Semester Project

Ecole Polytechnique Fédérale de Lausanne

Mathematics Department

Amenability of discrete groups through random walks: two major results

Author
Vincent Dumoncel

Supervisor
Prof. Nicolas Monod

May 19, 2023

Introduction

Since its introduction by Von Neumann in 1929, the notion of amenability became a central theme of research in geometric group theory. It admits a wide variety of equivalent characterizations, and links group theory to various areas of mathematics, such as geometry, functional analysis, or probability theory.

In 1959, Harry Kesten established [8] a major result in the field, proving that amenability of a group relates in a deep manner to the behaviour of random walks on the group. Namely, he proved that a group is amenable if and only if the return probabilities at the identity of the group decrease slower than an exponential.

In the following decade, Harry Furstenberg realized that amenability had also something to do with the behaviour of random walks "at infinity". This is the starting point of what is called *boundary theory*, a field in which a lot of progress have been made in the last fifty years. In the early 80's, Kaimanovich and Vershik proved another reformulation of what it means for a group to be amenable. More precisely, they showed that a (countable) group is amenable if and only if it carries a probability measure such that the associated *Poisson-Furstenberg boundary* is trivial.

The goal of this project is to provide and understand the proofs of these two results, by developing the adapted framework. In Section 1, we recall basic facts about bounded linear operators on Hilbert spaces. They will be widely used in Section 2, to obtain an intermediate characterization of amenability, and to study simple random walks on finitely generated groups. We then establish Kesten's result, and we compute explicitely the spectral radius for simple random walks on free groups. In Section 3, we introduce the basics of boundary theory. We define harmonic functions on groups and the Liouville property. We then introduce the Poisson boundary, whose existence and properties are however admitted. The bridge between this space and harmonic functions on the group is the Poisson transform, introduced in subsection 3.3.

Two appendices are added. The first one aims at complete Section 1, and provides a proof of Riesz representation theorem. The second establishes a theorem of convergence for bounded martingales, which will be a crucial tool at our disposal to construct an inverse to the Poisson transform.

Acknowledgments

I am grateful to Nicolas Monod for having proposed and investigated the topic with me, and for his numerous advices all along this project.

Contents

1. Hilbert spaces

In this first part, we introduce complex Hilbert spaces, and bounded linear operators on Hilbert spaces. We gather several useful results that will be used in our proof of Kesten's theorem.

1.1 Complex Hilbert spaces

Definition 1.1. Let $\mathcal{H}$ be a complex vector space.
A hermitian inner product on $\mathcal{H}$ is a map

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} &\longrightarrow \mathbb{C} \\ (x, y) &\longmapsto \langle x, y \rangle \end{aligned}$$

such that

(i) $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$, for all $x, y, z \in \mathcal{H}$, $\lambda, \mu \in \mathbb{C}$.

(ii) $\langle x, y \rangle = \overline{\langle y, x \rangle}$, for all $x, y \in \mathcal{H}$.

(iii) $\langle x, x \rangle \geq 0$ for all $x \in \mathcal{H}$, and $\langle x, x \rangle = 0$ implies $x = 0$.

When $\mathcal{H}$ is equipped with a hermitian inner product, the pair $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is called a *pre-Hilbert space*.

A priori, for $x \in \mathcal{H}$, $\langle x, x \rangle$ is a complex number, and its sign is undefined. However $\overline{\langle x, x \rangle} = \langle x, x \rangle$ by (ii), so $\langle x, x \rangle$ is in fact a real number. Also, the above properties together imply

$$\langle x, \lambda y + \mu z \rangle = \overline{\langle \lambda y + \mu z, x \rangle} = \overline{\lambda \langle y, x \rangle + \mu \langle z, x \rangle} = \overline{\lambda} \langle x, y \rangle + \overline{\mu} \langle x, z \rangle$$

for all $x, y, z \in \mathcal{H}$, $\lambda, \mu \in \mathbb{C}$. Lastly, for the special case $\lambda = \mu = 0$ in (i), we get $\langle 0, x \rangle = \langle x, 0 \rangle = 0$ for all $x \in \mathcal{H}$. In particular, $\langle x, x \rangle = 0$ if and only if $x = 0$.

The most important examples are the following.

Example 1.2. (i) The space of complex numbers $\mathcal{H} = \mathbb{C}$, equipped with the inner product $\langle x, y \rangle := x\overline{y}$, is a pre-Hilbert space. More generally, the space $\mathbb{C}^n$ with the inner product defined as

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i \overline{y_i}$$

for all $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{C}^n$, is a pre-Hilbert space.

(ii) Fix $(X, \mathcal{A}, \mu)$ a measure space, and let $\mathcal{H} = L^2(X, \mathcal{A}, \mu)$. For $f, g \in \mathcal{H}$, the formula

$$\langle f, g \rangle := \int_X f(x)\overline{g(x)}\, \mathrm{d}\mu(x)$$

defines a hermitian inner product on $\mathcal{H}$. If X is countable, we denote this space $\ell^2(X)$ rather than $L^2(X, \mathcal{A}, \mu)$.

Any space $\mathcal{H}$ with an inner product $\langle \cdot, \cdot \rangle$ can be turned into a normed space, by setting $\|x\| := \sqrt{\langle x, x \rangle}$. Indeed, the latter is well defined since $\langle x, x \rangle \geq 0$ for all $x \in \mathcal{H}$, and $\|x\| = 0$ if and only if $x = 0$. Moreover, for all $x \in \mathcal{H}$ and $\lambda \in \mathcal{H}$, we have

$$\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x \rangle} = \sqrt{\lambda \overline{\lambda} \langle x, x \rangle} = |\lambda| \|x\|.$$

Then, we are left to show the triangle inequality. This relies on the Cauchy-Schwarz inequality, of which it is difficult to underestimate the importance.

Lemma 1.3. Let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be a pre-Hilbert space.
Then, for all $x, y \in \mathcal{H}$, one has

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

where $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$.

