 8 June 13, 2025

We've been studying the Dirichlet form on the Sierpinski gasket – recall that we defined points in the set in terms of contractions at the three outer vertices, and we similarly defined energies $E^{(m)}$ in terms of simple random walk at

iterative levels and took limits. This yielded a limiting $\mathcal{E}^{(*)}$ which we can use to define our $(\mathcal{E}, \mathcal{F})$.

We saw last time that $\mathcal{E}$ is nonnegative definite and bilinear, and we showed that the domain $\mathcal{F}$ is an algebra under pointwise multiplication. We now need to check the rest of the properties (closed, regular, etc.) – let's review the construction for showing that we separate points. We actually only need to check this for two points $x, y \in K$ with $x \neq y$: we can always pick some sufficiently large $m \in \mathbb{N}$ and some words $u, v \in W_m$ so that $x \in F_u(K)$ and $y \in F_v(K)$ but $F_u(K) \cap F_v(K) = \varnothing$.

For any prescribed values $a, b \in \mathbb{R}$, our goal is now to define a function $g \in \mathbb{R}^{V_m}$ satisfying that $g|_{F_u(V_0)} = a$ and $g|_{F_u(V_0)} = b$. Such a function clearly exists (we can pick any of them), and now we can extend it harmonically to level $m+1$, then $m+2$, and so on – we proved that harmonic extensions don't increase the (rescaled) energy, and thus there will be some unique $f \in \mathcal{F}$ (existing at all levels) such that

$$f|_{V_m} = g, \quad \mathcal{E}(f, f) = \mathcal{E}^{(m)}(g, g),$$

which is what we wanted. So by Stone-Weierstrass, we prove that $\mathcal{F}$ is dense in the continuous functions $(C(K), ||\cdot||_{\sup})$, and that is half of the regularity condition – we also have to check that we're dense with respect to the inner product. For this, we have for all $x, y \in K$ and all $u \in \mathcal{F}$ that

$$\begin{aligned} \boxed{u(x)^2} &\leq 2\left(u(y)^2 + (u(x) - u(y))^2\right) \\ &\lesssim \boxed{u(y)^2 + \mathcal{E}(u, u)} \end{aligned}$$

because all distances are finite. We'll show that for any Radon measure $m$ with full support (in fact because $K$ is compact we can assume $m(K) = 1$) the limit of a Cauchy sequence under $\mathcal{E}_1 = \mathcal{E} + \langle\cdot\rangle_{L^2(m)}$ is also in the space $\mathcal{F}$. But what we do is integrate the boxed inequality over $m(dy)$ and find that

$$u(x)^2 \lesssim \int u(y)^2 m(dy) + \mathcal{E}(u, u) = \mathcal{E}_1(u, u),$$

where we're assuming that we've always chosen the continuous representative so that these functions are pointwise defined. This implies that for any $\mathcal{E}_1$-Cauchy sequence $f_n$, $f_n$ converges pointwise to $f$, and the limit has to be bounded because $\sup |f|$ is finite by our inequality. Furthermore, $f_n$ is also a Cauchy sequence in $L^2(m)$, so we have $f_n \to f$ in $L^2(m)$ as well, and it just remains to show that we have convergence with respect to the $\mathcal{E}_1$-inner product (so the problem is the $\mathcal{E}$ part). For this, it's useful that our energies are defined in terms of discrete energies: since $(f_n)$ is an $\mathcal{E}$-Cauchy sequence, we know that for any $\varepsilon > 0$ there is some $N$ such that

$$\mathcal{E}^{(m)}(f_k - f_\ell, f_k - f_\ell) \leq \mathcal{E}(f_k - f_\ell, f_K - f_\ell) < \varepsilon$$

for all $m$ and for all $k, \ell \geq N$. Now we have a **finite sum** for fixed $m$, so we can interchange limits and take $k \to \infty$ to get that $\mathcal{E}^{(m)}(f - f_\ell, f - f_\ell) \leq \varepsilon$ for all $\ell \geq N$ and all $m$. Finally, taking $m \uparrow \infty$ yields convergence under the $\mathcal{E}$-inner product, so we have proved that our form is closed and that we have a regular Dirichlet form on $L^2(K, m)$. (Note here that we need that $\mathcal{F}$ is a vector space, so $f$ is in the domain because both $f_\ell$ and $f - f_\ell$ are in the domain.)

But we want to define a diffusion process, not just a general Markov process, so we next want to show that our Dirichlet form is **strongly local**. For this, pick $f, g \in \mathcal{F}$ such that $f = c$ on all of $\operatorname{supp}(g)$ (we don't need to worry about the $m$-support since we're choosing everything to be pointwise defined here). Disjoint compact sets have a nonzero distance, so we can pick some $n$ such that

$$\operatorname{dist}(\operatorname{supp}(f - c1_K), \operatorname{supp}(g)) > 2^{-n}.$$

This implies that if we contract the triangle $n$ times, we can only intersect one of the two sets $\text{supp}(f - c1_K), \text{supp}(g)$. In other words, for all words $w \in W_n$, the set $F_w(K) \cap \text{supp}(f - c1_K) \cap \text{supp}(g)$ is always empty. And now we can use the self-similarity from Proposition 79:

$$\mathcal{E}(f, g) = \left(\frac{5}{3}\right)^n \sum_{W \in W_n} \mathcal{E}(f \circ F_w, g \circ F_w).$$

And each term of this sum is zero. Indeed, either $f \circ F_w$ or $g \circ F_w$ must be constant for each $w$, and the Dirichlet form involves a difference between values of the function and thus all terms in the contribution $(f(x) - f(y))(g(x) - g(y))$ are always zero. Thus the strongly local property is proved.

**Fact 80**

The point is that Fukushima's theorem now yields a Markov process with continuous paths on the Sierpinski gasket for all times, unique up to throwing away exceptional sets. And as an exercise, we can check that in this case any $\mathcal{E}$-polar set must actually be empty, for example using the Hölder continuity estimate Proposition 78 to show that 1-capacity is always strictly positive.

**Example 81**

We'll now use this to construct the Brownian motion itself. This process was first constructed in the 1980s by Goldstein, by Kusuoka, and also by Barlow-Perkins (and the last authors also got precise heat kernel bounds for the diffusion). We won't do the full version, but we'll give some upper bounds near the diagonal and explanation for the shape of the kernel overall via lower bounds.

Our strategy will be to prove a Nash inequality, though we'll have to use something completely different from $\mathbb{R}^d$. We'll instead use Poincaré inequalities, which we'll need some background to explain.

**Fact 82**

Strongly local regular Dirichlet forms are of the type

$$\mathcal{E}(f, f) = \int_X d\Gamma(f, f),$$

where $d\Gamma(f, f)$ is called the **energy measure**.

For example for the ordinary Dirichlet form we have $\mathcal{E}(f, f) = \int_{\mathbb{R}^N} |\nabla f|^2 dx$, where $\Gamma(f, f)(A) = \int_A |\nabla f|^2 dx$. In general, the way we construct such a measure is to specify its integrals on various functions.

**Definition 83**

Let $(\mathcal{E}, \mathcal{F})$ be a strongly local regular Dirichlet form on a general space $L^2(X, m)$. For all bounded functions $f \in \mathcal{F} \cap L^\infty$, there is a unique (nonnegative) measure $\Gamma(f, f)$ on $X$, called the **energy measure**, satisfying

$$\int g d\Gamma(f, f) = \mathcal{E}(f, fg) - \frac{1}{2}\mathcal{E}(f^2, g)$$

for all $g \in \mathcal{F} \cap C_c(X)$ (this is a dense subset by regularity). For arbitrary functions $f \in \mathcal{F}$, we similarly define the energy measure via truncation – that is, $\Gamma(f, f)(A) = \lim_{n \to \infty} \Gamma((-n \vee f) \wedge n, (-n \vee f) \wedge n)(A)$ (the measures on the right side converge in total variation so this is well-defined).

This defines a measure by the Riesz representation theorem – linearity is clear from bilinearity, and nonnegativity is not so clear but it is indeed true by approximating the Dirichlet form via the semigroup or resolvent (this is due to Le Jan). And we do have to work with the energy measure to avoid certain issues with singularity in settings like the Sierpinski gasket.

Note that the energy measure is not the same as the "equilibrium measure" (which we'll introduce later on), though they have the common feature that they do not give positive mass to polar sets.

Returning to the Sierpinski gasket, the Dirichlet form is defined for any reference measure $m$. But we'll want to pick a particular one now:

**Proposition 84** (Poincaré inequality)
For the Sierpsinski gasket, for any function $f \in \mathcal{F}$, any point $x \in K$, and any $r \in (0,1)$, there is some constant $A > 1$ such that

$$\int_{B(x,r)} \left|f(y) - f_{B(x,r)}\right|^2 dm \lesssim r^{d_W} \int_{B(x,Ar)} d\Gamma(f,f)$$

for $d_W = \frac{\log 5}{\log 2}$ and $m$ the measure satisfying self-similarity

$$m(F_w(K)) = 3^{-m} \text{ for all } m \text{ and all } w \in W_m.$$

Here we write $f_A$ or $\fint_A f\,dm$ for the averaged value $\frac{1}{m(A)}\int_A f\,dm$.

Note that the measure of any ball $B(x,r)$ satisfies

$$m(B(x,r)) \asymp r^{d_f} \text{ for all } x \in K, r \in (0,1),$$

where $d_f = \frac{\log 3}{\log 2}$ is the Hausdorff dimension of the set.

*Proof.* We want to control the variance of the function $f$ on a ball of radius $r$. It actually suffices to bound by the maximum deviation:

$$\int_{B(x,r)} \left(f(y) - f_{B(x,r)}\right)^2 dm(y) \le m(B(x,r)) \sup_{y,z \in B(x,r)} \left(f(y) - f(z)\right)^2$$

By the Hölder continuity estimate, we already know how to bound the energy, and we can use self-similarity to relate that to the following fact about the energy measure (exercise): for all $f \in \mathcal{F}$,

$$\Gamma(f,f) = \left(\frac{5}{3}\right)^m \sum_{w \in W_m} (F_w)_* \Gamma(f \circ F_w, f \circ F_w)$$

We need to control $(f(y) - f(z))^2$; to do this, let $n$ be the largest integer such that there are words $v, w \in W_n$ with $y \in F_v(K)$ and $z \in F_w(K)$ for $F_v(K) \cap F_w(K) \neq \varnothing$. (We can check that this actually implies that we must have $d(y,z) \asymp 2^{-n}$.) Taking a point $q$ in the intersection, we can thus write

$$\begin{aligned}(f(y) - f(z))^2 &\le (f(y) - f(q))^2 + (f(q) - f(z))^2 \\ &\lesssim \mathcal{E}(f \circ F_v, f \circ F_v) + \mathcal{E}(f \circ F_w, f \circ F_w) \\ &\le \left(\frac{3}{5}\right)^n \Gamma(f,f)(F_v(K) \cup F_w(K))\end{aligned}$$

this time by self-similarity of the energy measure. And our constant $d_W$ is exactly defined so that this is bounded by

$$(f(y) - f(z))^2 \le d(y,z)^{d_W - d_f} \Gamma(f,f)(F_v(K) \cup F_w(K)),$$

and now $F_v(K)$ and $F_w(K)$ each have diameter $2^{-n}$, but in fact this implies that for any points $x, x'$ in that ball we have $d(x, x') \leq 2^{-n} + r \leq Ar$ for some constant $A$. So plugging this in yields the Poincaré inequality: we end up with a bound $r^{d_W - d_f}\Gamma(f, f)(B(x, Ar))$, which is exactly what we want. □

Next, we'll prove a variant which will look similar to what just proved, except that the ball moves along with the point:

**Proposition 85** (Pseudo-Poincaré inequality)
For all $f \in \mathcal{F}$ and all $r \in (0, 1)$, we have

$$\int_K |f(x) - f_r(x)|^2 dm(x) \lesssim r^{d_W}\mathcal{E}(f, f),$$

where $f_r = \fint_{B(x,r)} f\, dm$.

*Proof.* We have

$$\int_K |f(x) - f_r(x)|^2 dm(x) \leq \int_K \fint_{B(x,r)} |f(x) - f(y)|^2 m(dy)m(dx)$$

by convexity and hence Jensen's inequality, and now we know by the Poincaré inequality that we have the bound

$$\boxed{\int_K |f(x) - f_r(x)|^2 dm(x) \lesssim r^{-d_f} \int_K \int_K (f(x) - f(y))^2 1\{d(x, y) \leq r\} m(dy)m(dx)}.$$

We now cover the Sierpinski gasket by balls of radius $r$ – that is, let $N$ be an $r$-net, meaning that it is an $r$-separated subset of $K$ which is maximal with respect to inclusion (the existence follows from Zorn's lemma). By maximality we have

$$\bigcup_{n \in N} B(n, r) = K \implies \sum_{n \in N} 1\{B(n, r)\} \geq 1_K,$$

which then further implies that (if $x, y$ are within $r$, then there must be some point in $N$ within $r$ of $x$, and then by the triangle inequality both $x, y$ are at most $2r$ away from the point $n$ in the net)

$$\sum_{n \in N} 1_{B(n,2r)}(x) 1_{B(n,2r)}(y) \geq 1_{d(x,y)<r}(x, y).$$

On the other hand, we also have the simple geometric fact that if $A$ is the constant from our Poincaré inequality, then for all $x$ we have

$$\sum_{n \in N} 1_{B(n,2Ar)}(x) \lesssim 1$$

since there are only finitely many points inside $B(n, 2Ar)$ that can all be at least $r$ apart from each other by a volume estimate. So we can return to the boxed inequality and write it as a sum over balls around points in our net:

$$\int_K |f(x) - f_r(x)|^2 dm(x) \lesssim r^{-d_f} \sum_{n \in N} \int_{B(n,2r)} \int_{B(n,2r)} (f(x) - f(y))^2 m(dy)m(dx),$$

and now we use that for any set $A$ we have $\frac{1}{2m(A)} \int_A \int_A (f(x) - f(y))^2 dm = \int (f(x) - f_A)^2 dm$ (this is the same trick as we saw in the proof of the Harris inequality!), so we can relate this back to something in the Poincaré inequality.

This yields

$$\int_K |f(x) - f_r(x)|^2 dm(x) \lesssim \sum_{n\in N} \int_{B(n,2r)} (f(x) - f_{B(n,2r)})^2 dm(x)$$
$$\lesssim \sum_{n\in N} r^{d_W} \int_{B(n,2Ar)} \Gamma(f,f)$$

where we've used the ordinary Poincaré inequality in the last step. But by the definition of the energy measure this is bounded by a constant times $r^{d_W}\mathcal{E}(f,f)$, as desired. □

We'll now use this to prove a Nash inequality:

**Theorem 86**

We have for all $f \in \mathcal{F} \cap L^1(K,m)$ (but actually $L^1(K,m)$ is a subset of $L^2(K,m)$ so this additional condition is not necessary)

$$||f||_2^{1+n/2} \lesssim ||f||_1 \left(||f||_2^2 + \mathcal{E}(f,f)\right)^{n/4}$$

for $n = \frac{2d_f}{d_W} = \frac{2\log 3}{\log 5} < 2$ (we call $n$ the "spectral dimension").

Once we prove this result, the existence of heat kernel and bounds on the diagonal then follow from our earlier discussion.

*Proof.* Just like on Euclidean space, we'll break up the $L^2$ norm into two pieces. Let $r \in (0,1)$ and write

$$||f||_2^2 \le 2\left(||f - f_r||_2^2 + ||f_r||_2^2\right).$$

We can control the first term by the psuedo-Poincaré inequality by $r^{d_W}\mathcal{E}(f,f) + ||f_r||_2^2$, and now we'll see how to control the second term. We have

$$||f_r||_2^2 \le ||f_r||_1 ||f_r||_\infty,$$

and now we can control the $L^\infty$ norm because $||f_r||_\infty \lesssim r^{-d_f}||f||_1$ (the integral on a ball cannot be larger than the total integral), meaning we just need to control the $L^1$ norm. We have

$$||f_r||_1 = \int |f_r(x)| m(dx)$$
$$\lesssim r^{-d_f} \int_K \int_K |f(y)| 1\{d(x,y) < r\} m(dy) m(dx).$$

We can now integrate over $x$ first by Fubini's theorem, but then that cancels out with the $r^{-d_f}$ factor in front and thus we find that $||f_r||_1 \lesssim ||f||_1$. Therefore

$$||f_r||2^2 \le ||f_r||_1 ||f_r||_\infty \lesssim r^{-d_f}||f||_1^2,$$

and now combining the two estimates of the two terms yields

$$||f||_2^2 = r^{d_W}\mathcal{E}(f,f) + r^{-d_f}||f||_1^2.$$

Now we want to optimize over $r$ to get the result, but we are constrained to $r \in (0,1)$ and that's why we have the extra term of $||f||_2^2$; otherwise everything is the same as usual. □

From here, we say that Nash implies ultracontractivity and thus existence of the heat kernel; in particular we get

the bound

$$\text{ess sup}_{x,y\in K} p_t(x,y) \lesssim 1 \wedge t^{-d_f/d_W}$$

for all $t > 0$. When $x = y$ we have a matching lower bound, so this is indeed sharp. It turns out that Barlow and Perkins managed to get a much better bound (we call these **sub-Gaussian heat kernel bounds**) of

$$p_t(x,y) \lesssim \frac{C}{t^{d_f/d_W}} \exp\left(-C'\left(\frac{d(x,y)^{d_W}}{t}\right)^{1/(d_W-1)}\right)$$

for all $t \in (0,1)$ and all $x, y \in K$, and there is also the corresponding matching lower bound. For $d_W = 1$ this is exactly the Gaussian we get on Euclidean space, but instead for the Sierpinski gasket here we have $d_W = \frac{\log 5}{\log 2} > 2$.

**Fact 87**

We know that we have a corresponding diffusion $(Y_t)$ for the objects we've been analyzing here. What we can compute from these estimates is that

$$\mathbb{E}_x[d(Y_0, Y_t)] \asymp t^{1/d_W} \text{ for all } t \in (0,1), x \in K,$$

and if we let $\tau_{B(x,r)}$ be the exit time of a ball of radius $r$,

$$\mathbb{E}_x\left[\tau_{B(x,r)}\right] \asymp r^{d_W} \text{ for all } x \in K, r \in \left(0, \frac{1}{2}\right).$$

This explains the name "walk dimension" for the constant $d_W$.

**Remark 88.** *We can always use Cauchy-Schwarz and symmetry of the semigroup to get $p_t(x,y) = \sqrt{p_t(x,x)p_t(y,y)}$. So the maximum value of the heat kernel will always be achieved on the diagonal.*

# 9 June 16, 2025

Last time, we saw how to obtain off-diagonal heat kernel upper bounds for the Sierpinski gasket last time (with scaling factor $\frac{1}{t^{d_f/d_W}}$ dependent on the Hausdorff and walk dimension). We'll first explain the exponential term which gives us matching lower bounds today – usually once we get an upper bound, we show lower bounds for $x$ near $y$ and then for arbitrary points.

The idea is that "by time $t$, we've spread out over a ball of roughly radius $t^{1/d_W}$." So we first try to get **near-diagonal lower bounds** of the form

$$p_t(x,y) \gtrsim \frac{1}{t^{d_f/d_W}} 1\left\{d(x,y) \le c t^{1/d_W}\right\}.$$

There's no exponential factor here yet, but it should then be exponentially unlikely to move much farther away than $t^{1/d_W}$. So what we want to understand is the "most likely mechanism by which we travel a long distance," and it turns out the answer is the following: break up the time $t$ into $n$ intervals, and for each interval $\frac{t}{n}$ we move a distance roughly $\left(\frac{t}{n}\right)^{1/d_W}$ in the same direction.

To make this precise, we'll pick the actual value of $n$ later, and consider a sequence of equally time-spaced points $x_0, \cdots, x_n$ with $x_0 = x$ and $x_n = y$. We choose these points satisfying the following two conditions:

$$d(x_i, x_{i+1}) \asymp \frac{d(x,y)}{n}$$

and also (here $c$ is the same constant as in the near-diagonal lower bound)

$$d(x_i, x_{i+1}) \lesssim \frac{c}{2}\left(\frac{t}{n}\right)^{1/d_W}$$

for all $i$. By the triangle inequality, we can then say that if $B_i$ is a ball of center $x_i$ and radius $\frac{c}{4}\left(\frac{t}{n}\right)^{1/d_W}$, then we know from our lower bound that

$$p_{t/n}(y_i, y_{i+1}) \gtrsim \frac{1}{(t/n)^{d_f/d_W}}$$

for all $y_i \in B_i, y_{i+1} \in B_{i+1}$. Therefore what we do is take (this is called the Chapman-Kolmogorov equation)

$$p_t(x, y) = \int \cdots \int p_{t/n}(x, x_1)p_{t/n}(x_1, x_2)\cdots p_{t/n}(x_{n-1}, y)dx_1 dx_2 \cdots dx_{n-1}$$

and bound it from below by **only integrating over the specified balls** $B_1, \cdots, B_{n-1}$. It's now not so bad to estimate the integral from below, since the measure of each ball is roughly $\left(\frac{t}{n}\right)^{d_f/d_W}$ and inside the ball we have $p_{t/n}(y_i, y_{i+1}) \sim \left(\frac{t}{n}\right)^{-d_f/d_W}$. So we end up with an exponential constant $\exp(-c'n)$ because we have a (possibly less than 1) constant dependence for each of the $n$ integrals, and the polynomial term $\left(\frac{t}{n}\right)^{-d_f/d_W}$ comes from having $n$ terms that we multiply with this factor and $(n-1)$ integrals with that factor canceling out. Since the exponential term dominates for large $n$ we thus see that

$$p_t(x, y) \gtrsim \exp(-c''n)t^{-d_f/d_W},$$

and now we pick $n$ as small as possible so that the conditions we specified above can still be satisfied. We can check by triangle inequality that we must in fact have $n \gtrsim \left(\frac{d(x,y)^{d_W}}{t}\right)^{1/(d_W-1)}$, as desired. So if we know that the diffusion process spreads uniformly over a small ball, we just chain the estimate using Chapman-Kolmogorov. (The upper bound uses a different kind of chaining argument but similar calculations, using tails of exit times.)

