Random Walks on countable groups

written\(^1\) by Antoine Gournoy

The aim of this note is to present some results on the random walks on groups. Throughout the text, we will restrict to countably generated groups and measures which are countably supported and symmetric.

A random walk on a graph \( G \) is a sequence \( \{W_n\}_{n=0}^{\infty} \) of random variables taking value in \( G \) defined as follows. First, fix a measure \( P : S \to [0, 1] \) (which defines how the random walk moves) which is countably supported on a set \( S \subset G \) and such that \( P(g) = P(g^{-1}) \) for any \( g \in G \) (in particular, \( S = S^{-1} \)). Next, fix an initial state: \( W_0 \) follows some probability distribution \( P^0 \) at the start (most of the time a Dirac mass, more precisely, a Dirac mass at the identity).

\[
\mathbb{P}(W_{n+1} = y | W_n = x) = P(x^{-1}y).
\]

Hence, one can compute inductively the law of the \( W_n \). The important point about formulating this inductively is that these random variables are dependant. This data which allows to deduce the \((n+1)\)th variable from the \(n\)th is called the kernel of the Markov process. Here, \( K(h, g) = P(h^{-1}g) \).

The classical picture comes from the Cayley graph of the group for the generating set \( S \). It is the graph whose vertices are elements of \( G \) and \( g \sim h \) if \( \exists s \in S \) such that \( g = hs \). The edges may be decorated by the label \( s \) \(^3\). When the random walker is at a vertex, the label of the edges determine the probability of walking along this edge for his next move.

A convenient notation is

\[
P^n_e(y) = \mathbb{P}(W_n = y | W_0 = x).
\]

Most of the time, one only considers \( P^n := P^n_e \). Note that \( P = P^1 = P^1 \). From there one can check that

\[
P^{n+1}_e(g) = \sum_{s \in S} P^n_e(gs^{-1})P(s) = \sum_{s \in S} P(s)P^n_e(g) = \sum_{h \in G} P^k_e(h)P^n_s(g)
\]

where \( k + \ell = n + 1 \).

This allows us to describe \( P^n \) as a convolution. Given \( f, g : G \to \mathbb{R} \), one defines their convolution as \( f \ast g(\gamma) = \sum_{\eta \in G} f(\eta)g(\eta^{-1}\gamma) \) (one needs to assume the sum is convergent for any \( \gamma \)). Then the law of \( W_n \) is given by \( P^n = P^0 \ast P \ast P \ast \cdots \ast P \) (where \( P \) appears \( n \) times). Note that \( P^n = P^0 \ast P^n_e \) which is why it is, in general, sufficient to understand \( P^n \).

There is another useful and alternative description of \( P^n_e \) via the push-forward. Formally, look at the space \( G^\mathbb{N} \) and endow it with the product probability measure \( P^\mathbb{N} \). Consider the shift

\[
\sigma : \quad G^\mathbb{N} \to G^\mathbb{N}
\]

\((g_0, g_1, g_2, \ldots) \mapsto (g_0 \cdot g_1, g_2, \ldots) \)

and the evaluation at the first coordinate: \( \text{ev} : (g_0, g_1, \ldots) \mapsto g_0 \). Then (writing \( \sigma^n \) for \( \sigma \) applied \( n \) times), one has \( P^n_e \) is the push forward \( (\text{ev} \circ \sigma^n)^\ast P^\mathbb{N} \), i.e. for \( A \subset G \), \( P^n_e(A) = P^\mathbb{N}((\text{ev} \circ \sigma^n)^{-1}(A)) \).

\(^1\)Please email comments / corrections / improvements / references / insults / etc... to first-name.name@gmail.com

\(^2\)In fact, a Markov process.

\(^3\)It may happen that \( s = s^{-1} \), but this is no problem. It may also happen that \( P(e) \neq 0 \), in which case one may draw a loop or a dead-end.
Some motivations

There are many possible sources of motivations to study these objects.

**Heat kernel and the Laplacian.** A first one would be the Laplacian, or more precisely the heat kernel. To make the bridge from random walks to heat kernel, one first needs to pass from a discrete time process to a continuous time process. The typical way to do this is to “use Poisson clocks”. Here is the idea.

A Poisson clock is a clock which rings at random times. It also has the property that even if one knows it did not ring for a while, the probability distribution is the same as if the clock just rang. In mathematical words, the time lapse between two times the clock rings is dictated by an exponential law. This also means that the number of times a clock has rung after time \( T \) is dictated by a Poisson distribution.

Now to pass to continuous time, the reader should do as follows: take his Poisson clock (no particular manufacturer is recommended), and every time it rings, execute one step of the [discrete time] random walk. Those interested can check that this process will still satisfy the Markov property.

Of course, the reader might be avid for a formula. The convolution argument still goes through, so that it suffices to describe \( \hat{P}_e^t \) (with now \( t \in \mathbb{R}_{\geq 0} \))

\[
\hat{P}_e^t(x) = e^{-t} \sum_{k \geq 0} \frac{t^k}{k!} P_e^k(x).
\]

It turns out \( \hat{P}_e^t \) is an extremely good analogue of the heat kernel (in the manifold setting). In fact, most of what is said in the present text about \( P^n \) on a group \( G \) can be said about the heat kernel on \( \widetilde{M} \) the universal covering of a compact manifold \( M \) with \( \pi_1(M) = G \). Indeed, another way to interpret the above formula is:

\[
f * \hat{P}_e^t = e^{-t (\text{Id} - R_P)} f
\]

where \( R_P \) is the operator defined by \( R_P f = f * P \) [and the exponential is to be interpreted in terms of functional calculus]. This is basically as “hot” (i.e. as close to the heat kernel) as you get since \( \text{Id} - R_P \) is the discrete analogue of the Laplacian. The reader will of course notice some discrepancies at some point of the exposition. For example, the isoperimetry on a manifold depends at small scale on the dimension of the manifold. The graph and manifold isoperimetry only agree at the “large scale”.

Some authors prefer to argue directly with \( \hat{P}_e^t \) because one can directly apply the familiar language of differentiation. We will try to avoid it here, because we want to focus on finitely supported \( P \) and it is easy to check that \( \hat{P}_e^t \) is not finitely supported. It decays very fast at infinity, so that most arguments go through without significant problems. One can also check that, if \( P(e) \geq 1/2, \|P_e^{2n} - \hat{P}_e^{2n}\|_{\infty} \leq K t e^{-K^2 t} \) for some constants which depend only on \( P \) (and \( G \)).

From estimates on the heat kernel, one may get a lot of information on the Laplacian \( \Delta \). Here is why one gets information about the inverse of \( \Delta = \text{Id} - R_P \) from \( P^n \). Note that (if defined) \( K_y(x) = \sum_{i \geq 0} P_i^y(x) \) has the property that \( \Delta K_y = \delta_y \), i.e. it is a “fundamental solution”. This sum can also be obtained by writing \( (\text{Id} - R_P)^{-1} \) as a power series in \( R_P \).

This means that \( \Delta^{-1} \) can be expressed directly as a convolution in term of \( G \) (called the Green’s kernel): \( \Delta(f * K_y) = f \). Of course the space of functions on which convolution by \( K \) is defined will depend on how well does the sum \( P_e^i \) converges, which is why it’s a good idea to look at the long term asymptotic of this heat kernel.
It turns out that the “on-diagonal” estimate (i.e. the behaviour of \( n \mapsto P^n_e(e) \)) is already sufficient to get a satisfactory (albeit not complete) answer. Indeed, if \( P(e) \geq 1/2 \), \( \|P^n\|_\infty \leq P^n_e(e) \) hence

\[
\|K_e\|_p \leq \sum_{i \geq 0} \|P^i\|_p \leq \sum_{i \geq 0} \|P^i\|_{\infty^{(p-1)/p}} \|P^i\|^{1/p}_1 \leq \sum_{i \geq 0} P^i_e(e)^{(p-1)/p}.
\]

So if one can prove that the sum on the right is finite, Young’s inequality on convolution will say that \( \Delta^{-1} \) is bounded from some \( l^p \) to some \( l^q \).

Characterising groups. It turns out random walks have been useful to study groups. For example, the Basilica group (and a bunch of other groups) has been shown to be amenable (by Bartholdi & Virág) using random walks.

It is also a difficult problem to classify groups up to quasi-isometry. It turns out that some properties of random walks are invariants of quasi-isometry.

Definition 0.1. For two monotone functions \( u, v : \mathbb{R}_{>0} \to \mathbb{R} \), one says they are dilatationally equivalent if \( \exists a, b, C > 0 \) such that \( u(t) \leq C v(bt) \) and \( v(t) \leq C u(at) \).

In other words, up to scaling the axis, the graph of one of the function bounds the other (and vice-versa).

Example 0.2. Two polynomials are dilatationally equivalent if their degree is the same. The same goes if the function is the [multiplicative] inverse of a polynomial.

If \( u(t) = c^t \) for some \( c \in \mathbb{R}_{>0} \setminus \{1\} \), then it is dilatationally equivalent to \( e^{\pm t} \) where the signs only depend on whether \( c < 1 \) or \( c > 1 \).

Pittet & Saloff-Coste have shown that the return probability is dilatationally invariant under a quasi-isometry. In particular, it does not [up to dilatational equivalence] really depend on \( P \). The volume growth (growth of \( \sup_{P^n} \)) is also a dilatational invariant of quasi-isometries. However, it is still unknown whether the entropy (which lie somehow between volume and return probability) and the speed are.

These four quantities (volume, return probability, speed and entropy) all behave nicely when one uses a surjective homomorphism, and can hence be interpreted as an algebraic obstruction. There has been a growing industry of examples of groups where those quantities behave in various ways which stay in the limit of the “obvious” constraints. However, the obvious constraints are not always that obvious, and it seems that a some of them are not yet known. One of the aims here is mostly to present links between those four quantities and to attract attention on the gray areas of the current knowledge.

1 Return probability

1.1 A few examples

Polyá’s theorem is usually seen as the one of the first result (for infinite groups!). It states that a random walker walking for an infinite time in \( \mathbb{Z}^d \) will, with probability 1, visit the origin infinitely many times if and only if \( d \leq 2 \). It will “go to infinity” with probability 1 if \( d \geq 3 \). This theorem relies on a good knowledge for the asymptotics of the function \( P^n_e(e) \) as \( n \to \infty \) (namely \( P^n_e(e) \asymp Kn^{-d/2} \)).
Example 1.1. The simplest example is the random walk on the group \( \mathbb{Z} \) with \( S = \{ \pm 1 \} \). Since the walker either moves to the left or the right, one can easily see that the law will be, up to cosmetic differences, a Binomial distribution with \( n \) trials and \( p = 1/2 \). More precisely, if \( B_n \) is a random variable with \( n \)-trials Binomial law, then

\[
\mathbb{P}(W_{2n+1} = k) = \begin{cases} \mathbb{P}(B_{2n+1} = k + \frac{2n+1-k}{2}) & \text{if } k \text{ is odd,} \\ 0 & \text{otherwise;} \end{cases}
\]

\[
\mathbb{P}(W_{2n} = k) = \begin{cases} \mathbb{P}(B_{2n} = k + \frac{2n-k}{2}) & \text{if } k \text{ is even,} \\ 0 & \text{otherwise.} \end{cases}
\]

Note that if one is only interested in "rough" properties of this walk, many of these properties can be easily computed via the approximation of the binomial law by the normal law: in this case we get an approximation by \( \mathcal{N}(0, n) \) (the normal law with mean 0 and variance \( n \)).

One can also get very good asymptotics using Stirling’s formula:

\[
P^{2n}(e) = \left( \frac{2n}{n} \right)^{2n} \left( \frac{1}{2} \right)^n \approx \frac{\sqrt{4\pi n(2n/e)^{2n}}}{2\pi n(n/e)^{2n}} \frac{1}{2^{2n}} = \frac{1}{\sqrt{\pi n}}.
\]

For \( \mathbb{Z}^2 \), the same argument (for \( P \) uniform on the standard basis vectors) yields \( P^{2n}(e) \approx \frac{1}{n^2} \).

Let us treat all cases for \( \mathbb{Z}^d \). Only the upper bound will be given (as it is easier).

Example 1.2. Let us try to compute the return probability in \( \mathbb{Z}^d \). Pick some \( P \) symmetric and finitely supported on \( S \). Recall that \( P^n = P \ast \cdots \ast P \) (with \( n \) appearances of \( P \)). Because the Fourier transform turns convolution in multiplication, let, for \( \Theta \in [-\pi, \pi]^d \),

\[
\phi(\Theta) = \sum_{s \in S} e^{is \Theta} P(s).
\]

Fourier analysis (because it turns convolution into multiplication) tells us

\[
P^n(x) = (2\pi)^{-d} \int_{[-\pi, \pi]^d} e^{-is \Theta} \phi(\Theta)^n d\Theta.
\]

To simplify some computations, assume \( P = \frac{1}{2}(\delta_e + P') \) for some other measure \( P' \) (i.e. assume \( P(e) \geq 1/2 \)). Now, pick \( S_+ \subset S \) so that \( S = S_+ \cup S_- \) (actually \( S_- = -S_+ \) because \( \mathbb{Z}^d \) is usually written additively). Then

\[
\phi(\Theta) = \sum_{s \in S} e^{is \Theta} P(s) = \sum_{s \in S_+} \cos(s \cdot \Theta) P(s) = \frac{1}{2} + \sum_{s \in S_+} \cos(s \cdot \Theta) P'(s)
\]

This shows \( \phi \) is real, \( \phi \) is \( \geq 0 \) (since \( \sum_{s \in S_+} P'(s) = 1/2 \)) and \( \phi(\Theta) = 1 \) if and only if \( \Theta = 0 \). This can be made even more visible by writing

\[
\phi(\Theta) = 1 - \sum_{s \in S_+} \left( 1 - \cos(s \cdot \Theta) \right) P(s).
\]

Since \( S_+ \) contains a basis of \( \mathbb{R}^d \),

\[
\phi(\Theta) = 1 - \sum_{s \in S_+} \left( 1 - \cos(s \cdot \Theta) \right) P(s) \leq 1 - C|\Theta|^2
\]
for some $C > 0$ depending on the basis inside $S^+$ and the values $P$ takes there. This last expression is very useful, for example one gets, using $1 - C|\Theta|^2 \leq e^{-C|\Theta|^2}$:

$$P^n(\varepsilon) = (2\pi)^{-d} \int_{[-\pi,\pi]^d} \phi(\Theta)^n d\Theta \leq (2\pi)^{-d} \int_{[-\pi,\pi]^d} e^{-nC|\Theta|^2} d\Theta \leq K/n^{d/2}$$

for some $K > 0$.