Proof. The result is clear if $x = 0$ or $y = 0$. Then we may assume that $x, y \neq 0$ and, up to scaling, we can take $\|x\| = \|y\| = 1$. We start by observing that

$$\langle x - \langle x, y \rangle y, y \rangle = \langle x, y \rangle - \langle x, y \rangle \|y\|^2 = 0$$

and it follows that

$$\begin{aligned} 0 &\leq \|x - \langle x, y \rangle y\|^2 \\ &= \langle x, x - \langle x, y \rangle y \rangle \\ &= \langle x, x \rangle - \overline{\langle x, y \rangle} \langle x, y \rangle \\ &= 1 - |\langle x, y \rangle|^2. \end{aligned}$$

Hence $|\langle x, y \rangle| \leq 1 = \|x\| \|y\|$, and this proves the lemma. □

As announced, this gives the triangle inequality for the map $\| \cdot \|$ defined above.

Corollary 1.4. For any $x, y \in \mathcal{H}$, we have $\|x + y\| \leq \|x\| + \|y\|$.
In particular, the pair $(\mathcal{H}, \| \cdot \|)$ is a $\mathbb{C}-$normed vector space.

Proof. Let $x, y \in \mathcal{H}$. Expanding the square of the norm of $x + y$ and using Cauchy-Schwarz, we get

$$\begin{aligned} \|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \|x\|^2 + 2\mathrm{Re}\langle x, y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2 \end{aligned}$$

and so $\|x + y\| \leq \|x\| + \|y\|$. This yields the desired claim. □

The norm induced by an inner product has many useful properties. The following identities are known, respectively, as the *parallelogram law* and the *Pythagore's theorem*.

Proposition 1.5. For any $x, y \in \mathcal{H}$, we have $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$. Moreover, if $\langle x, y\rangle = 0$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Proof. On one hand, we compute that

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2 = \|x\|^2 + 2\mathrm{Re}\langle x, y\rangle + \|y\|^2$$

while on the other hand,

$$\|x-y\|^2 = \langle x-y, x-y\rangle = \|x\|^2 - \langle x, y\rangle - \langle y, x\rangle + \|y\|^2 = \|x\|^2 - 2\mathrm{Re}\langle x, y\rangle + \|y\|^2.$$

Adding these two lines, the first claim follows. Pythagore's theorem is a consequence of

$$\|x+y\|^2 = \|x\|^2 + 2\mathrm{Re}\langle x, y\rangle + \|y\|^2$$

since the middle term of the right hand side vanishes if $\langle x, y\rangle = 0$. □

We can now define Hilbert spaces. For this, recall that a metric space X is called *complete* if any Cauchy sequence in X converges. Recall also that a normed space $(V, \|\cdot\|)$ is automatically a metric space for the metric d defined by $\mathrm{d}(v, w) := \|v - w\|$, $v, w \in V$. We say that $(V, \|\cdot\|)$ is a *Banach space* if the metric space (V, d) is complete.

Definition 1.6. A complex Hilbert space is a pre-Hilbert space $\mathcal{H}$ which is a Banach space for the norm $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$.

Example 1.7. (i) $\mathbb{C}$ is complete, so it is a Hilbert space. More generally, $\mathbb{C}^n$ is a Hilbert space for all $n \geq 1$.

(ii) If $(X, \mathcal{A}, \mu)$ is a measure space, $L^2(X, \mathcal{A}, \mu)$ is a complex Hilbert space.

We close this subsection with the statement of the Riesz representation theorem, which will be the key to construct adjoint operators in the next part.

Theorem 1.8. Let $\mathcal{H}$ be a complex Hilbert space, and $f \in \mathcal{H}^*$.
Then there exists a unique $y \in \mathcal{H}$ such that

$$f(x) = \langle x, y\rangle$$

for all $x \in \mathcal{H}$. Moreover, $\|f\| = \|y\|$.

Proof. See Appendix A. □

1.2 Self-adjoint operators

Throughout this section, unless stated otherwise, $\mathcal{H}$ is a complex Hilbert space, and $A\colon \mathcal{H} \longrightarrow \mathcal{H}$ is a bounded linear operator on $\mathcal{H}$. We denote $\mathcal{B}(\mathcal{H})$ the Banach space of bounded linear operators on $\mathcal{H}$.

Fix $y \in \mathcal{H}$. Consider the linear functional φ defined as $\varphi(x) := \langle Ax, y\rangle$, for any $x \in \mathcal{H}$. Since A and the first variable of the inner product are linear, φ is linear, and the Cauchy-Schwarz inequality tells us it is bounded, as

$$|\varphi(x)| = |\langle Ax, y\rangle| \leq \|Ax\|\|y\| \leq \|A\|\|x\|\|y\|$$

for any $x \in \mathcal{H}$. Thus $\|\varphi\| \leq \|A\|\|y\|$. Therefore, Riesz representation theorem gives the existence of a unique element A^*y of $\mathcal{H}$ so that $\varphi(x) = \langle x, A^*y\rangle$ for all $x \in \mathcal{H}$, *i.e.*

$$\langle Ax, y\rangle = \langle x, A^*y\rangle$$

for all $x \in \mathcal{H}$. Moreover, $\|\varphi\| = \|A^*y\|$. This correspondence defines a map

$$\begin{aligned} A^*\colon \mathcal{H} &\longrightarrow \mathcal{H} \\ y &\longmapsto A^*y \end{aligned}$$

and one easily checks that A^* is in fact linear. This motivates the next definition.

Definition 1.9. The operator $A^*\colon \mathcal{H} \to \mathcal{H}$ defined above, such that

$$\langle Ax, y\rangle = \langle x, A^*y\rangle$$

for all $x, y \in \mathcal{H}$, is called the adjoint operator of A.

As a consequence of Riesz representation theorem, the adjoint A^* of A is the unique bounded linear operator satisfying the equality of Definition 1.9.

Here are the first general properties for computations with adjoint operators.