**Fact 89**

The main geometric feature needed in this argument was that given any pair of points and any $n$, we can connect them with some constraints on the distance. This doesn't always work – it means our metric space needs to be close to a geodesic space – but we'll come back to this later on.

So far, we've only used the Nash inequality (which implies ultracontractivity) to derive our bounds. But this property may not always hold even when the heat kernel does not exist:

**Example 90**

Consider a Dirichlet form on the space $L^2(\mathbb{R}^n, m_\alpha)$, where $m_\alpha(dx) = (1+|x|^2)^{\alpha/2}$ is a weighted version of the Lebesgue measure ($\alpha = 0$ recovers the original Lebesgue measure). Specifically, take the case $n \geq 2$, $\alpha \in (-n, \infty)$ (otherwise we fail to be a doubling measure), and the form defined by

$$\mathcal{E}(f, f) = \int |\nabla f(x)|^2 m_\alpha(dx).$$

We can calculate the generator here by using integration by parts: we have that (here $\Delta$ denotes the Laplacian)

$$\begin{aligned} L_\alpha f &= (1+|x|^2)^{-\alpha/2}\text{div}\left((1+|x|^2)^{\alpha/2}\nabla f\right) \\ &= \Delta f + \frac{\alpha x \cdot \nabla f}{1+|x|^2} \end{aligned}$$

So compared to Brownian motion, we have a drift either towards infinity or towards the origin depending on the sign of $\alpha$. We can calculate that for all $x \in \mathbb{R}^n$ and all $r > 0$, the measure of the ball satisfies the bounds

$$m_\alpha(B(x,r)) \asymp r^\alpha (1+|x|+r)^\alpha.$$

**Theorem 91** (Grigor'yan, Saloff-Coste 2005)
We have in the setting above that for all $t > 0$ and all $x, y \in \mathbb{R}^n$, we have the heat kernel bounds

$$p_t(x,y) \asymp \frac{c}{m_\alpha(B(x,\sqrt{t}))} \exp\left(-c' \frac{d(x,y)^2}{t}\right).$$

Notably, this example does not satisfy ultracontractivity, because if we plug in our estimate for the volume we get

$$p_t(x,y) \asymp \frac{\exp\left(-c\frac{d(x,y)^2}{t}\right)}{t^{n/2}(1+|x|+\sqrt{t})^\alpha}.$$

So for fixed $t$, if $\alpha < 0$ then we can make the denominator arbitrary large by choosing $x$ far from the origin. Thus $||P_t||_{1\to\infty} = \text{ess sup}_{x,y} p_t(x,y) = \infty$ even though the heat kernel exists. It turns out that **the converse of "Nash implies ultracontractivity" also holds**, and so we would need a different method for heat kernels that may behave very differently in different parts of the space even if we get a Gaussian bound. We'll now see some methods that do work.

The main idea is to work with local versions of ultracontractivity instead of global ones. We'll start with **Harnack inequalities**:

**Theorem 92** (Elliptic Harnack inequality)
Fix some dimension $n$. There exists some constant $C \geq 1$ such that for any $x \in \mathbb{R}^n$, any $r > 0$, and any nonnegative harmonic function $h : B(x,2r) \to (0,\infty)$, we have

$$\sup_{y \in B(x,r)} h(y) \leq C \inf_{y \in B(x,r)} h(y).$$

Here $C$ depends only on the dimension of the Euclidean space $n$, though if we want to talk about harmonic functions with respect to other parameters it could have dependence there as well (for example the $\alpha$ in our example above). Notice that it's not true if we only have the harmonic function defined on $B(x,r)$, since we could then solve any Dirichlet problem so that $h$ goes to zero near some point on the boundary and then the sup and inf are no longer comparable.

**Remark 93.** *Probabilistically, the way to interpret this is that some harmonic functions come from harmonic measures (for example the probability of a Brownian motion exiting at a subset of its boundary); the idea is then that this probability is comparable among all points in a smaller ball.*

Moser showed in 1964 that this inequality also holds if we replace $\Delta$ with some other uniformly elliptic operator in divergence form

$$\mathcal{A}f = \sum_{i,j=1}^n \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial f}{\partial x_j}\right)$$

for $(a_{ij}(x))_{i,j=1}^n$ a positive-definite matrix-valued function with eigenvalues uniformly bounded from above and below.

The elliptic Harnack inequality is not enough on its own to prove bounds, because it only talks about harmonic functions and not solutions to the heat equation. It turns out this inequality is usually useful for obtaining continuity estimates which help us get better regularity for solutions instead of just living in $L^2$ or a Sobolev space:

**Theorem 94** (Cheng, Yan 1975)
The elliptic Harnack inequality holds on manifolds of nonnegative Ricci curvature. Furthermore, the elliptic Harnack inequality implies Hölder regularity of the solution to the associated operator.

The idea here is called **Moser's oscillation lemma**: we define the oscillation $\operatorname{osc}_A h = \sup_A h - \inf_A h$, and for any harmonic $h$ on $B(x,2r)$ we can apply the elliptic Harnack inequality to the functions $h - \inf_{B(x,2r)} h$ and $\sup_{B(x,2r)} h - h$. This yields

$$\sup_{B(x,r)} h - \inf_{B(x,2r)} h \leq C\left(\inf_{B(x,r)} h - \inf_{B(x,2r)} h\right)$$

$$\sup_{B(x,2r)} h - \inf_{B(x,2r)} h \leq C\left(\sup_{B(x,2r)} h - \sup_{B(x,r)} h\right),$$

so if we add the two inequalities together we find that

$$\operatorname{osc}_{B(x,r)} h \leq \frac{C-1}{C+1}\operatorname{osc}_{B(x,2r)} h.$$

and thus the oscillation actually decreases as we go to smaller balls: for all $y_1, y_2 \in B(x,r)$ we have

$$|h(y_1) - h(y_2)| \lesssim \left(\frac{d(y_1,y_2)}{r}\right)^{\alpha} ||h||_{\infty}, \quad \alpha = \frac{\log\left(\frac{C+1}{C-1}\right)}{\log 2}.$$

Notice that such a bound also implies the **Liouville property** (that is, if we have a function $h : \mathbb{R}^n \to \mathbb{R}$ which is harmonic and bounded, then $h$ must be constant). Indeed, we can fix $y_1, y_2$ and let $r \to \infty$, and the right-hand side will go to zero.

We've been working with the Laplace operator so far, but now we'll switch to the heat operator. The idea is to work with spacetime cylinders for the heat equation which have the right spacetime scaling:

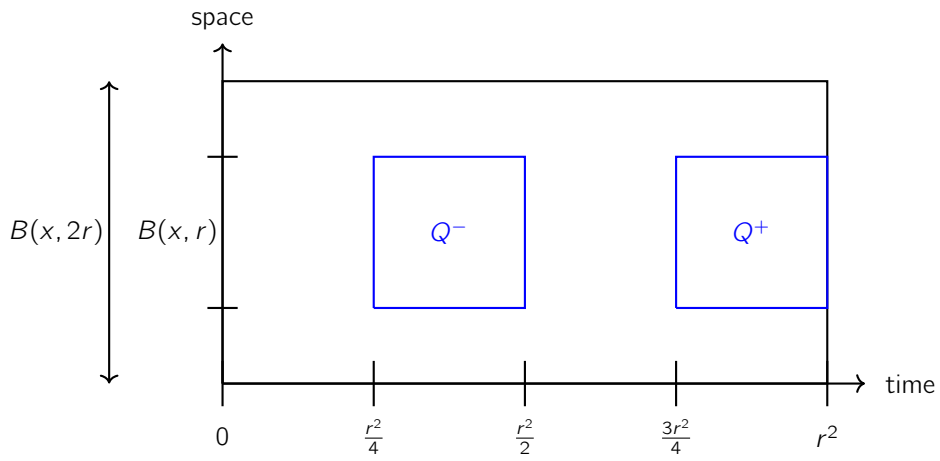
**Theorem 95** (Parabolic Harnack inequality)
Fix some dimension $n$. There exists some constant $C \geq 1$ such that for any $x \in \mathbb{R}^n$, any $r > 0$, and any solution to the heat equation $u : (0,r^2) \times B(x,2r) \to (0,\infty)$ (meaning that $\partial_t u = \Delta u$ on this domain $Q = (0,r^2) \times B(x,2r)$),

$$\sup_{Q_-} u \leq C \inf_{Q^+} u,$$

where we define the spaces

$$Q^- = \left(\frac{r^2}{4}, \frac{r^2}{2}\right) \times B(x,r), \quad \left(\frac{3r^2}{4}, r^2\right) \times B(x,r).$$

The space-time visualization of the sub-domains is shown below:




We'll now see how we can use such inequalities – what we have is basically an ultracontractivity statement. Think of $u$ as a solution to the heat equation, and the left-hand side $\sup_{Q^-}$ is some kind of a bound on the $L^\infty$ norm but only on a small part of the space. And we're controlling via the infimum over $Q^+$, which is better than controlling over $L^1$. In words, "supremum in the past is controlled by infimum in the future."

Everything here is for Euclidean spaces, but we can define an analog using the Dirichlet form (we just need a notion of balls and of harmonic functions). And Moser showed similarly that parabolic Harnack inequalities hold for uniformly elliptic operators (so we have stability under perturbation), and Li and Yau showed these inequalities also for manifolds with nonnegative Ricci curvature.

We want to get a notion of harmonic functions on strongly local regular Dirichlet forms on a metric space – from now on, we'll also abbreviate this as **MMD space**. The idea is to use a weak formulation much like when we solve PDEs:

**Definition 96**

On an MMD space with Dirichlet form $(\mathcal{E}, \mathcal{F})$, we say that a function $h \in \mathcal{F}$ is **harmonic** (also $\mathcal{E}$-harmonic) on an open set $U \subset X$ if $\mathcal{E}(h, f) = 0$ for all $f \in \mathcal{F} \cap C_c(X)$ with support contained in $U$.

In other words, the generator applied to $h$ should give us zero in a weak sense.

**Definition 97**

Let $I \subset \mathbb{R}$ be an open interval. We say that a function $u : I \to L^2(X, m)$ is **weakly differentiable** if for any function $f \in L^2(X, m)$, the function $t \mapsto \langle u(t), f \rangle_{L^2(m)}$ is differentiable on $I$.

For any weakly differentiable function $u$ and any time $t_0$, there is a unique $w \in L^2(X, m)$ such that

$$\lim_{t \to t_0} \left\langle \frac{u(t) - u(t_0)}{t - t_0}, f \right\rangle = \langle w, f \rangle_{L^2(m)}.$$

So we know how to make sense of the time-derivative and also the Laplacian, and we can combine these to get the notion of a solution to the heat equation:

**Definition 98**

Let $\Omega \subset X$ be an open set. A function $u : I \to \mathcal{F}$ is **caloric** in $I \times \Omega$ if for any $f \in \mathcal{F} \cap C_c(\Omega)$ and any $t \in I$, we have

$$\langle u'(t), f \rangle + \mathcal{E}(u(t), f) = 0.$$

(The reason the Dirichlet form is on the left-hand side is that $\mathcal{E}(u(t), f)$ is playing the role of $-\Delta u$.) The parabolic Harnack inequalities turn out to be localized enough to help us study the heat kernel (because they impose strong geometric conditions on the measure and so on), and we'll soon see ways to make that precise.

**Definition 99**

Let $(X, d)$ be a metric space and $\mu$ a Borel measure on $X$. We say that $\mu$ satisfies the **volume doubling property** (also **VD**) if there is some constant $C_D$ such that

$$0 < \mu(B(x, 2r)) \leq C_D \mu(B(x, r))$$

for all $x \in X$ and all $r > 0$.

Everything we've talked about so far satisfies this property except for $m_\alpha$ for $\alpha < -n$. Indeed, as an exercise, we can check that $(1 + |x|^2)^{\alpha/2} dx$ satisfies the volume doubling property if and only if $\alpha > -n$.

**Proposition 100**

As an exercise, we can check that if $\mu$ satisfies the volume doubling property, then by repeated iterations we have

$$\frac{\mu(B(x, R))}{\mu(B(x, r))} \lesssim \left(\frac{R}{r}\right)^\alpha, \quad \alpha = \frac{\log C_D}{\log 2}$$

for all $x \in X$ and all $0 < r < R$.

We can also sometimes reverse this inequality and put $\gtrsim$ instead of $\lesssim$:

**Definition 101**

The measure $\mu$ satisfies the **reverse volume doubling property** if there are constants $C_1, C_2 \in (1, \infty)$ and some $\beta > 0$ such that

$$\mu(B(x, R)) \gtrsim C_1^{-1} \left(\frac{R}{r}\right)^\beta B(x, r)$$

for $0 < r < R < \frac{\operatorname{diam}(X,d)}{C_2}$.

**Proposition 102**

If the metric space $(X, d)$ is connected, then volume doubling implies reverse volume doubling.

Indeed, the intuition is that if we want to compare a small radius $r$ with a large radius $10r$, there is some disjoint ball also of radius $r$ within the large ball, and those neighboring balls have similar volumes by volume doubling.

We're now ready to state the more general heat kernel bounds, modeled on what we saw with the Sierpinski gasket. We had a spacetime scaling exponent in both that case and in Euclidean space, and in fact we don't necessarily need to have power-law scaling: if we have something like the triangular lattice with the Sierpinski gasket inside each cell, we will have something like $\psi(r) \sim r^{\log_2(5)}$ for $r \leq 1$ and $\psi(r) \sim r^2$ for $r \geq 1$, since the behavior is Euclidean globally but fractal locally.

**Definition 103**

A **scale function** is a homeomorphism $\psi : [0,\infty) \to [0,\infty)$ such that there exist constants $1 < \beta_1 < \beta_2 < \infty$ and $C > 1$ with

$$C^{-1}\left(\frac{R}{r}\right)^{\beta_1} \lesssim \frac{\psi(R)}{\psi(r)} \lesssim C\left(\frac{R}{r}\right)^{\beta_2}$$

for all $0 < r < R$.

**Definition 104**

An MMD space $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies the **elliptic Harnack inequality** if there exist some constants $C, A > 1$ such that for any $x \in X$, any $r > 0$, and any $\mathcal{E}$-harmonic function $h : B(x, Ar) \to (0,\infty)$, we have

$$\text{ess sup}_{x\in B(x,r)} h \le C \text{ ess inf}_{x\in B(x,r)} h.$$

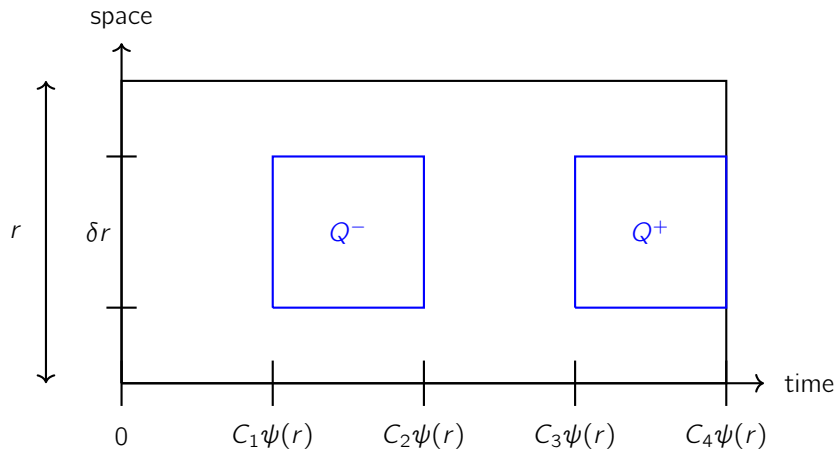
We didn't need spacetime scaling for the elliptic Harnack inequality because there is no time-dependence in the equation $\Delta h = 0$. But we'll need one now:

**Definition 105**

An MMD space $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies the **parabolic Harnack inequality** with space-time scaling $\psi$ if there exist some constants $C > 1$, $\delta \in (0,1)$, $0 < C_1 < C_2 < C_3 < C_4$ such that for any $x \in X$, any $r > 0$, and any caloric function $u$ which is nonnegative on $(0, C_4\psi(r)) \times B(x, r)$, we have

$$\sup_{Q^-} u \le C \inf_{Q^+} u$$

with $Q^-$ and $Q^+$ analogously defined as shown in the diagram below.




**Fact 106**

Note that parabolic Harnack automatically implies elliptic Harnack, since for any harmonic function $h$ the function $u(t, x) = h(x)$ is caloric. So in that sense the spacetime scaling doesn't really matter, and so people were curious whether elliptic Harnack inequalities imply parabolic Harnack inequalities with the usual Gaussian spacetime scaling. But diffusions on fractals provide the first example where we do get something different, and we can even go from fractal counterexamples to manifold counterexamples.

Indeed, take some graphs that approximate the Sierpinski gasket and replace each edge with a cylinder. Gluing those cylinders together yields a two-dimensional surface whose global geometry is fractal but local geometry is $\mathbb{R}^2$, so we'll get $r^2$ scaling at small scales but something weirder at large scales.

We'll see next time how to relate all of these inequalities to heat kernel estimates!

# 10 June 17, 2025

Last time, we introduced the elliptic and parabolic Harnack inequality on MMD spaces (corresponding to the Laplace and heat equations, respectively). We mentioned that the parabolic inequality implies some version of contractivity, which then gets us bounds on the heat kernel. We'll see that in more detail now.

**Remark 107.** *In the definition of a scale function, the lower bound on growth is currently specified to be $\beta_1 > 1$. It will turn out when the parabolic Harnack inequality holds, we actually have $\beta_1 \geq 2$, and we'll see that later on.*

We want to use the parabolic Harnack inequality at some time $t_1$ satisfying $c_1\psi(x) < t_1 < c_2\psi(x)$. We know that if $f \in L^2(X, m)$ is a nonnegative function with $||f||_1 = 1$, then $P_t f$ is caloric and nonnegative with $||P_t f||_1 \leq 1$. The Harnack inequality then tells us that

$$\sup_{B(x,\delta r)} P_{t_1}(f) \leq C \inf_{B(x,r)} P_{t_1} f \lesssim \inf_{B(x,\delta r)} P_{t_2} f \lesssim \frac{1}{m(B(x,r))}$$

(since the infimum is less than the average value on the ball). Since $t_1 \asymp \psi(r)$, we have $r \asymp \psi^{-1}(t_1)$, and thus we get ultracontractivity when we look at a small ball

$$\sup_{B(x,\delta r)} P_{t_1}(f) \lesssim \frac{1}{m(B(x,\psi^{-1}(t_1)))} \quad \text{for all nonnegative } f \in L^2 \cap L^\infty \text{ with } ||f||_1 = 1,$$

and then taking the supremum over all balls yields a heat kernel bound

$$\operatorname{ess\,sup}_{\substack{y \in B(x,\delta r) \\ z \in X}} p_{t_1}(y,z) \lesssim \frac{1}{m(B(x,\psi^{-1}(t_1)))}.$$

Once we know the existence of a heat kernel, we can fix $x$ and consider $P_t(x,y)$ as a function of $t, y$ and use the parabolic Harnack inequality on that function. Bounding inf from below therefore yields a lower bound on the heat kernel.

**Definition 108**

We say that an MMD space satisfies **sub-Gaussian heat kernel bounds** (also written **HKE($\psi$)** with spacetime scale function $\psi$ if the following holds. Define

$$\Phi(s) = \sup_{r>0}\left(\frac{s}{r} - \frac{1}{\psi(r)}\right)$$

(so if $\psi(r) = r^\beta$, then $\{\Phi(s) \asymp s^{\beta/(\beta-1)}$). Then $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies HKE($\psi$) if the heat kernel exists and satisfies the upper bound

$$p_t(x,y) \lesssim \frac{1}{m(B(x,\psi^{-1}(t)))} \exp\left(-ct\Phi\left(\frac{d(x,y)}{t}\right)\right)$$

for all $x, y \in X$ and all $t > 0$, as well as the near-diagonal lower bound

$$p_t(x,y) \gtrsim \frac{c}{m(B(x,\psi^{-1}(t)))} \text{ for all } x, y \text{ such that } d(x,y) \leq \delta\psi^{-1}(t).$$

The form of this expression comes from the same optimization argument we did previously in terms of chaining balls together, but now no longer with the ordinary Gaussian scaling. We can also ask for bounds for $x, y$ far apart, but we need to distinguish that because we don't always have this exact lower bound (for example if we use the square root of the Euclidean metric for Brownian motion on $\mathbb{R}^d$).

**Definition 109**

We say that $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies **full sub-Gaussian heat kernel bounds** (denoted **HKE$_f(\psi)$**) if we also have a lower bound of the form

$$p_t(x,y) \gtrsim \frac{1}{m(B(x, \psi^{-1}(t)))} \exp\left(-c' t \Phi\left(\frac{d(x,y)}{t}\right)\right)$$

for all $x, y \in X$ and for all $t > 0$.

**Theorem 110** (Murugan, 2020)

The following are equivalent:

1. $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies HKE$_f(\psi)$.
2. $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies HKE$(\psi)$, the volume doubling property, and the following **chain condition** on $(X, d)$: there is some constant $C \geq 1$ such that for all $n \in \mathbb{N}$ and all $x, y \in X$, there is some sequence of points $x_0 = x, x_1, \cdots, x_n = y$ with $d(x_i, x_{i+1}) \leq \frac{Cd(x,y)}{n}$ for all $i$.

In particular, recall that we got a hands-on lower bound via chaining with a sequence of points; this equivalence says that this is actually essentially a necessary condition.

The proof of this result isn't so difficult by doing a different chaining argument with another metric (defined in terms of minimum lengths of chains). But we'll focus instead on general conditions for when these inequalities hold on general spaces. We've already encountered the Nash inequality, and along the way we saw some variants (the Poincaré inequality and pseudo-Poincaré inequality) which will turn out to be relevant. Poincaré is particularly nice because it is local and thus we can work on balls instead of the whole space.