We used $P(\varepsilon) \geq 1/2$ to have $\phi \geq 0$, so as to cheaply get that $\phi(\Theta) \leq 1 - C|\Theta|^2 \implies \phi(\Theta)^{2n} \leq (1 - C|\Theta|^2)^{2n}$. This computation can be seriously improved to show that $P^n(\varepsilon) \simeq Cn^{-d/2}$, compute the constant $C$, and show that the first correction term is $C'n^{-1-d/2}...$ ♣

Much more can be computed using Fourier analysis, e.g. the spectrum of $P$ as an operator $l^2 G \to l^2 G$. In a general group, such methods require a very good knowledge of the representation theory (look for the keywords "Plancherel measure"). Here is a slightly different example:

**Example 1.3.** Let us show that if the Cayley graph of $G$ is a tree $T$ of valency $v \geq 3$ (e.g. $G$ is a free group on at least two generators, e.g. $G = \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$) then the probability of return decay exponentially. To do this pick an infinite [geodesic] path $R$ from $e$ to somewhere at infinity. Define the level of a vertex $v$ in the tree by $L(v) = 2d(v, R) - d(v, e)$.

Note that $|L(v)| \leq d(v, e)$. $L$ also defines a map $L : T \to \mathbb{Z}$. Let $Z_t = L(W_t)$. Note that $Z_t$ is going to be a random walk on $\mathbb{Z}$ but with a preference for a direction:

$$\mathbb{P}(Z_{t+1} = k+1 \mid Z_t = k) = \frac{v-1}{v} \quad \text{and} \quad \mathbb{P}(Z_{t+1} = k-1 \mid Z_t = k) = \frac{1}{v}.$$ 

As in the previous example, the law of $Z_t$ will be a (scaled and translated) binomial law (with probability of success $\frac{v-1}{v^2}$). Simple computations give

$$\mathbb{P}(Z_{2n} = 0) = \left(\frac{v-1}{v^2}\right)^n {2n \choose n}.$$ 

Since $|Z_n| \leq d(W_n, e)$, one concludes that

$$P^n_1(e) = \mathbb{P}(W_{2n} = e) \leq \left(\frac{v-1}{v^2}\right)^n {2n \choose n}.$$ 

Recall from 1.1 that ${2n \choose n} \simeq \frac{2^{2n}}{\sqrt{\pi n}}$, then

$$P^n_1(e) \leq \frac{1}{\sqrt{\pi n}} \left(\frac{2\sqrt{v-1}}{v}\right)^{2n}.$$ 

This estimate is surprisingly sharp: the exponential term is correct but the correct polynomial power is of the form $n^{-3/2}$. ♣
In all the above examples, one actually has some form of central limit theorem. To state this slightly more precisely, let \( |g|_s \) be the graph distance from \( g \) to \( e \) in the Cayley graph generated by \( S \). In other words,

\[
|g|_s = \text{minimal number of letters from } S \text{ required to write } g.
\]

**Exercise 1:** Show that for \( S \) and \( S' \) two finite symmetric generating sets, there is a \( K \geq 1 \) such that

\[
K^{-1}|g|_s \leq |g|_{S'} \leq K|g|_s.
\]

In other words, the identity map is bi-Lipschitz between different Cayley graphs.

### 1.2 A few general features

In order to present a few general facts about \( P^n \), it is best to cast things in \( \ell^2 G \). Note that, for any \( f, g \in \ell^2 G \),

\[
\langle \delta_x * f \mid g \rangle = \sum_{\gamma \in G} f(x^{-\gamma}g(\gamma)) = \sum_{\gamma \in G} f(\gamma)g(x\gamma) = \langle f \mid \delta_x * g \rangle.
\]

Similarly, \( \langle f \mid \delta_x \rangle = \langle f \mid g * \delta_x \rangle \). This implies that for a symmetric \( P \in \ell^2 G \subset \ell^2 G \),

(SA) \[
\langle P * f \mid g \rangle = \langle f \mid P * g \rangle \quad \text{and} \quad \langle f \mid P \rangle = \langle f \mid g \rangle.
\]

If \( R_P : \ell^2 G \to \ell^2 G \) is defined by \( R_Pf = f * P \), the second equality is exactly saying that \( R_P \) is self-adjoint. Let us explore some easy consequences.

**Lemma 1.4**

For any \( P \in \ell^2 G \) symmetric, one has

1. \( P^{2n}(e) = \|P^n\|^2 \leq \|\delta_x \|P^n\|^2 \); 
2. \( \sup_{x \in G} P^{2n}(x) = P^{2n}(e) \); 
3. \( n \mapsto P^{2n}(e) \) is non-increasing; 
4. the operator norm of \( R_P : \ell^2 \to \ell^2 \) is \( \leq 1 \); 
5. \( P^{2(n+m)}(e) \geq P^{2n}(e)P^{2m}(e) \), i.e. \( n \mapsto -\ln P^{2n}(e) \) is sub-additive; 
6. The function \( n \mapsto \frac{P^{2n+2}(e)}{P^{2n}(e)} \) is increasing, i.e. \( n \mapsto -\ln P^{2n}(e) \) is concave.

**Proof:** For (a), just note that

\[
P^{2n}(e) = \langle P^{2n} \mid \delta_x \rangle = \langle P^n \mid P^n \rangle = \|P^n\|^2.
\]

Then remark that \( \|\delta_x * f\|_{\ell^2}^2 = \|f\|_{\ell^2}^2 \) for any \( f \in \ell^2 \). (b) comes from a slight variation:

\[
P^{2n}(x) = \langle P^{2n} \mid \delta_x \rangle = \langle P^n \mid \delta_x * P^n \rangle \leq \|P^n\|_{\ell^2} \|\delta_x * P^n\|_{\ell^2} = \|P^n\|^2_{\ell^2} = P^{2n}(e).
\]

(c) can be obtained as a corollary:

\[
P^{2n+2}(e) = \sum_x P^{2n}(x)P^2(x^{-1}) \leq P^{2n}(e) \sum_x P^2(x^{-1}) = P^{2n}(e).
\]
(d) is a simple consequence of Young’s inequality on convolutions (or, in this case, it can be obtained using Cauchy-Schwarz). (e) is an elementary statement: the probability on the right is the probability of returning at time $2(n + m)$ and $2n$ (which is smaller than the probability of returning at time $2(n + m)$).

(f) is another easy consequence of (SA):

$$||P^n||_2^2 = \langle P^n | P^n \rangle = \langle P^{n-1} | P^{n+1} \rangle \leq ||P^{n+1}||_2 ||P^{n-1}f||_2.$$  

Remember that if a ratio test sequence is increasing, then the root test sequence also increases (to the same value). Hence, $n \mapsto P^{2n}(e)^{1/2n}$ is also increasing. Here, (e) would be enough to know (thanks to Fekete’s lemma) that $\lim_{n \to \infty} P^{2n}(e)^{1/2n}$ exists.

The simplest lower bound to $P^{2n}(e)$ is given by looking at volume growth. Let $S = \text{Supp} P$ and $B_n$ be the ball of radius $n$ in the Cayley graph generated by $S$. Then $B_n \supseteq S^n = \text{Supp} P^n$. Because of (b), one gets

$$P^{2n}(e) \geq 1/|B_n|.$$  

For $\mathbb{Z}^d$, one has $B_n \approx n^d$ so this is far from the sharp constant. On the tree the exponential behaviour is correct but the exponent is much bigger.

### 1.3 The operator $R_P$ and amenability

Recall that

$$||R_P||_{p \to q} = \sup\{||R_Pf||_q \text{ s.t. } ||f||_p = 1\}.$$  

Those with some knowledge of operator theory will see fairly easily that:

**Lemma 1.5**

$$\rho := \lim_{n \to \infty} P^{2n}(e)^{1/2n}$$  

is the spectral radius of the operator $R_P$. Also, $||R_P||_{1 \to 2} = P^{2n}(e)^{1/2}$

The proof is left as an exercise:

**Exercise 2:** To prove the previous lemma, one can proceed as follows:

1. Show that for any $f \in \mathcal{E} G$, the map $n \mapsto \frac{||R^{n+1}_Pf||_2}{||R^nf||_2}$. (Hint: see the proof of Lemma 1.4.)

2. Deduce that, for any $f$, $\rho_f := \lim_{n \to \infty} \frac{||R^{n+1}_Pf||_2}{||R^nf||_2}$ exists.

3. Show that if $\forall f, \rho_f \leq \rho$, then $||R_P||_{2 \to 2} \leq \rho$.

4. Show that $\rho_f = \lim_{n \to \infty} ||R_Pf||_1^n$. (Hint: it’s a Lemma from basic analysis: if the ratio test converges, then the root test has converges to the same limit.)

5. Show that for any $x \in G$, $\rho_{\delta_x} = P^{(2n)}(e)^{1/2n}$ and deduce that, if $f$ has finite support, $\rho_f \leq P^{(2n)}(e)^{1/2n}$.

6. Conclude the first statement using 5. (Hint: finitely supported functions are dense in $\mathcal{E} G$.)

7. Show that $||R_P||_{1 \to 2} = P^{2n}(e)^{1/2}$. (Hint: re-read carefully your proof of 5.)

Note that because $R_P$ is self adjoint, $||R_P||_{1 \to 2} = ||R_P||_{2 \to \infty} = ||R_P||_{1 \to \infty}^{1/2}$. 


Theorem 1.6

Assume \( G \) is finitely generated. The following are equivalent:

(i) If the support of \( P \) generates \( G \), then \( \| R_P \|_{\ell^2} = 1 \), i.e. the probability of return decays subexponentially: \( P^{(2n)}(e) 1/2n \to 1 \). [Kesten 1959]

(ii) There is a sequence \( f_n \in \ell^2 G \) of finitely supported elements so that \( \| f_n \|_{\ell^2} = 1 \) and, \( \forall \gamma \in G, \| f_n * \delta_\gamma - f_n \|_{\ell^2} \to 0 \). [Dixmier 1950]

(iii) There is a sequence \( \xi_n \) of finitely supported probability measures such that, \( \forall \gamma \in G, \| \xi_n * \delta_\gamma - \xi_n \|_{\ell^1} \to 0 \). [Reiter 1964-6]

(iv) There exists a Følner sequence, i.e. \( \{ F_n \} \) finite sets such that \( \forall \gamma \in G, \lim_{n \to \infty} \frac{|\partial \gamma \cap F_n|}{|F_n|} = 0 \). [Følner 1955]

If one of the above holds, the group is called amenable\(^4\). A corollary of this theorem is that the above conditions are also equivalent to

\( (i') \) For any \( P \) with finite support, \( \| R_P \|_{\ell^2} = 1 \).

An advantage of \( (i') \) is that it can be formulated on a group which is not necessarily finitely generated. It also shows that amenability is really a property of the group.

There is also a variant of \( (iv) \) which is more elegant on finitely generated groups. Let \( \partial A \) be the edges between \( A \) and \( A^c \).

Lemma 1.7

Assume \( G \) is finitely generated. A sequence \( \{ F_n \} \) is Følner if and only if \( \lim_{n \to \infty} \frac{|\partial F_n|}{|F_n|} = 0 \). Furthermore, this property does not depend on the choice of set \( S \) to draw the Cayley graph.

Exercise 3: Prove the previous lemma. (Hint: use that a Cayley graph always has bounded degree.)

The proof of the theorem will go as follows:

\( (ii) \implies (ii) \implies (iv) \implies (i) \implies (ii) \)

But before moving let us just say a few words about the meaning of these properties. \( (ii) \) comes from representation theory and is there stated as “the right-regular representation weakly contains the trivial representation”. Indeed, when one looks at the vectors \( f_n \), these are almost not change by the translation, so that inside the right-regular representation one “almost” sees the trivial one. From \( (iii) \) one can go very quickly to the original definition of amenability, namely, one can define a measure on infinite subsets which is invariant under translating the subset. \( (iv) \) carries the idea that the group can be well-approximated by finite parts of the group: multiplying by \( x \) an element of \( F_n \) one almost never leaves \( F_n \).

As a last remark, it would also be great if \( P^n \) would be a suitable sequence \( \xi_n \) in \( (iii) \). Indeed, this would mean that an element produced by the random walk is “fairly” generic

\(^4\)The original definition is due to von Neumann (1929) and was introduced in relation to the Banach-Tarski paradox.
(since the distribution obtained by applying some translation is very similar). However, this is a much stronger property (called the Liouville property) which will be discussed afterwards.