Proposition 1.10. (i) $\mathrm{Id}_{\mathcal{H}}^* = \mathrm{Id}_{\mathcal{H}}$, and $(A^*)^* = A$ for all $A \in \mathcal{B}(\mathcal{H})$.

(ii) $(A + \lambda B)^* = A^* + \overline{\lambda} B^*$ for all $A, B \in \mathcal{B}(\mathcal{H})$ and $\lambda \in \mathbb{C}$.

(iii) $(B \circ A)^* = A^* \circ B^*$ for all $A, B \in \mathcal{B}(\mathcal{H})$.

(iv) $\|A^*\| = \|A\|$, and $\|A^*A\| = \|A\|^2$ for all $A \in \mathcal{B}(\mathcal{H})$.

The proofs rely on the following strategy: since the adjoint of an operator is the *unique* operator satisfying a certain property, to show a given operator coincides with an adjoint, it suffices to show that the given operator satisfies the same property as the adjoint.

Proof. (i) For any $x, y \in \mathcal{H}$, we have $\langle x, \mathrm{Id}_{\mathcal{H}}(y)\rangle = \langle x, y\rangle = \langle \mathrm{Id}_{\mathcal{H}}(x), y\rangle$, so necessarily $\mathrm{Id}_{\mathcal{H}}^* = \mathrm{Id}_{\mathcal{H}}$. In the same way, we compute that

$$\langle x, Ay\rangle = \overline{\langle Ay, x\rangle} = \overline{\langle y, A^*x\rangle} = \langle A^*x, y\rangle$$

which implies $A = (A^*)^*$.

(ii) Fix $x, y \in \mathcal{H}$, and observe that

$$\langle x, (A^* + \overline{\lambda} B^*)y\rangle = \langle x, A^*y\rangle + \lambda\langle x, B^*y\rangle = \langle Ax, y\rangle + \lambda\langle Bx, y\rangle = \langle (A + \lambda B)x, y\rangle$$

by using properties of the inner product. Therefore, $A^* + \overline{\lambda} B^* = (A + \lambda B)^*$.

(iii) Here again, we have

$$\langle x, A^*(B^*y)\rangle = \langle Ax, B^*y\rangle = \langle B(Ax), y\rangle$$

for all $x, y \in \mathcal{H}$, implying $(B \circ A)^* = A^* \circ B^*$.

(iv) The paragraph preceding Definition 1.9 shows that $\|A^*y\| \leq \|A\|\|y\|$ for all $y \in \mathcal{H}$, giving the upper bound $\|A^*\| \leq \|A\|$. On the other hand, the same inequality with A^* instead of A provides

$$\|(A^*)^*\| \leq \|A^*\|$$

so by (i) we get in fact $\|A\| \leq \|A^*\|$. Henceforth, $\|A^*\| = \|A\|$.
For the last claim, let $x \in \mathcal{H}$ with $\|x\| = 1$. The definition of the operator norm provides

$$\|A^*Ax\| \leq \|A^*\|\|Ax\| \leq \|A^*\|\|A\|\|x\| = \|A\|^2$$

using $\|A^*\| = \|A\|$ in the last step. On the other hand, an application of Cauchy-Schwarz inequality shows that

$$\|Ax\|^2 = \langle Ax, Ax\rangle = \langle x, A^*Ax\rangle \leq |\langle x, A^*Ax\rangle| \leq \|A^*Ax\| \leq \|A^*A\|$$

providing the other bound $\|A\|^2 \leq \|A^*A\|$. This finishes the proof. □

For our purposes, we will be interested in a special class of operators.

Definition 1.11. If $A \in \mathcal{B}(\mathcal{H})$ satisfies $A^* = A$, then A is called self-adjoint.

Self-adjoint operators play a crucial role in finite dimensional linear algebra, mainly through the famous spectral theorem. For us, the important property they carry is an other way of computing their norms.

Theorem 1.12. Let $A \in \mathcal{B}(\mathcal{H})$ be a self-adjoint operator on $\mathcal{H}$. Then, one has

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x\rangle|.$$

Proof. Let us denote $C := \sup_{\|x\|=1} |\langle Ax, x\rangle|$. Note that C also equals $\sup_{x \neq 0} \dfrac{|\langle Ax, x\rangle|}{\|x\|^2}$. If $\|x\| = 1$, the Cauchy-Schwarz inequality and the definition of the norm of A gives

$$|\langle Ax, x\rangle| \leq \|Ax\|\|x\| \leq \|A\|\|x\|\|x\| = \|A\|$$

leading $C \leq \|A\|$. On the other hand, a direct computation proves that

$$\langle A(x+z), x+z\rangle - \langle A(x-z), x-z\rangle = 2(\langle Ax, z\rangle + \langle Az, x\rangle)$$

for all $x, z \in \mathcal{H}$, and since A is self-adjoint, $\langle Az, x\rangle = \overline{\langle x, Az\rangle} = \overline{\langle Ax, z\rangle}$. Thus

$$\langle A(x+z), x+z\rangle - \langle A(x-z), x-z\rangle = 2(\langle Ax, z\rangle + \overline{\langle Ax, z\rangle}) = 4\mathrm{Re}\langle Ax, z\rangle.$$

This way, we can estimate

$$|\mathrm{Re}\langle Ax, z\rangle| \leq \frac{C}{4}(\|x+z\|^2 + \|x-z\|^2) \leq \frac{C}{2}(\|x\|^2 + \|z\|^2)$$

by using the paralellogram law. Now let $x \in \mathcal{H}$ with $\|x\| = 1$, and suppose $Ax \neq 0$. Set $z := \frac{Ax}{\|Ax\|}$. Then $\mathrm{Re}\langle Ax, z\rangle$ reduces to $\|Ax\|$, and it follows from the last estimate that

$$\|Ax\| = \mathrm{Re}\langle Ax, z\rangle \leq \frac{C}{2}\left(\|x\|^2 + \left\|\frac{Ax}{\|Ax\|}\right\|^2\right) = C$$

which provides the upper bound $\|A\| \leq C$. This concludes the proof. □

It turns out the norm of a bounded self-adjoint operator can be recovered as the limit of a sequence involving $n-$th powers of the operator. To prove this, we need a useful lemma from analysis, usually known as *Fekete's lemma*. To state it, recall that a sequence of real numbers $(a_n)_{n\geq 0}$ is *subadditive* if $a_{n+m} \leq a_n + a_m$ for all $n, m \geq 0$. Similarly, $(a_n)_{n\geq 0}$ is *submultiplicative* if $a_{n+m} \leq a_n a_m$ for all $n, m \geq 0$.