**Theorem 111** (Poincaré)

$(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies the **Poincaré inequality** (denoted **PI($\psi$)**) if there are constants $C, A \geq 1$ such that

$$\int_{B(x,r)} (f - f_{B(x,r)})^2 dm \leq C\psi(r) \int_{B(x,Ar)} d\Gamma(f,f)$$

for all $f \in \mathcal{F}, x \in X$, and $r > 0$. T

Geometrically, the idea is that we have a bottleneck through a very thin part of the space, then this creates a large variance but small energy because the function only changes in a very small part of the space. So the Poincaré inequality is telling us that there are actually lots of curves between any two points, and we'll make this intuition precise soon. To understand the relationship between functional inequalities and subgaussian heat kernels, we'll state the following result:

**Theorem 112** (Grigor'yan and Saloff-Coste, 1992)

For Brownian motion on on a Riemannian manifold, the following are equivalent:

1. The volume doubling property, along with the Poincaré inequality PI(2) with the usual spacetime scaling $|\psi(r)| = r^2$,
2. The parabolic Harnack inequality PHI(2),
3. The full heat kernel bounds $\text{HKF}_f(2)$.

The first of these conditions is useful because it is stable under perturbations. Indeed, if $g, \tilde{g}$ are two different Riemannian metrics on some manifold $M$ with tangent vectors $\xi \in TM$ satisfy $g(\xi,\xi) \asymp \tilde{G}(\xi,\xi)$. And what we're saying is that if volume-doubling and PI(2) are still preserved, then so is $\text{HKE}_f(2)$ and PHI(2).

This result was further extended to the setting of MMD spaces later on, and the question is whether these results extend beyond Gausian scaling as well. To understand how (3) implies volume doubling, we can use that

$$1 \geq \int_{B(x,2\sqrt{t})} p_T(x,y) \geq \frac{m(B(x,2\sqrt{t})}{m(B(x,\sqrt{t}))},$$

using that when $d(x,y) \leq 2\sqrt{t}$, we can use our full heat kernel bounds to get $p_t(x,y) \gtrsim \frac{1}{m(B(x,\sqrt{t}))}$ for all $x, y$ of distance at most $2\sqrt{t}$ apart. And the Poincaré inequalities are harder to see – if we look at the reflected process, we are uniformly dispersed through a ball of radius $R$, and we can relate the $L^2$ norms of the function along the heat semigroup using a similar strategy to what we saw in the Nash inequality.

We'll state a version of this stability result for general spacetime scaling now. It turns out we do not just replace 2 with $\beta$ and we do need another condition going under the name of **cutoff energy inequality**:

**Definition 113**

Suppose $U, V$ are open sets with $U \subset V$. We say that $\phi \in \mathcal{F}$ is a **cutoff function** for $U \subset V$ if $0 \leq \phi \leq 1$, $\phi$ is 1 on a neighborhood of $\overline{U}$, and $\text{supp}_m(\phi) \subset V$. An MMD space satisfies the **cutoff energy inequality** (denoted **CS($\psi$)** if there are constants $A_1, A_2$, such that for all $x, \in X$ and all $0 < r < \frac{\text{diam}(X,d)}{A_2}$, there exists a cutoff function $\phi$ for $B(x,r) \subset B(x,A_1 r)$ with

$$\int_{B(x,A_1 r)} \tilde{f}^2 d\Gamma(\phi,\phi) \lesssim C_1 \int_{B(x,A_1 r)} d\Gamma(f,f) + \frac{C_1}{\psi(r)} \int_{B(x,A_1 r)} f^2 dm.$$

We often multiply our functions by these cutoffs before applying Sobolev inequalities. In the Euclidean case, these cutoff functions take a form where they are 1 on a ball $B(x,r)$ and 0 outside a ball $B(x,Ar)$, and the problem was that it wasn't clear what the analog was for general spacetime scalings. We don't have to remember this exactly, but the point is that all of these involve integrals of a function over energies, and hence they are stable under perturbations.

Also, notice that if a function $\phi$ is constant on $B(x,A_1 r)$, then by strong locality the energy is zero on this set and we find that we have an upper bound on energy

$$\mathcal{E}(\phi,\phi) \lesssim \frac{m(B(x,r))}{\psi(r)}$$

that is, we get an upper bound on capacity between two balls. Poincaré typically gives lower bound on energies – that is, PI($\psi$) gives a matching lower bound of the form for any cutoff function $\phi$

$$\mathcal{E}(\phi,\phi) \gtrsim \frac{m(B(x,r))}{\psi(r)}$$

(because the variance term is of constant order), so this is giving us something in the other direction.

If $A, B$ are disjoint, we can consider the set of cutoff functions

$$\mathcal{F}(A,B) = \{f \in \mathcal{F} : f = 1 \text{ on a neighborhood of } A \text{ and } 0 \text{ on a neighborhood of } \},$$

and define the **capacity**

$$\text{Cap}(A,B) = \inf\{\mathcal{E}(f,f) : f \in \mathcal{F}(A,B)\}.$$

We now introduce another condition:

**Definition 114**

We say that $(X,d,m,\mathcal{E},\mathcal{F})$ satisfies the **Cap($\psi$)** condition if for all $x, y \in X$ and all $\alpha < \frac{\text{diam}(X)}{A_2}$, we have

$$\text{Cap}\left(B(x,r), B(x,A_1 r)^c\right) \asymp \frac{m(B(x,r))}{\psi(r)}.$$

**Theorem 115** (BB, 2004; BBK, 2006)

The following are equivalent for a space $(X,d,m,\mathcal{E},\mathcal{F})$:

1. The parabolic Harnack inequality PHI($\psi$),
2. Volume doubling along with HKE($\psi$),
3. (Stable characterization) Volume doubling, Poincaré inequalities PI($\psi$) and cutoff energy inequality CS($\psi$),
4. Volume doubling, the elliptic Harnack inequalities, and the capacity bound Cap($\psi$).

We saw last time that the parabolic Harnack inequality implies the elliptic one, and we now have some kind of converse: some additional capacity bounds give us the other direction too. It's currently open to show whether condition (3) can be replaced with volume doubling, PI($\psi$), and the capacity upper bound(the lower bound is already implied by Poincaré); in fact there are various situations where it's actually possible to show this weaker condition but where it's difficult to establish the cutoff energy inequality. So resolving this in the positive would have some applications!

We'll see some applications of all of this in the next few lectures – it will turn out to be useful to obtain bounds for transformations of a Dirichlet form (for example reflection or killing at a boundary).

**Example 116**

First, we'll show how to establish a Poincaré inequality on $\mathbb{R}^n$, and it'll turn out to work for a lot more examples as well.

*Proof of Poincaré on $\mathbb{R}^n$.* We saw a proof for the Sierpinski gasket using the self-similarity energy – our proof here will be more geometric. We wish to show that

$$\int_{B(x,r)} |f(y) - f_{B(x,r)}|^2 dy \le C r^2 \int_{B(x,Ar)} |\nabla f(y)|^2 dy.$$

We'll use the usual trick about writing variance as an expectation over the product measure

$$\int |f(y) - f_{B(x,r)}|^2 dy = \frac{1}{2m(B(x,r))} \int_{B(x,r)} \int_{B(x,r)} (f(y) - f(z))^2 dy dz.$$

We want to control this integrand using the gradient, and the idea is the following **pencil of curves argument**. For a pair of points $x, y$, we join them by some curve. One such example would be a straight line segment, but here we will think about all possible points $w$ on the perpendicular bisector between $y$ and $z$, and consider the curve $\gamma$ which draws a straight line segment from $y$ to $w$ and then $w$ to $z$, as long as the distance $d(w,z) = d(w,y)$ is at most $d(y,z)$ (so in other words, a cone of angle 60 degrees of spread from the vertical).

Let $\gamma_{y,z}$ be a **random** such curve with $w$ uniformly among the allowed possibilities.

For any fixed such curve, we have by the fundamental theorem of calculus and the triangle inequality that

$$f(y) - f(z) \leq \int_{\gamma_{y,z}} |\nabla f(\gamma_{y,z}(s)| ds,$$

and so squaring and taking expectation we get

$$\begin{aligned}(f(y) - f(z))^2 &\leq \mathbb{E}\left[\left(\int_{\gamma_{y,z}} |\nabla f(\gamma_{y,z}(s)| ds\right)^2\right] \\ &\lesssim \mathbb{E}\left[L(\gamma)\int |\nabla f(\gamma(s))|^2 ds\right]\end{aligned}$$

by Cauchy-Schwarz, where the length $L(\gamma)$ is of order $r$. Choosing uniformly among the allowed point $p$ along the curve is then similar to choosing a uniform angle up to constant factors within our allowed region, so we can further write this as (here note that

$$(f(y) - f(z))^2 \lesssim r \int |\nabla f(p)|^2 (|y-p|^{-n+1} + |z-p|^{-n+1}) dp.$$

We now plug this back into our expression. By triangle inequality, we know that if $y, z$ are distance $r$ from $x$, then $p$ can be at most distance $2r$ away from them, and thus

$$\int_{B(x,r)}\int_{B(x,r)} (f(y) - f(z))^2 \lesssim r \int_{B(x,3r)}\int_{B(x,r)}\int_{B(x,r)} |\nabla f(p)|^2 (|y-p|^{-n+1} + |z-p|^{-n+1}) dy dz dp$$

We can now do these integrals in polar coordinates, and we get a bound $r^n r^2 \int_{B(x,3r)} |\nabla f|^2 dm$, as desired because the $r^n$ cancels out with the $m(B(x,r))$ term. □

Our next result talks about what possible joint values of volume-growth and exit time behavior are possible for a space.

**Theorem 117** (Murugan, 2025+)

Let $V, \psi : [0,\infty) \to [0,\infty)$ be two homeomorphisms such that $\psi$ is a scale function and $V$ satisfies (this should be the measure of a ball of radius $r$) $\frac{V(R)}{V(r)} \lesssim \left(\frac{R}{r}\right)^\alpha$ for all $0 < r < R$. Suppose we have an MMD space $(X, d, m, \mathcal{E}, \mathcal{F})$ of infinite diameter satisfying full sub-Gaussian heat kernel bounds $\mathrm{HKE}_f(\psi)$, and we have $m(B(X,r)) \asymp V(r)$ for all $r > 0$. Then $V$ and $\psi$ must have some mutual constraints; specifically,

$$\left(\frac{R}{r}\right)^2 \lesssim \frac{\psi(R)}{\psi(r)} \lesssim \frac{R}{r} \cdot \frac{V(R)}{V(r)}.$$

(In particular if $V(r) = r^{d_f}$ and $\psi(r) = r^{d_W}$, this implies that $2 \leq d_W \leq 1 + d_f$.) Conversely, if $V$ and $\psi$ satisfy these mutual constraints, then there is some space $(X, d, m, \mathcal{E}, \mathcal{F})$ with full sub-Gaussian heat kernel bounds and the specified volume growth.

(Barlow previously proved this in the special case of random walks on graphs.) Let's understand where those mutual constraints come from:

*Proof of upper bound.* The reason we require $\frac{\psi(R)}{\psi(r)} \lesssim \frac{R}{r}\frac{V(R)}{V(r)}$ is that if $\phi$ is a cutoff function for $B(x,r) \subset B(x,Ar)$, by Poincaré inequality we have

$$\mathcal{E}(\phi,\phi) \gtrsim \frac{m(B(x,R))}{\psi(R)} \asymp \frac{V(R)}{\psi(R)},$$

and we'll now use a chain of balls $B_0, \cdots, B_n$ at the smaller scale $r$ to get from inside $B(x,r)$ to outside the larger ball $B(x,R)$. The number of steps we must take is on the order of $\frac{R}{r}$ (since having full heat kernel bounds implies conditions on these chains), and then the average of the function on the chained balls satisfies

$$\boxed{|1-0|} = |\phi_{B_0} - \phi_{B_n}| \lesssim \left|\sum_{i=0}^{n-1} (\phi(B_i) - \phi(B_{i+1}))\right|^2$$
$$\lesssim \frac{R}{r}\sum_{i=0}^{n-1} |\phi_{B_i} - \phi_{B_{i+1}}|^2.$$

But now we can enclose both balls $B_i, B_{i+1}$ in a single ball $B_i'$ of order $2r \sim r$ as well, and by the Poincaré inequality we thus get an upper bound on this quantity of

$$\frac{R}{r}\sum_{i=0}^{n-1}\int_{AB_i'} d\Gamma(\phi,\phi)$$

These balls can overlap with each other but not by much, and thus we can use this last expression to bound something with the energy of $\phi$: we get $\frac{\psi(r)}{V(r)}\frac{R}{r}\mathcal{E}(\phi,\phi) \le \boxed{\frac{\psi(r)}{V(r)}\frac{R}{r}\frac{V(R)}{\psi(R)}}$, which is exactly what we wanted. □

Notice that this bound is tight when we have "only one path from the small ball to the large ball," but on the other hand, the lower bound is exactly what happens in the Euclidean case. So what we're saying is that "all spaces are somewhere between Euclidean and tree-like!"

# 11 June 19, 2025

We stated Theorem 117 last time, explaining that a space with the same volume growth around all points must have spacetime scaling growing at least quadratic (so Gaussian spacetime is an extreme situation) and at most linear times the volume growth. The idea behind the proof is to use estimates at a smaller scale $r$ to control the larger scale $R$.

*Proof of Theorem 117, lower bound.* We know that the sub-Gaussian heat kernel bounds $\mathrm{HKE}_f(\psi)$ imply capacity bounds (here originally the numerator reads $m(B(x,r))$

$$\mathrm{Cap}(B(x,r), B(x,Ar)^c) \asymp \frac{V(r)}{\psi(r)}.$$

We want to use this at smaller scales to get bounds at larger scales. For this, consider an annulus of inner and outer radius $R$ and $R+r$, respectively; we will show that

$$\mathrm{Cap}(B(x,R), B(x,R+r)^c) \lesssim \frac{m(B(x,R+r)\setminus B(x,R))}{\psi(r)}.$$

This next point is one of the key ideas in this area: we cover with balls of the smaller scales. Fix an $N = \frac{A^{-1}r}{2}$-net of $B(x,R)$ (meaning that we have some maximal $\frac{A^{-1}r}{2}$-separated subset of $B(x,R)$); given such a family of points). Maximality implies that the balls $\bigcup_{n\in N} B\left(n, \frac{A^{-1}r}{2}\right)$ cover $B(x,R)$. For each such point $n$, pick a cutoff function $\psi_n$

for the annulus between $B(n, \frac{A^{-1}r}{2})$ and $B(n, \frac{r}{2})$, with energy controlled as

$$\mathcal{E}(\psi_n, \psi_n) \lesssim \frac{m(B(n, \frac{r}{2}))}{\psi(r)}.$$

(by the doubling property we can assume that balls of different radii are comparable).

We can now define the function $\psi = \max_{n \in N} \psi_n$; this is a cutoff function for $B(x, R)$ which is zero outside of $B(x, R + \frac{r}{2})$. It's not so clear that this function belongs in the domain of the form or how to estimate its energy, but a useful fact about the energy measure in terms of the associated process $(Y_t)$ is the following:

$$\int_X f d\Gamma(\phi, \phi) = \lim_{t \downarrow 0} \frac{1}{t} \mathbb{E}_{f,m}[\phi(Y_t) - \phi(Y_0)]$$

(here $\mathbb{E}_{f,m}$ means we have some starting distribution with density $f$ relative to $m$). Indeed, we can decompose $\phi(Y_t)$ into a martingale and finite-variation part by Ito's formula, with the growth of the quadratic variation part depending on $|\nabla \phi|^2$. The left-hand side becomes the energy when $f = 1$, and if we plug in $\phi = \psi$ we can also use

$$|\psi(Y_t) - \psi(Y_0)|^2 \leq \max_{n \in N} |\psi_n(Y_t) - \psi_n(Y_0)|^2 \leq \sum_{n \in N} |\psi_n(Y_t) - \psi_n(Y_0)|^2,$$

so the energy of the maximum is less than the sum of the energies and we have the same inequality for energy measures. And this gives us a probabilistic interpretation of the energy measure in terms of the quadratic variation. But the point is that

$$\Gamma(\psi, \psi) \leq \sum_{n \in N} \Gamma(\psi_n, \psi_n),$$

and because $\psi$ is identically equal to 1 on the ball $B(x, R)$ (as the maximum of cutoff functions) we don't have to worry about energy contributions by strong locality: $\Gamma(\psi, \psi)(B(x, R)) = 0$. (Similarly there are no energy contributions outside $B(x, R + r)$.) Thus we get the estimate

$$\begin{aligned} \mathcal{E}(\psi, \psi) &= \Gamma(\psi, \psi)(B(x, R + r) \setminus B(x, R)) \\ &\leq \sum_{n \in N} \Gamma(\psi_n, \psi_n)(B(x, R + r) \setminus B(x, R)). \end{aligned}$$

Notice also that most of the $\psi_n$s don't contribute to this sum, because any $\psi_n$ that is not near the boundary will have support inside the smaller ball. And with the remaining ones we can use our estimate on $\mathcal{E}(\psi_n, \psi_n)$, and we know that the balls will not overlap too much. Thus the result is that we get the bound

$$\mathcal{E}(\psi, \psi) \lesssim \frac{m(B(x, R + r) \setminus B(x, R))}{\psi(r)}.$$

We'll now build a cutoff function at the scale $R$: if we have balls $B(x, R)$ and $B(x, AR)$, we will break that up into $k$ concurrent annuli of depth $r$, meaning that $k \sim \frac{R}{r}$. That is, for all $1 \leq i \leq k$, we define balls $B_i = B(x, R + ir)$ and let $\psi_i$ be a cutoff function for $B(x, R + ir) \subset B(x, R + (i + 1)r)$ satisfying that previous bound

$$\mathcal{E}(\psi_i, \psi_i) \lesssim \frac{m(B(x, R + (i + 1)r) \setminus B(x, R + ir))}{\psi(r)}.$$

We then average these functions together by defining

$$\psi_k = \frac{1}{k} \sum_{i=0}^{k-1} \psi_i,$$

which will be a cutoff function for $B(x,R) \subset B(x,AR)$. Again, we want to estimate the energy, but by bilinearity

$$\mathcal{E}(\psi,\psi) = \frac{1}{k^2}\sum_{i=0}^{k-1}\mathcal{E}(\psi_i,\psi_i) \lesssim \frac{1}{k^2}\frac{m(B(x,R))}{\psi(r)}$$

because cross-terms are zero (one function is always constant in the support of the other for $\mathcal{E}(\psi_i,\psi_j)$ when $i \neq j$). Here in the last inequality we use that by the doubling property, the total sum of all the volumes of the annuli is of the same order as $B(x,R)$. But this quantity is proportional to $\left(\frac{r}{R}\right)^2 \frac{V(R)}{\psi(r)}$, and therefore what we've proved is that

$$\mathrm{Cap}(B(x,R), B(x,AR)^c) \lesssim \left(\frac{r}{R}\right)^2 \frac{V(R)}{\psi(r)},$$

but also we previously found the lower bound $\frac{V(R)}{\psi(R)} \lesssim \mathrm{Cap}(B(x,R), B(x,AR)^c)$. So putting these inequalities together yields the Gaussian lower bound, as desired. □

Working with the capacity makes the calculations simpler here; we could also interpret from the point of view of exit times from balls, but it's not so clear how to do the appropriate bounds in that case.

**Example 118**

We'll now give some ideas for how to construct spaces given some specified volume growth $V$ and spacetime scaling $\psi$, assuming they satisfy the bounds above. These ideas go under the name **Laakso spaces** (and they were constructed in 2000); these spaces are constructed so it's easy to compute the spacetime and volume profile.

Even for relatively simple looking fractals, it can be hard to compute the exponents. Indeed, as we said last time, we know that the Sierpinski carpet has some exponent which we can estimate numerically, but for which the actual value is unknown. But the other direction (given exponents, construct a space) is much more manageable.

We'll do the special case $\psi(r) = r^2$ first. The idea is to glue many copies of the line $[0,\infty)$, indexed by a certain ultrametric space. Define a uniformly bounded function $g : \mathbb{Z} \to \mathbb{N}$ (meaning that $1 \leq \inf g \leq \sup g \leq \infty$ which we should think of as a **gluing function**. Define the set of functions

$$\mathcal{U}(g) = \left\{ s : \mathbb{Z} \to \mathbb{Z} : s(k) \in [0, g(k)-1] \text{ for all } k \text{ and } \lim_{k\to\infty} s(k) = 0 \right\}.$$

So the $k$th component has $g(k)$ different possibilities, and we're taking products but restricting to eventually-zero functions so that $\mathcal{U}(g)$ can be made into a metric space. Specifically, we can set

$$d_{\mathcal{U}(g)}(s,t) = 2^{\inf\{k\in\mathbb{Z}: s(\ell)=t(\ell) \text{ for all } \ell\geq k\}},$$

which will always be finite because both sequences are eventually zero. This satisfies a version of the triangle inequality with sum replaced by maximum, and we now define the **Laasko space**

$$\mathcal{L}(g) = \frac{\mathcal{U}(g)\times[0,\infty)}{\sim},$$

with $\sim$ the smallest equivalence relation which glues points

$$(s,x) \sim (s',x') \text{ if } x = x' = (2k-1)2^n \text{ for some } k \in \mathbb{N}, n \in \mathbb{Z} \text{ and } s|_{\mathbb{Z}\setminus\{n\}} = |s'_{\mathbb{Z}\setminus\{n\}}.$$

To explain what's happening, suppose $g(n) = 3$. Then at every odd multiple of $2^n$, we glue three different copies of the positive real line together, and otherwise those copies are disjoint. These sets $W_k = \{(2k-1)2^n : k \in \mathbb{N}\}$ are typically called **level-$n$ wormholes**.

The metric on $\mathcal{L}(g)$ is then the shortest path (geodesic) metric such that each copy of $[0,\infty)$ has the usual metric. Since there are wormholes at all scales $n$ (which can be arbitrarily small), they are dense in our space. So our diffusion will behave like a Brownian motion within each real line, but each time it encounters a wormhole it goes to each of the branches with equal probability (some version of this is also called the **Walsh Brownian motion**). But of course it's hard to visualize this because we always encounter wormholes at arbitrarily small scales.