**Proof of (ii) \(\implies\) (iii):** Take \(\xi_n = f_n^2\). Then
\[
||\xi_n \ast \delta - \xi_n||_{\ell^1} = \sum_{x \in G} |f_n(x \gamma) - f_n(x)|^2 \\
= \sum_{x \in G} |f_n(x \gamma) - f_n(x)| \cdot |f_n(x \gamma) + f_n(x)| \\
\leq \sum_{x \in G} |f_n(x \gamma) - f_n(x)| \cdot (|f_n(x \gamma)| + |f_n(x)|).
\]

Split in two sums and use Cauchy-Schwarz to get
\[
||\xi_n \ast \delta - \xi_n||_{\ell^1} \leq \|f_n \ast \delta - f_n\|_{\ell^1} (\|f_n\| + \|f_n \ast \delta\|_{\ell^1}) = 2\|f_n \ast \delta - f_n\|_{\ell^1}.
\]
This tends to 0 as \(n \to \infty\).

**Proof of (iii) \(\implies\) (iv):** The idea is to find Folner sets in the level sets of \(\xi_n\). Define for \(\mu \in \ell^1(G)\) a probability measure, \(F(\mu, t) = \{x \in G \mid \mu(x) > t\}\). Then for a fixed \(x\)
\[
|\xi(x) - \mu(x)| = \int_0^1 |1_{F(\xi, t)}(x) - 1_{F(\mu, t)}(x)| \, dt.
\]
Summing over \(x\) one gets,
\[
||\xi_n \ast \delta - \xi_n||_{\ell^1} = \int_0^1 |F(\xi_n \ast \delta, t) - F(\xi_n, t)| \, dt.
\]
Assume, there is no way to find a Folner sequence in \(F(\xi_n, t)\) (as \(t\) and \(n\) vary). This implies that there is some \(\epsilon > 0\) such that \(|F(\xi_n \ast \delta, t) - F(\xi_n, t)| \geq \epsilon|F(\xi_n, t)|\). But then
\[
||\xi_n \ast \delta - \xi_n||_{\ell^1} \geq \epsilon \int_0^1 |F(\xi_n, t)| \, dt = \epsilon||\xi_n||_{\ell^1} = \epsilon.
\]
This contradicts the fact that \(\xi_n\) is a Reiter sequence.

In fact, looking closer at the previous argument shows that “most” level sets are good in the following sense. Let \(v_1 > v_2 > \ldots > v_h > v_{h+1} = 0\) the values taken by \(\xi_n\) and \(F_i = \{x \in G \mid \xi_n(x) \geq v_i\}\). Assume \(||\xi_n \ast \delta - \xi_n||_{\ell^1} \leq \epsilon\), then
\[
\epsilon \geq ||\xi_n \ast \delta - \xi_n||_{\ell^1} = \sum_{i=1}^{n} \beta_i \frac{|F_i \gamma \Delta F_i|}{|F_i|}
\]
where \(\beta_i = (v_i - v_{i+1})|F_i|\) (so that \(\sum \beta_i = ||\xi_n||_{\ell^1}\)). Hence \(\beta\) can be seen as a measure on the indices. Then \(\beta(\{i \mid |F_i \gamma \Delta F_i| \geq C\epsilon|F_i|\}) \leq 1/C\).

**Proof of (iv) \(\implies\) (i):** Take \(f_n = 1_{F_n} / \sqrt{|F_n|}\) (the \(\ell^1\)-normalized characteristic function of \(F_n\)). Suppose the measure \(P\) is supported on \(S \subset G\) (finite). Then a direct computation gives
\[
||R_P f_n||_{L^2} \geq \frac{|\partial^{-S}F_n|}{|F_n|} \to 1,
\]
where \(\partial^{-S}F_n = \{\gamma \in F_n \mid \gamma U \subset F_n\}\). (To see why this should converge to 1, prove lemma 1.7!) By definition, of the operator norm, one has \(||R_P||_{L^2} = 1\).
Proof of (i) ⇒ (ii): Just assume that there is some $P$ with $S = \text{Supp} P$ generating $G$ and $\|R_P\|_{2 \to 2} = 1$. Recall that $R_P$ is contracting, i.e. its operator norm is $\leq 1$ (see 1.4.(d)). Next, $R_P$ has no eigenvector of eigenvalue 1, i.e. no $f \in \ell^2 G$ such that $f * P = f$. Indeed, such a function would be $P$-harmonic and, hence, satisfy a maximum principle. But since $f \in \ell^2 G$ it must decrease to 0 at infinity. This implies $f$ must be 0 everywhere.

If you are not familiar with harmonic functions, just think of it as follows. Pick some element $x$. The equation $f = f * P$ tells you you can express the value at $x$ in terms of a convex combination of values taken on the set $S = \text{Supp} P$. Now repeat this for each element of $S$. This might make $f(x)$ pop-up again but you can cancel it using the first step. Repeating this process over and over again, you will have an expression of $f(x)$ in term of a convex combination of values of $f$ which are as far away as you like. However, these need to tend to 0 because $f \in \ell^2 G$. Hence, at any point the value is 0.

As before, it is convenient to consider $P' = \frac{1}{2} (\text{id} + P)$ instead of $P$. Indeed, $P'$ has positive spectrum. Repeating the argument above, one sees that 1 is in the spectrum\(^5\) (since $\|R_P\|_{\ell^2 \to \ell^2} = 1$) but there are no such eigenvectors. This means there are almost eigenvectors: for any $\epsilon > 0$ there is a $f$ of norm 1 which decomposition\(^6\) is made of “eigenvectors” of eigenvalue $\geq 1 - \epsilon$, i.e. $\langle R_P f_n, f_n \rangle \geq 1 - \epsilon$. One may also assume this $f$ is finitely supported by cutting off very far away.

Thus

$$\sum_x (1 - \langle f * \delta_x | f \rangle) P(x) = 1 - \langle R_P f, f \rangle \leq \epsilon$$

But for any $x$, $\|f * \delta_x\|_{\ell^2} = \|f\|_{\ell^2} = 1$ so $(1 - \langle f * \delta_x | f \rangle) \geq 0$. This implies that, for any $x$ in the support of $P$, $1 - \langle f * \delta_x | f \rangle \leq \epsilon / \inf_{s \in S} P(s)$.

Recall that $\|a - b\|_{\ell^2} = \langle a - b, a - b \rangle = 2 - 2 \langle a, b \rangle$. Hence, for all $x$ in the support of $P'$,

$$\frac{1}{2} \|f * \delta_x - f\|_{\ell^2}^2 = 1 - \langle f * \delta_x | f \rangle \leq \epsilon / \inf_{s \in S} P'(s)$$

To conclude, let $F_k$ be an increasing sequence of finite sets so that $\cup F_k = G$. Define $f_k$ by looking at $P^k$ (the support of $P^k$ is $B_k$, the ball of radius $k$) and then take $\epsilon = \frac{1}{k} \inf_{x \in B_k} P^k(s)$. This gives a sequence $f_k$ which satisfies $\|f * \delta_x - f\|_{\ell^2} \leq 2/k$ for any $x \in F_k$ and concludes the proof. \[\square\]

Before moving on, it is worth to mention that a concatenation of the above proofs to get (iii) and some extra efforts, show that (assuming $P(x) = 1/2$) $x \mapsto P^n(x)^2 / \|P^n\|^2$ is a sequence of almost-invariant measures. The reason it is “computationally inferior” to having $P^n$ a sequence of almost-invariant measure is that squaring (and computing the $\ell^2$-norm) requires to know the law of $P^n$. This requires to identify all the outcomes; in particular, to identify when two “words” with letter in $S$ give the same group element. This can be impossible (in groups where the word problem is not solvable) and at the very least cumbersome (in more explicit groups). On the other hand (say in a group where the elements are explicit, e.g. a subgroup of matrices) sampling directly from $P^n$ is a simple process.

Example 1.8. Here are example of amenable groups:

- finite groups;
- Abelian, nilpotent, polycyclic and solvable groups;

\(^5\)With more efforts, one can show contractions on $\ell^2$ always have vectors which are “almost” of eigenvalue 1.
\(^6\)This can, of course, be said more rigorously with spectral decomposition of self-adjoint operators.
if $B_n$ is the ball of radius $n$ (around $e$) in the Cayley graph and $\liminf \frac{1}{n} \log |B_n| = 0$ then the group is amenable (exercise!). [Such groups are called of “subexponential” growth.]

Here are examples of non-amenable groups:

- free groups on at least 2 generators;
- hyperbolic groups (non-elementary ones, i.e. except virtually-$\mathbb{Z}$ groups$^7$);
- some infinite torsion groups (“Burnside groups”).

Given a few amenable groups there are many ways to build new ones:

**Theorem 1.9 (“The closure properties”)**

Let $\Gamma$, $N$ and $\{\Gamma_i\}_{i \geq 0}$ be amenable groups.

(a) If $H$ is a subgroup$^8$ of $\Gamma$ then $H$ is amenable  

(b) If $H$ is an extension of $N$ by $\Gamma$ (i.e. $1 \to N \to H \to \Gamma \to 1$ is an exact sequence) then $H$ is amenable

(c) If $N \vartriangleleft \Gamma$ then $H = \Gamma/N$ is amenable

(d) If $H$ is a direct limit of the $\Gamma_i$ then $H$ is amenable

Note that any group containing a non-amenable group is non-amenable (by (a)). The previous properties (with the eventual exception of (d)) were first shown in von Neumann [24] (together with the first definition of amenability).

Lastly, to make a bridge with the motivations, one could ask: when is convolution by $K_e$ bounded from $\ell^p \to \ell^p$?

**Lemma 1.10**

The following are equivalent

1. $\| R_P \|_{2 \to 2} < 1$.
2. $\exists p \in ]1, \infty[, \| R_P \|_{p \to p} < 1$.
3. $\forall p \in ]1, \infty[, \| R_P \|_{p \to p} < 1$.

Also, $Id - R_P$ has a bounded inverse $\ell^p \to \ell^q$ if and only if $\| R_P \|_{2 \to 2} \leq 1$.

Hence, if one of the above holds then the formal inverse $(Id - R_P)^{-1} := \sum_{i \geq 0} R_P^i$ is bounded from $\ell^p \to \ell^p$.

**PROOF:** The statements after the equivalences are classical operator theory.

The rest of the proof is essentially present in Lohoué [17] (but may be anterior). It is quite easy to check that $\| R_P \|_{1 \to 1} = 1 = \| R_P \|_{\infty \to \infty}$ (we only need that they are $\leq 1$ which is proven in Lemma 1.4). By Riesz-Thorin interpolation, if $q \in ]p_0, p_1[$ (with $p_0, p_1 \in [1, \infty]$ then

$$\| R_P \|_{q \to q} \leq \| R_P \|_{p_0 \to p_0} \| R_P \|_{1 \to 1}^{1 - t} \| R_P \|_{p_1 \to p_1}^t$$

$^7$If $P$ is a property of groups (e.g. being $\mathbb{Z}$, being Abelian, being nilpotent, ...), a group group is said to be virtually-$P$ is it contains a subgroup of finite index which has the property $P$. 

$^8$If $\Gamma$ is a subgroup of $G$, we note $\Gamma \vartriangleleft G$ the $\Gamma$ is normal in $G$. The quotient group $H = G/\Gamma$ is the factor group of $G$ by $\Gamma$. If $\lambda$ (resp. $\lambda'$) is a homomorphism of $G$ into $\mathbb{C}$ (resp. $\mathbb{C}^*$), then the pull back (resp. push forward) of $\lambda$ (resp. $\lambda'$) by $\Gamma$ is the homomorphism $\lambda \Gamma : \Gamma \rightarrow \mathbb{C}$ (resp. $\lambda' \Gamma : \Gamma \rightarrow \mathbb{C}^*$) defined by $(\lambda \Gamma)(\gamma) = \lambda(\gamma \Gamma)$.
where $t$ satisfies $\frac{1}{q} = \frac{1}{p_0} + \frac{1-t}{p_1}$. The conclusion follows easily. Assuming $\|R_P\|_{p \to p} < 1$, for any $q < p$ apply the above to $p_0 = 1$ and $p_1 = p$ gives $\|R_P\|_{q \to q} < \|R_P\|_{p \to p}^{-1} < 1$ (because $t \neq 1$).

For any $q > p$, apply the above to $p_0 = p$ and $p_1 = \infty$.

\section*{1.4 Some results about $P^n$ and $|B_n|$}

Before moving on, let us just mention kind of behaviours one can expect from the functions $n \mapsto P^n(e)$ and $n \mapsto |B_n|$. This will be helpful to know whether what we are going to prove is sharp or not.

Let us start with $|B_n|$ and a nice exercise:

**Exercise 4:** A group is said to have the volume doubling property if there is a $C$ such that $|B_{2n}| \leq C|B_n|$. Show that such a group satisfies $|B_n| \leq (2m)^k$ where $k = \ln C/\ln 2$ (better estimates are possible!). (Hint: start with $n = 2^l$, then use $|B_{n+m}| \leq |B_{2n}|$ if $m \leq n$)

1. If $|B_n| \leq n^a$ for some $a \in \mathbb{R}$, then there is a $d \in \mathbb{N}$ and constants $K_1, K_2$ so that $K_1n^d \leq |B_n| \leq K_2n^d$. Furthermore, the group contains a finite index subgroup which is nilpotent 9 (this is a famous, and famously nontrivial, theorem of Gromov). The canonical example of a nilpotent group is the group of $k \times k$ matrices with integer coefficients whose lower diagonal entries are all 0 and diagonal entries are all 1. Subgroup of this group are also nilpotent.

2. Say a group is linear if it can represented as a matrix group. It is either solvable 10 (a condition which implies amenability) or it contains a subgroup isomorphic to the free group (Tits alternative). A group which is solvable but does not have polynomial growth is exponentially growing (a theorem of Milnor & Wolf 1968). If the coefficients of a linear group are integral, then solvable implies polycyclic 11 (a result of Malcev). And if any group is polycyclic then it is a subgroup of the integral matrices (a result of Auslander and Swan).