Lemma 1.13. (i) Let $(a_n)_{n\geq 1}$ be subadditive. Then $\lim\limits_{n\to\infty} \dfrac{a_n}{n}$ exists and equals $\inf\limits_{n\geq 1} \dfrac{a_n}{n}$.

(ii) Let $(a_n)_{n\geq 1}$ be a submultiplicative sequence of positive real numbers. Then $\lim\limits_{n\to\infty} a_n^{\frac{1}{n}}$ exists and equals $\inf\limits_{n\geq 1} a_n^{\frac{1}{n}}$.

Proof. (i) Let $\varepsilon > 0$. Denote $\ell := \inf\limits_{n\geq 1} \dfrac{a_n}{n}$. There exists $N \in \mathbb{N}$ such that $\frac{a_N}{N} \leq \ell + \varepsilon$. By euclidean division, any $n \geq 1$ can be written as $n = kN + q$, with $0 \leq q < N$. Subadditivity of $(a_n)_{n\geq 1}$ now implies $\ell n \leq a_n = a_{kN+q} \leq k a_N + a_q$ and dividing by n it follows that

$$\ell \leq \liminf_{n\geq 1} \frac{a_n}{n} \leq \limsup_{n\geq 1} \frac{a_n}{n} \leq \frac{a_N}{N} \leq \ell + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, this yields the announced claim.

(ii) This is a consequence of the previous point, since if $(a_n)_{n\geq 1}$ is submultiplicative, then $(\log a_n)_{n\geq 1}$ is subadditive. □

This result is very useful to establish easily the existence of some limits. Here, we employ it as follows.

Proposition 1.14. Let $A \in \mathcal{B}(\mathcal{H})$.
Then, the number

$$r(A) := \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}}$$

is well-defined, and equals $\inf\limits_{n\geq 1} \|A^n\|^{\frac{1}{n}}$. Moreover, $r(A) \leq \|A\|$.

Proof. Since $\|\cdot\|$ is submultiplicative, the sequence $(\|A^n\|)_{n\geq 1}$ is submultiplicative, and Lemma 1.13(ii) gives the existence and the value of $r(A)$. For the second claim, we just note that $\|A^n\| \leq \|A\|^n$ for all $n \geq 1$, so that $\|A^n\|^{\frac{1}{n}} \leq \|A\|$ for all $n \geq 1$, and thus $r(A) \leq \|A\|$. □

For self-adjoint operators, the last inequality is in fact an equality.

Proposition 1.15. Let $A \in \mathcal{B}(\mathcal{H})$ be a self-adjoint operator on $\mathcal{H}$.
Then, one has $r(A) = \|A\|$.

Proof. Using Proposition 1.10(iv) and applying the equality $\|A^*A\| = \|A\|^2$ with $A^* = A$ leads to $\|A^2\| = \|A\|^2$. Applying it with the operator A^*A instead of A leads $\|A^4\| = \|A\|^4$, and by induction we get

$$\|A^{2^n}\| = \|A\|^{2^n}$$

for all $n \geq 0$, so $\|A^{2^n}\|^{\frac{1}{2^n}} = \|A\|$ for all $n \geq 0$. Hence $(\|A^n\|^{\frac{1}{n}})_{n\geq 0}$ has a constant, thus convergent, subsequence. By Proposition 1.14, $(\|A^n\|^{\frac{1}{n}})_{n\geq 0}$ converges, so its limit is the limit of any of its subsequences. This forces

$$r(A) = \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}} = \lim_{n\to\infty} \|A^{2^n}\|^{\frac{1}{2^n}} = \|A\|$$

and the proof is over. □

1.3 Uniform convexity in Hilbert spaces

In a Hilbert space, the norm is induced by an inner product, making computations easier than in a standard normed space. In this part, we use its properties at our advantage to understand the behaviour of the mean of unit vectors in terms of distance that separates those vectors.

Lemma 1.16. Let $u, v \in \mathcal{H}$ be such that $\|u\| = \|v\| = 1$.
Then, it holds that $\|\frac{u+v}{2}\|^2 = 1 - \frac{1}{4}\|u - v\|^2$.

Proof. This follows from the parallelogram law, since

$$\left\|\frac{u+v}{2}\right\|^2 + \frac{1}{4}\|u-v\|^2 = \frac{1}{4}(\|u+v\|^2 + \|u-v\|^2) = \frac{1}{4}(2\|u\|^2 + 2\|v\|^2) = 1$$

using that $\|u\| = \|v\| = 1$. □

This identity has the following consequence.

Lemma 1.17. Let $\delta > 0$.
There exists $\varepsilon > 0$ such that, for any pair of unit vectors $u, v \in \mathcal{H}$ satisfying $\|u - v\| \geq \delta$, we have $\|\frac{u+v}{2}\| \leq 1 - \varepsilon$.