We get a measure on $\mathcal{U}(g)\times[0,\infty)$ by taking the product measure $m_{\mathcal{U}(g)}\times \mathrm{Leb}([0,\infty))$, and then the pushforward of this under the quotient map yields a measure on $\mathcal{L}(g)$. Here we're defining the measure $m_{\mathcal{U}(g)}$ to be the restriction to $\mathcal{U}(g)$ of the following measure:

$$\prod_k \begin{cases} \text{uniform probability measure on } [0,g(k)-1]\cap\mathbb{Z}, & k\le 0,\\ \text{counting measure}, & k>0. \end{cases}$$

So if we fix a function up to some negative scale $k$ and choose the rest uniformly, we can estimate the volume (because it's uniform after that on some space). We find that

$$m_{\mathcal{U}(g)}\left(B_{\mathcal{U}(g)}(s,r)\right)\asymp \begin{cases} \left(\prod_{k=1}^n g(k)\right)^{-1} & 2^{n-1}\le r\le 2^n, n\le 0,\\ 1 & 1\le r\le 2,\\ \prod_{k=1}^{n-1} g(k) & n>0, 2^{n-1}\le r<2^n. \end{cases}$$

Defining the function $V_{\mathcal{U}(g)}(r)=\prod_{k=1}^{n-1} g(k)$, the volume growth of the Laasko space is then of order $rV_{\mathcal{U}(g)}(r)$ (since we also have linear growth along the Lebesgue measure part). And we won't write down the Dirichlet form explicitly, but we can think of it as $\int (f')^2 dm_{\mathcal{L}}$ with $f'$ the derivative along the real line. Indeed, this diffusion has the property that it becomes the usual Brownian motion if we project down onto $[0,\infty)$, and we just end up with some additional randomness about how to jump at each scale.

It turns out that by varying the function $g$, we can produce any space with the specified Gaussian volume/spacetime scaling. (Indeed, making $g$ larger essentialy corresponds to having a "higher-dimensional space.") We do this by proving a Poincaré inequality (using the "pencil of curves" approach from last lecture), and that's what Laasko does in the original paper.

**Fact 119**

Generalizing beyond the Gaussian case, instead of gluing different copies of the real line, we can also glue different copies of an $r$-tree. This fixes the spacetime scaling behavior, and then we can choose a similar gluing $g$ to get the correct volume growth behavior.

Also, recall from some of our afternoon talks that the Heisenberg group doesn't embed into Euclidean space in any bilipschitz fashion – the same holds for Laasko spaces.

A probabilistic construction of diffusion on Laasko spaces was first given by Barlow and Evans in 2001, but it was hard to get heat kernel bounds until these more analytic techniques came in.

**Theorem 120** (Carron and Tewodrose, 2022)
Let $(X, d, m, \mathcal{E}, \mathcal{F})$ be an MMD space, and let $\alpha \in [1, \infty)$ be such that the corresponding diffusion process has Euclidean-like heat kernel (that is, "like Brownian motion on $\mathbb{R}^\alpha$")

$$p_t(x, y) = \frac{1}{(2\pi t)^{\alpha/2}} \exp\left(-\frac{d(x, y)^2}{2t}\right)$$

for all $x, y \in X$ and all $t > 0$. Then in fact we have $\alpha = n$ for some $n \in \mathbb{N}$, and $X = \mathbb{R}^n$, $d$ is the Euclidean metric, and the corresponding process is ordinary Brownian motion.

So the point is that if we restrict ourselves to equality rather than just having upper and lower bounds, we can't choose any $\alpha$ and in fact we need to be exactly Euclidean. Thus what we have is a result about **rigidity of the heat kernel**.

*Proof idea.* We'll see now why the volume doubling and Poincaré inequality characterization is useful. The proof uses a result of Colding and Minicozzi from 1997, which states that volume doubling and PI(2) implies the space of harmonic functions with polynomial growth (meaning that $h(x) \lesssim (1 + d(0, x))^k$) has finite dimension. (For comparison, in Euclidean space, every harmonic function of polynomial growth is actually polynomial.) So the space of harmonic functions with linear growth is finite-dimensional; in Euclidean space this dimension is $n + 1$, so what we can do is throw out the constant function and use the remaining space to get an isometry to $\mathbb{R}^n$. □

**Remark 121.** *There is a similar heat kernel rigidity result for the n-sphere which has also been proved. The analogous result for hyperbolic space is conjectured but still open.*

**Example 122**
In the rest of this lecture and the next, we'll see some more applications of characterizations of heat kernel bounds. The theme from now on is **transformations**, getting new things from old things – we'll study reflection, killing/absorbing, and time change. The idea is that once we reach the boundary of our domain, we want to either be able to kill the process or send it back inside.

We'll begin with **reflected diffusions**. For reflected Brownian motion on $\mathbb{R}^n$, one strategy is the SDE approach (also called the **Skorokhod equation**), where we define a process via the equation

$$Y(t) = Y(0) + B(t) + \int \vec{n}(Y(s)) dL_s,$$

where $\vec{n}$ is the inward-point normal vector and $L_s$ is called the **boundary local time** (we'll come back to this next week). And the other strategy (due to Fukushima) is to use Dirichlet forms directly:

$$\mathcal{E}_D(f, f) = \frac{1}{2} \int_D |\nabla f(x)|^2 dx,$$

where the domain here is the set of functions

$$\mathcal{F} = W^{1,2}(D) = \left\{ f \in L^2(D) : |\nabla f| \in L^2(D) \right\}.$$

These approaches yield the same result if $D$ is a smooth domain and we want an "ordinary" reflection. But the first method can also handle oblique reflections at some other angle (the second cannot because we no longer get a

symmetric process and thus Dirichlet forms do not apply), while the second can also handle non-smooth domains (the first can do so with an approximation via smooth domains, but that takes more work to do).

**Example 123**

We can understand the construction above in the simple one-dimensional case where the domain is $D = [0, \infty)$. In this case, we should actually have that $Y_t = |B_t|$ for $B_t$ an ordinary one-dimensional Brownian motion.

In this case, we can write down the heat kernel explicitly: we have

$$p_t^D(x, y) = p_t(x, y) + p_t(x, -y)$$

for $p_t$ the ordinary heat kernel. So we know the heat semigroup, meaning that we can calculate the Dirichlet form by taking limits of that semigroup; it's a good exercise to check that we end up getting the same $\mathcal{E}_D$ as above.

If we now move to more general spaces, we need to make sense of the condition that "$|\nabla f|$ is in $L^2$." In general, the form $(\mathcal{E}_D, W^{1,2}(D))$ that we construct will **not be a regular Dirichlet form on** $L^2(\overline{D})$:

**Example 124**

Consider the space $\mathbb{R}^2 \setminus (-\infty, 0] \times \{0\})$, which is the plane with a slit. Defining a reflected Brownian motion on this space means we cannot take the space to be $L^2(\mathbb{R}^2)$, because it would mean a single point on the negative real line has to correspond to two points in the state space.

To get around this, Fukushima defines a Dirichlet form on a certain abstract closure of $D$, called the **Martin-Kuramochi compactification**:

**Definition 125**

Let $(X, d, m, \mathcal{E}, \mathcal{F})$ be an MMD space, and let $U \subset X$ be a **domain** (here we mean a nonempty open proper connected subset of $X$). Define the function space $\mathcal{F}_{\text{loc}}(U)$ to be the $m$-equivalence class of functions $f$ on $U$, such that for every relatively compact subset $V$ of $U$, we have $f1_V = f^\sharp f 1_V$ for some $f^\sharp \in \mathcal{F}$. We then define

$$\mathcal{F}(U) = \left\{ f \in \mathcal{F}_{\text{loc}}(U) : \int_U f^2 dm + \int_U \Gamma_U(f, f) < \infty \right\},$$

where $\Gamma_U(f, f)$ is the energy measure for $\mathcal{F}_{\text{loc}}(U)$ which restricts to any $V$ via

$$\Gamma_U(f, f)|_V = \Gamma(f^\sharp, f^\sharp)|_V .$$

The distribution of the gradient only depends on the local behavior of the functions, so it makes sense that this may yield a sensible Dirichlet form. We have to check (via strong locality) that even though there are multiple different $f^\sharp$s that we can take, $\Gamma_U(f, f)$ is still well-defined.

**Remark 126.** *We need to be able to pick different $f^\sharp$s for different relatively compact subsets $V$, because otherwise we would be specifying that $f$ extends to a single global function in $W^{1,2}$; that turns out to be a strictly smaller space in general. For example we may want to allow functions $f$ which are 1 on one side of the slit plane and 0 on the other.*

We thus get a bilinear form $(\mathcal{E}_U, \mathcal{F}(U))$ which will correspond to reflected diffusion by setting $\mathcal{E}_U(f, f) = \Gamma_U(f, f)(U)$. And next time, we'll see some conditions to get us a regular Dirichlet form on $\overline{U}$.

# 12 June 20, 2025

We discussed two approaches last time for discussing reflection at a domain – let's continue elaborating on the Dirichlet form approach today. Since the form is strongly local, we can define an energy measure for any function in $\mathcal{F}_{\text{loc}}(U)$ (this is kind of like "agreeing on charts for manifolds"), and the questions we're interested in are the following:

- When is $(\mathcal{E}_U, \mathcal{F}(U))$ a regular Dirichlet form on the closure $L^2(\overline{U}, m)$?
- When does the corresponding reflected diffusion inherit behavior of diffusion on the ambient space? (For example, can we get sub-Gaussian heat kernel bounds for the reflected diffusion as well?)

We'll be able to give a positive answer for a certain class of domains:

**Definition 127**

A **uniform domain** is a domain $D$ such that there is a constant $A$ where the following condition holds: for any pair of points $x, y \in U$, we can connect them with a curve $\gamma_{x,y}$ such that

$$\text{diam}(\gamma) \lesssim A d(x,y), \quad \text{dist}(z, U^c) \geq \frac{1}{A}(d(x,z) \wedge d(z,y)) \text{ for all } z \in \gamma_{x,y}.$$

In other words, our curve does not stray too far from $x$ and $y$, and we can avoid getting too close to the boundary of the domain.

Generally domains that do not satisfy these conditions have certain (outward or inward) cusps or slits, but for example the graph of a Lipschitz function will be the boundary of a uniform domain. And even things like the snowflake domain are uniform even though they are very fractal.

**Theorem 128** (Rajala)

Consider a metric space $(X, d)$ satisfying the doubling property (so every ball of radius $r$ can be covered by some constant $N$ number of balls of radius $r/2$) and such that there exists some $C \geq 1$ such that for all $x, y \in X$ there is a curve $\gamma_{x,y}$ whose length is at most $C d(x,y)$. Then for any bounded domain $\Omega$ of $X$ and any $\varepsilon > 0$, there exist uniform domains $\Omega_i, \Omega_o$ ("inside" and "outside") such that $\Omega_i \subset \Omega \subset \Omega_o$, but if we take $\varepsilon$-neighborhoods we have $\Omega_o \subset \Omega^\varepsilon$ and $\Omega_i^c \subset \Omega_\varepsilon^c$.

So any bounded domain can be approximated by uniform domains from inside and outside.

**Theorem 129** (Murugan, 2024)

Suppose $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies HKE($\psi$) heat kernel bounds for some scale function $\psi$ and has volume doubling. Then $(\mathcal{E}_U, \mathcal{F}(U))$ is a regular, strongly local Dirichlet form on $L^2(\overline{U}, m)$, and the corresponding reflected diffusion also satisfies HKE($\psi$).

The case where $\psi(r) = r^2$ was already known (and due to Gyrya and Saloff-Coste from 2011), and in fact it was proven more generally for inner uniform domains $U$ (that is, uniform with respect to the intrinsic metric). But the proof here for more general scale functions doesn't work for inner uniform domains, and it's natural to conjecture that the theorem should also be true for inner uniform domains as well (but now using the intrinsic metric).

To understand why Gaussian and sub-Gaussian cases can be rather different, consider the following:

**Theorem 130** (Kajino, Murugan 2020)

Suppose that $(X, d, m, \mathcal{E}, \mathcal{F})$ is an MMD space that satisfies HKE($\psi$) with $\psi(r) = r^{d_W}$ for some $d_W \geq 2$.

1. If $d_W = 2$, then we have absolute continuity $\Gamma(f, f) \ll m$ for all $f \in \mathcal{F}$, and additionally $m \ll \Gamma(f, f)$ for a dense set of functions $f \in \mathcal{F}$.
2. However, if $d_W > 2$, then the measures are singular: $\Gamma(f, f) \perp m$ for all $f \in \mathcal{F}$.

*Proof idea of Theorem 129.* The main idea is to use an extension of functions in $\mathcal{F}(U)$ to $\mathcal{F}$; the original idea is due to Jones (in 1981), where it's shown that $f \in W^{k,p}(U)$ can be extended to $W^{k,p}(\mathbb{R}^n)$. (So already this part of the proof will not work for inner regular domains.) The generalization is the following:

**Theorem 131**

Let $(X, d, m, \mathcal{E}, \mathcal{F})$ be as in the previous theorem, and let $U$ be a uniform domain. Let $(\mathcal{E}_U, \mathcal{F}(U))$ denote the quadratic form corresponding to reflected diffusion. Then there is an extension map $E : \mathcal{F}(U) \to \mathcal{F}$, such that $E(f)|_U = f$ for all $f$. Furthermore, there exist constants $C, K \in (1, \infty)$ such that we get the boundedness properties for the energy

$$\Gamma(E(f), E(f))(B(x, r)) \leq C\Gamma_U(f, f)(U \cap B(x, Kr))$$

for all $0 < r < c \operatorname{diam}(U)$, and we also bound the $L^2$ norms via

$$\int_{B(x,r)} |E(f)|^2 dm \leq C \int_{U \cap B(x,Kr)} f^2 dm$$

for all $r > 0$.

In particular, taking $r \to \infty$, this implies that

$$\int_X |E(f)|^2 dm \leq C \int_U f^2 dm, \quad \mathcal{E}(E(f), E(f)) \leq C \left( \mathcal{E}_U(f, f) + \frac{1}{\psi(\operatorname{diam}(U))} \int f dm \right),$$

and these kinds of global bounds are what Jones was originally interested in (extending continuously to a function on the whole space). But we have to care about local inequalities and thus we need extension properties at various scales.

We'd like to get pointwise bounds on things like the gradient, but because the energy measure is singular with respect to the reference measure there isn't much hope of doing that. Even in the Euclidean setting, this is an interesting question: for some function $f \in W^{1,2}(\mathbb{R}^n)$, can we estimate $\int |\nabla f|^2 dx$ without any pointwise estimates on $|\nabla f|$? The answer (due to Korevaar and Schoen in 1993) is the following:

**Fact 132**

A function $f \in L^2(\mathbb{R}^n)$ belongs to $W^{1,2}(\mathbb{R}^n)$ if and only if the quantity

$$\limsup_{r \downarrow 0} \int_{\mathbb{R}^n} \fint_{B(x,r)} \frac{(f(x) - f(y))^2}{r^2} dy dx < \infty,$$

and in such a setting the left-hand side is asymptotic to $\int |\nabla f|^2 dx$.

More generally if we have HKE($\psi$) estimates, we can then extend this naturally (this is due to Grigor'yan, Hu, and Lau) by replacing $dydx$ by $m(dy)m(dx)$, replacing $r^2$ with $\psi(r)$, integrating over the metric space instead of $\mathbb{R}^n$, and replacing $\int |\nabla f|^2 dx$ with $\mathcal{E}$. So up to a constant factor, we can characterize purely off the metric space structure!

**Remark 133.** *The extension map itself is constructed using* **Whitney covers**. *The definition is as follows: for any* $\varepsilon \in (0, \frac{1}{2})$ *and any open set* $U$*, a collection of balls* $\{B(x_i, r_i) : x_i \in U, r_i > 0\}$ *is called an* ***ε*-Whitney cover** *if (1) the balls are pairwise disjoint, (2)* $r_i = \frac{\varepsilon}{1+\varepsilon} dist(x_i, U^c)$ *(so the radius is essentially proportional to the distance to the boundary), and (3) slight dilations cover the space in the sense that* $\bigcup_{i\in I} B(x_i, 2(1+\varepsilon)r_i) = U$.

*So now to explain the construction, for simplicity assume* $U$ *and* $U^c$ *are unbounded domains (for example if our domain is above the graph of a Lipschitz function) and construct Whitney covers* $\mathcal{W}(U), \mathcal{W}(V)$ *for* $U$ *and* $(U^c)^\circ$*. We then construct some map* $Q : \mathcal{W}(V) \to \mathcal{W}(U)$ *so that* $diam(Q(B)) \asymp diam(B) \asymp dist(x_B, x_{Q(B)})$*. For a function* $f$ *only defined on* $U$*, we can then define*

$$E(f) = f1_U + \sum_{B\in\mathcal{W}(V)} f_{Q(B)}\psi^B,$$

*where* $f_{Q(B)}$ *is the average of the function* $f$ *on the image* $Q(B)$*, and* $\{\psi^B\}$ *is a partition of unity of* $V = (U^c)^\circ$ *(so that* $\sum \psi^B = 1_V$*); we can think of those as basically being cutoff functions except that they should have small energy:* $\mathcal{E}(\psi^B, \psi^B) \asymp \frac{m(B)}{\psi(r)}$ *for* $r = r(B)$.

To see now why the extension theorem should imply good heat kernel bounds for the reflected diffusion, first we'll see why it defines a regular Dirichlet form.

- To show that it is closed, first pick a Cauchy sequence $(f_n)$ under $(\mathcal{E}_U)_1$. Then the extended functions $(E(f_n))$ also form a Cauchy sequence (by the energy and $L^2$ bounds in our extension theorem), which will converge to some $f$ which we can restrict on $U$. In fact, this tells us that for uniform domains $\mathcal{F}(U)$ can be viewed simply as $\{f_U : f \in \mathcal{F}\}$, and that is not always true for things like slit domains.
- To show that it is regular, we must further use the fact that $E$ maps continuous functions $C(U)$ to continuous functions $C(X)$ (where continuous means existence of a continuous representative). Indeed, any function on the whole space can be approximated by continuous functions, and so every function on the local space is the limit of continuous functions with compact support.

From here, Poincaré inequalities turn out to be easy to prove from the extension property as well. Pick some $f \in \mathcal{F}(U)$, and bound the variance via the integral on the whole space:

$$\begin{aligned}\int_{U\cap B(x,r)} \left|f - f_{U\cap B(x,r)}\right|^2 dm &= \inf_\alpha \int_{U\cap B(x,r)} |f-\alpha|^2 dm \\ &\le \inf_\alpha \int_{B(x,r)} |E(f)-\alpha|^2 dm.\end{aligned}$$

Now using the Poincaré inequality for the original diffusion and then using one of the bounds from the extension map yields the result: we can bound by $\psi(r)\int_{B(x,Ar)} \Gamma(E(f), E(f))d\mu \lesssim \psi(r)\int_{U\cap B(x,KAr)} \Gamma(f,f)$. This is only true up to order of the diameter, but then a covering argument gets us the rest. □

From a probabilistic perspective, it is a bit unnatural to have to look outside $U$ to do the proof, so it'd be nice to have a strategy which avoids this.

**Fact 134**

If we knew that HKE($\psi$) is equivalent to volume doubling, PI($\psi$) and Cap($\psi$), then we would be able to obtain heat kernel bounds HKE($\psi$) for reflected diffusions as well. Indeed, capacity bounds are easy to show for the reflected diffusion, but cutoff energy inequalities are more difficult.

**Example 135**

We'll now turn to diffusions that are killed upon exiting; first we need to explain how to define the Dirichlet form. Let $Y_t$ be a process corresponding to $(X, d, m, \mathcal{E}, \mathcal{F})$, and let $\tau_U$ be the first time the process goes outside $U$. our goal is to define the **killed** or **absorbing diffusion**

$$Y_t^U = \begin{cases} Y_t & t < \tau_U, \\ \Delta & \tau \geq \tau_U, \end{cases}$$

where recall that $\Delta$ is the cemetery state.

Much like with reflected diffusions, the general characterization of using functional inequalities like Poincaré and cutoff energy is very useful, but there is an additional step involved here. By standard convention we know that we tend to take functions to be zero on the cemetery state, or more precisely we set $P_t f(x) = \mathbb{E}_x [f(Y_t) 1\{t < \zeta\}]$. Thus we can define the class of functions which are "zero outside $U$:"

$$\mathcal{F}^0(U) = \left\{ f \in F : \tilde{f} = 0 \ \mathcal{E}\text{-q.e. on } X \setminus U \right\}$$

(Indeed, this is the right definition to make – if we're just removing a single point from something like 3D Brownian motion, we shouldn't end up with anything different because almost surely we won't visit that point.) We can thus define (note the tricky notation here, with $\mathcal{E}_U$ meaning reflected and $\mathcal{E}^U$ meaning killed)

$$\mathcal{E}^U = \mathcal{E}|_{\mathcal{F}^0(U) \times \mathcal{F}^0(U)},$$

which yields a regular Dirichlet form $(\mathcal{E}^U, \mathcal{F}^0(U))$ in $L^2(U, m)$ which corresponds to $Y_t^U$. So no additional conditions on the domain are necessary – we can just define the Dirichlet form in this straightforward way.

The main idea for analyzing this is to use the **Doob $h$-transform** – for this, we need to introduce boundary conditions for functions only defined on $U$.

**Definition 136**

Let $V \subset U$ be two open sets. Define the "functions on $V$ with zero boundary conditions with respect to $U$" $\mathcal{F}_{\text{loc}}^0(U, V)$ as follows: it is the set of $m$-equivalence classes of functions $f$ such that for every set $A$ where $A$ is open, $\overline{A}$ is compact, and $\overline{A} \cap \overline{U \setminus V} = \varnothing$, we have $f = f^\sharp$ $m$-a.e. on $A$ for some $f^\sharp \in \mathcal{F}^0(U)$.

In our previous definitions of this kind, we always looked at compact subsets of $V$, but now we're also allowing boundary conditions as long as they're fully inside $V$. So $\mathcal{F}_{\text{loc}}^0(U, V)$ is some strictly smaller subset of the $\mathcal{F}_{\text{loc}}(V)$ we defined before.

**Fact 137**

We'll restrict to unbounded uniform domains $U$ (though with some modifications things also work for the bounded case). Then up to constants, there exists a unique harmonic function $h > 0$ on $U$, such that $h \in \mathcal{F}_{\text{loc}}^0(U, U)$.