3. Grigorchuk was the first to show that there are groups (even finitely presented ones) which have growth of type $e^{an}$ \leq |B_n| \leq e^{bn}$ for $0 < a < b < 1$. A group is finitely presented if and only if it the fundamental group of a compact manifold (a result of Seifert). Bartholdi & Erschler have constructed groups with $|B_n| \simeq e^{an}$ with $a \in [3/4, 1]$ (and some other behaviours between those two).

4. It is still unknown if other behaviours are possible. It is known that if a group has superpolynomial $|B_n|$, then $|B_n| \geq n^{\ln n}$ (Shalom & Tao).

See [11] for more details! It is still unknown if there is a group with growth \leq e^{n^{3/4}}

---

9Nilpotent groups contain Abelian groups and some groups which fail to be Abelian "by little". To be precise, let $[A, B] = \{aba^{-1}b^{-1} \mid a \in A, b \in B\}$. Let $G_0 = G$ and $G_{i+1} = [G, G_i]$. The $G_i$ are normal subgroups of $G$. $G$ is nilpotent if and only if $G_n = \{e\}$ for some $n$.

10Solvable groups is a larger class than nilpotent groups. This time, let $G_0 = G$ and $G_{i+1} = [G_i, G_i]$. $G_{[i]}$ is a normal subgroups of $G_{[i-1]}$ (and hence of $G$). $G$ is solvable if and only if $G_{[n]} = \{e\}$ for some $n$. There is a world of solvable but not nilpotent groups.

11nilpotent $\subseteq$ polycyclic $\subseteq$ solvable... A group is polycyclic if it is solvable and $G_{[i]}/G_{[i+1]}$ is finitely generated for every $i$. There is again a world of difference between polycyclic and solvable, as well as between polycyclic and nilpotent.
For $P_n^e(e)$, we require $P(e) \geq 1/2$ (to avoid some bipartiteness issue) and $\sim$ is to be understood as dilatational equivalence.

1. $|B_n| \sim n^d$ is equivalent to $P_n^e(e) \sim n^{-d/2}$.

2. If a group is polycyclic, then $P_n^e(e) \sim e^{-n^{1/3}}$. Some solvable but not polycyclic groups (lamplighter on $\mathbb{Z}$) also present this behaviour (Pittet & Salo-Coste).

3. There are solvable groups with $P_n^e(e) \sim e^{-n^{d/2}}$. Some solvable but not polycyclic groups (lamplighter on a nilpotent group) also present this behaviour (Pittet & Salo-Coste).

4. There are solvable groups with $P_n^e(e) \lesssim e^{-f(n)}$ for any function $f(n)$ which is $o(n)$ (Erschler). Finitely presented examples are possible if one restricts to $f(n) \sim n^\alpha$ (Pittet & Salo-Coste).

5. There are amenable groups with $P_n^e(e) \gtrsim e^{-f(n)}$ for any function $f(n)$ which is not $o(n^{1/3})$ (implicit in Kotowski & Virag). Finitely presented examples are possible if one restricts to $f(n) \sim n^\alpha$.

However, there are still no known way to construct groups with some behaviours. There is no known group with $P_n^e(e) \gtrsim e^{-f(n)}$ where $f(n)$ is $o(n^{1/3})$ and $P_n^e(e)$ is not polynomially decreasing.

1.5 Transience and recurrence.

2 Functional inequalities

References: most of this section is taken from Coulhon & Salo-Coste [7] or Woess [27]. Although Nash [and many others] already saw links between heat kernels decay, the main breakthrough of this theory was established by Varopoulos around 1985 (e.g. see [23]).

Consider the edges of our Cayley graph $E \subset G \times G$ as directed edges, namely

$$E = \{(x, y) \in G \times G \mid x^{-1} y \in S\} = \{(x, xs) \mid x \in G, s \in S\}.$$ 

Define $\nabla$ of any function $f : G \to \mathbb{R}$ by

$$\nabla f(x, y) = f(y) - f(x).$$

Define $\ell^2(E, \frac{1}{P})$ as follows, if $f : E \to \mathbb{R}$ then

$$\|f\|^2_{\ell^2(E, \frac{1}{P})} = \sum_{s \in G} \sum_{x \in S} \frac{1}{P(s)} f(x, xs)^2 = \sum_{x \in G} \sum_{y \sim x} \frac{1}{P(x^{-1} y)} f(x, y)^2.$$

Remark that $\nabla^*$

$$\nabla^* : \{E \to \mathbb{K}\} \to \{G \to \mathbb{K}\}$$

$$f \mapsto \nabla^* f(x) = \sum_{y \sim x} P(y^{-1} x) g(y, x) - P(x^{-1} y) g(x, y)$$

Note that $\nabla^* \nabla = \text{Id} - R_P$. 

The main idea goes as follows:

\[ P_{2n}(e) - P_{2n+2}(e) = \langle P^n - P^{n+2} \mid P^n \rangle = \| \nabla_P P^n \|_2^2 \geq \mathcal{A}(\| P^n \|_2^2) \]

where \( \mathcal{A} \) is some function to be determined yet. It turns out that if \( \mathcal{A} \) is known one can deduce some things about the behaviour of \( P_{2n}(e) \). Basically, before doing the \( \ell^2 \) case, one might one ot see what such an inequality means for \( \ell^1 \) norms, i.e. \( \| \nabla f \|_1 \geq \) some function of \( \| f \|_1 \). It turns out, this reduces to the case of the isoperimetric problem (how large is the boundary of a set).

2.1 The isoperimetric and pseudo-Poincaré inequalities

The boundary of a set \( A \subset G \) is the set of edges \( \partial A \) between \( A \) and \( A^c \).

The isoperimetric profile (w.r.t. \( P \)) is defined as

\[ I(n) = \inf \{|\partial A| \mid A \subset G \text{ and } |A| = n\} = \inf \{|\nabla 1_X| \mid X \subset G \text{ and } |X| = n\} \]

This is saying that the boundary of a set \( A \) has size at least \( I(|A|) \).

The main result about isoperimetry is

Corollary 2.1 (The isoperimetric inequalities)

Let \( \phi(\lambda) = \inf\{n \mid |B_n| \geq \lambda\} \). Then

\[ \frac{|A|}{\phi(2|A|)} \leq 2|\partial A| \]

i.e. \( I(n) \geq n/2\phi(n/2) \).

\( \phi \) is a “generalised inverse” to \( n \mapsto |B_n| \)

Example 2.2. If \( |B_n| \geq Kn^d \), one gets \( I(n) \geq K'n^{(d-1)/d} \). This is called the \( d \)-dimensional isoperimetric profile.

If \( |B_n| \geq K \exp(n^\nu) \), one gets \( I(n) \geq K'n/(2+\ln n)^{1/\nu} \). Let us call this a \( \nu \)-intermediate isoperimetric profile (non-standard terminology).

If \( I(n) \geq Kn \), this is called a strong isoperimetric profile.

Exercise 5: What about the other direction?

1. If one has a \( d \)-dimensional isoperimetric profile, then \( |B_n| \geq Kn^d \).

2. If \( I(n) \geq Kn/(2+\ln n)^{1/\nu} \), show that \( |B_n| \geq K \exp(n^{(\nu+1)/\nu}) \).

3. Show that \( G \) is not amenable if and only if \( I(n) \geq Kn \) for some \( K \).

4. Show that if \( G \) is not amenable, then \( |B_n| \) grows exponentially.

(Hint: \( |\partial B_n| \) is (up to constants) \( |B_{n+1}| - |B_n| \).)

Changing \( S \) only changes the isoperimetric profile by a dilatational equivalence. Hence, having a isoperimetric profile of the above form may be considered a property of the group, and not of \( P \).

Corollary 2.1 follows directly [by taking \( f = 1_A \) and \( \lambda = 1 \)] from
Theorem 2.3

Let \( f \in \ell^1 G \). For \( \lambda \in \mathbb{R} \), let \( \{ f \geq \lambda \} = \{ x \mid f(x) \geq \lambda \} \). Then

\[
|\{ f \geq \lambda \}| \leq \phi \left( \frac{\lambda}{2} \| \nabla f \|_1 \right) \| \nabla f \|_1
\]

PROOF: Define the averaging of \( f \) on a ball of radius \( n \) by

\[
f_n(x) = \left( f * 1_{B_n}/|B_n| \right)(x) = \frac{1}{|B_n|} \sum_{y \in B_n} f(xy)
\]

Note that

\[
|\{ f \geq \lambda \}| \leq |\{ f - f_n \geq \lambda/2 \}| + |\{ f_n \geq \lambda/2 \}|
\]

However, \( \| f_n \|_\infty \leq |B_n|^{-1} \| f \|_1 \). Hence if \( n_0 \) is the smallest integer such that \( |B_{n_0}|^{-1} \| f \|_1 < \lambda/2 \) (i.e. \( n_0 = \phi \left( \frac{\lambda}{2} \| f \|_1 \right) \)), then \( |\{ f_n \geq \lambda/2 \} | = 0 \) and

\[
|\{ f \geq \lambda \}| \leq |\{ f - f_{n_0} \geq \lambda/2 \}| \leq \frac{2}{\lambda} \| f - f_{n_0} \|_1
\]

The last inequality uses the Markov inequality. The proof is complete if one knows the pseudo-Poincaré inequality:

\[
\| f - f_n \|_1 \leq n\| \nabla f \|_1
\]

which we will prove in a few moments. \( \square \)

Lemma 2.4 (The pseudo-Poincaré inequality)

For any \( f \in \ell^1 \) (in particular of finite support), \( \| f - f_n \|_1 \leq n\| \nabla f \|_1 \)

PROOF: Note that \( \| f - f * \delta_y \|_1 \leq \| \nabla f \|_1 \) if \( y \) is in \( S \). If \( y \in B_n \), then apply the triangle inequality: write \( y = s_1 \ldots s_n \) with \( s_i \in S \) and

\[
\| f - f * \delta_y \|_1 \leq \| \sum_{i=1}^n f * \delta_{s_1 \ldots s_{i-1}} - f * \delta_{s_1 \ldots s_i} \|_1
\]

\[
\leq \sum_{i=1}^n \| f(s_1 \ldots s_{i-1}) - f(s_1 \ldots s_i) \|_1
\]

\[
\leq n\| \nabla f \|_1
\]

and then

\[
\| f - f_n \|_1 = \| f - \frac{1}{|B_n|} \sum_{y \in B_n} f * \delta_y \|_1 = \| \frac{1}{|B_n|} \sum_{y \in B_n} f - f * \delta_y \|_1
\]

\[
\leq \frac{1}{|B_n|} \sum_{y \in B_n} \| f - f * \delta_y \|_1 \leq n\| \nabla f \|_1
\]

This concludes the proof. \( \square \)

This inequality (and the proof) generalises without problem if one changes \( 1 \) to \( p \in [1, \infty] \). Also, in some groups, much better inequalities may hold (e.g. a Poincaré inequality).

Question 2.5: Is \( I \) dilatationally equivalent to \( I'(n) := \inf \{ |\partial A| \text{ s.t. } |A| \geq n \} \)?

Question 2.6: Is \( J(n) := \sup \{ |A| \text{ s.t. } |A| \leq n \} \) dilatationally equivalent to \( I_n \)?

In a group \( J(n) := \sup \{ |A| \text{ s.t. } \frac{n}{2} \leq |A| \leq n \} \). It is known that \( I \) and \( J \) are both dilatationally invariant under changing \( S \).
2.2 Nash inequalities

There are many (other) inequalities that one can obtain from Corollary 2.1. For example, if \( f \) is finitely supported,
\[
\|f\|_{P/(d-P)} \leq K \|\nabla f\|_P
\]
which can be interpreted as a Sobolev embedding. We will here need the Nash inequality. Note that the Nash inequality can be derived from the Sobolev inequality (together with a careful use of Hölder’s inequality) if \( d > 2 \).

The Nash inequality can be seen as a desire to get an upper bound on \( \|f\|_2 \) solely depending on \( \|\nabla f\|_2 \). It turns out this is impossible in a linear fashion:

**Exercise 6:** Show that if, for any finitely supported \( f \), \( \|f\|_2 \leq K \|\nabla f\|_2 \), then \( \|R_p\|_{2 \to 2} < 1 \). (Hint: Use that \( \|\nabla f\|_2^2 = \langle (I-R_p)f \mid f \rangle \).)

So in the non-amenable case, the relation has to be more complicated.

**Lemma 2.7**

Assume \( f \) is finitely supported, and let \( \eta(t) = t/I(t) \). Assume \( \eta \) is increasing, then
\[
\|f\|_2 \leq |S|2\sqrt{2}\eta \left( \frac{4\|f\|_2^2}{\|f\|_2^2} \right) \|\nabla f\|_2
\]

**Proof:** WLOG, one may assume \( f \) is positive (it does not change the norms of \( f \) but reduces \( \|\nabla f\| \)). Then
\[
\|f\|_1 = \int_0^\infty |\{ f > t \}| \mathrm{d}t \leq \eta(|\text{Supp } f|) \int_0^\infty |\partial \{ f > t \}| \mathrm{d}t
\]
\[
= \eta(|\text{Supp } f|) \|\nabla f\|_1.
\]

Next, using Cauchy-Schwarz (write the first term as a sum of \( a^2 - b^2 = (a-b)(a+b) \)),
\[
\|\nabla(f^2)\|_1 \leq K \|\nabla f\|_2 \|f\|_2,
\]
where \( K = 2|S| \). So
\[
\|f\|_2 = \frac{\|f^2\|_1}{\|f\|_2} \leq Kg(|\text{Supp } f|) \|f\|_2.
\]

Now let \( f_{[t]}(x) = \max \{ f(x) - t, 0 \} \). Then \( f^2 \leq f_{[t]}^2 + 2tf \), \( \|\nabla f_{[t]}\|_2 \leq \|\nabla f\|_2 \) and \( |\text{Supp } f_{[t]}| \leq \|f\|_1/t \) (Markov’s inequality). Putting this together gives:
\[
\|f\|_2^2 \leq \|f_{[t]}\|_2^2 + 2t\|f\|_1 \leq K^2g(|\text{Supp } f_{[t]}|)^2\|\nabla f_{[t]}\|_2^2 + 2t\|f\|_1 \leq K^2g(\|f\|_1/t)^2\|\nabla f\|_2^2 + 2t\|f\|_1
\]
Pick \( t = \frac{\|f\|_1^2}{4\|f\|_1} \) to conclude.