Proof. Let $\varepsilon := 1 - \frac{1}{2}\sqrt{4 - \delta^2}$. This is a well-defined quantity since $\|u - v\| \leq \|u\| + \|v\| = 2$, so $\delta \leq 2$. Moreover $\delta > 0$ so $\varepsilon > 0$ as well. Now, using the previous lemma, one has

$$\left\|\frac{u+v}{2}\right\|^2 = 1 - \frac{1}{4}\|u-v\|^2 \leq 1 - \frac{1}{4}\delta^2 = (1-\varepsilon)^2$$

proving that $\|\frac{u+v}{2}\| \leq 1 - \varepsilon$. □

Remark 1.18. The heart of the statement lies really in the interaction between δ and ε, and the explicit formula $\varepsilon = 1 - \frac{1}{2}\sqrt{4-\delta^2}$ quantifies this interaction. In words, the more u and v are far from each other, the more their mean has norm far from 1. On the other hand, Lemma 1.16 tells also that if u and v are very closed from each other, then their mean has norm closed to 1.

This idea, and the previous lemma, can be generalized to an arbitrary large family of vectors, provided that at least two of them are far from each other.

Proposition 1.19. Let $n \geq 2$ and $\delta > 0$.
There exists $\varepsilon > 0$ such that, for any family of unit vectors $u_1, \dots, u_n \in \mathcal{H}$ with $\max\limits_{1 \leq i < j \leq n} \|u_i - u_j\| \geq \delta$, we have

$$\left\| \frac{u_1 + \dots + u_n}{n} \right\| \leq 1 - \varepsilon.$$

Proof. Up to relabeling, we can assume that $\|u_1 - u_2\| \geq \delta$. Then, by Lemma 1.17, there is $\varepsilon' > 0$ such that $\|\frac{u_1+u_2}{2}\| \leq 1 - \varepsilon'$. Since the norm of the mean of the $n-2$ remaining vectors is bounded by 1, it follows that

$$\begin{aligned}
\left\| \frac{u_1 + \dots + u_n}{n} \right\| &= \left\| \frac{2}{n}\frac{u_1 + u_2}{2} + \frac{n-2}{n}\frac{u_3 + \dots + u_n}{n-2} \right\| \\
&\leq \frac{2}{n}(1 - \varepsilon') + \frac{n-2}{n} \\
&= 1 - \frac{2\varepsilon'}{n}
\end{aligned}$$

and we get the desired result by setting $\varepsilon := \frac{2\varepsilon'}{n} > 0$. □

2. Amenability and random walks on groups

We now define the main notion of this work, namely that of *amenability* for groups, and we introduce the model of random walks on finitely generated groups.

A group G is called *finitely generated* if there exists a finite subset $S \subset G$ such that $G = \langle S \rangle$, *i.e.* every $g \in G$ can be written as a composition of finitely many elements of S or their inverses. Those are called *generators* for the group G.

2.1 Reiter properties

Initially, Von Neumann defined the concept of amenability for groups in 1929, after his study of the famous Banach-Tarski paradox. The latter claims it is possible, in the euclidean space $\mathbb{R}^3$, to cut the unit ball into finitely many pieces and move these pieces around to get two disjoint copies of the initial ball. Von Neumann realized this paradox comes from a feature of the underlying group $\mathrm{Isom}(\mathbb{R}^3)$ we use to move the pieces. He then proposed a first definition of amenability, in terms of the existence of *invariant means* on the group.

In 1955, Følner showed this definition of amenability could be restated as the existence of *almost invariant sets* in the group [3]. This characterization is in some sense more analytical, and led a third criterion of amenability, established by Hans Reiter [10]. This is the definition we will use below.

Definition 2.1. Let $(V, \|\cdot\|)$ be a normed space. Let $S \subset G$ be finite and $\varepsilon > 0$. A vector $v \in V$ is called $(S, \varepsilon)-$invariant if $\|sv - v\| < \varepsilon\|v\|$ for all $s \in S$.

Once a group G is fixed, we have directly in our hands a collection of normed spaces on which G acts naturally, namely all the $(\ell^p(G), \|\cdot\|_p)$ spaces, for $p \geq 1$. For $f \in \ell^p(G)$, its $p-$norm is

$$\|f\|_p = \left(\sum_{h \in G} |f(h)|^p \right)^{\frac{1}{p}}$$

and if $g \in G$ is fixed, the function gf is defined as $(gf)(h) := f(g^{-1}h)$ for all $h \geq 1$. Note that this action is isometric, *i.e.* $\|gf\|_p = \|f\|_p$ for all $f \in \ell^p(G)$ and $g \in G$, since the left multiplication by g^{-1} is a bijection on G.

Definition 2.2. Let $1 \leq p < \infty$.
We say that G has the Reiter property (R_p) if the action $G \curvearrowright \ell^p(G)$ has $(S, \varepsilon)-$invariant vectors, for all $S \subset G$ finite and $\varepsilon > 0$.

Example 2.3. (i) If G is a finite group, then the function $v(g) := \frac{1}{|G|} \in \ell^1(G)$ is (S, ε)–invariant for all $S \subset G$ and $\varepsilon > 0$. Hence G has (R_1). More generally, the constant function $v(g) = \frac{1}{|G|^{1/p}}$ is always invariant, so any finite group has (R_p), for all $p \geq 1$.

(ii) The group $\mathbb{Z}$ has (R_1). To see this, fix $S \subset \mathbb{Z}$ and $\varepsilon > 0$. Without restriction, we may assume that S is symmetric. Let $m := \max_{s \in S} |s|$, $n := \lfloor \frac{2m}{\varepsilon} \rfloor + 1$ and $v(k) := \frac{1}{n}\mathbf{1}_{\{1,\ldots,n\}}(k)$, for $k \in \mathbb{Z}$. Clearly, $\|v\|_1 = 1$ and for any $s \in S$, one has

$$\|sv - v\|_1 = \frac{2s}{n} \leq \frac{2m}{n} < \varepsilon$$

so $v \in \ell^1(\mathbb{Z})$ is (S, ε)–invariant.

We can now define what an amenable group is.

Definition 2.4. A group G is amenable if it has the Reiter property (R_1).

The two previous examples show then that finite groups, and $\mathbb{Z}$, are amenable.

Among all $\ell^p(G)$–spaces, $\ell^2(G)$ carries naturally the structure of a Hilbert space, distinguishing it from the others. We are then willing to relate the two properties (R_1) and (R_2). This is done by the following result.