**Example 138**

If we consider Brownian motion on $\mathbb{R}^n$ and $U$ is the "first orthant" $(0, \infty)^n$, then the harmonic function turns out to just be $h(x_1, \cdots, x_n) = \prod_{i=1}^n x_i$. And if $U = \{x \in \mathbb{R}^n : |x| > 1\}$, then we instead have $h(x) = \log |x|$ for $n = 2$ and $h(x) = 1 - |x|^{2-n}$ for $n \geq 3$ (for $n = 1$ our domain is no longer connected).

So in any of these domains, a Brownian motion is almost surely going to exit the domain. Doob wanted to condition on the Brownian motion to not exit for a very long time, and that's the idea of the $h$-transform – we can describe the process in a very nice way.

**Definition 139**

Let $M_h$ be the multiplication-by-$h$ operator $L^2(U, h^2 m) \to L^2(U, m)$; with our definition this operator is unitary. The semigroup we get from conditioning is the following: let $P_t^U$ be the semigroup corresponding to the process killed upon exiting. We can then define a new semigroup via conjugation

$$P_t^{U,h} = M_h^{-1} \circ P_t^U \circ M_h.$$

We can check that this is also a strongly continuous contractive semigroup on $L^2(U, h^2 \cdot m)$, and (because $h$ is harmonic) that the Markov property holds. Indeed, $P_t^{U,h}\mathbf{1} = \frac{P_t^U h}{h}$, and now this is less than $\mathbf{1}$ by harmonicity. So now if $p_t^U(\cdot,\cdot)$ is the heat kernel of the killed semigroup $P_t^U$, then the heat kernels are also related via conjugation relations

$$p_t^U(x,y) = h(x)h(y)p_t^{U,h}(x,y).$$

Thus estimating the heat kernel of the original group is more or less the same as estimating the $h$-transform semigroup.

**Example 140**

Consider Brownian motion on the half-line $U = (0,\infty)$ exited at 0. We know that $h(x) = x$, $h^2(x)dx = x^2 dx$, so the $h$-transformed process takes the form

$$\mathcal{E}^{(U,h)}(f,f) = \int_0^\infty (f'(x))^2 x^2 dx,$$

which is the familiar three-dimensional Bessel process.

Once we know the Dirichlet form, we can compute

$$\mathcal{E}^{U,h}(f,f) = \mathcal{E}(M_h(f), M_h(f)) \text{ for all } f \in \mathcal{F}^{U,k} = M_H^{-1}(\mathcal{F}^{(0)}(U)).$$

Whenever we have Gaussian or sub-Gaussian heat kernel bounds and a uniform domain whose complement is big enough, this $h$-transform process has infinite lifetime and will not be killed. So we have a relation between the $h$-transform semigroup and the ordinary semigroup; instead of trying to directly estimate our heat kernel **we can estimate** $P_t^{U,h}(x,y)$ **instead**. To do this, we have the following:

**Theorem 141** (Gyrya, Saloff-Coste 2011, Lierl 2021)

Let $U$ be uniform and unbounded. Then if the MMD space $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies HKE($\psi$), then $(\mathcal{E}^{U,h}, \mathcal{F}^{U,h})$ on $L^2(U, h^2 \cdot m)$ satisfies HKE($\psi$).

**Remark 142.** *In the bounded case, instead of looking at the harmonic functions, we can similarly define the h-transform using the first eigenvector even though it's not harmonic. This then yields an extra exponential term in the estimates coming from the nonzero eigenvalue.*

The proof proceeds by showing PI($\psi$) and CS($\psi$) for $(\mathcal{E}^{(U,h)}, \mathcal{F}^{(U,h)})$ on $L^2(U, h^2 \cdot m)$. The idea is that we stay far enough away from the boundary that we can prove Poincaré using balls in a Whitney cover using Harnack inequalities, and then with covering we can get it for everything.

**Fact 143**

To understand the relationship between the heat kernel $p_t$ for the whole space and $p_t^U$ for the whole process, clearly $p_t \geq p_t^U$ but we want to know "by how much." We have the Dynkin-Hunt formula

$$p_T(x,y) = p_t^0(x,y) + \mathbb{E}_x\left[p_{t-\tau_u}(Y_{\tau_u}, y)1\{\tau_u < t\}\right]$$

corresponding to the exit time being larger and smaller time $t$, respectively, but this isn't very useful for estimating compared to the $h$-transform techniques mentioned before.

# 13 June 23, 2025

In the last two lectures, we saw how characterizations of the heat kernel can be used to analyze reflected and killed diffusions. We mentioned in particular the $h$-transform process, but we didn't mention how we actually obtain the function $h$. It turns out to be not so explicit, coming out of a subsequential limit construction:

**Example 144**

We've been looking at MMD spaces $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfying HKE$(\psi)$ and looking at what happens when we kill our process outside a uniform domain $U$. We'll now restrict our attention to $\mathbb{R}^n$ for simplicity, and we'll assume that the killed process $(Y_t^U)$ is transient. (If it's recurrent, then the complement of $U$ is a polar set and thus the heat kernel will be the same as $Y_t$ itself.)

In this setting, the **Green's function** is the function $g_U$ such that $\Delta g_U(x,\cdot) = -\delta_x$ for all $x \in U$ and $g_U(x,\cdot) = 0$ on $\partial U$ (we can think of this as the occupation density for the associated process). We see that $g_U(x,y) = g_U(y,x)$ is nonnegative for all $x, y \in U$.

We're assuming that $U$ is unbounded (otherwise things depend on the first Dirichlet eigenfunction), so we can pick some sequence of points $x_n \uparrow \infty$ and fix some base point $x_0 \in U$. Normalize the function and define

$$h_n(\cdot) = \frac{g_U(x_n, \cdot)}{g_U(x_n, x_0)}.$$

By the elliptic Harnack inequality and Moser's oscillation lemma, the functions $h_n$ are uniformly bounded and equicontinuous on compact subsets of $U$ (since they all have value 1 at $x_0$ and then we can use the Harnack inequality on a chain of balls to any other point), so by Arzela-Ascoli we can pass to a subsequence to get a limit $h = \lim_{k\to\infty} h_{n_k}$ which is uniformly convergent on compact subsets. We can check that $h$ is harmonic everywhere (because each $g_U$ is harmonic except at a single point $x$, but those points are going to infinity) and that it is zero on the boundary; in particular we find that $h \in \mathcal{F}^0_{\text{loc}}(U)$.

This isn't so satisfactory if we want a concrete estimate (since everything in the Doob $h$-transform has estimates dependent on $h$), and it's an interesting question to see how we might do this on a specific domain.

To show uniqueness, we need the **boundary Harnack principle**, which will play an important role when talking about time-changes later.

**Definition 145**

Let $(X, d, m, \mathcal{E}, \mathcal{F})$ be an MMD space and $U$ an open subset of $X$. We say that $U$ satisfies the **boundary Harnack principle (BHP)** if there are constants $A_0, A_1, c_1 \in (1, \infty)$ such that for any point $\xi \in \partial U$, any $r \in (0, \frac{\text{diam}(U)}{A_1})$, and any pair of nonnegative $\mathcal{E}$-harmonic functions $u, v$ on $U \cap B(\xi, A_0 r)$ with zero boundary conditions relative to $U$ (that is, $u, v \in \mathcal{F}^0_{\text{loc}}(U, U \cap B(\xi, A_1 r))$), we have

$$\sup_{B(\xi,r)\cap U} \frac{u}{v} \leq C_1 \inf_{B(\xi,r)\cap U} \frac{u}{v}.$$

This estimate controls how the harmonic function (which is zero at the boundary) decays near the boundary, and it says that two such functions always decay at the same rate. This particular definition is due to Aikawa in 2001, though there's been lots of work in more specific settings before that.

This kind of fact helps to prove uniqueness of harmonic functions with zero boundary conditions, since one way to think about uniqueness of a nonnegative harmonic function is a Liouville property for ratios. In the Harnack inequality case, we saw that Moser's oscillation lemma implies Hölder regularity, which was enough to show the Liouville property. A similar argument works here – we have uniqueness of $h$ up to a positive multiplicative constant, since in this case the oscillation lemma tells us that for $u, v$ nonnegative on $B(\xi, R) \cap U$ and zero along $\partial U$, we have

$$\text{osc}_{B(\xi,r)\cap U} \frac{u}{v} \leq C \left(\frac{r}{R}\right)^{\alpha} \text{osc}_{B(\xi,R)\cap U} \frac{u}{v}.$$

Indeed, as $R \to \infty$ we see that the oscillation of $\frac{u}{v}$ approaches zero.

**Remark 146.** *The rate at which we decay can still depend on the particular boundary point on the domain, though. For example if $U$ has a sharp 90-degree corner, conformal mapping techniques tell us that near most points on the boundary we decay as $dist(x, U^c)$, but at the corner point we decay instead as $dist(x, U^c)^2$.*

The resulting $h$-transform semigroup is exactly the same no matter what this multiplicative constant is, so everything we care about is still unique. And we'll see more applications of this soon as we move now into time-changes. We'll begin this last topic of the course with an example of finite Markov chains:

**Example 147**

Recall Example 8 of the state space $X = \{1, 2, \cdots, n\}$ equipped with a conductance matrix $c : X \times X \to [0, \infty)$, which defines a discrete-time Markov chain with transition probabilities

$$P(x, y) = \frac{c(x, y)}{m(x)} = \frac{c(x, y)}{\sum_y c(x, y)}.$$

We also defined a function $\lambda : X \to (0, \infty)$ which governs the jump rate by specifying an exponential wait time at $x$ with rate $\lambda(x)$. This then gets us a continuous-time process.

As we saw in the first week, it's generally not so easy to write down the semigroup, but it's easy to write down the generator (we'll keep the $\lambda$ in the notation now)

$$L^{\lambda} f(x) = \lim_{t \downarrow 0} \frac{P^{\lambda}_t f(x) - f(x)}{t} = -\lambda(x)(I - Q)f(x),$$

where $Q$ is the operator such that $Qf(x) = \sum_y P(x, y) f(y)$. Indeed, if $\lambda$ is high we jump out of the state faster and thus the derivative will be larger. This generator $L^{\lambda}$ is then $\tilde{m}^{\lambda}$-symmetric, where we define $\tilde{m}^{\lambda}(x) = \frac{1}{\lambda(x)} m(x)$. We

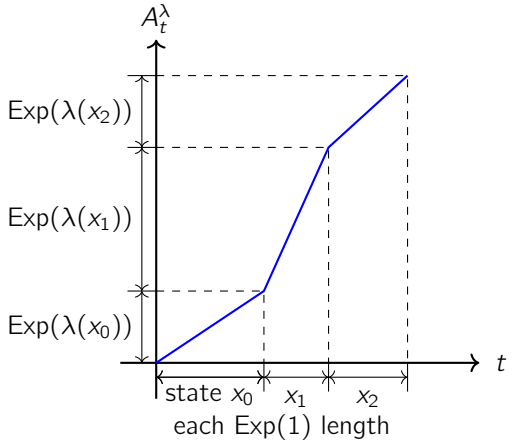
can then compute the Dirichlet form

$$\mathcal{E}(f,f) = \left\langle -L^\lambda f, f \right\rangle_{L^2(\tilde{m}^\lambda)} = \frac{1}{2}\sum_{x,y}(f(x)-f(y))^2 c(x,y),$$

and in particular **this Dirichlet form doesn't depend on $\lambda$ at all**. Indeed, $\lambda$ is just specifying the time parameterization, and other than that the process behaves exactly the same way. So we can express all of these processes $(Y_t^\lambda)$ in terms of one where $\lambda$ is identically 1: we'll drop $\lambda$ from the notation in that case and just write $(Y_t)$. For this, we'd like to define a "clock"

$$A_t^\lambda = \int_0^t \frac{1}{\lambda(Y_s)}\,ds.$$

This is an increasing process which starts at zero at time 0, and it is piecewise linear. Each of those pieces has (time) length independently distributed as Exp(1), and the slope of those pieces depends on the value of $\lambda$ at the current state (so if $\lambda$ is small, the process is quite large).




But the point now is that the value gained on each interval for $A_t^\lambda$ is Exp($\lambda(x)$)-distributed, and those amounts are again independent! Thus if we define

$$\tau_t^\lambda = \text{right-continuous inverse of } A_t^\lambda = \inf\{s \geq 0 : A_s > t\}$$

(in this case it's continuous because we have an increasing process), the description above tells us that $(Y_{\tau_t^\lambda})$ is a Markov process with generator $L^\lambda$. So we can get from $Y_t$ to $Y_t^\lambda$ by reparametrizing using this function $A_t$. The key properties of $A_t$ are the following: let $(\mathcal{F}_t)$ be the filtration and $(\theta_t)$ the time-shift operators corresponding to $(Y_t)$.

- $(A_t^\lambda)$ is $[0,\infty$-valued, non-decreasing, and $(\mathcal{F}_t)$-adapted with $A_0 = 0$.
- The process is additive with respect to the shift operators, meaning that

$$A_{t+s}^\lambda(w) = A_t^\lambda(w) + A_s^\lambda(\theta_t(w)).$$

**Example 148**

Now suppose we slightly change our setting so that we **now allow $\lambda$ to be $\infty$-valued** (meaning that we spend zero time at some particular states, since we automatically jump away at infinite rate). We should then not consider those states as part of the state space anymore. An example is if we have a $4 \times 4$ grid of vertices in a lattice, where the jump rates in the middle $2 \times 2$ grid are infinite. The point is then that we can now form new edges between states that weren't already there, and it's not so clear how we compute those rates.

In the graph of $A_t^\lambda$ above, some of those piecewise linear segments can now be horizontally flat, which means that the right-continuous inverse will make a jump. For this part, now suppose that the support of $\tilde{m}^\lambda$ is now some nonempty proper subset $F$ of $X$, and also assume that $(Y_t)$ is irreducible. Define

$$T_F = \inf\{t \geq 0 : Y_t \in F\},$$

and now for any function $g : F \to \mathbb{R}$, define the harmonic extension $H_f(g) : X \to \mathbb{R}$ via

$$H_F(g)(x) = \mathbb{E}_x\left[g(Y_{T_F})\right].$$

**Proposition 149**

Let $X, c, \lambda$ be as above, and consider any function $g : F \to \mathbb{R}$ with extension $\tilde{g} : X \to \mathbb{R}$. Then we have the following equivalent characterizations:

1. Harmonicity in the sense above; that is, $\tilde{g}(x) = H_F(g(x)) = \mathbb{E}_x[g(Y_{T_F})]$,
2. $\mathcal{L}\tilde{g}(x) = 0$ on $X \setminus F$, for $\mathcal{L}$ a variant of the Laplacian,
3. (Weak formulation of the point above) $\mathcal{E}(\tilde{g}, \phi) = 0$ for all $\phi : X \to \mathbb{R}$ with support contained in $X \setminus F$.
4. (Harmonic functions are energy minimizers) $\mathcal{E}(\tilde{g}, \tilde{g}) = \min_{h:X\to\mathbb{R} \text{ extension of } g} \mathcal{E}(h, h)$.

From here, we'll just do the same computation as we did before. $F$ will be the state space of the new time-changed process, so for $x \in F$ we have

$$\begin{aligned} P_t^\lambda g(x) &= \mathbb{E}_x(g(Y_t^\lambda)) \\ &= e^{-\lambda(x)t} g(x) + \left(1 - e^{-\lambda(x)t}\right)\left(\sum_{y\in F} P(x,y)g(y) + \sum_{y\in F^c} P(x,y)H_F(g)(y)\right) + o(t^2). \end{aligned}$$

This is basically the same logic as usual, where we make either zero jumps, one jump, or more with vanishingly small probability as $t \downarrow 0$. But for the one-jump case, we have to separate into possibilities of whether we get to another state in $F$ or something else, and in the latter case we need to use the harmonic extension to get the probabilities of ending up at various states. Remembering now that the function $g$ is exactly $H_F g$ for any $x \in F$, so we can actually write this more simply as

$$P_t^\lambda g(x) = e^{-\lambda(x)t} H_F g(x) + \left(1 - e^{-\lambda(x)t}\right) P(x,y) H_F^g g(y) + o(t^2),$$

meaning that the generator is

$$L^\lambda g(x) = -\lambda(x)(I - Q)H_F(g)(x)$$

for all functions $g : F \to \mathbb{R}$ and all $x \in F$. And again, we can now write down an inner product to compute the Dirichlet form (notably on the space $F$, not on the space $X$)

$$\tilde{\mathcal{E}}(g, g) = \left\langle -L^\lambda g, g\right\rangle_{L^2(\tilde{m}^\lambda)};$$

we can write this out using that $g = H_F g$ on $F$ and noting that again the factors of $\lambda$ cancel out between the generator and the measure to get

$$\langle (I - Q)H_F(g), H_F(g)\rangle_{L^2(m)} = \mathcal{E}(H_F(g), H_F(g)).$$

A similar story will occur in general now beyond finite state spaces, and the Dirichlet form of the time-change will

now typically be defined on some subset of the whole state space.

**Definition 150**

Let $(\mathcal{E}, \mathcal{F})$ be a regular Dirichlet form on $L^2(X, m)$, $(Y_t)$ the corresponding process, $(\mathcal{F}_t)$ the filtration, and $(\theta_t)$ the time-shifts. A **positive continuous additive functional (PCAF)** is a collection of random variables $(A_t)_{t\geq 0}$ satisfying the following properties:

- $(A_t)$ is $(\mathcal{F}_t)$-adapted.
- There exists some $\Lambda \in \mathcal{F}_\infty$ and a properly exceptional set $\mathcal{N} \subset X$ for $(Y_t)$, such that $\mathbb{P}_x(\Lambda) = 1$ for all $x \in X \setminus \mathcal{N}$. Furthermore, $\theta_t(\Lambda) \subset \Lambda$ for all $t \in (0, \infty)$.
- For all points $w \in \Lambda$ in this "almost-sure set," the function $t \mapsto A_t(w)$ is a continuous function from $[0, \infty]$ to $[0, \infty]$, such that $A_0(w) = 0$, $A_t(w) < \infty$ for all $t < \zeta(w)$, $A_t(w) = A_{\zeta(w)}(w)$ for all $t \geq \zeta(w)$ (the process stops doing anything after the lifetime ends), and $A_{t+s}(w) = A_t(w) + A_s(\theta_t(w))$ for all $s, t > 0$.

We call $\Lambda$ the **defining set** for the PCAF, and we call $\mathcal{N}$ the **exceptional set**. We also further say that $(A_t)$ is a PCAF **in the strict sense** if $\mathcal{N} = \varnothing$. The **support** of $A$ is then defined by

$$F = \{x \in X \setminus \mathcal{N} : \mathbb{P}_x(R = 0) = 1\},$$

where $R = \inf\{t \geq 0 : A_t > 0\}$.

The $F$ here should be thought of as the same $F$ as in our finite state space case.

**Theorem 151** (Revuz correspondence, special case)

For any PCAF $A$, there is a unique Borel measure $\mu$ on $X$, called the **Revuz measure** of $A$, such that

$$\int_X f \, d\mu = \lim_{t \downarrow 0} \frac{1}{t} \mathbb{E}_m \left[ \int_0^t f(Y_s) dA_s \right],$$

where the right-hand side is defined via the Stieltjes integral. This measure further satisfies $\mu(A) = 0$ for any $\mathcal{E}$-polar set $A$. Conversely, for any Radon measure $\mu$ which assigns zero mass to polar sets, there is some PCAF whose Revuz measure is $\mu$.

We can verify that $\tilde{m}$ from before plays the role here of $\mu$, and the idea is that we should not be able to see sets that were originally negligible before the time-change.

**Example 152**

Suppose $\mu(dx) = g(x)m(dx)$, where $g : X \to (0, \infty)$ is a "nice function" so that $\mu$ is a Radon measure. Then the associated PCAF has an explicit description in the same way as in the discrete space case, with

$$A_t = \int_0^t g(Y_s) ds.$$

It turns out that every Revuz measure of every PCAF can be approximated (in some sense) by measures of this form; this is actually one of the steps in the proof of the correspondence.

**Example 153**

Now specialize to the case of 1-dimensional Brownian motion. In this case, looking at the measure $\mu = \delta_x$ (the Dirac mass at $x$), the associated PCAF $A_t^\mu$ is the **local time at $x$**

$$L_t^x = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t 1\{B_s \in (x-\varepsilon, x+\varepsilon)\} ds.$$

For a more general Radon measure $\mu$ on $\mathbb{R}$, we then get an averaged combination of these local times

$$A_t^\mu = \int_{\mathbb{R}} L_t^x \mu(dx).$$

This description doesn't make sense in high dimensions (because local time isn't defined), but it's still giving us some intuition about what's going on.

If we now consider a PCAF $A$ with Revuz measure $\mu$, and we let $F$ be the support of that PCAF, it's typically the case that $F$ will be the topological support of $\mu$, but there are some counterexamples. Either way, we can define the **time-changed process corresponding to $A$**

$$\check{Y}_t = Y_{\tau_t}, \quad \tau_t = \inf\{s \geq 0 : A_s > t\}.$$

The process $(\check{Y}_t)$ is then a $\mu$-symmetric process on $F$ with Dirichlet form $(\check{\mathcal{E}}, \check{\mathcal{F}})$ given by (similar proof but more technical, in fact given by resolvents)

$$\check{\mathcal{F}} = \left\{ \tilde{f}|_F : f \in \mathcal{F}_e, \quad \tilde{f}|_F \in L^2(F, \mu) \right\}.$$

To explain where these "quasi-continuous versions" come from, we'll be curious about things like reflected Brownian motion, and it doesn't make sense to restrict $f$ to be zero on the boundary (of measure zero) unless we impose quasi-continuity. And unsurprisingly we have, for any $u \in \check{\mathcal{F}}$,

$$\check{\mathcal{E}}(u,u) = \mathcal{E}(H_f(\tilde{u}), H_f(\tilde{u}));$$

where the harmonic extension is given by

$$H_F(\tilde{u}(x)) = \mathbb{E}_x \left[ \tilde{u}(Y_{T_F}) 1\{T_F < \infty\} \right]$$

(it's possible for the time-changed process to have finite lifetime if the hitting time is infinite).