**Example 2.8.**
1. So in the \( d \)-dimensional case, this gives:
\[
\|f\|_2^{1+2/d} \leq K \|f\|_1^{2/d} \|\nabla f\|_2.
\]

2. In the \( v \)-intermediate case, one has
\[
\|f\|_2 \left( \ln(2 + 4\frac{\|f\|_2}{\|f\|_2}) \right)^{-1/v} \leq K \|\nabla f\|_2.
\]
3. Lastly, in the strong isoperimetric case, one gets
\[ \|f\|_2 \leq K \|\nabla f\|_2 \]

**Exercise 7:** What about converses? Use characteristic functions to

1. show that 2.8.3 gives back the strong isoperimetric inequality: \(|A| \leq K'|\partial A|\).

2. show that 2.8.2 implies the \(\frac{v}{2}\) -intermediate isoperimetry: \(|A| \ln(2 + |A|^{-v}) \leq K'|\partial A|\).

3. show that 2.8.1 returns only the \(d\)-dimensional isoperimetry: \(|A|^{1-2/d} \leq K'|\partial A|\).

It turns out that, for groups, there is a much more direct proof of these estimates.

**Exercise 8:** Obtain 2.9.1 and 2.9.2 directly from volume and the \(\ell^2\)-pseudo-Poincaré inequality. First, argue that
\[ \|f\|_2 \leq \|f - f_n\|_2 + \|f_n\|_2 \leq n\|\nabla f\|_2 + |B_n|^{-1/2}\|f\|_2 \]
and then make an optimisation on \(n\).

The proofs using isoperimetry is however better adapted to a more general context.

### 2.3 Upper bound on the heat kernel

The main goal of this subsection is to prove:

**Corollary 2.9**

1. The \(d\)-dimensional isoperimetry implies \(P^{2n}_e(e) \leq K'n^{-d/2}\).

2. The \(v\)-intermediate isoperimetry implies \(P^{2n}_e(e) \leq K_1 \exp(-K_2 v^{v/(2+v)})\).

3. The strong isoperimetry implies \(P^{2n}_e(e) \leq K_1 \exp(-K_3 n)\).

2.9.1 and 2.9.3 are sharp. 2.9.2 is also sharp in the case that \(v = 1\) [it is sharp for polycyclic groups]. Groups of growth neither exponential nor polynomial remain too poorly understood to decide whether 2.9.2 can sharp for other \(v\)...

Before showing Corollary 2.9, the following further technical properties of \(P^n\) are required.

**Lemma 2.10**

Let \(P' = \frac{1}{2}(\delta_e + P)\) and \(f\) be any finitely supported function. Then

1. \(P^{2n}_e(e) \leq 2P^{2n}_e(e)\)

2. \(\|\nabla_P f\|_2^2 \leq 2(\|f\|_2^2 - \|R Pf\|_2^2)\)

3. \(n \mapsto \|R^nf\|_2^2 - \|R^{n+1}f\|_2^2\) is decreasing.
Proof: Using that \( n \mapsto P^{2n}_e(e) \) is decreasing,
\[
P^{2n}_e(e) = \frac{1}{2^{2n}} \sum_{i=0}^{2n} \binom{2n}{i} P^i_e(e) \geq \frac{1}{2^{2n}} \sum_{k=0}^{n} \binom{2n}{2k} P^{2k}_e(e) = \frac{1}{2^{2n}} P^{2n}_e(e) \sum_{k=0}^{n} \binom{2n}{2k} = \frac{1}{2} P^{2n}_e(e)
\]

For the second inequality,
\[
\|f\|^2 - |P^{2}_f| f ||^2 = ((I - P^2_0) f | f) = \|\nabla P^2 f \|^2 \geq \frac{1}{2} \|\nabla P f \|^2
\]
where the last inequality comes from the fact that \( P^2_0 = \frac{1}{4} \delta_e + \frac{1}{2} P + \frac{1}{4} P^2 \geq \frac{1}{2} P \) element-wise.

The last statement is similar, it uses the existence of a square root for \( I - P^{2}_0 \) (which commutes with \( P \)) and \( \|P\|_{\infty} \leq 1 \): 
\[
\|P^2 f \|^2 - |P^{2+1}_f| f ||^2 = ((I - P^2_0)^{1/2} P^2 f | f) \leq \|(I - P^2_0)^{1/2} P^2 f \|^2 = \|P^{2^{-1}} f \|^2 \leq \|P^2 f \|^2
\]

Theorem 2.11

Assume the Nash inequality

\[
\|f\|^2 \leq \mathcal{A} \left( \frac{\|f\|^2}{\|f\|^2} \right) \|\nabla f \|^2
\]
holds for some non-increasing \( \mathcal{A} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \). Then \( P^{a}_e(e) \leq c(n) \) where \( c : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \) satisfies
\[
2a'(t) = - \frac{c(t)}{\mathcal{A}(1/c(t))}.
\]
and \( c(0) = 1 \)

Proof: Thanks to Lemma 2.10, apply the Nash inequality to any \( f \) of finite support with \( \|f\|_1 = 1 \):
\[
\|f\|^2 \leq 2\mathcal{A}(1/\|f\|^2)(\|f\|^2 - \|P' f \|^2)
\]
Set \( u(n) = \|P^n\|^2 \) and use this inequality on \( f = P^n_e \):
\[
u(n) \leq 2\mathcal{A}(1/u(n)) \left( u(n) - u(n+1) \right)
\]
Take a piecewise linear interpolation of \( u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}, \) and rewrite the above as:
\[
u(t) \leq -2\mathcal{A} \left( 1/u(t) \right) v'(t+).
\]
Since \( u(0) = 1 \), one can obtain \( u(t) \leq c(t) \) by traditional calculus. The result follows by Lemma 2.10.

Corollary 2.9 is then obtained by further [long] computations (see Woess [27, p.151]).

One can define a \( \ell^2 \)-isoperimetric profile by \( J_2(n) = \sup \{ \|f\|_2^2 | \|\text{Sup} f\| = n \} \). A similar definition leads to a Nash profile. It is known (under very very general hypothesis, though not in full generality) that the \( \ell^2 \)-isoperimetric profile, the Nash profile and the “return profile” (i.e. \( n \mapsto P^{2n}_e(e) \)) are all related by some formulas.

Question 2.12 (hard): Are there groups with dilatationally equivalent \( I \) but where the \( J_2 \) are not dilatationally equivalent?

Before moving on, let us also go back to one of our motivations.
Corollary 2.13

Assume $G$ has $d$-dimensional isoperimetry, then the Laplacian $\Id - R_P$ has a bounded inverse from $\ell^q$ to $\ell^r$ for any $r > \frac{dq}{d-2q}$.

In particular, if $G$ does not have polynomial growth (i.e. is not virtually nilpotent) then $(\forall \epsilon > 0)$ the inverse is bounded from $\ell^q \to \ell^{q+\epsilon}$.

Proof Proof: First note that if $P' = \frac{1}{2}(\delta_e + P)$, then $P_{2n+2}(e) \geq \frac{1}{2}P_{2n+1}(e)$. So the behaviour of even times dominate that of odd times. Furthermore, $\Id - R_{P'} = \frac{1}{2}(\Id - R_P)$ so that having an inverse of $\Id - R_{P'}$ also gives an inverse of $P$.

So WLOG, we can assume that the result of Corollary 2.9 holds also for odd times. Assuming the $d$-dimensional isoperimetry (and writing $p'$ for the Hölder conjugate exponent of $p$),

$$||K_n||_p \leq \sum_{i \geq 0} P_i^p(e)^{(p-1)/p} \leq 2\sum_{i \geq 0} i^{-d/2p}.$$ 

The last sum converges if $p' > d/2$, i.e. if $p > \frac{d}{d-2}$. Now Young's inequality gives that $||f * K_n||_r \leq ||f||_q ||K_n||_p$ for $1 + \frac{1}{p} = \frac{1}{q} + \frac{1}{r}$. Assume $q$ is fixed, then this reads

$$\frac{1}{r} = \frac{1}{q} + \frac{1}{p} - 1 > \frac{1}{q} - \frac{2}{d} = \frac{d - 2q}{dq}.$$ 

This completes the first claim. If $G$ is not virtually nilpotent, then it has $d$-dimensional isoperimetric profile for all $d$ so that $r$ can be taken arbitrarily close to $q$.

2.4 Some other implications

More bounds on the heat kernel. There are also methods for the lower bound estimates. They will not be presented here in full details. These may be summed up by:

**Corollary 2.14**

1. $|B_n| \leq Kn^d$ implies $P_{2n}(e) \geq K n^{-d/2}.$

2. $|B_n| \leq \exp(-n^v)$ implies $P_{2n}(e) \geq K_1 \exp(-K_2 n^{v/(2-v)}).$

2.14.1 is clearly sharp (combine with 2.9.1). 2.14.2 is also sharp (for $v = 1$, in the non-amenable case).

The equivalence. There is a complete equivalence between the following:

$$I(n) \simeq n^{1-1/d} \iff |B_n| \simeq n^d \iff P^n(e) \simeq n^d \iff \nabla(t) \simeq t^{d/4}.$$ 

However this breaks up as soon as one looks at a larger space. Note that the "best possible" off-diagonal estimate is true in groups with $|B_n| \sim n^d$: there are $K_1, K_2 > 0$ such that

$$P^n(g) \leq K_1 P^n(e) \exp(-K_2 |g|^2/n).$$

See next section for more on such estimates.

Further links. Note that 2.8.1 only returned a $\frac{d}{2}$-isoperimetric profile. However, one can show that it implies the volume growth $|B_n| \geq n^d$ (i.e. if one does not use simply the characteristic functions like in exercise 7, but uses more cleverly designed functions).

**Exercise 9:** Warning: The following exercises contains longish designed computations.
1. Apply 2.8.1 to the functions $\sum_{i=0}^{n} 1_{B_i}$ to get that $|B_i| \geq K'n^d$. (Hint: use that $|\partial B_i| = |B_{i+1}|-|B_i|$ up to constants, then use piecewise linear extensions and some calculus.)

2. Using 2.8.2 to functions of the type $\sum_{i=0}^{n} a_i 1_{B_i}$ (with $a_i \in \mathbb{R}$) what lower bounds on $|B_n|$ are obtained? (Disclaimer: the author did not actually try this one!)

However, homogeneity of the space is crucial to go from volume to isoperimetry (or Nash or return estimates). Indeed, if one is not in a group (i.e. one considers the simple random walk on a graph), the best thing one can get from $|B_n| \geq n^d$ is $P^n(e) \leq n^{-(d-1)/d}$ (for any choice of the "root vertex" $e$). Also it is easy to construct an example of a graph with exponential growth which is recurrent\(^{12}\) (in particular $P^n(e)$ does not decrease faster than $n^{-1-\epsilon}$). This is constructed as follows. Take two copies of the regular tree (as in example 1.3), delete all vertices with $L(v) \geq 1$, and then identify the pair of vertices (one for each copy of the tree) which have $L(v) = 0$. It is easy to check that the resulting graph will have exponential volume growth (and a very bad isoperimetry: there are very large sets with only two edges in their boundary). With some more efforts one may show the probability of return to the initial position decays very slowly.

It is also relatively simple to show that $P^n(e) \leq Kn^{-d/2}$ implies the Nash inequality 2.8.1. There is a more general results saying that

$$\|f\|_2 \leq |S|2^{\sqrt{2}}(4\|f\|_2^2)\|\nabla f\|_2 \iff P^n(e) \leq \psi(n)$$

where $\psi : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is define by $t = \int_t^{1/\psi(t)} \mathfrak{g}(s)^2 \frac{ds}{s}$.

As mentioned before, there is also an equivalence for $P^n(e) \geq \psi(n)$ involving the Nash profile.

### 2.5 Carne’s proof of the Gaussian estimate

Due to Varopoulos (1989), but this proof comes from Carne (also 1989).

**Theorem 2.15**

Let $Z^n_k$ be the probability that a simple random walk on $\mathbb{Z}$ starting at 0 arrives at $k$ in $n$ steps. Let $T_k$ be the $k^{th}$-Chebyshev’s polynomial. Then

1. $R^n_P = \sum_{i=-n}^{n} Z^n_i T_i(R_P)$
2. $\|T_k(P)\|_{2 \to 2} \leq 1$
3. $\sum_{i \geq |k|} Z^n_i \leq e^{-k^2/2n}$
4. $P^n_e(g) \leq \exp(-|g|^2/2n)$.

The Chebyshev polynomials may be defined as satisfying $T_k(\cos x) = \cos kx$ or

$$T_k(z) = \frac{1}{2} \left( (z + \sqrt{1-z^2})^k + (z - \sqrt{1-z^2})^k \right).$$

\(^{12}\)A possible definition of recurrence is that $\sum_{n \geq 0} P^n_e(e) = +\infty$. 

Proof: Write \( \cos x = \frac{1}{2}(e^{ix} + e^{-ix}) =: \frac{1}{2}(w + w^{-1}) =: z \) where \( w = e^{ix} \). Use the binomial formula and \( Z_n^R = Z_{-n}^R \) to get
\[
z^n = \sum_{i=-n}^{n} Z_i^R w_i = \sum_{i=-n}^{n} Z_i^R \frac{w_i + w^{-i}}{2} = \sum_{i=-n}^{n} Z_i^R T_i(z).
\]
This is an identity of polynomials, so one is allowed to plug in \( z = R_P \).