Proposition 2.5. A group G has (R_1) if and only if it has (R_2).

Proof. Suppose G has the property (R_1), and fix $S \subset G$ finite, $\varepsilon > 0$.
By assumption, there is $\psi \in \ell^1(G)$ which is (S, ε^2)–invariant, *i.e.*

$$\|s\psi - \psi\|_1 < \varepsilon^2 \|\psi\|_1$$

for all $s \in S$. Let then $\varphi := |\psi|^{\frac{1}{2}}$. First, it is an element of $\ell^2(G)$, because

$$\|\varphi\|_2^2 = \sum_{g \in G} |\varphi(g)|^2 = \sum_{g \in G} |\psi(g)| = \|\psi\|_1$$

and $\|\psi\|_1 < \infty$. Now, if $s \in S$, one has

$$\begin{aligned}
\|s\varphi - \varphi\|_2^2 &= \sum_{g \in G} |(s\varphi)(g) - \varphi(g)|^2 \\
&= \sum_{g \in G} |\varphi(s^{-1}g) - \varphi(g)|^2 \\
&= \sum_{g \in G} \left| |\psi(s^{-1}g)|^{\frac{1}{2}} - |\psi(g)|^{\frac{1}{2}} \right|^2
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{g\in G} \big| |\psi(s^{-1}g)| - |\psi(g)| \big| \\
&\leq \sum_{g\in G} |\psi(s^{-1}g) - \psi(g)| \\
&= \|s\psi - \psi\|_1 \\
&< \varepsilon^2 \|\psi\|_1 = \varepsilon^2 \|\varphi\|_2^2
\end{aligned}$$

using the inequality $|a-b|^2 \leq |a^2 - b^2|$ if $a, b \geq 0$, the fact that ψ is (S, ε)–invariant, and our previous computation. Now taking the square root yields $\|s\varphi - \varphi\|_2 < \varepsilon\|\varphi\|_2$, and this holds for all $s \in S$. Thus G has (R_2) as well.

Conversely, assume G has (R_2). Let $S \subset G$ be finite, and $\varepsilon > 0$. By hypothesis, there is $\psi \in \ell^2(G)$ which is $(S, \frac{\varepsilon}{2})$–invariant. Let us define $\varphi := \psi^2$. Note to begin that

$$\|\varphi\|_1 = \sum_{g\in G} |\varphi(g)| = \sum_{g\in G} |\psi(g)|^2 = \|\psi\|_2^2 < \infty$$

so indeed $\varphi \in \ell^1(G)$. Proceeding as above, for all $s \in S$ we get

$$\begin{aligned}
\|s\varphi - \varphi\|_1 &= \sum_{g\in G} |(s\varphi)(g) - \varphi(g)| \\
&= \sum_{g\in G} |\varphi(s^{-1}g) - \varphi(g)| \\
&= \sum_{g\in G} |\psi(s^{-1}g)^2 - \psi(g)^2| \\
&= \sum_{g\in G} |\psi(s^{-1}g) - \psi(g)||\psi(s^{-1}g) + \psi(g)| \\
&\leq \left(\sum_{g\in G} |\psi(s^{-1}g) - \psi(g)|^2 \right)^{\frac{1}{2}} \left(\sum_{g\in G} |\psi(s^{-1}g) + \psi(g)|^2 \right)^{\frac{1}{2}} \\
&= \underbrace{\|s\psi - \psi\|_2}_{<\frac{\varepsilon}{2}\|\psi\|_2} \underbrace{\|s\psi + \psi\|_2}_{\leq 2\|\psi\|_2} \\
&< \varepsilon\|\psi\|_2^2 = \varepsilon\|\varphi\|_1
\end{aligned}$$

and the first inequality follows from Cauchy-Schwarz. The bound $\|s\psi + \psi\|_2 \leq 2\|\psi\|_2$ comes from the triangle inequality and the fact that the action of s on $\ell^2(G)$ is isometric. Hence we found a (S, ε)–invariant vector $\varphi \in \ell^1(G)$, which shows G has (R_1). This concludes the proof. □

As a matter of fact, we will need a third equivalent characterization of amenability, that we formulate now. We say that G *has the property* (C) if, for all $S \subset G$ finite and $\varepsilon > 0$, there exists $\varphi \in \ell^2(G)$ such that

$$\left\| \frac{1}{|S|} \sum_{s\in S} s\varphi \right\|_2 > (1-\varepsilon)\|\varphi\|_2.$$

Thanks to our work above in Hilbert spaces, we can derive the following proposition.

Proposition 2.6. A group G has (R_2) if and only if G has (C).

Proof. First, let's assume G has (R_2). Fix $S \subset G$ finite and $\varepsilon > 0$. Then there exists $\varphi \in \ell^2(G)$ such that $\|s\varphi - \varphi\|_2 < \varepsilon\|\varphi\|_2$ for all $s \in S$. The two triangle inequalities provide

$$\begin{aligned}
\|\varphi\|_2 - \left\|\frac{1}{|S|}\sum_{s\in S} s\varphi\right\|_2 &\leq \left\|\varphi - \frac{1}{|S|}\sum_{s\in S} s\varphi\right\|_2 \\
&\leq \frac{1}{|S|}\left\|\sum_{s\in S} \varphi - s\varphi\right\|_2 \\
&\leq \frac{1}{|S|}\sum_{s\in S} \underbrace{\|\varphi - s\varphi\|_2}_{\leq \varepsilon\|\varphi\|_2} \\
&\leq \varepsilon\|\varphi\|_2
\end{aligned}$$

which is equivalent to $\left\|\frac{1}{|S|}\sum_{s\in S} s\varphi\right\|_2 \geq (1-\varepsilon)\|\varphi\|_2$. Hence G has also the property (C).