**Example 154**

Consider $(\mathcal{E}, \mathcal{F})$ the Dirichlet form for two-dimensional Brownian motion on $L^2(\mathbb{R}^2, m = \text{Leb})$. Consider the time-change formally defined by

$$M = \exp(\gamma X) dx, \quad X \text{ a Gaussian free field.}$$

For $\gamma \in (0,2)$ this is an LQG measure defined via rescaling, and it turns out that $M$ defines a PCAF $A$ in the strict sense, and it has full support $\mathbb{R}^2$. The corresponding time-changed process is called the **Liouville Brownian motion**.

Recall that in a previous mini-course, it was stated as an exercise that conformality can be characterized by invariance of Brownian motion. We'll see how to prove that now:

**Example 155**

We'll begin by showing how we can compute an "image process" given a bijection between two spaces. Let $(\mathcal{E}, \mathcal{F})$ be a regular Dirichlet form on $L^2(U, m)$ and $\Phi : U \to V$ be a bijection, and suppose $(Y_t)$ is the process corresponding to the given Dirichlet form $(\mathcal{E}, \mathcal{F})$. We would like to compute the Dirichlet form of the process $\tilde{Y}_t = \Phi(Y_t)$.

The reference measure to look at must be the pushforward of the measure $m$ – that is, $\tilde{m} = \Phi_*(m)$ – and we'll try to compute the semigroup. We have, for any function $f \in L^2(V, \tilde{m})$ and any $y \in V$, that

$$\begin{aligned}\tilde{P}_t f(y) &= \tilde{\mathbb{E}}_y \left[ f(\tilde{Y}_t) \right] \\ &= \tilde{\mathbb{E}}_y \left[ f(\Phi(Y_t)) \right] \\ &= \mathbb{E}_{\Phi^{-1}(y)} \left[ (f \circ \Phi)(Y_t) \right] \\ &= P_t (f \circ \Phi)(\Phi^{-1}(y)).\end{aligned}$$

So then taking limits as $t \downarrow 0$ lets us compute the Dirichlet form. There are two things that are easy to check, namely that (1) $(\tilde{P}_t)$ is an $\tilde{m}$-symmetric Markovian semigroup on $L^2(V, \tilde{m})$, and that (2) the corresponding Dirichlet form $(\tilde{\mathcal{E}}, \tilde{\mathcal{F}})$ on $L^2(V, \tilde{m})$ is given by

$$\tilde{\mathcal{F}} = \{ f \in L^2(V, \tilde{m}) : f \circ \Phi \in \mathcal{F} \},$$

$$\tilde{\mathcal{E}}(f, f) = \mathcal{E}(f \circ \Phi, f \circ \Phi) \text{ for all } f \in \tilde{\mathcal{F}}.$$

Next time, we'll restrict to the case where $\Phi$ is a nice $C^1$ map and use that to get from reflected or killed Brownian motion to ordinary Brownian motion. And we'll further characterize when this image process corresponds to a time-change of a diffusion generated by a uniform elliptic operator (rather than just Brownian motion) – that's given by what are called **quasi-conformal maps**.

# 14 June 24, 2025

We started talking about time-change of symmetric Markov processes last time – we take a PCAF $A$ with Revuz measure $\mu$, and we look at the right-continuous inverse of $(A_t)$. This defines a time-change process which will be symmetric on the support $F$, and its Dirichlet form can be defined in terms of quasi-continuous versions and harmonic extensions.

We were in the middle of an example last time where we start with a process $Y_t$ on a state space $U$ and want to look at the image process under some bijection $\Phi : U \to V$. We computed the semigroup and got a description for the new Dirichlet form $\tilde{\mathcal{E}}(f, f) = \mathcal{E}(f \circ \Phi, f \circ \Phi)$.

**Example 156**

We will specialize now to the case where $U, V$ are domains in $\mathbb{R}^2$ and $\Phi : U \to V$ is a $C^1$-diffeomorphism which is orientation-preserving (meaning $\det(D\Phi) > 0$ on $U$). Consider the case where $(\mathcal{E}, \mathcal{F})$ is reflected or killed Brownian motion on $U$ (we won't worry too much about the boundary conditions).

In coordinates, we can write $\Phi = (\Phi_1, \Phi_2)$ and $D\Phi = \begin{bmatrix} \frac{\partial \Phi_1}{\partial x_1} & \frac{\partial \Phi_2}{\partial x_1} \\ \frac{\partial \Phi_1}{\partial x_2} & \frac{\partial \Phi_2}{\partial x_2} \end{bmatrix}$. The Dirichlet form is then

$$\tilde{\mathcal{E}}(f,f) = \int_V |\nabla(F \circ \Phi)|^2 dx = \int_U \nabla f(x)^T \mathcal{A}(x) \nabla f(x) dx$$

by a change of variable and the chain rule, where $\mathcal{A}$ is the matrix

$$\mathcal{A}(x) = \det(D\Phi(\Phi^{-1}(x)))^{-1} D\Phi(\Phi^{-1}(x))^T D\Phi(\Phi^{-1}(x)).$$

(it's some $2 \times 2$ matrix which is always positive definite, so we have an elliptic operator). Then asking when we get Brownian motion is the same as asking for this matrix to be the identity matrix, and that's the same as saying that we have a rotation matrix for $\Phi$ – But that's exactly the same as requiring the Cauchy-Riemann equations to hold, so

$$\mathcal{A}(x) = I_2 \iff \Phi \text{ is a conformal map,}$$

and $\mathcal{A}(x)$ is a uniformly elliptic operator if and only if the singular values are uniformly bounded from above or equivalently from below (because the determinant of $\mathcal{A}$ is always 1, no matter what the derivative is – indeed, the constant $\det(D\Phi(\Phi^{-1}(x)))^{-1}$ will be squared when we evaluate the determinant). Thus

$$\mathcal{A}(x) \text{ uniformly elliptic} \iff ||D\Phi(x)||^2 \leq K \det(D\Phi(x)) \text{ for all } x$$

for some constant $K$. Geometrically, conformal maps map infinitesimal circles to infinitesimal circles, and $\mathcal{A}$ maps infinitesimal circles to infinitesimal ellipses of eccentricity at most $K$ – we say that $\Phi$ is **$K$-quasiconformal** in such a case.

It turns out it's possible to construct uniformly elliptic operators even on the upper half-space so that the associated harmonic measure is singular:

**Fact 157** (Caffarelli, Fabes, Kenig '81)
For any smooth domain $D$, there exists a uniformly elliptic operator whose corresponding diffusion has singular harmonic measure with respect to the surface measure on $D$. For example, we can map the half-space to itself in a way that is continuous on the boundary, such that we can make the pushforward of the Lebesgue measure singular on the real line boundary.

The point, though, is that the image of a conformal or quasiconformal map will be the timechange of a Brownian motion or a diffusion generated by a uniformly elliptic operator.

**Example 158** (Kigami '06)
Let $K$ be the Sierpinski gasket and $m$ the corresponding normalized Hausdorff measure. Let $(\mathcal{E}, \mathcal{F})$ be the Dirichlet form corresponding to Brownian motion on the Sierpinski gasket; we can construct a time-change via the **Kusuoka measure** $\nu = \Gamma(h_1, h_1) + \Gamma(h_2, h_2)$, where $h_1, h_2$ are the functions with boundary condition $\frac{\sqrt{2}}{3}(1, -\frac{1}{2}, -\frac{1}{2})$ and $\frac{1}{\sqrt{6}}(0, 1, -1)$ on the outer three vertices. We can check that the total energy is 1 and that the cross-energy $\mathcal{E}(h_1, h_2) = 0$, so we're choosing some kind of orthonormal basis. (What we said turns out to still be true if we just pick only one harmonic function as well, but we'll stick to our current example.)

Letting $\nu$ be a Revuz measure for the PCAF $A$, this functional will have full support and thus the time-change is also a diffusion process. We've proven previously that with respect to the Euclidean metric, the original Brownian motion

$(\mathcal{E}, \mathcal{F})$ on $L^2(K, m)$ satisfies heat kernel bounds HKE $\left(\frac{\log 5}{\log 2}\right)$. But the time-changed process $(\mathcal{E}, \mathcal{F})$ on $L^2(K, \nu)$ will actually satisfy Gaussian heat kernel bounds HKE(2) with respect to the metric

$$d_{\text{int}}(x, y) = \sup\{f(x)_f(y) : f \in \mathcal{F} \cap C(K)\}.$$

which generates the same topology as the Euclidean topology. (And this is not just about the choice of metric – because of singularities, we can't change the metric for the original Brownian motion on the Sierpinski gasket to get Gaussian heat kernel bounds.) So things can be pretty complicated under time-change!

So far in all of our examples, we started with processes which were already diffusions. But we can also end up with processes that go into a smaller space:

**Example 159** (Boundary trace of reflected diffusions)
Suppose we have a reflected diffusion $Y_t$ on a nice domain $D$. Let $A$ be a PCAF whose support is the boundary $\partial D$, so the timechange will live on the boundary only. We can think of this as having a reflected Brownian motion which jumps around from point to point, and we want to calculate the Dirichlet form of the resulting jump process. To do this, we start with the simple situation of a **smooth bounded domain and reflected Brownian motion**, and we'll talk about extensions afterward.

Let $U \subset \mathbb{R}^n$ be such a domain; the **Green's function** can be defined via the equation

$$\begin{cases} \Delta_y g_U(x, y) = -\delta_x, & y \in U \\ g_U(x, y) = 0, & x \in \partial u. \end{cases}$$

We'll come back to the probabilistic interpretation of this when we work on more general domains later, but for now the idea is that we start with a boundary function, extend it harmonically, and compute the Dirichlet energy on the interior to get the trace Dirichlet form on the boundary. Given some compactly supported smooth $f : \partial U \to \mathbb{R}$, we need to compute the harmonic extension $H_{\partial U}(f) : \overline{U} \to \mathbb{R}$ (meaning that we're harmonic in $U$ with boundary condition $f$ on $\partial U$). By Green's formula, when we have smooth domains we can use the Green's identity (basically integration by parts)

$$\int_U (v_2 \Delta v_1 - u_1 \Delta u_2) = \int_{\partial U} \left( v_1(\xi) \frac{\partial v_2}{\partial \vec{n}_\xi}(\xi) - v_2(\xi) \frac{\partial v_1}{\partial \vec{n}_\xi} \right) d\sigma,$$

where $d\sigma$ is surface measure on $\partial U$ and $\vec{n}$ is the inward-pointing normal derivative. We apply this to

$$v_1 = H_{\partial U}(f), \quad v_2 = g_U(x, \cdot),$$

to get

$$H_{\partial U}(f)(x) = \int_{\partial U} f \frac{\partial g_U(x, \xi)}{\partial \vec{n}_\xi} d\sigma,$$

and thus we get an expression for the harmonic measure of the domain $U$: for the Brownian motion started at $x$, we have the classical expression

$$\omega_x^U = \frac{\partial g_U(x, \xi)}{\partial \vec{n}_\xi} d\sigma.$$

We now want to get a formula for the Dirichlet energy expressed in terms of the boundary values (since that's what we started with). Again we use integration by parts via the identity

$$\int_U \nabla v_1 \cdot \nabla v_2 dx = -\int_U v_1 \Delta v_2 - \int_{\partial U} v_1 \frac{\partial v_2}{\partial \vec{n}_\xi} \sigma(d\xi).$$

Taking $v_1, v_2$ to both be the harmonic extension $H_{\partial U}(f)$, we thus have

$$\int_U |\nabla H_{\partial U}(f)|^2 dx = -\int_{\partial U} f(\xi) \frac{\partial H_{\partial U}(f)}{\partial \vec{n}_\xi} \sigma(d\xi),$$

and now we can interchange the integral and derivative and plug in our expression of $H_{\partial U}(f)$ from above to get a second derivative

$$\int_U |\nabla H_{\partial U}(f)|^2 dx = -\int_{\partial U}\int_{\partial U} f(\xi) f(\eta) \frac{\partial^2 g_U(\eta,\xi)}{\partial \vec{n}_\xi \partial \vec{n}_\eta} \sigma(d\eta)\sigma(d\xi).$$

We can rewrite this in a form more familiar to us from the Beurling-Deny decomposition as

$$\frac{1}{2}\int_{\partial U}\int_{\partial U} (f(\xi) - f(\eta))^2 \frac{\partial^2 g_U(\eta,\xi)}{\partial \vec{n}_\xi \partial \vec{n}_\eta} \sigma(d\eta)\sigma(d\xi).$$

(Indeed, the point is that the cross-terms give us what we want, and the non-cross-terms are the derivative of a constant $\int_{\partial U} \frac{\partial g(x,\xi)}{\partial \vec{n}_\xi} \sigma(d\xi) = 1$, hence zero.) So what this tells us is that both the diffusion and the killing part are zero in that Beurling-Deny decomposition, and so we do have a pure jump process.

This calculation actually tells us something nice – the jump measure is just the second derivative $\frac{\partial^2 g_U(\eta,\xi)}{\partial \vec{n}_\xi \partial \vec{n}_\eta}\sigma(d\eta)\sigma(d\xi)$, which we will express in terms of the harmonic measure – fixing some $x_0 \in U$, we can write it as

$$\frac{\partial^2 g_U(\eta,\xi)}{\partial \vec{n}_\xi \partial \vec{n}_\eta} \left(\frac{\partial g_U(x_0,\xi)}{\partial \vec{n}_\xi}\right)^{-1} \left(\frac{\partial g_U(x_0,\eta)}{\partial \vec{n}_\eta}\right)^{-1} \omega^U_{x_0}(d\xi)\omega^U_{x_0}(d\eta).$$

**Definition 160** (Naim kernel, '57)

Define

$$\Theta^U_{x_0}(x,y) = \frac{g_U(x,y)}{g_U(x_0,x)g_U(x_0,y)} \text{ for } x \neq y \in U.$$

Naim showed that $\Theta^U_{x_0}$ admits a continuous extension to $(\overline{U} \setminus \{x_0\}) \times (\overline{U} \times \{x_0\}) \setminus \text{diag}$, where $\overline{U}$ is something called the Martin compactification and we have continuity in the "fine topology," which is finer than the usual Euclidean topology.

The point is that the blue part in the expression above looks a lot like the Naim kernel, and indeed Doob wrote down the following formula:

**Proposition 161** (Doob, '62)

We have for Brownian motion that the jump process satisfies

$$J(d\xi, d\eta) = \Theta^U_{x_0}(\xi,\eta)\omega^U_{x_0}(d\xi)\omega^U_{x_0}(d\eta).$$

The nice thing here is that every term on the right-hand side has a probabilistic interpretation (in contrast to something like the surface measure, which doesn't necessarily have such an interpretation). We'd like to extend this to arbitrary diffusions and metric spaces, but first we'll give one more example of a boundary trace process which has been quite influential:

**Definition 162**

The **fractional Laplacian operator** with parameter $\alpha \in (0,2)$ is the operator $-(-\Delta)^{\alpha/2}$.

We can make sense of this using the spectral theorem, since $-\Delta$ is a nonnegative self-adjoint operator. And we restrict to this particular range of values of $\alpha$ because larger $\alpha$ makes the corresponding quadratic form no longer a

Dirichlet form (it fails the Markov property). There are also some other definitions: we can think of it in the Fourier transform sense via

$$\widehat{(-\Delta)^{\alpha/2} f} = |\xi|^\alpha \hat{f}(\xi),$$

or we can also define it via the singular integral

$$-(-\Delta)^{\alpha/2} f(x) = \lim_{r \downarrow 0} c_{n,\alpha} \int_{\mathbb{R}^n \setminus B(x,r)} \frac{f(y) - f(x)}{|x-y|^{n+\alpha}} dy.$$

These definitions are all equivalent, though it does take some work to show that. And the point is that $-(-\Delta)^{\alpha/2}$ is the **generator of a symmetric $\alpha$-stable Lévy process**, with jump kernel given by

$$J(x,y) = c_{n,\alpha} \frac{1}{||x-y||^{n+\alpha}}.$$

There's some work talking about regularity of solutions even for these non-local operators:

**Example 163**

Molchanov and Ostrowski showed in 1969 that $\alpha$-stable processes can be viewed as the boundary trace process of the following process: let $(B_t)$ be a Brownian motion on $\mathbb{R}^n$, and let $Z_t$ be an independent $(2-\alpha)$-dimensional Bessel process on $[0,\infty)$, and define $(B_t, Z_t)$. We then get the Dirichlet form

$$\mathcal{E}(u,u) = \int_{\mathbb{R}^n} \int_0^\infty |\nabla u(x,y)|^2 |y|^{1-\alpha} dy\, dx$$

on $L^2(\mathbb{H}^{n+1}, m)$, and the boundary trace is an $\alpha$-stable process.

This was also independently rediscovered by Caffarelli and Silvestre in 2007, who found a harmonic extension on the upper half-space and computed a Dirichlet energy in terms of the boundary conditions. The reason they were interested in this procedure is that the generator is (via integration-by-parts) $\Delta_x + \partial_{yy} + \frac{1-\alpha}{y}\partial_y$ (we can think of this as a weighted Laplacian). They were then able to obtain things like Harnack inequalities for the fractional Laplace operator using the corresponding properties for the local operator, and this led to lots of other results in regularity theory.

We'll show part of those calculations now. Let $f : \mathbb{R}^n \to \mathbb{R}$ be some boundary condition (viewed as the boundary of the upper half-space in $\mathbb{R}^{n+1}$); our goal is to solve the equation

$$\Delta_x u + \frac{1-\alpha}{y} u_y + u_{yy} = 0.$$

When $\alpha = 1$ this is very easy to solve, but when $\alpha \neq 1$ it's best to convert this using the Fourier transform in the $x$-direction only: we thus look at

$$\hat{u}(\xi, y) = \int u(x,y) e^{-ix\cdot\xi} dx,$$

and we end up with

$$-|\xi|^2 \hat{u}(\xi,y) + \frac{1-\alpha}{y} \hat{u}_y(\xi,y) + \hat{u}_{yy}(\xi,y) = 0,$$

with boundary condition $\hat{u}(\xi,0) = \hat{f}(\xi)$, which is an ODE that was can solve. For fixed $\xi$ this is linear in $y$, and it suffices for us to solve

$$-\phi + \frac{1-\alpha}{y}\phi_y + \phi_{yy} = 0 \text{ on } (0,\infty), \text{ where } \phi()) = 1.$$

This is called the **Bessel differential equation**, and it has a unique solution with $\lim_{y\to\infty} \phi(y) = 0$ (we want this

because the function should be in $L^2$). The solution to the original problem is then

$$\hat{u}(\xi, y) = \hat{f}(\xi)\phi(|\xi|y),$$

and we're interested in computing the energy now. Absorbing all multiplicative constants involving gamma functions, we find by Parseval in the $x$-coordinate and then a change of variables that

$$\begin{aligned}\int_{\mathbb{R}^n}\int_0^\infty |\nabla u(x,y)|^2 |y|^{1-\alpha} dy dx &= c\int_0^\infty \int_{\mathbb{R}^n} \left(|\xi|^2|\hat{u}(\xi,y)|^2 + |\hat{u}_y(\xi,y)|^2\right) y^{1-\alpha} d\xi dy \\ &= c\int_0^\infty \int_{\mathbb{R}^n} |\xi|^2|\hat{f}(\xi)|^2 \left(\phi(|\xi|y)^2 + (\phi'(|\xi|y))^2\right) y^{1-\alpha} d\xi dy \\ &= c'\int_0^\infty \int_{\mathbb{R}^n} |\xi|^\alpha |\hat{f}(\xi)|^2 (\phi(z)^2 + \phi'(z)^2)|z|^{1-\alpha} d\xi dz.\end{aligned}$$

We can now do the integral in the $z$-coordinate, which yields

$$c''\int_{\mathbb{R}^n} |\xi|^\alpha |\hat{f}(\xi)|^2 d\xi,$$

which again by Parseval can be viewed as $\langle (-\Delta)^{\alpha/2} f, f\rangle_{L^2}$. So that proves the result and explains the appearance of the $\alpha$-stable process.

**Example 164**

What we want to explain now is that everything we're seeing is a special case of the Doob-Naim formula (even when the diffusion is not a Brownian motion). And we can think about Brownian motion on something like the Sierpinski carpet – even in a complicated setting like that, Doob-Naim does give us a jump measure. For this, we need to introduce some assumptions.

Return now to the setting of an MMD space $(X, d, m, \mathcal{E}, \mathcal{F})$, and assume this space satisfies sub-Gaussian heat kernel bounds $\mathrm{HKE}_f(\psi)$. Further assume $(X, d)$ is a geodesic space (just to simplify things, though it's not needed), and that $U$ is a uniform domain satisfying the **capacity density condition**

$$\mathrm{Cap}(B(\xi, r)\setminus U, B(\xi, 2r)^c) \lesssim \mathrm{Cap}(B(\xi, r), B(\xi, 2r)^c)$$

for all $\xi \in \partial U$ and all $r \lesssim \mathrm{diam}(U)$. (For example, removing a line from three-dimensional Brownian motion will not do anything and thus we won't get an interesting trace process, so we want to exclude that case.) Such a condition ensures good behavior of the harmonic measure, and the fact that this condition can be used to get good estimates was discovered by Aikawa and Hirata in 2007: as long as $0 < r < \frac{d(x_0,\xi)}{A}$, we have that

$$\omega_{x_0}^U(B(\xi, r), \partial U) \asymp g_U(x_0, \xi_r)\mathrm{Cap}(B(\xi, r), B(\xi, 2r)^c),$$

so we almost have the same formula as for the Green's function. Here $\xi_r \in U$ is some point such that $d(\xi, \xi_r) \asymp r \asymp \mathrm{dist}(\xi_r, U^c)$ (it doesn't matter exactly which point we choose because of the Harnack inequalities). This result also applies to a more general setting than just Brownian motion. And that explains why we deal with this class of domains (which in particular includes all Lipschitz domains and also some domains with fractal boundaries).

In particular, these estimates on the harmonic measure imply that the support of $\omega_{x_0}^U$ is $\partial U$, and $\omega_{x_0}^U$ defines a PCAF $A$ whose support is in $\partial U$ (technical detail here that it may be smaller than the support of the corresponding Veruz measure). We want to introduce the Naim kernel in this context, but we first need to state the Green's function and its properties.