For the second point, note that the operator \( \text{Id} - R_P^2 \) has a square root (being \( \geq 0 \)). Thinking shortly of \( L^2 G \) as a complex Hilbert space, consider \( R = R_P + i(\text{Id} - R_P^2)^{1/2} \). Since \( R_P \) is self adjoint, one can check that \( R^* = R \) and \( R'R = RR^* = \text{Id} \) so that \( R \) is an isometry. This implies that \( T_i(R_P) = \frac{1}{2}(R^i + R^i) \) is a contraction.

The third is a standard estimate: for \( t > 0 \) and \( W_n \) the random variable of the random walk on \( Z \),
\[
\sum_{i \geq |g|} Z_i^n = \mathbb{P}(|W_n| \geq k \mid W_0 = 0) = \mathbb{P}_0(e^{i\hat{W}_n} \geq e^{it}) \leq e^{-tk} \mathbb{E}[e^{i\hat{W}_n}] = e^{-tk} (\cosh t)^n \leq e^{-k+n^2/2}.
\]
The conclusion follows by putting \( t = k/n \).

The last point is obtained by computing:
\[
P^n_e(g) = \langle R_P^n \delta_e \mid \delta_g \rangle = \sum_{i=-n}^{n} Z_i^n \langle T_i(R_P) \delta_e \mid \delta_g \rangle
\]
Because \( T_i \) is of degree \( i \), note that the scalar product is trivial if \( i < |g| \). Combine this with the bound on the norm of \( T_i(R_P) \) to get
\[
P^n_e(g) = |\sum_{i \geq |g|} Z_i^n \langle T_i(R_P) \delta_e \mid \delta_g \rangle| \leq T.I. and 2 \sum_{i \geq |g|} Z_i^n \leq \exp(-|g|^2/2n) \quad \blacksquare
\]

A natural question from there is to ask, how close are we to an ideal bound (which is inspired from the case of \( Z^d \)): for some constants \( K_1, K_2 > 0 \),
\[
(SOD) \quad P^n_e(g) \leq K_1 P^n_e(e) \exp(-K_2 |g|^2/n)
\]
The answer is known to be positive in nilpotent groups (this includes Abelian groups, see Hebisch & Saloff-Coste [15]) and on the tree. Dungey has proved that there are constants \( K_1 \) and \( K_2 \) such that, for any \( \epsilon \in [0,1] \),
\[
P^n_e(g) \leq P^n_e(e) e^{(K_1 e^{-K_2 |g|} |g|^2/n)^{1-\epsilon}}.
\]
However, recent example have been by Brieussel & Zheng [5] given where \( (SOD) \) fails. The proof is very indirect (nobody computes the actual values of these return probabilities).

Exercise 10: Let \( \rho = ||R_P||_{2 \rightarrow 2} \). Show that
\[
P^n_e(g) \leq \rho^n \exp(-|g|^2/2n).
\]
(Hint: note that in Theorem 2.15.3 we only used that \( R_P \) is self-adjoint and has operator norm \( \leq 1 \).)

Question 2.16: Could \( (SOD) \) hold for polycyclic groups? What is the best term on can put in front of the estimate in 2.15.4?

The methods of Theorem 2.15 can also be used to prove Corollary 2.14 (modulo a logarithmic term in the polynomial case, see [27]).
3 Entropy and speed

3.1 Entropy

The entropy of a measure \(m\) on \(G\) (here \(G\) could really be any set) is defined by

\[
H(m) = - \sum_{g \in G} m(g) \ln m(g),
\]

"with the convention that \(0 \ln 0 = 0\)" (better said: restrict the sum to \(\text{Supp} m\)). In Physics, \(m\) measures the probability of the state of a given system to happen. The entropy of \(m\) counts the [logarithm] number of states that really have a chance of happening. It satisfies countless many nice properties, which go well beyond the scope of the present text. Let's start with a simple example.

Example 3.1. Take \(u_k\) to be the uniform distribution on \(k\) elements. Then

\[
H(u_k) = - \sum_{i=1}^{k} \frac{1}{k} (-\ln k) = \ln k
\]

Now construct \(u_{k,\ell}\) by splitting the mass on one of the \(k\) elements uniformly over \(\ell\) elements.

\[
H(u_{k,\ell}) = - \sum_{i=1}^{k-1} \frac{1}{k} (-\ln k) - \sum_{j=1}^{\ell} \frac{1}{k\ell} (-\ln k\ell) = \ln k + \frac{1}{k} \ln \ell.
\]

Notice that in order that \(H(u_{k,\ell}) \geq H(u_{2k})\) one needs that \(\ell \geq e^{k}\). This very large if \(k\) is large. This is to be expected, because the very many states we are creating, have a small chance of occurring.

The original introduction of entropy comes from the following setup. One has a random variable \(X\) with \(\mathbb{P}(X = a) = p_a\).

Given \(x_1, \ldots, x_n\) sampled i.i.d. from this distribution, what is the probability that these statistics fit the distribution? In other words, let

\[
N_a = |\{i \mid x_i = a\}| \quad \text{and} \quad \pi_a = N_a/n.
\]

What's the chance that (for \(n >> 1\)) \(\pi_a \simeq p_a\), i.e. \(N_a = np_a\).

Using Stirling's formula, \(k! \simeq \sqrt{2\pi k(k/e)^k}\), the probability of having found those \(N_i\)'s is

\[
\left(\begin{array}{c}
N_1 N_2 \ldots N_n \\
N_n \ldots N_2 N_1
\end{array}\right) \frac{\prod_i p_i^{N_i} \sim (2\pi n)^{\frac{1-j}{2}} \prod_i \pi_i^{N_i} \sim (2\pi n)^{1-j} / \prod_i \pi_i^{1/2}}{\prod_i \pi_i^{N_i} \sim \exp(N[H(x) - \sum_i \pi_i (-\ln p_i)])}.
\]

The term inside the exponential is the relative entropy of \(p\) w.r.t. \(x\). This quantity is quite useful to consider in:

Lemma 3.2

The entropy of a measure supported on a set \(F\) is maximal for the uniform distribution.

Proof: Let \(m'\) be some other measure with \(\text{Supp} m' = \text{Supp} m\), then

\[
H(m) - \sum_{g \in F} m(g) \ln \left(\frac{1}{m'(g)}\right) = \sum_{g \in F} m(g) \left( - \ln \frac{m(g)}{m'(g)} \right) \leq 0
\]
using $- \ln t \leq \frac{1}{t} - 1$. Now, if $m'$ is the uniform measure on the support of $m$, then the above reads

$$H(m) \leq - \sum_{g \in F} m(g) \ln m'(g) = - \sum_{g \in F} m(g) \ln m'(g) = H(m').$$

In fact, the same proof will show that if $m$ is supported on $F$ and $E \subseteq F$,

It seems at first sight that the entropy might not have much to do with the return probability or the volume growth. To make this link clear, consider for $p \neq 1$ and $m$ a measure supported on $F$,

$$\Phi_p(m) = \frac{\ln \|m\|_p^p}{1 - p} \text{ where } \|m\|_p^p = \sum_{g \in F} |m(g)|^p$$

For example, $\Phi_0(P^n_e) = \ln |B_n|$ (actually, $\ln |\text{Supp} P^n_e|$ which might differ by an additive factor of $\ln |S|$) and $\Phi_2(P^n_e) = - \ln P^n_{2n}(e)$ (two sub-additive functions). The main reason to introduce the above is that

$$\lim_{p \to 1} \Phi_p(m) = H(m).$$

Recall

Lemma 3.3 (Lyapounov’s interpolation inequality)

If $q, r \in \mathbb{R}$ and $p = \lambda r + (1 - \lambda)q$ (with $\lambda \in [0, 1]$), then

$$\|m\|_p^p \leq \|m\|_r^\lambda \|m\|_q^{1-\lambda},$$

Exercise 11:

1. Prove Lyapounov’s interpolation (Hint: write $|m|^p = |m|^\lambda r |m|^{1-\lambda} q$) and use Hölder with exponent $1/\lambda > 1$)

2. Show that $p \mapsto \Phi_p(m)$ is a decreasing function.

3. Show that the entropy is maximal for the uniform distribution using the fact that $\Phi_p$ is monotonic. Hint: Show that $\Phi_2(m')$ is maximal (where $m'$ is the uniform distribution) by looking at the derivative of $t \mapsto \Phi_2(m' + td)$ at $t = 0$ for all $d : F \to \mathbb{R}$ such that $\sum_{g \in F} d(g) = 0$. Also use that $\Phi_0(m') = \Phi_2(m')$.

A simple corollary that $\Phi_2(n) \leq \Phi_1(n) \leq \Phi_0(n)$ is

Corollary 3.4

$$- \ln P^n_{2n}(e) \leq H(P^n_e) \leq \ln |B_n|.$$ 

Proof: The proof (without using $\Phi_p$) goes as follows. The second inequality is obvious from the fact that the entropy is maximal for the uniform distribution (and Supp $P^n_e \subset B_n$).

For the first, use the concavity of $\ln$ and Jensen’s inequality to get

$$- \sum_{g \in G} m(g) \ln m(g) \geq - \ln \sum_{g \in G} m(g)^2$$

and conclude with Lemma 1.4.

These three quantities are readily checked to be sub-additive (as functions of $n$).
Lemma 3.5

\[ n \mapsto H(P^n) \] is a sub-additive function, i.e. \( H(P^{n+m}) \leq H(P^n) + H(P^m). \)

**Proof:** It is a simple but longish computation:

\[
H(P^{n+m}) = -\sum_{g \in G} P^n(g) \ln P^n(g) - \sum_{k \in G} P^m(k) \ln P^m(k)
\]

Inside the second sum (where \( g \) is fixed), put \( g = k\ell \). Since \( P^m(k\ell) = P^m(\ell) \),

\[
H(P^{n+m}) = -\sum_{k \in G} P^n(k) \sum_{\ell \in G} P^m(\ell) \ln P^m(k\ell)
\]

Since \( P^{n+m}(k\ell) \geq P^n(k)P^m(\ell) = P^n(k)P^m(\ell) \),

\[
H(P^{n+m}) \leq -\sum_{k \in G} P^n(k) \sum_{\ell \in G} P^m(\ell) (\ln P^n(k) + \ln P^m(\ell))
\]

\[
= -\sum_{k \in G} P^n(k) \ln P^n(k) \sum_{\ell \in G} P^m(\ell)
\]

\[
-\sum_{k \in G} P^n(k) \sum_{\ell \in G} P^m(\ell) \ln P^m(\ell)
\]

\[
= -\sum_{k \in G} P^n(k) \ln P^n(k) - \sum_{\ell \in G} P^m(\ell) \ln P^m(\ell)
\]

\[
= H(P^n) + H(P^m)
\]

which shows the claim. \( \square \)

In fact, \( n \mapsto H(P^n) \) is a concave function, i.e. \( H(P^{n+1}) - H(P^n) \) is decreasing.

**Exercise 12:** Show that the volume growth is sub-additive: \( \Phi_0(P^{n+m}) \leq \Phi_0(P^n) + \Phi_0(P^m) \)

Note that if \( n \mapsto \Phi(n) \) is sub-additive, one may show the limit \( \lim_{n \to \infty} \frac{\Phi(n)}{n} \) exists and equals \( \inf_{n \to \infty} \frac{\Phi[n]}{n} \) (Fekete’s Lemma).

If one changes \( P \), it is known that \( n \mapsto \Phi_0(P^n) \) (easy) and \( n \mapsto \Phi_2(P^n) \) (see Pittet & Saloff-Coste [21] or Woess [27, pp. 161-164]) only change up to affine equivalence\(^{13}\).

**Question 3.6 (hard):** Is \( \Phi_1 \) affinely invariant under changing \( P \)?

In fact it not known whether \( \Phi_1 \) is affinely invariant under quasi-isometries (between Cayley graphs). This question earns the tag “hard” because quite a few mathematicians have tried to tackle it in the last 10-15 years.

Given the above presentation, it would be tempting to go through the limit as \( p \to 1 \) of estimates on \( \Phi_p \).

**Remark 3.7.** Lemma 1.4 implies that \( \Phi_\infty(2n) = \Phi_2(n) \). Simple considerations shows that, if \( p, p' \in [1, \infty] \) then \( \Phi_p \) and \( \Phi_{p'} \) are affinely equivalent. However, the constants in these equivalence diverge as \( p \to 1 \).

It is not clear to the author whether all \( \Phi_p \) are also affinely equivalent for \( p \in [0, 1] \) or \( p \in [0, 1] \).

The following conjecture is roaming in some corridors, but the original perpetrator has still not been identified.

\(^{13}\) i.e. two monotone functions \( f \) and \( g \) are affinely equivalent if \( \exists K_1, K_2, K_3 > 0 \) such that \( g(t) \leq f(K_1 t) + K_2 \) and \( f(t) \leq g(K_2 t) + K_3 \)
Question 3.8: “Assume that for some constants \( K_1, K_2 \in \mathbb{R} \) one has \( |B_n| \geq K_1 \exp(K_2 n) \). Is it true that \( H(n) \geq Kn^{1/2} \)?”

A more general question would be:

Question 3.9: “If \( |B_n| \geq K_1 \exp(K_2 n) \), what is the best lower bound on \( H(n) \)? Can one use isoperimetry to get this lower bound?”