Conversely, suppose G does not have (R_2). We will show it does not have property (C) either. By hypothesis, there is a finite subset S of G and a constant $\varepsilon > 0$ such that, for all $\varphi \in \ell^2(G)$, there exists $s \in S$ with $\|s\varphi - \varphi\|_2 \geq \varepsilon\|\varphi\|_2$. Write $S = \{s_1, \dots, s_{n-1}\}$, and let us define $S' := S \cup \{e_G\}$. Fix $\varphi \in \ell^2(G)$ with $\|\varphi\|_2 = 1$. By hypothesis, in the family of unit vectors $u_1 = \varphi, u_2 = s_1\varphi, \dots, u_n = s_{n-1}\varphi$, we can find an index $i \in \{1, \dots, n-1\}$ such that

$$\|u_i - u_1\| \geq \varepsilon.$$

By Proposition 1.19, we therefore find $\varepsilon' > 0$ such that

$$\left\|\frac{1}{|S'|}\sum_{s'\in S'} s'\varphi\right\|_2 = \left\|\frac{1}{n}\sum_{i=1}^{n} u_i\right\|_2 \leq 1 - \varepsilon'.$$

Note that the constant ε' does not depend on φ. Now, if $\varphi \in \ell^2(G) \setminus \{0\}$ doesn't have norm 1, we can apply what we just proved to $\psi := \frac{\varphi}{\|\varphi\|}$ to get

$$\left\|\frac{1}{|S'|}\sum_{s'\in S'} s'\psi\right\|_2 \leq 1 - \varepsilon'$$

which is equivalent to $\left\|\frac{1}{|S'|}\sum_{s'\in S'} s'\varphi\right\|_2 \leq (1-\varepsilon')\|\varphi\|_2$. Lastly, if $\varphi = 0$ the inequality clearly holds. This proves that G does not have (C), and finishes the proof. □

Therefore we will make use of the following corollary to prove Kesten's theorem.

Corollary 2.7. A group G is amenable if and only if it has property (C).

Proof. Combine Proposition 2.5 and 2.6. □

2.2 Random walks on finitely generated groups

Let G be a finitely generated group, with a finite symmetric generating set S. Denote e its neutral element, and assume $e \notin S$.

Let μ be a symmetric probability measure on G, *i.e.* a map $\mu\colon G \longrightarrow [0,1]$ such that

$$\sum_{g\in G} \mu(g) = 1$$

and $\mu(g) = \mu(g^{-1})$ for any $g \in G$.

A *random walk* on G with distribution μ is a Markov chain consisting of $G-$valued random variables $(X_n)_{n\geq 0}$, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, of the form

$$X_n = X_0\xi_1 \dots \xi_n$$

where $(\xi_n)_{n\geq 1}$ are independent and identically distributed $G-$valued random variables with distribution μ, and $X_0 = x$ with probability one, where $x \in G$ is an arbitrary group element. The underlying probability space is $\Omega = G^{\mathbb{N}}$, $\mathcal{F}$ is the cylindrical $\sigma-$algebra, generated by *cylinder sets*, *i.e.* subsets of Ω consisting of sequences with a finite number of fixed coordinates. More precisely, for $k \geq 0$ and $g \in G$, let

$$C_g^k := \{(x_n)_{n\geq 0} \in \Omega \mid x_k = g\}$$

and for finitely many group elements $g_0, \dots, g_n \in G$, let $C_{g_0,\dots,g_n} := \bigcap_{k=0}^{n} C_{g_k}^k$. Denoting by $\mathcal{C}$ the collection of all cylinder sets, we then have $\mathcal{F} = \sigma(\mathcal{C})$. Finally, $\mathbb{P} = \mu^{\mathbb{N}}$ is the product measure, *i.e.* the unique measure on Ω with $\mathbb{P}(C_{g_0,\dots,g_n}) = \prod_{i=0}^{n} \mu(g_i)$.

The *step distribution* of the random walk is the common distribution μ of the increments $(\xi_n)_{n\geq 1}$, that is

$$\mathbb{P}(\xi_i = y) = \mu(y)$$

and in general $\mathbb{P}(X_n = y) = \mu^{*n}(y)$, where μ^{*n} denotes the $n-$fold convolution[1] of μ with itself. The transition probabilities are then given by $p(x,y) = \mu(x^{-1}y)$, and more

[1] If $f\colon (E, \mathcal{A}) \longrightarrow (F, \mathcal{B})$ is a measurable map between two measure spaces, and if μ is a measure on E, its *push-forward* under f is the measure $f_*\mu$ on F defined by $f_*\mu(B) := \mu(f^{-1}(B))$ for all $B \in \mathcal{B}$. Now, for two measures μ and ν on G, their *convolution* $\mu * \nu$ is the push-forward of the product measure $\mu \otimes \nu$ under the map $G \times G \longrightarrow G$, $(x,y) \longmapsto xy$.

generally

$$p^{(n)}(x, y) = \mu^{*n}(x^{-1}y)$$

for all $n \geq 1, x, y \in G$. Since S generates G, this Markov chain is irreducible, meaning that for all $x, y \in G$, there exists $k \geq 1$ such that $p^{(k)}(x, y) > 0$.

Lastly, when the random increments take values in the generating set S, we say the model is a *nearest-neighbour* random walk, and it is the *simple* random walk if $\mu(s) = \frac{1}{|S|}$ for all $s \in S$. We will restrict ourselves to that case in what follows. Note that in this setting, $p(x, y) \neq 0$ if and only if $x^{-1}y \in S$, *i.e.* we pass from x to y by right multiplication by a generator $s \in S$.

To visualize a random walk concretely, it is common to use *Cayley graphs* of finitely generated groups. First, let us state the definition of a graph we will use.

Definition 2.8. A graph Γ is a pair (V, E) of sets together with three maps

$$\begin{aligned} o &: E \longrightarrow V \\ t &: E \longrightarrow V \\ \bar{\cdot} &: E \longrightarrow E \end{aligned}$$

such that $o(\bar{e}) = t(e)$ for all $e \in E$, and $\bar{\cdot}: E \longrightarrow E$ is an involution without fixed points.