**Definition 165**

Consider a metric measure space with sub-Gaussian heat kernel bounds, and suppose we have a **transient domain** $D$ (meaning that the process killed on exit is transient). Then there is some **Green's function** $g_D : D \times D \to [0, \infty]$ such that the following properties are true:

- (Symmetry) We have $g_D(x, y) = g_D(y, x)$ for all $x, y \in D$.
- (Continuity, which is natural from the Harnack inequalities) $g_D(\cdot, \cdot)$ is $[0, \infty)$-valued and continous on $D \times D \setminus \text{diag}$.
- (Harmonicity) We have $g_D(x, \cdot) \in \mathcal{F}^0_{\text{loc}}(D, D \setminus \{x\})$ and that $g_D$ is $\mathcal{E}$-harmonic except at $x$ – think of this as saying that $g_D(x, \cdot)$ is zero on the boundary.
- (Occupation density) Let $Y_t$ be the corresponding diffusion process. Then for any nonnegative measurable function, we have
$$\mathbb{E}_x \left[ \int_0^{\tau_D} f(Y_s) ds \right] = \int_D g_D(x, y) f(y) m(dy)$$
for all $x \in D \setminus \mathcal{N}$ (outside some properly exceptional set).

We will use this last property to define the **Green operator** $G^D f(x)$; if $f$ is a nonnegative function which satisfies $\int f G^D(f) dm < \infty$, then $G^D(f) \in \mathcal{F}^0(D)_e$ lives in some space on which we have the energy

$$\mathcal{E}(G^D(f), G^D(f)) = \int f G^D(f) dm < \infty.$$

We then have $\mathcal{E}(G^D(f), v) = \int f v dm$ for all $v \in \mathcal{F}^0(D)_e$, which justifies the Green operator and Laplacian being inverses. We now would like to write the harmonic measure in terms of the Green's function so that we can generalize the equation

$$\omega^U_{x_0} = \frac{\partial g(x, \xi)}{\partial \vec{n}_\xi} d\sigma(\xi).$$

One way to understand the connection is that the Laplacian of the Green's function satisfies $\Delta_y g(x, \cdot) = -\delta_x$ inside the domain. But if we try to compute the same quantity on the boundary, it turns out we get the harmonic measure, and thus we can instead specify

$$\Delta_y g(x, \cdot) = -\delta_x + \omega^U_x.$$

This then has hope of being generalized, and this is the result we will prove next time:

**Lemma 166**

For all $x \in U$ and all functions $u \in \mathcal{F}(U) \cap L^\infty(U)$, such that $x$ is not in the support of $u$, the reflected Dirichlet form satisfies

$$\mathcal{E}_U(g_U(x, \cdot), u) = -\int_{\partial U} \tilde{u} d\omega^U_x,$$

where $\tilde{u}$ is again the quasicontinuous version of $u$ so that we have things well-defined up to sets of harmonic measure zero.

We'll make use of this in our generalization of Doob's formula soon!

# 15 June 26, 2025

We'll continue discussing today a generalization of the Doob-Naim formula (this is Professor Murugan's joint work with Kajino). We saw last time that Doob-Naim gives an expression for the jump kernel of the boundary trace process by starting with a boundary function, extending it harmonically inside, and then computing the Dirichlet formula in terms of the boundary values via integration by parts. But in general integration by parts is not useful because the harmonic measure may be singular with respect to the surface measure, and we should end up with something absolutely continuous to the product of the harmonic measure with itself.

So we need to first show how to define the Naim kernel in our setting – last time we gave a definition which uses things like the fine topology, but we'd like to avoid that. We previously defined the Green's function and Green operator on a transient domain $D$ – recall that the Green operator can be viewed as the inverse of the generator, meaning that $G^D f$ is in the extended Dirichlet space corresponding to the killed diffusion $\mathcal{F}^0(D)_e$ for any $f : D \to [0,\infty)$ where $\int f G^D(f) dm$ is finite.

**Example 167**

The setup we have is the following: $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies full sub-Gaussian heat kernel estimates $\mathrm{HKE}_f(\psi)$, $d$ is geodesic (though this isn't strictly necessary), and $U$ is a uniform domain satisfying the bound

$$\mathrm{Cap}(B(\xi, r) \setminus U, B(\xi, 2r)^c) \lesssim \mathrm{Cap}(B(\xi, r), B(\xi, 2r)^c)$$

for all $\xi \in \partial U$ and all $r \lesssim \mathrm{diam}(U)$.

**Lemma 168**

For all $x \in U$ and all $u \in \mathcal{F}(U) \cap L^|infty(U)$ with $x \notin \mathrm{supp}_m(u)$, we have

$$\mathcal{E}_U(g_U(x, \cdot), u) = -\int_{\partial U} \tilde{u} d\omega_x^U.$$

(This should be thought of as an analog of how the harmonic measure $\omega_x^U$ is given by $\frac{\partial g_U(x,\xi)}{\partial \vec{n}_\xi} d\sigma$; more generally we should think of $\Delta g_U(x, \cdot) = -\delta_x + \omega_x^U$ as having a Dirac mass inside and a harmonic measure on the boundary.)

*Proof.* The idea is that the complement has positive capacity, so the process killed upon exit will be transient. So using the semigroup definition of transience, we know that for any positive function $f_1 \in L^1(U, m)$ such that $G^U(f_1) < \infty$ $m$-almost-everywhere, we can replace it with a better-behaved function $f = \frac{f_1}{G(f_1) \vee 1}$ which satisfies that $\int f G^U(f) dm < \infty$. Replacing $f$ with the function $f 1_{B(x,r)}$ which is supported on a small ball, we see that $\mathrm{supp}_m(f)$ and $\mathrm{supp}_m(u)$ do not intersect. We can now further normalize $f$ so that $\int f dm = 1$.

The point is that after all of this, $G^U(f) \in \mathcal{F}^0(U)_e$ by the property of the Green's function. Now we claim that

$$\mathcal{E}_U(G^U(f), u) = \mathcal{E}_U(G^U(f), u) - \mathcal{E}_U(G^U(f), H_{\partial U}\tilde{u})$$

(that is, the last term is zero), where $H_{\partial U}(\tilde{u})$ is the harmonic extension – indeed, $G^U(f)$ has zero boundary conditions on $U$. But now we can apply the characterizing formula for the Green's function since $u - H_{\partial U}(\tilde{u})$ has zero boundary conditions: we find that

$$\mathcal{E}_U(G^U(f), u - H_{\partial U}(\tilde{u})) = \int f u dm - \int f H_{\partial U}(\tilde{u}) dm.$$

Now the first term is always zero because we have disjoint supports $f, u$, and we can take $r \downarrow 0$ (approaching a Dirac mass at $x$) so that the second term approaches $\mathcal{E}_U(g_U(x,\cdot),h) = -\int \tilde{u} d\omega_x^U$, as desired. □

The next lemma concerns "classical objects" of equilibrium measures:

**Lemma 169**

Suppose $A, B \subset X$ are disjoint closed sets such that $A$ and $\overline{B^c}$ are compact. (The picture to have in mind is that $A$ is a small ball and $B$ is the complement of a larger ball containing it.) Let $e_{A,B} \in \mathcal{F}_e \cap L^\infty$ such that the quasicontinuous version satisfies $\tilde{e}_{A,B} = 1$ $q$-a.e. on $A$ and $\tilde{e}_{A,B} = 0$ $q$-a.e. on $B$ and is harmonic on $(A \cup B)^c$. Then there are finite Borel measures $\lambda^1_{A,B}$ and $\lambda^0_{A,B}$ supported on $\partial A$ and $\partial B$, respectively, with the same total mass

$$\lambda^1_{A,B}(X) = \lambda^0_{A,B}(X) = \mathcal{E}(e_{A,B}, e_{A,B}),$$

and such that

$$\mathcal{E}(e_{A,B}, u) = \int \tilde{u} d\lambda^1_{A,B} - \int \tilde{u} d\lambda^0_{A,B}.$$

Furthermore, $\lambda^1_{A,B}$ and $\lambda^0_{A,B}$ do not charge polar sets (so everything here is indeed well-defined) – that is, the measure of any polar set is zero, so these measures can be Revuz measures for PCAFs.

In the classical case, we would have $\Delta e_{A,B} = -\lambda^1_{A,B} + \lambda^0_{A,B}$, and we're just doing a weak formulation of that here because. (The flipped signs are because the Dirichlet form involves the negative Laplacian.) We won't need this probabilistic interpretation, but $\tilde{e}_{A,B}(x)$ can be thought of as the probability of hitting $A$ before $B$ when started at $x$.

*Proof sketch.* The idea is that $e_{A,B}$ is an energy minimizer for the class of functions

$$\mathcal{L}_{A,B} = \left\{ f \in \mathcal{F}^0(B)_e : \tilde{f} \geq 1 \ q\text{-a.e. on } A \right\}.$$

Then $(\mathcal{F}^0(B^c)_e, \mathcal{E})$ is a Hilbert space and $\mathcal{L}_{A,B}$ is a closed convex subset of it, so there is a unique energy minimizer on $\mathcal{L}_{A,B}$ and we can compute it with Euler-Lagrange type arguments. For any nonnegative $v \in C_c(B^c) \cap \mathcal{F}$, we see that $e_{A,B} + tv$ belongs in $\mathcal{L}_{A,B}$ as well (since we're only increasing the function on $A$), and so the minimizer should satisfy

$$\mathcal{E}(e_{A,B} + tv, e_{A,B} + tv) \geq \mathcal{E}(e_{A,B}, e_{A,B}).$$

Expanding out both sides and taking $t \to 0$ shows that $\mathcal{E}(e_{A,B}, v) \geq 0$, so $v \mapsto \mathcal{E}(e_{A,B}, v)$ is a nonnegative linear functional on $C_c(B^c)$. Riesz representation then tells us that this must be of the form $\int v d\lambda^1_{A,B}$, which gives us the positive part. And for the negative part we do something similar but perturb with negative functions on the other boundary. The reason they should be supported on the boundary is strong locality – if we perturb by some function in the interior of $A$, then we should have the exact same energy. And they should have the same mass because we can plug in a function $u$ which is constant within the regions we care about. □

We next define the Naim kernel. This actually has a predecessor called the Martin kernel, which we will introduce first:

**Definition 170**

Define the **Martin kernel**

$$K^U_{x_0}(x,\xi) = \begin{cases} \frac{g_U(x,\xi)}{g_U(x_0,\xi)} & x,\xi \in U, x \neq \xi, \\ \lim_{y\to\xi, y\in U} \frac{g_U(x,y)}{g_U(x_0,y)} & x \in U, \xi \in \partial U. \end{cases}$$

(The limit in the second case is defined due to the boundary Harnack principle, along with a version of Moser's oscillation lemma.)

(This is actually slightly different from Martin's original definition from 1941 – he defined an 'ideal boundary," which is an abstract compactification which allows us to account for all possible limits of ratios. But in our case the Martin boundary will be the same as the topological boundary because of boundary Harnack.)

We saw the definition of the Naim kernel in Definition 160 last time, but now we'll do it not in the fine topology:

**Definition 171** (Naim kernel, revisited)

For $x, y \in (U \setminus \{x_0\}) \times (U \setminus \{x_0\}) \setminus \text{diag}$, define

$$\Theta^U_{x_0}(x,y) = \frac{g_U(x,y)}{g_U(x_0,x)g_U(x_0,y)}$$

and extend it continuously using the boundary Harnack principle. In fact, the interesting case for us will be where $x, y$ are both on the boundary.

This is a more symmetric definition than the Martin kernel, and we need to explain now why we can do this: the boundary Harnack principle should only look at two harmonic functions, not three. But for fixed $x$ we have a ratio of harmonic functions in $y$ and vice versa, so we can use boundary Harnack in the two different coordinates and run a two-variable version of Moser's oscillation lemma, and we can still show that we get a jointly continuous extension away from the diagonal.

Another way to define the Martin kernel is to have it be the density of the harmonic measure at $x$ with respect to the harmonic measure started at $x_0$ – in some papers that is what is used. But when we define things in terms of densities, we lose the ability to talk about finer properties like continuity. Nevertheless, we can see that these two notions are equivalent in the following way:

**Proposition 172**

For all $x_0, x \in U$, we have (With our definition)

$$K^U_{x_0}(x,\cdot) = \frac{d\omega^U_x}{d\omega^U_{x_0}}.$$

Indeed, we know that both of the measures on the right-hand side are absolutely continuous with respect to the surface measure in the smooth case, and so we should formally think of this as $\frac{\partial g(x,\xi)/\partial \vec{n}_\xi)}{\partial g(x_0,\xi)/\partial \vec{n}_\xi)}$.

*Proof.* Consider the sets $A = \partial U \cap B(\xi, r)$ and $B^c = U \cap B(\xi, 2r)$, where we choose $r$ small enough so that $2r < d(\xi, x)$ (so we have a small set on the boundary and then a large set away from it inside the domain containing $x$). We want to look at some kind of capacity bound and integrate that against the harmonic measure. We have, because the

harmonic measure has full support, that

$$0 < \int \tilde{e}_{A,B} d\omega_x^U = -\mathcal{E}_U(g_U(x,\cdot), e_{A,B})$$
$$= \int_{S(\xi,2r)} g_U(x,y) d\lambda^0_{A,B}(y)$$

where we've applied our equilibrium measure lemma and used that when the equilibrium potential is 1, the Green's function is zero so the $\lambda^1_{A,B}$ term goes away. We also have a similar type of expression for $x_0$ as well:

$$0 < \int \tilde{e}_{A,B} d\omega_{x_0}^U = \int_{S(\xi,2r)} g_U(x_0,y) d\lambda^0_{A,B}(y).$$

Taking $r \downarrow 0$, we see that the ratio $\frac{\int \tilde{e}_{A,B} d\omega_x^U}{\int \tilde{e}_{A,B} d\omega_{x_0}^U}$ approaches the Martin kernel $K^U_{x_0}(x,\xi)$ (since the Martn kernel is continuous). More carefully, we can express $g_U(x,y)$ as $K^U_{x_0}(x,y)g_U(x_0,y)$ and then take limits of that equation; then we need a version of the Lebesgue differentiation theorem to conclude that in fact this implies $\frac{d\omega_x^U}{d\omega_{x_0}^U} = K^U_{x_0}(x,\xi)$ as well. □

We'll use a very similar argument for getting the Doob-Naim formula, though we'll need two variables instead. The idea is to get the jump kernel of the trace process via the Beurling-Deny decomposition, and whenever we want such a kernel we want to plug in functions with disjoint supports. Indeed, we have by definition of the trace Dirichlet form that

$$\check{E}(u,v) = \mathcal{E}_U(H_{\partial U}(u), H_{\partial U}(v)),$$

and if $u, v$ have disjoint support then the strongly local and killing part of $\mathcal{E}_U$ will vanish and we just get the jump part. Specifically we will have two points $\eta, \xi$ on the boundary and let $u = H_{\partial U} e_{A,B}$ around $\xi$ (as in the proof above) and $v$ similar around $\eta$. We will end up with an expression of the form $-\int uv d\overline{U}$, which is just like how we plugged in $u = 1_A$ and $v = 1_B$ for disjoint $A, B$ back when we introduced Beurling-Deny.

Let's do those computations in detail now. For distinct points $\xi, \eta \in \partial U$ and any $r < \frac{d(\xi,\eta)}{4}$, we will pick

$$A = B(\xi, r) \cap \partial U, \quad B = B(\xi, 2r)^c \cap \overline{U},$$

and choose

$$v = e_{A,B} \in \mathcal{F}^0(B^c)$$

and $u$ to be any function in $\mathcal{F}(U) \cap C_c(\overline{U})$ satisfying $1_{B(\eta,r)} \leq u \leq 1_{B(\eta,2r)}$. We will compute the energy of the trace process in two ways, using the harmonic measure expression and the Beurling-Deny decomposition. **On the one hand**, we have

$$\mathcal{E}_U(H_{\partial U}(u), H_{\partial U}(\tilde{e}_{A,B})) = \mathcal{E}_U(H_{\partial U}(u), \tilde{e}_{A,B}),$$

since the difference between $\tilde{e}_{A,B}$ and its harmonic extension is zero on the boundary (so we're using the same orthogonality property as before). But now we can use the properties of the equilibrium measure to rewrite this expression as

$$\int_A H_{\partial U}(u) d\lambda^1_{A,B} - \int_{S(\xi,2r)} H_{\partial U}(u) d\lambda^0_{A,B}.$$

The first term is actually just zero – $u$ has zero value on $A$ because the support of $\lambda^1_{A,B}$ is some region close to $v$ on the boundary, which is far away. So we're just left with

$$-\int_{S(\xi,2r)} H_{\partial U}(u) d\lambda^0_{A,B} = -\int_{S(\xi,2r)} \int_{\partial U} u(z) d\omega_y^U(z) d\lambda^0_{A,B}(y).$$

Now we want to normalize everything in terms of one base point, which we do using the Martin kernel: we end up with

$$-\int_{S(\xi,2r)}\int_{\partial U} u(z)K_{x_0}(y,z)d\omega^U_{x_0}(z)d\lambda^0_{A,B}(y).$$

But this should **also** be equal to $-\int u(x)\tilde{e}_{A,B}(y)J(dx,dy)$ by the Beurling-Deny decomposition of the trace Dirichlet form (only the cross-terms matter because we have disjoint support). Our goal is to show, much like Doob did for Brownian motion in Proposition 161, that

$$J(dx,dy)=\Theta^U_x(\xi,\eta)d\omega^U_{x_0}(\xi)d\omega^U_{x_0}(\eta).$$

We have a good expression for the jump kernel, so what we should try to do now is get

$$\begin{aligned}-\mathcal{E}_U(H_{\partial U}(u),H_{\partial U}(\tilde{e}_{A,B}))&=\int u(x)\tilde{e}_{A,B}(y)dJ(x,y)\\ \implies \frac{-\mathcal{E}_U(H_{\partial U}(u),H_{\partial U}(\tilde{e}_{A,B}))}{\int u(x)d\omega^U_{x_0}(x)\int \tilde{e}_{A,B}(y)d\omega^U_{x_0}(y)}&=\frac{\int u(x)\tilde{e}_{A,B}(y)dJ(x,y)}{\int u(x)d\omega^U_{x_0}(x)\int \tilde{e}_{A,B}(y)d\omega^U_{x_0}(y)},\end{aligned}$$

so that we can take $r\to 0$ on the right-hand side and use Lebesgue differentiation to conclude much like before. Indeed, the left-hand side can be rewritten (numerator by our calculation above) as

$$\frac{\int_{S(\xi,2r)}\int_{\partial U} u(z)K_{x_0}(y,z)d\omega^U_{x_0}(z)d\lambda^0_{A,B}(y)}{\int ud\omega^U_{x_0}\int g_U(x_0),y)d\lambda^0_{A,B}(y)}$$

where we use the same formula as before to replace the second term in the denominator. The blue and red parts now "cancel out" (by Lebesgue differentiation details), and the point now is that the ratio of Martin kernel to Green's function left

$$\frac{K_{x_0}(y,z)}{g_U(x_0,\cdot)}=\frac{g_U(y,z)}{g_U(x_0,z)g_U(x_0,y)}=\Theta(y,z)$$

is exactly the Naim kernel! So as the radius approaches zero, we will have $y\to\eta$ and $z\to\xi$, and continuity in all arguments means we're looking at convergence to the correct boundary points. This completes the idea of the proof. (We also have to compute the other two parts of the decomposition, which turn out to be zero – we'll skip that part of the computation.)

**Remark 173.** *Again, the point to remember is that harmonic measure is the right thing to use instead of surface measure even on smooth domains, since diffusions on the upper half-space can still be singular with respect to the surface measure.*

**Example 174**

We now want to understand how to choose the time parameterization of the boundary trace process (or equivalently, the Revuz measure) – the most natural probabilistic choice to use is the harmonic measure, but that still has a choice of base point that needs to be made.

For bounded domains it doesn't really matter – any $x_0\in U$ with $\text{dist}(x,U^c)\asymp \text{diam}(U)$ will have comparable harmonic measure to any other one by the elliptic Harnack inequality. But for unbounded domains, this is no longer true – far-away points can have very different harmonic measures, and so we want to pick a "harmonic measure from infinity." We'll explain that procedure now.

The hint comes from the formula

$$\mathcal{E}_U(g_U(x_0,\cdot),u)=-\int \tilde{u}d\omega^U_{x_0} \text{ if } x_0\notin \text{supp}_m(u).$$

We know that the ratio $\frac{g_U(x_n,\cdot)}{g_u(x_0,\cdot)}$ should converge as $x_n \to \infty$ to some specific function $h^U_{x_0}(\cdot)$; recall that this is how we constructed the function for the Doob $h$-transform, which we found is unique by boundary Harnack. So what we should do is consider the sequence of rescaled harmonic measures

$$\frac{\omega^U_{x_n}}{g(x_0,x_N)},$$

which should converge to some limit which we call the **elliptic measure at infinity** $\nu^U_{x_0}(\cdot)$. We actually get something even better than the bounded case, in that the thing we end up with is just defined up to a constant regardless of $x_0$. That is, we get the characterization

$$\mathcal{E}_U(h^U_x, u) = -\int \tilde{u} d\nu^U_{x_0},$$

and we don't even have to worry about support – this formula is true for all $u \in \mathcal{F}(U) \cap L^\infty$ in the reflected form, since the singularity of the Green's function goes to infinity now. (That is, we have $\Delta h^U_{x_0} = \nu^U_x$ with no Dirac mass). We also get the doubling property (due to Aikawa-Hirata).

**Fact 175**

Once we have a good understanding of all three parts of the Beurling-Deny decomposition, characterization of stable-like heat kernels yields good heat kernel bounds as well. Chen, Kumagai, and Wang obtained (in 2018) characterizations of stable-like heat kernel bounds

$$p_t(x,y) \asymp \frac{1}{t^{n/\alpha}} \wedge \frac{t}{|x-y|^{n+\alpha}},$$

corresponding to a space-time scaling where by time $t$ we travel distance approximately $t^{1/\alpha}$. One can compute the spacetime scaling for the boundary trace process, but we'll skip those details now.

Another approach due to Cao and Chen estimates these kinds of heat kernels for boundary trace processes, but they do not exactly calculate the jump kernel and just get estimates – that turns out to be enough for the estimates of Chen, Kumagai, and Wang. The Doob-Naim formula is more direct and implies the necessary estimates while getting the jump kernel exactly.