Remark 3.10. Indeed, it is possible to show that the \( v \)-intermediate isoperimetry is equivalent to: \( \forall f \) of finite support with \( \|f\|_1 = 1 \) and \( \|f\|_\infty \leq 1/3 \), one has (for some \( K > 0 \))

\[
|\nabla f|_1 \geq KH(f)^{-1/v}
\]

As promised in the motivations, let us show that the entropy behaves well under epimorphisms (surjective homomorphisms). This follows from some general features:

Lemma 3.11

Let \( F, A \) be sets, assume \( m \) is a measure on \( F \) and \( f : F \to A \) is a map. Let \( \mu = f^*m \) (i.e. for \( C \subseteq A \), \( \mu(C) = m(f^{-1}C) \)). Then

\[
H(m) \geq H(\mu)
\]

Proof: To make the proof slightly more useful, introduce a normalised measure on each preimage of \( f \): for \( a \in A \), let \( m_a(x) = m(x)/m(f^{-1}\{a\}) \). Then define the entropy of the map by the [weighted] average of those entropies:

\[
H(f) = \sum_{a \in A} m(f^{-1}\{a\}) H(m_a).
\]

Obviously, \( H(f) \geq 0 \) since \( H(m_a) \geq 0 \) for any \( a \in A \). Now

\[
H(f) = \sum_{a \in A} -m(f^{-1}\{a\}) \sum_{x \neq f^{-1}\{a\}} \frac{m(x)}{m(f^{-1}\{a\})} \ln \frac{m(x)}{m(f^{-1}\{a\})}
\]

\[
= \sum_{a \in A} \sum_{x \neq f^{-1}\{a\}} m(x) \ln \frac{m(x)}{m(f^{-1}\{a\})}
\]

\[
= -\sum_{a \in A} \sum_{x \neq f^{-1}\{a\}} m(x) \ln m(x) + \sum_{a \in A} \sum_{x \neq f^{-1}\{a\}} m(x) \ln m(f^{-1}\{a\})
\]

\[
= -\sum_{x \in F} m(x) \ln m(x) + \sum_{a \in A} m(f^{-1}\{a\}) \ln m(f^{-1}\{a\})
\]

\[
= -H(m) + \sum_{a \in A} m(f^{-1}\{a\}) \ln m(f^{-1}\{a\})
\]

\[
= H(m) - H(\mu). \tag*{\blacksquare}
\]

So \( H(m) = H(f) + H(\mu) \geq H(\mu) \).

As a corollary:

Corollary 3.12

Let \( \psi : G \to H \) be a surjective homomorphism and let \( P' = \psi^*P \). Then

\[
H(P'^n) \geq H(P^n).
\]
The proof relies on the fact that $\psi^n G_e = P^n_e$ (because $\psi$ is a homomorphism).

**Exercise 13:** Show that if $\psi: G \rightarrow H$ is a surjective homomorphism, then

1. $\Phi_2(P^n_e) \geq \Phi_2(P^n_H)$.
2. $\Phi_0(P^n_e) \geq \Phi_0(P^n_H)$.

As we will need it afterwards, note there is only a very small difference between subadditive and concave:

**Lemma 3.13**

Let $f: \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ be a sub-additive, non-decreasing function with $f(0) = 0$. If $g$ is the concave hull of $f$ then $f(x) \leq g(x) \leq 2f(x)$.

Parent thesis on concavity... Nevertheless, for some purposes concavity is much more desirable. Basically, sub-additivity ensures that a “root test” converges monotonously whereas concavity ensures monotonic convergence of the “ratio test”. For example, from sub-additivity, $|B_n|^{1/n}$ is an upper bound to its limit [and lower bounds are somehow easier to construct]. However, the monotonicity of the ratio would give a much better numerical upper bound.

There are many groups (and generating sets) where the limit is not known. Nevertheless, the monotonicity of the ratio would give a much better numerical upper bound.

There are examples where $n \mapsto \Phi_0(n)$ is not concave$^{14}$ and examples where $n \mapsto \Phi_\infty(P^n_e)$ is not concave$^{15}$. These examples are consequence of some form of “bipartiteness”. However, $n \mapsto \Phi_\infty(P^n_e) = \Phi_2(P^n_e)$ is always concave and $n \mapsto \Phi_1(P^n_e)$ is also always concave. This leads one to wonder if $n \mapsto \Phi_0(P^n_e)$ is concave, or more realistic:

**Question 3.14** (hard if true): Are $|B_{2n+2}|/|B_{2n}|$ and $|B_{2n+1}|/|B_{2n-1}|$ monotone sequences?

The reason the previous question has the tag “hard” is that it would imply a conjecture about the sequence of balls being Følner sets in groups of subexponential growth$^{16}$. A more reasonable question is

**Question 3.15:** Find a large class of groups where $|B_{2n+2}|/|B_{2n}|$ converges or a group where it does not.

A similar question (on self-avoiding walks on $\mathbb{Z}^d$) has a positive answer, so that one could hope for something in this direction. A self-avoiding walk on a graph is a path which only visits each vertex at most once. Fix a root vertex $e$ and let $A_n$ be the number of such paths starting at $e$. Then $A_{n+m} \leq A_n + A_m$. As above, the limit $A_n^{1/n}$ exists, but its value is not known even for $\mathbb{Z}^2$. A theorem of Kesten shows that $A_{2n+2}/A_{2n}$ converges.

### 3.2 Speed

Recall $|g| = d(e, g)$ is the distance between $g$ and $e$ in the Cayley graph. It can also be interpreted as the number of letters from $S$ required to write $g$ as a word.

---

$^{14}$These examples are $G = \ast_{i=1}^n G_i$ where each $G_i$ is either finite or $\mathbb{Z}$ and $S = \cup S_i$ where $S_i \subset G_i$ is either $G_i \setminus \{e\}$ if $G_i$ is finite or $\pm 1$ else.

$^{15}$Any bipartite graph if $P(e) = 0$.

$^{16}$It is known that a subsequence of $B_n$ is a Følner set in a group of subexponential growth, but it not known whether this holds for the full sequence.
**Definition 3.16.** Let \( r \in \mathbb{R} \). The \( r \)-th-moment of a distribution \( m \) is

\[
M_r(m) = \sum_{g \in G} m(g) \left| g \right|^r.
\]

The speed [or drift] function of the random walk is the function

\[
n \mapsto M_1(P^n_e) = \sum_{g \in G} P^n_e(g) \left| g \right|
\]

In other words, the speed is the expected distance to identity. This is a measure of the expected distance after \( n \)-steps to the starting point. A first thing to check is

**(Exercise 14)** \( M_1(P^n_{e+m}) \leq M_1(P^n_e) + M_1(P^n_m) \).

The speed is tricky to compute. Although it is relatively easy to check that \( M_1(P^n_e) \sim n^{1/2} \) for Abelian groups, it was not known until 2005 (an argument of Virág, see Lee & Peres [16]) that \( M_1(P^n_e) \geq Kn^{1/2} \) for any infinite group\(^{17}\). This was expected since Abelian groups are somehow the “smallest” infinite groups.

**(Question 3.17 (hard):** Is \( n \mapsto M_1(P^n) \) affinely invariant under quasi-isometries (between Cayley graphs)\)? or under change of \( P \)?

The tag “hard” is due to same reasons as Question 3.6.

**(Example 3.18.** Let us go back quickly to our example with \( \mathbb{Z} \). After \( n \) steps, the distribution is well approximated by a centred normal distribution with variance \( n \). For example, writing \( N \) for a random variable with law the standard centred Gaussian,

\[
\mathbb{P}(|W_n| \leq A) \simeq \mathbb{P}(|\sqrt{n}N| \leq A) = \int_{-A/\sqrt{n}}^{A/\sqrt{n}} e^{-x^2/2}dx \simeq 2A/\sqrt{n}.
\]

where the \( \simeq \) should be read as lower and upper bounds up to constants. In this particular case, one can easily compute the expectation directly:

\[
\mathbb{E}|W^n| \simeq 2 \int_0^{\infty} \sqrt{n}xe^{-x^2/2} = 2\sqrt{n}.
\]

The speed exponent also satisfies a monotonicity for surjective homomorphism.

**(Lemma 3.19**)

Assume \( \psi : G \to H \) is a surjective homomorphism. Let \( S = \text{Supp} \ P \) be generating for \( G \) (hence \( \psi(S) \) generates \( H \)). Let \( P' = \psi^*P \), i.e. \( P'(A) = P(\psi^{-1}A) \). Then \( M_1(P^n_{e_0}) \geq M_1(P^n_{e_H}) \).

In particular, since it is relatively easy to show \( M_1(P^n) \sim K_P n^{1/2} \) for any symmetric \( P \) on the group \( \mathbb{Z} \) (where \( K_P \in \mathbb{R} \)), one gets that \( M_1(P^n) \geq K_P n^{1/2} \) for any \( G \) with non-trivial homomorphism to \( \mathbb{Z} \). As mentioned above, there is a way of proving that \( M_1(P^n) \geq Kn^{1/2} \) for any group (due to Virág, see [16]).

**(Proof:** Let \( d_H \) be the distance of the Cayley graph with respect to \( S_H = \text{support of } P' \). Define the function \( d' : G \to \mathbb{N} \) by \( d'() = d_H(\psi(), e_H) \). Note that \( d'(\gamma) \leq d_G(\gamma, e) \): indeed \( d_H(h_1, h_2) = d_G(\psi^{-1}(h_1), \psi^{-1}(h_2)) \), so that \( d'(\gamma) = d_G(\gamma N, N) \) where \( N = \ker \psi \). Let \( W^n_g \) be

\(^{17}\)The constant \( K \) depends on \( P \), but for amenable groups it depends linearly on \( \sqrt{\min \{P(s) | s \neq e\}} \).
the random walker on $G$ and $W^H_n$ be the random walker on $H$ (which moves according to $P'$ as in the statement). Note that $\mathbb{P}(d_H(W^H_n, e_H) = i) = \mathbb{P}(d(W^G_n) = i)$. This implies

$$M_1(P^n) = \mathbb{E}(d_H(W^H_n, e_H)) = \mathbb{E}(d(W^G_n)) = M_1(P^n)$$

Note that, when considering simple random walks on graphs, the statements made so far about epimorphisms may be generalised to coverings of graphs. However, it is sometimes more convenient to consider something slightly more general. Indeed, the previous lemma generalises to weighted graph morphisms.

A graph morphism is a map between graphs, so that each edge is mapped to an edge. The “weighted” condition one needs to check is the following: if $(y, y')$ is an edge in the target graph, then for any $x \in \psi^{-1}(y)$ the probability to move to $\psi(y')$ [in one step] does not depend on $x$. This condition is necessary so that the random process on the target graph does not depend on the vertex [Markov property].

Let us do an example of this phenomenon.

Example 3.20. Let us go back to Example 1.3. We had the following “weighted graph morphism” $L$ (not a homomorphism!):

\[
\begin{array}{c}
\text{e} \\
\text{R} \\
L
\end{array}
\]

Note that $L(v) \leq d(v, e_G)$. Recall that $Z_t = L(W_t)$ is a random walk on $Z$ but with a preference for a direction:

$$\mathbb{P}(Z_{t+1} = k + 1 \mid Z_t = k) = \frac{v - 1}{v} \quad \text{and} \quad \mathbb{P}(Z_{t+1} = k - 1 \mid Z_t = k) = \frac{1}{v}.$$  

The law of $Z_t$ is a (scaled and translated) binomial law (with probability of success $\frac{v - 1}{v}$). The normal approximation will be $\mathcal{N}(n\frac{v - 2}{v}, n\frac{(v - 1)}{v^2})$. Using this\(^\text{18}\), it is very easy to show that, for any $\epsilon > 0$

$$\mathbb{P}((1 - \epsilon)n\frac{v - 2}{v} \leq Z_t \leq (1 + \epsilon)n\frac{v - 2}{v}) \rightarrow 1.$$

Since $Z_n \leq d(W_n, e_G)$, one concludes that $\forall \epsilon, \exists N_\epsilon$ so that $\forall n > N_\epsilon$

$$\mathbb{E}(d(W_n, e_G)) \geq n(1 - \epsilon_n)\frac{v - 2}{v}.$$

One can also formulate it as $M_1(P^n) \geq n\frac{v - 2}{v}(1 - \epsilon_n)$ (where $\epsilon_n \to 0$ can be very precisely estimated). \hfill \clubsuit

Lemma 3.21

\[2n\Phi_1(P^n) \geq M_1(P^n)^2.\]

\(^{18}\)If you are into probability, it’s probably more natural for you to use the Hoeffding inequality [or Chernoff, or Azuma-Hoeffding, or …] here.
\textbf{Proof:} Just write $\Phi_1(P^n_e) = H(P^n_e) = \sum_g P^n_e(g)(-\ln P^n_e(g))$ and use 2.15 that $-\ln P^n_e(g) \geq \ln |g|^2/2n$:

$$H(P^n_e) \geq \sum_g P^n_e(g)|g|^2/2n \geq \frac{1}{2n} \left( \sum_g P^n_e(g)|g| \right)^2$$

where we used Jensen's inequality on the convex function $x \mapsto x^2$.  

\begin{lemma}
For some $K > 0$, $\Phi_1(P^n_e) \leq 4\Phi_0 \circ M_{1}(P^n_e) + K$.
\end{lemma}

\textbf{Proof:} We will only do a “easy” proof which adds a logarithmic term to the inequality instead of a constant.