**Example 176**

One more small caveat here is that we can also apply these results if we have the **weak capacity density condition**, where instead of having $\mathrm{Cap}(B(\xi,r)\setminus U, B(\xi,2r)^c) \lesssim \mathrm{Cap}(B(\xi,r), B(\xi,2r)^c)$ for all $0 < r \le \mathrm{diam}(U)$ and all $\xi \in \partial U$, **it's also okay to just have** $0 < r < \mathbf{diam}(\partial U)$ (for example if the domain is unbounded but the boundary is bounded). The Doob-Naim formula still works, but the killing measure now might be nonzero if the reflected diffusion is transient.

It's not so difficult to understand what's happening probabilistically here – the complement of a bounded set in Euclidean space is a uniform domain satisfying this condition, so reflected Brownian motion in the complement of a ball will hit the boundary of the ball some number of times and then exit to infinity. The PCAF will then only increase when we hit the boundary and then after a while approach some limit $A_\infty$ which is finite. Thus the inverse function $\tau_{A_\infty} = \infty$, so the time-change process will go to infinity in finite time, which is the same as killing.

To actually compute this, recall that we interpreted the killing measure by letting one of our functions be the constant function 1. We have the nice probabilistic interpretation

$$1 - H_{\partial U}(\mathbf{1}) = \mathbb{P}_x(T_{\partial U} = \infty)$$

(that is, the probability that our process started at $x$ doesn't hit the boundary at all). But the left-hand side is harmonic on $U$ and has zero boundary conditions on $\partial U$, and there's a unique such function where this works: it's exactly $\mathbb{P}_{x_0}(T_{\partial U} = \infty)h^U_{x_0}(\cdot)$, where $h$ is the function we used for the Doob $h$-transform. So plugging in, we find after some calculation that

$$\mathcal{E}(H_{\partial U}(u), H_{\partial U}(\mathbf{1})) = \int_{\partial U} u d\kappa, \quad \boxed{\kappa = \mathbb{P}_{x_0}(T_{\partial U} = \infty)\nu^U_{x_0}},$$

giving us an exact equality which in particular implies various heat kernel estimates.

# 16 June 27, 2025

We discussed the boundary trace process last time as an instance of timechange – we found that we were able to calculate the Dirichlet form in an explicit way using the Beurling-Deny decomposition, which let us analyze the jump kernel. We'll now go back to time-changes that are diffusion processes, seeing how we can relate **elliptic and parabolic Harnack inequalities**.

The idea is that elliptic Harnack inequalities are preserved under (1) timechange by admissible measures (we'll explain what this means later) and (2) quasisymmetric change of metric. The symmetric measure doesn't actually play a role in the sense that timechange still yields the same family of harmonic functions (since the extended Dirichlet space doesn't vary).

**Definition 177**

For an MMD space $(X, d, m, \mathcal{E}, \mathcal{F})$, an **admissible measure** is a measure $\mu$ such that $\mu$ is a Revuz measure for some PCAF $A$ which has **full support**.

(Examples that we've already defined include the Kusuoka measure for the Sierpinski gasket or the Liouville measure for the Brownian motion.) The associated time-changed Dirichlet form $(\mathcal{E}^\mu, \mathcal{F}^\mu)$ is then given by

$$\mathcal{F}^\mu = \mathcal{F}_e \cap L^2(\mu), \quad \mathcal{E}^\mu(f, f) = \mathcal{E}(f, f),$$

In particular, $\mathcal{F}_e = (\mathcal{F}^\mu)_e$. In some sense the process doesn't change very much, so we won't have a situation where a diffusion process becomes a jump process. Furthermore because the extended Dirichlet space doesn't change, $(X, d, m, \mathcal{E}, \mathcal{F})$ and $(X, d, \mu, \mathcal{E}^\mu, \mathcal{F}^\mu)$ have the same family of harmonic functions, and thus the elliptic Harnack inequality must be invariant.

The second family of transformations comes from trying to extend the notion of conformal maps to arbitrary metric spaces:

**Definition 178**

A homeomorphism $f : (X_1, d_1) \to (X_2, d_2)$ between two metric spaces is a **quasisymmetry (QS)** if there exists a homeomorphism $\eta : [0, \infty) \to [0, \infty)$ such that distances are controlled in the fashion

$$\frac{d_2(f(x), f(y))}{d_2(f(x), f(z))} \leq \eta\left(\frac{d_1(x, y)}{d_1(x, z)}\right)$$

for any three points $x, y, z \in X$ with $x \neq z$.

We can think of $\eta$ as typically being something like $\eta(t) = t^{\alpha_1} \wedge t^{\alpha_2}$. This tells us that, for example, if $d_1(x, y) = d_1(x, z)$ then we can't have $d_2(f(x), f(y))$ and $d_2(f(x), f(z))$ very far apart; thus circles must be taken to at most ellipses of some specified eccentricity $\eta(1)$.

**Definition 179**

Two metrics $d, \rho$ on a space $X$ are **quasisymmetric** if the identity map $\mathrm{id} : (X, d) \to (X, \rho)$ are quasisymmetric. (In particular, both metrics must have the same topology.)

The notion of quasisymmetry is an equivalence relation (symmetry holds because we can always invert the homeomorphism). And there's a relation between the global quasisymmetry relation and something more infinitesimal:

**Definition 180**

A homeomorphism $f : (X_1, d_1) \to (X_2, d_2)$ is **$k$-quasiconformal ($k$-QC)** if we have the infinitesimal eccentricity condition

$$\limsup_{r \to 0} \frac{\sup \{d_2(f(x), f(y)) : d_1(x, y) \leq r\}}{\inf \{d_2(f(x), f(y)) : d_1(x, y) \geq r\}} \leq k$$

for all $x \in X_1$. (We can think of the numerator as the major axis and the denominator as the minor axis.)

Clearly every quasisymmetric map is quasiconformal by taking $k = \eta(1)$, and we also have a local-to-global property which explains why this is the right generalization:

**Theorem 181** (Gehring, '60s)

Fix some $n \geq 2$. If $f : \mathbb{R}^n \to \mathbb{R}^n$ is a $k$-quasiconformal map, then it is $\eta$-quasisymmetric for some $\eta : [0, \infty) \to [0, \infty)$ depending only on $k$.

We do in fact need $n \geq 2$ here – $n = 1$ is false, and we can check that $f(x) = x + e^x$ is a quasiconformal map which is not quasisymmetric. This result was generalized to a large class of spaces in the 1990s (for example we know that it also holds on the Heisenberg group).

This next fact is a simple consequence of the definition:

**Lemma 182**

If $\mathrm{id} : (X, d_1) \to (X, d_2)$ is an $\eta$-quasisymmetric map, then for any $A > 1$, any center $x \in X$, and any radius $r > 0$, there is some other radius $s > 0$ such that (letting $B_1, B_2$ be the balls under $d_1, d_2$

$$B_2(x, s) \subset B_1(x, r) \subset B_1(x, Ar) \subset B_2(x, \eta(A)s).$$

The picture is that if we have a ball of radius $r$ and of $Ar$ in the first metric, then we can have a ball of radius $s$ inside the small ball and a ball of radius $\eta(A)s$ outside the big ball in the second metric. And we can swap the roles of the two metrics to find that for all $A, x, r$, there is some $s = s(r)$ such that

$$B_1(x, r) \subset B_2(x, s) \subset B_2(x, As) \subset B_1(x, A_1 r),$$

where $A_1 = \frac{1}{\eta^{-1}(A^{-1})}$. The point is that Harnack inequalities only care about these kinds of annuli, and so because the shapes are preserved up to constants we get the following:

**Lemma 183** (Kigami)

Suppose $(X, d_1, m, \mathcal{E}, \mathcal{F})$ and $(X, d_2, m, \mathcal{E}, \mathcal{F})$ are such that $d_1, d_2$ are quasisymmetric. Then if $(X, d_1, m, \mathcal{E}, \mathcal{F})$ satisfies the elliptic Harnack inequality, so does $(X, d_2, m, \mathcal{E}, \mathcal{F})$.

*Proof.* This follows immediately from the estimates discussed above. Suppose that for any nonnegative harmonic function $h$, we know that

$$\sup_{B_1(x,r)} h \leq C \inf_{B_1(x,\tilde{A}r)} h.$$

By our inclusions, we thus have

$$\sup_{B_2(x,s)} h \leq \sup_{B_1(x,r)} h \leq C \inf_{B_1(x,\tilde{A}r)} h \leq C \inf_{B_2(x,\eta(\tilde{A})s)} h.$$

□

**Example 184**

We can in fact build some fractal surfaces with this idea, which we call **snowballs** (they are generalizations of the Koch snowflake). We start with the surface of a three-dimensional cube, divide each face into nine parts, and remove the middle part and add a bump also of height $\frac{1}{3}$. (This turns a single square into a surface with 13 squares of side length $\frac{1}{3}$.) We then iterate this repeatedly: at level $n$ we will have a polygonal surface with $6 \times 13^n$ faces, which are all squares of side length $3^{-n}$.

The resulting surface $S_n$, called the level-$n$ snowball, is homeomorphic to $\mathbb{S}^2$ and comes with the geodesic metric $d_n$ (the idea is that we have lots of "small bumps" and thus the shortest path can have many parts). It turns out that we have convergence in the Gromov-Hausdorff sense

$$(S_n, d_n) \overset{\text{GH}}{\to} (\mathcal{S}, d_{\mathcal{S}})$$

for a surface which we call the snowball; it has Hausdorff dimension $\frac{\log 13}{\log 3}$.

Every polygonal surface can also be viewed as a Riemann surface (a manifold, but the change of coordinates must be holomorphic – we just need to worry about the "corner bump vertices" at which there are three or five right angles. At those points we need to make things flat and thus we need a map $z \mapsto z^{4/5}$ or $z \mapsto z^{4/3}$ instead. But the point is that by the uniformization theorem, we get something conformally equivalent to $\mathbb{S}^2$ and thus unique up to Möbius transformations, and then we need to do an appropriate normalization (for example sending the three corners of a square to $0, 1, \infty$ for consistency of the scaling limit) to get something well-defined.

**Theorem 185** (Meyer '01)

Let $d_{\mathbb{S}^2}$ denote the round metric on $\mathbb{S}^2$. There is a sequence of conformal maps $f_n : (S_n, d_n) \to (\mathbb{S}^2, d_{\mathbb{S}^2})$, as well as a map $f : (\mathcal{S}, d_{\mathcal{S}}) \to (\mathbb{S}^2, d_{\mathbb{S}^2})$, such that we have the following. Letting $\rho_n$ be the metric on $\mathbb{S}^2$ such that $f_n : (S_n, d_n) \to (\mathbb{S}^2, \rho_n)$ is now an **isometry**. Then $\rho_n$ converges pointwise on $\mathbb{S}^2 \times \mathbb{S}^2$ to some function $\rho$, which is itself a metric on $\mathbb{S}^2$, and $\rho$ is quasisymmetric to the round metric $d_{\mathbb{S}^2}$.

In particular, this means that the pushforward of Brownian motion on the snowball should converge to ordinary Brownian motion but just appropriately time-changed.

**Theorem 186**

If we let $\mu_n$ be the uniform surface area measure on $(S_n, d_n)$, then $\mu_n$ converges to the **normalized Hausdorff measure** on $(\mathcal{S}, d_{\mathcal{S}})$. Furthermore, if $Y_n$ is the canonical diffusion on $S_n$, then (with $f_n$ defined in the theorem above) $f_n(Y_n)$ is the timechange of Brownian motion on $\mathbb{S}^2$ with the Revuz measure $(f_n)_* \mu_n$. Then the limiting object $f_*(\mu)$ is an admissible measure for Brownian motion on $\mathbb{S}^2$.

We'd like to understand everything with respect to the original measure, such as the heat kernel. Since elliptic Harnack inequalities are preserved under both of these operations that we've performed, we see that we still have the elliptic Harnack inequality for the diffusion on $(\mathcal{S}, d_{\mathcal{S}}, \mu)$ (via pullback of the timechange).

Notice that we have capacity across annuli of constant order in two dimensions:

$$\mathrm{Cap}(B(x,r), B(x,Ar)^c) \asymp 1 \text{ for } r \lesssim \frac{\mathrm{diam}}{c}.$$

Thus we actually get two-sided capacity bounds of order equal to the volume growth: $\psi(r) = r^{d_f}$ for $d_f = \frac{\log 13}{\log 3}$. Volume doubling is also preserved under these operations, so putting everything together we actually get sub-Gaussian heat kernel bounds for the snowball with $d_W = d_f$.

It turns out that in any situation like we have where $d_W = d_f$ and we have all of these other conditions, the limiting metric is quasisymmetric to the 2-sphere metric. Interestingly, we get a phenomenon with $d_W > 2$ even in a deterministic setting! This is somehow kind of a toy model for Liouville Brownian motion, though it is not useful because the LQG measure is not doubling and the metric is not quasisymmetric to the Euclidean one (it fails to satisfy the metric doubling property).

**Remark 187.** *If we try to do this same story with the snowball but with a $5 \times 5$ grid and put two bumps instead of one, the limit actually is not quasisymmetric anymore! So we do need some amount of symmetry for this argument to work; we need the Brownian motion to not be obstructed in some dimensions more than others. It's possible that there's some weaker version of quasisymmetric which yields distortion bounds only at most scales at most points rather than all scales at all points, which still gives connections to heat kernel bounds.*

So we have two transformations that preserve the elliptic Harnack inequality, and now we want to optimize them in some way, finding the best possible metric to express the diffusion process. This leads to the following notion (which is helping us get a converse of the form "elliptic Harnack implying parabolic Harnack," which is not true in general because we don't have doubling measures):

**Definition 188**

Let $(X, d, m, \mathcal{E}, \mathcal{F})$ be an MMD space. The **conformal walk dimension**, denoted $d_{\mathrm{CW}}$, is defined as follows: define the **conformal gauge**

$$\mathcal{J}(X,d) = \{\rho : \rho \text{ is a metric on } X \text{ quasisymmetric to } d\}$$

and set

$$d_{\mathrm{CW}} = \inf\left\{\beta > 0 : \text{there is some admissible } \mu \text{ and metric } \rho \in \mathcal{J}(X,d) \text{ s.t. } (X, \rho, \mu, \mathcal{E}^\mu, \mathcal{F}^\mu) \text{ satisfies PHI}(\beta)\right\},$$

where PHI($\beta$) means the parabolic Harnack inequality with the function $\psi(r) = r^\beta$.

Observe that if the elliptic Harnack inequality were to fail to hold for some space $(X, d, m, \mathcal{E}, \mathcal{F})$, then the conformal walk dimension must be infinite (since no timechange will make the parabolic Harnack inequality hold, since that would have to imply the elliptic Harnack inequality which is preserved under change of admissible measure). Also, we know that $\frac{\psi(R)}{\psi(r)} \gtrsim \left(\frac{R}{r}\right)^2$ and thus in fact we always have $d_{\mathrm{CW}} \geq 2$.

We will also need a weak condition on the metric, where any two points should be connected by a good curve:

**Definition 189**

We say that $(X, d)$ is **quasi-arc-connected** if there is some homeomorphism $\eta : [0, \infty) \to [0, \infty)$ such that for any $x \neq y \in X$, there is some curve $\gamma_{xy} : [0, 1] \to X$ with $\gamma_{xy}$ a quasisymmetry between $[0, 1]$ and $\gamma_{xy}([0, 1])$.

In other words, any pair of points is connected by an arc with not too much distortion – in particular for a geodesic space we can take the function to be linear. And if we replace the metric $d$ with $d^\alpha$ for some $\alpha \in (0, 1)$, we still have a metric and we still preserve the quasi-arc-connected property. But this replacement turns PHI($\beta$) into PHI($\beta/\alpha$), so it's easy to increase the conformal walk dimension as much as we want (we'll indeed stay within the conformal gauge) – what we're really curious about is what the minimum possible conformal walk dimension is.

**Theorem 190**

Assume that $(X, d, m, \mathcal{E}, \mathcal{F})$ is an MMD space where $d$ is quasi-arc-connected. Then the following are equivalent:

1. The metric space $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies the elliptic Harnack inequality,
2. The conformal walk dimension $d_{\mathrm{CW}}$ is finite,
3. The conformal walk dimension is equal to 2.

The equivalence between the first two properties was shown for geodesic spaces by Barlow, Chen, and Professor Murugan, and the equivalence between the latter two was obtained by Kajino and Professor Murugan. The implications (3) implies (2) implies (1) are immediate from what we've already said, and what's difficult is to go in reverse.

**Example 191**

One question we may ask is whether the infimum in the definition of $d_{\mathrm{CW}}$ is actually attained, and in fact it need not be. One such example is actually the snowball from before, but another one which is illustrative is the following. Recall Kigami's Gaussian heat kernel bound on the two-dimensional Sierpinski gasket with respect to the intrinsic metric (which we've also already shown)

$$d_{\mathrm{int}}(x, y) = \sup \{u(x) - u(y) : u \in C(X) \cap \mathcal{F}, \Gamma(u, u) \leq \mu\}, \quad \mu \text{ the Kusuoka measure.}$$

We in fact have $d_{\mathrm{int}} \in \mathcal{J}(K, d_{\mathrm{Euc}})$, and so we can attain the infimum for the conformal walk dimension in this case. However, Kajino and Professor Murugan showed that Brownian motion on the $n$-dimensional Sierpinski gasket for $n \geq 3$ does **not** attain the infimum of $d_{\mathrm{CW}}$. (And everything is rather case-by-case here – it's not well understood what happens on the Sierpinski carpet instead of the Sierpinski gasket.)

One motivation for obtaining these results above is the following fact:

**Corollary 192**

The elliptic Harnack inequality is stable under perturbations. More concretely, the Laplace-Beltrami operator on a Riemaninan manifold satisfies elliptic Harnack inequalities if and only if any uniformly elliptic operator does.

So this is a generalization of Moser's theorem to all manifolds, and it comes from the previously-known stability of the parabolic Harnack inequality.

*Heuristics for why $d_{CW}$ finite implies $d_{CW} = 2$.* Assume that $(X, d, m, \mathcal{E}, \mathcal{F})$ satisfies the parabolic Harnack inequality for some $\gamma$. We wish to get $\gamma$ down to $2+\varepsilon$, and we do so by "optimistically" trying to construct an admissible measure $\mu$ and a metric $\theta \in \mathcal{J}(X, d)$ such that the time-changed space $(X, \theta, \mu, \mathcal{E}^\mu, \mathcal{F}^\mu)$ satisfies PHI(2).

The basic idea is to view this metric space at different scales: consider an annulus at scale $(r, 2r)$ and discretize at a much finer mesh scale $s$ than $r$. That is, choose an $s$-net $N$ for the metric space $X$ (that is, a maximal $s$-separated subset), and view $N$ as a graph where two points $x_1, x_2 \in N$ are adjacent if $B_d(x_1, \lambda s) \cap B_d(x_2, \lambda s)$ is nonempty (for some parameter $\lambda$). We'll only look at balls of radius $s$ centered at the points in our net (which cover $X$ but do not overlap too much); we now define the "relative diameter"

$$\tilde{\rho}(v) = \frac{\operatorname{diam}_\theta(B_d(v, s))}{\operatorname{diam}_\theta(B_d(x, r))}.$$

That is, we look at a ball of scale $s$ in the original metric and see how much it's distorted under the new metric. This function satisfies the property of **no shortcuts**, meaning that for any path $\gamma$ in the graph (think of this as a sequence of balls) with first vertex $v$ satisfying $B_d(v, s) \subset B_d(x, r)$, last vertex $w$ satisfying $B_d(w, s) \subset B_d(x, 2r)^c$, and all other vertices contained in $B_d(x, 2r)$, we have

$$\sum_{u \in \gamma} \tilde{\rho}(u) \gtrsim 1.$$

This turns out to be true because if we enlarge distances by a factor we should still get the same order for the diameter, and so if we add up the diameters along the path

$$\sum_{u \in \gamma} \frac{\operatorname{diam}_\theta(B_d(u, s))}{\operatorname{diam}_\theta(B_d(x, 2r))} \asymp \sum_{u \in \gamma} \frac{\operatorname{diam}_\theta(B_d(u, \lambda s))}{\operatorname{diam}_\theta(B_d(x, 2r))} \gtrsim \sum_{u \in \gamma} \frac{\operatorname{dist}(B_d(x, r), B_d(x, 2r)^c)}{\operatorname{diam}_\theta(B_d(x, 2r)}$$

and quasisymmetry roughly preserves ratios of distances so if this quantity was of order 1 in $d$ then it also will under $\theta$.

The other condition (which actually uses that parabolic Harnack PHI(2) is satisfied) is called a **mass estimate** on $\mu$. We know by the doubling property that

$$\mu(B_d(x, r)) \asymp \mu(B_d(x, 2r)) \asymp \sum_{v \in N(B_d(x, 2r)) : B(v,s) \cap B(} \mu(B_d(v, s))$$

since the balls roughly don't intersect. Balls in the two metrics are approximately the same, so we can replace this with $B_\theta(v, \operatorname{diam}_\theta(B_d(v, s))$, and the measure of such balls can be related to the capacity and spacetime scaling: capacity across an annulus is roughly the volume divided by the spacetime scaling, so

$$\mu(B_d(x, r)) \asymp \sum_{v \in N(B_d(v, 2r))} \operatorname{diam}_\theta(B_d(v, s))^2 \operatorname{Cap}(B_d(v, s), B_d(v, 2s)^c).$$

But if we've seen the notion of "extremal length" on graphs, that's exactly the admissibility condition we have: the function $u$ will end up being the optimal equilibrium potential (energy minimizer) for the capacity between $B_d(x, 2r)$ and $B_d(x, 2r)^c$. We want $\tilde{\rho}(v)$ to be roughly the same as a discrete gradient $\sum_{w \sim v} \left| \fint_{B_d(v,s)} u dm - \fint_{B_d(w,s)} u dm \right|$. This satisfies both properties we mentioned above (using the Poincaré inequality). Thus somehow using the gradient of energy minimizers is the right way to construct optimal metrics and measures, much like in the Riemann mapping theorem. □