Denote the “spheres” by $S_n = B_n \setminus B_{n-1}$, then compare $P^n_e$ to a measure which is uniform on the $S_n$ (so has bigger entropy by a simple variant of Lemma 3.2). Note that in this “uniformisation”, the measure of $g \in S$ is $P^n_e(S_i)/|S_i|$.

$$H(P^n_e) \leq \sum_{g \in G} P^n_e(g)(-\ln P^n_e(S_{|g|})/|S_{|g|}|)$$

$$= \sum_{i=0}^n P^n_e(S_i)(-\ln P^n_e(S_i)/|S_i|)$$

$$\leq -\sum_{i=0}^n P^n_e(S_i) \ln P^n_e(S_i) + \sum_{i=0}^n P^n_e(S_i) \ln |S_i|.$$

$$\leq \ln(n+1) + \sum_{i=0}^n P^n_e(S_i) \ln |B_i|.$$ 

where the $\ln(n+1)$ comes from the fact that the first term was the entropy of some measure on $n+1$ elements, so it is bounded by $\ln(n+1)$. Now, let $g$ be the concave hull of $n \mapsto \Phi_0(P^n_e) =: f(n)$. Recall $f(n) \leq g(n) \leq 2f(n)$ by Lemma 3.13 and Exercise 12 ($f$ is sub-additive). Then,

$$H(P^n_e) \leq \ln(n+1) + \sum_{i=0}^n P^n_e(S_i) \ln |B_i|$$

$$\leq \ln(n+1) + \sum_{i=0}^n P^n_e(S_i) g(i)$$

$$\leq \ln(n+1) + g(\sum_{i=0}^n P^n_e(S_i) i)$$

$$\leq \ln(n+1) + 2f(M_{1}(P^n_e))$$

which, up to the $\ln(n+1)$, is what was claimed.  

\begin{corollary}
If $|B_n| \leq \exp(Kn^\nu)$ then $M_{1}(P^n_e) \leq K'n^{1/(2-\nu)}$. Also $\Phi_2(P^n_e) \leq H(P^n_e) \leq K_1n^{\nu/(2-\nu)} + K_2$.
\end{corollary}

\textbf{Proof:} Combine Lemmas 3.21 and 3.22 to get

$$M_{1}(P^n_e)^2 \leq 2n\Phi_1(P^n_e) \leq 4n\Phi_0(M_{1}(P^n_e)) \leq 2n(4KM_{1}(P^n_e)^\nu + K'').$$

where the last inequality comes from the hypothesis on the volume. Solving for $M_{1}(P^n_e)$ yields the bound on speed (with $K' = 8K + 2K''/M_{1}(P^n_e)$). Then use Lemma 3.22 again to get

$$\Phi_1(P^n_e) \leq 2KM_{1}(P^n_e)^\nu + K_2 \leq K_1n^{\nu/(2-\nu)} + K_2.$$
For \( v = 1 \), the above bound are all sharp (in the non-amenable case). Lemma 3.21 goes back to Varopoulos (1989) and was used to show that linear speed and linear entropy are equivalent conditions. The previous corollary as well as Lemma 3.22 are from \( \delta \) [9].

Note that the corollary only gives \( M_1(P_n^v) \leq K n^{1/2} \ln n \) for groups with \( |B_n| \leq K n^d \) (the correct speed is \( \simeq n^{1/2} \)). The bound on entropy is also unsharp (and one must be much more careful with the constants).

A result of Salo-Coste & Zheng [26] shows that the entropy can be bounded in terms of return probability: if \( \Phi_2(P_n^v) \leq K_1 n^\gamma + K_2 \) then \( H(P_n^v) \leq K_1 n^\gamma (1-\gamma) + K_2 \) [this produces a non-trivial upper bound on the entropy only if \( \gamma < 1/2 \)]. Their proof requires "good" embeddings of the Cayley graphs into Hilbert spaces.

When \( v = 1 \), Brieussel & Zheng [5] showed that there is no other constraint than the above between speed and entropy, except for Question 3.8.

**Gap.** The corollary also gives some restrictions to the answer of 3.9. Indeed, based on the naive belief that entropy is a form of volume growth, one could speculate that in Exercise 5, the \( v \)-intermediate isoperimetry should yield back \( H(P_n^v) \geq n^{\nu/(1+\nu)} \) (and not just \( \Phi_0(P_n^v) \geq n^{\nu/(1+\nu)} \)). This would be consistent with the expected positive answer to Question 3.8. However, it would also imply that there are no groups with \( |B_n| \simeq \exp n^v \) for \( v \in ]0, 1/2[ \). This is a weak form of the so called "gap conjecture" on growth of groups.

Unique geodesics. Say a group \( G \) together with a generating set \( S \) has unique geodesics if there is a unique shortest combinatorial path between any two points of the Cayley graph of \( G \).

**Question 3.24:** Assume \( G \) has unique geodesics for \( S \) and let \( P = \frac{1}{2}(\delta_s + \frac{1}{|S|} \delta_S) \). If \( G \) is not \( \mathbb{Z} \) or the infinite dihedral group, can one show that \( M_1(P_n^v) \geq n^{1/2} \ln n \)?

The actual conjecture is that \( G \) is a free product of groups \( G_i \) generated by \( S_i \) with each pair \( (G_i, S_i) \) yielding a Cayley graph which is a line, an odd cycle, or a complete graph. Note that it is not even clear whether the only finite Cayley graphs with unique geodesics are odd cycles and complete graphs.

Note that the only known results about groups with unique geodesics is that: 1- they have no even cycles (obvious); 2- they cannot surject on \( \mathbb{Z}^2 \) unless they have exponential growth (exercise!).

## 4 Harmonic functions

Let \( \mathcal{P}_c(G) \) be the space of finitely supported measures on \( G \). For \( P \in \mathcal{P}_c(G) \), one says that a function \( f : G \to \mathbb{R} \) is \( P \)-harmonic if it satisfies

\[
\forall x \in G, \quad \sum_{s \in G} P(s^{-1}) f(xs) = f \ast P(x) = f
\]

i.e. if it is fixed by \( R_P \).

There are three conditions of great interest about such functions:

1. what are the bounded harmonic functions?
2. what are the harmonic functions \( f \) with gradient in \( \ell^2 \)?
3. what are the harmonic functions \( f \) with gradient in \( \ell^\infty \) (i.e. Lipschitz)?
The first is related to a compactification of the group (the Poisson boundary) or alternatively to the dynamics of the random walks (tail events). The second is related to percolation and to the reduced $\ell^2$-cohomology of the group. The third plays a important role in the elementary proof of Gromov’s theorem on groups of polynomial growth. Other conditions on the gradient pop up in the cohomology of Hilbertian representations of groups, see Ø & Jolissaint [?].

4.1 Bounded harmonic functions

As mentioned before, one could think that some $P^n$ could be good choices for the (iii) of Theorem 1.6 sequence. Recall that a function $f$ is $P$-harmonic (for some $P \in \mathcal{P}_r(G)$) if $f \ast P = f$. $G$ is said to be $P$-Liouville if there are no $P$-harmonic functions in $\ell^\infty G$ except constant functions.

Before moving on to the proof, a small lemma will be required:

**Lemma 4.1**

In $\ell^1 \mathbb{N}$, weak and strong convergence coincide.

**Proof:** It is classical that strong convergence implies weak convergence. Assume the sequence $x_n$ converges weakly but not strongly. WLOG, one may assume it converges to 0 (if it converges to $x$ look at the sequence $x_n - x$ instead). WLOG, one may also assume that $\|x_n\|_\ell^1 \geq 1$ (because it does not converge strongly, this is $\geq \delta$ for some subsequence; just restrict to this subsequence and multiply by a scalar to get $\geq 1$).

Since all function with values $\pm$ on a finite set belong to $\ell^\infty \mathbb{N}$, weak convergences implies that the $x_n$ tend to 0 on finite sets. On the other hand, for each $x_n$ there is a finite set $A$ so that $\|x_n\|_\ell^1(A) < \epsilon$ (for some fixed and small $\epsilon \to 0$. Hence one may find sequences $n_i$ and $N_i$ so that

$$\forall i, \|x_{n_i}\|_\ell^1(N_{i-1}, N_i) < \epsilon$$

Pick $y \in \ell^\infty \mathbb{N}$, given by

$$y(n) = \text{sgn} x_{n_i}(n)$$

for $n \in [N_{i-1}, N_i]$, with $\text{sgn} r = \frac{r}{|r|}$ if $\neq 0$ and 0 if $r = 0$. Then, notice that, $\forall i$,

$$\langle y, x_{n_i} \rangle \geq \|x_{n_i}\|_\ell^1 N \geq 2\epsilon \geq 1 - 2\epsilon.$$

Hence, the sequence cannot converge weakly to 0. \qed

Recall that $\mathcal{P}_r(G)$ is the space of finitely supported probability measures. $\mathcal{P}_r(G)$ is the norm-closure of $\mathcal{P}_r(G)$ in $\ell^1 G$, i.e. probability measures in $\ell^1 G$.

**Proposition 4.2**

Let $G$ be finitely generated and let $P \in \mathcal{P}_r(G)$ be symmetric and irreducible. $G$ is $P$-Liouville if and only if the sequence $\xi_N = \frac{1}{N} \sum_{n=0}^{N-1} P^n$ is a Reiter sequence.

**Proof:** ($\Rightarrow$) Assume $f$ is $P$-harmonic and $\xi_n$ is Reiter. Since $\xi_n$ are convex combinations of $P^n$, $f \ast \xi_n = f$. So for any $x, s \in G$

$$|f(x s) - f(x)| = |f \ast \delta_s(x) - f(x)| = |f \ast \xi_n \ast \delta_s(x) - f \ast \xi_n(x)| = |f \ast (\xi_n \ast \delta_s - \xi_n)(x)| \leq \|f\|_{\ell^\infty} \|\xi_n \ast \delta_s - \xi_n\|_{\ell^1}$$
Since \( n \) is arbitrary, \( f(xs) = f(x) \). But \( x \) and \( s \) are also arbitrary so \( f \) is constant.

\((\Rightarrow)\) Assume \( \xi_N = \frac{1}{N} \sum_{i=0}^{N-1} P_e^i \) (or any of its subsequence) is not almost invariant as in (iii) of Theorem 1.6. Pick some \( s \in G \). Then\(^{19} \exists \epsilon > 0 \) so that, \( \forall N, ||\delta_x * \xi_N - \xi_N||_\nu > \epsilon. \) Since this is norm-separated from 0, there is an element in the dual of \( \ell^1 G \) (i.e. \( \ell^\infty G \)) which also does\(^{20} \). So, there exists \( g \in \ell^\infty G \) such that, \( \forall n, \)

\[ (*) \quad \int_G g(d\xi_N) - \int_G g(d\xi_N) > \epsilon. \]

Put \( f(x) = \lim_n |g(y)| \delta_x \ast \xi_n)(y) \) where the limit is given by some diagonal subsequence\(^{21} \) or a weak* accumulation point.

Then \( f \) is bounded (obvious) and harmonic:

\[ |(f - f \ast P)(x)| = |\lim_N |g(y)| \delta_x \ast \xi_N - \delta_x \ast \xi_N \ast P(y)| = |\lim_N |g(y)| \delta_x \ast \xi_N - \delta_x \ast \xi_N \ast P^N(y)| \leq ||g||_{\ell^\infty} \lim_N \|\delta_x - \delta_x \ast P^N\|_\nu \leq ||g||_{\ell^\infty} \lim_N \|\delta_x - \delta_x \ast P^N\|_\nu = 0. \]

Note that it is crucial that the limit tends to 0 whatever the subsequence chosen. Lastly, \( |f(s^{-1}) - f(e)| > \epsilon \) thanks to equation (**), so we have that \( f \) is not a constant.

The first proof that \( \frac{H(P_e^n)}{n} \to 0 \) (the entropy) gives a Reiter sequence is due to Avez \([2]\).

When the graph is transient, one can define a metric on the graph thanks to Green’s kernel \( K_0 = \sum_{i \geq 0} P_i^e \) as follows:

\[ d_K(x, y) = -\ln \frac{K_e(y)}{K_e(x)} \]

**EXERCISE 15:** Check this is a metric. \((\text{Hint: interpret it as the probability to hit } y \text{ if starting a random walk at } x.\)\)

Define the Green balls as the balls for this metric.

**Question 4.3:** “Does the sequence of Green balls forms a Følner sequence if and only if \( G \) is P-Liouville?”

The reason is that Green balls are level sets of \( K_0 \). In “nice” cases, one expects these to be level sets of the partial sums \( \sum_{i=0}^{N} P_i^e \). But then, these are also level sets of \( \frac{1}{N} \sum_{i=0}^{N} P_i^e \). By the proof of \((\text{iii}) \implies (\text{iv})\) in Theorem 1.6, one expects such sets to be invariant.

**Question 4.4:** “Try to compute the Green balls on the lamplighter on \( \mathbb{Z}^3 \) (\( G = \mathbb{Z}^3 \ast C_2 \)).”

---

\(^{19}\) We use the absence of invariance on the left instead of the right. But since \( P \) is symmetric, this is equivalent.

\(^{20}\) This is where Lemma 4.1 comes in: the sequence is also weakly separated from 0.

\(^{21}\) Since \( g \) is bounded by \( c = ||g||_{\ell^\infty} \), this is a sequence of functions in \([-c, c]^2\). Enumerate the vertices and take a first subsequence \( n_k^{[1]} \) which converges at the first vertex. Refine this subsequence at a second vertex to get a subsequence \( n_k^{[2]} \) where the function converges on both vertices, etc... One obtains a sequence of subsequences \( n_k^{[n]} \). The diagonal subsequence \( n_k = n_k^{[n]} \) will converge everywhere.
4.2 Tail events

\((G^N, \mathcal{P}^N)\) path space and \(\mathcal{A}\) the tail \(\sigma\)-algebra.

For “nice” probability spaces there is a correspondence between \(\sigma\)-subalgebras and quotient probability spaces. One direction is easy: a measure preserving map \((X, \mathcal{P}) \rightarrow (X, \mathcal{P}')\) gives rise to a \(\sigma\)-subalgebra (given by the preimage of measurable sets). The Rokhlin lemma provides the other direction, i.e. a \(\sigma\)-subalgebra gives rise to a quotient map.

The Poisson boundary can be seen as a the quotient corresponding to the tail \(\sigma\)-subalgebra.

4.3 Martin boundary

....

References

[6] Coulhon ....