An element $v \in V$ is called a *vertex* of the graph Γ, an element $e \in E$ is an (oriented) *edge*, and $\bar{e} \in E$ is the *reversed* edge to e. Above, the letters "o" and "t" were not chosen randomly : think to the "origin" and the "terminus" of an edge. The third map is the one reversing the orientation of an edge, and it is therefore natural to require it is an involution. Changing the orientation twice does nothing. To say it does not have fixed points means each edge $e \in E$ is different from its reversed $\bar{e} \in E$.

The *degree* of a vertex $v \in V$ is the cardinality of the set $o^{-1}(v)$, and is denoted $\deg(v)$. If $\deg(v) = k$ for all $v \in V$, then Γ is said to be $k-$*regular*.

Given two vertices $u, v \in V$, a *path* between u and v is a sequence $e_1, \ldots, e_n$ of edges such that $o(e_1) = u$, $t(e_n) = v$ and $t(e_i) = o(e_{i+1})$ for all $i = 1, \ldots, n-1$. The graph Γ is called *connected* if for every $u, v \in V$, there exists a path between u and v.

A *loop* is an edge $e \in E$ such that $o(e) = t(e)$. Note that if e is a loop, then so is $\bar{e}$. A *cycle* in Γ is a path connecting a vertex $v \in V$ to itself without repetitions (unless v which starts and ends the cycle), *i.e.* we cannot have $o(e_j) = t(e_i)$ if $|i - j| \geq 1$. Lastly, a non-empty connected graph without cycles is called a *tree*, and a graph without loops and multiple edges is *simple*.

Remark 2.9. The above definition, due to Serre [12], is a bit more involved and heavier than other ones, and perhaps less intuitive compared to the way we draw graphs in practice. However, it is by far the best, and the ambiguities encountered with others definitions are not an issue anymore.

Let's adapt this framework in the particular case of a finitely generated group.

Definition 2.10. The Cayley graph of G with respect to S is the graph specified by $V = G$, $E = G\times S$, and the maps $o(g,s) = g$, $t(g,s) = gs$ and $\overline{(g,s)} = (gs, s^{-1})$.

We shall check these satisfy Definition 2.8. Indeed, we have

$$o(\overline{(g,s)}) = o(gs, s^{-1}) = gs = t(g,s)$$

and $\overline{\overline{(g,s)}} = \overline{(gs,s^{-1})} = (gss^{-1}, s) = (g,s)$ so $\overline{\cdot}$ is an involution. The fact it does not have fixed points comes from the assumption that $e \notin S$. Consequently, the sets $V = G$ and $E = G \times S$, together with these three maps, form a graph. We will denote it $\mathrm{Cay}(G,S)$.

In a Cayley graph, each edge carries a label, given by a generator $s \in S$. Around each vertex, there are $|S|$ outgoing edges, and $|S|$ incoming edges. In particular, the degree of each vertex is $2|S|$. However, in the sequel we will replace a pair of edges corresponding to a generator and its inverse by a single edge. In this way, $\mathrm{Cay}(G,S)$ is $|S|$−regular.

Finally, observe that since S generates G, $\mathrm{Cay}(G,S)$ is connected, and since $e \notin S$, $\mathrm{Cay}(G,S)$ is simple.

Below are shown Cayley graphs for two cyclic groups, one finite and one infinite.

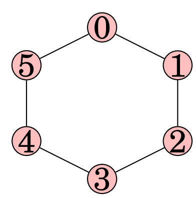

Figure 1: Cay$(\mathbb{Z}/6\mathbb{Z}, \{\pm 1\})$

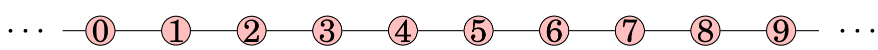

Figure 2: Cay$(\mathbb{Z}, \{\pm 1\})$

For further examples, especially the ones we obtain by changing the generating set S, we refer to [5, chapter IV.A].

Let's go back to random walks. To study their long-run behavior, a useful object to consider is the *Green function*. It encodes probabilities transition of the model.

Definition 2.11. Let $x, y \in G$, $z \in \mathbb{C}$.
The generating function $G(x, y|z)$ of the sequence $(p^{(n)}(x, y))_{n\geq 0}$, defined as

$$G(x, y|z) := \sum_{n\geq 0} p^{(n)}(x, y)z^n$$

is called the Green function at $x, y \in G$.

As the definition suggests it, we must do a computation for every pair of elements x, y in G. However, it is not the case if G is finitely generated.

Lemma 2.12. Let $x_1, y_1, x_2, y_2 \in G$, and $z \in \mathbb{C}$.
$G(x_1, y_1|z)$ converges absolutely if and only if $G(x_2, y_2|z)$ converges absolutely.

Proof. Fix $x_1, y_1, x_2, y_2 \in G$, and $n \in \mathbb{N}$. Since S generates G, we find $s_1, \dots, s_i \in S$ such that $x_2 = x_1 s_1 \dots s_i$, so $p^{(i)}(x_1, x_2) > 0$. Likewise, y_1 can be obtained from y_2 by right multiplication by j generators, so $p^{(j)}(y_2, y_1) > 0$, for some $i, j \geq 1$. It follows that

$$p^{(n)}(x_1, y_1) \geq p^{(i)}(x_1, x_2)p^{(n-(i+j))}(x_2, y_2)p^{(j)}(y_2, y_1).$$

Suppose now that $G(x_1, y_1|z)$ converges absolutely. Then we get

$$\sum_{n\geq i+j} p^{(n)}(x_2, y_2)|z|^n \leq \frac{1}{|z|^{i+j}p^{(i)}(x_1, x_2)p^{(j)}(y_2, y_1)} \sum_{n\geq 0} p^{(n)}(x_1, y_1)|z|^n < \infty$$

which implies that $G(x_2, y_2|z)$ also converges absolutely. A symmetric reasoning gives the other implication, finishing the proof. □

Thus, for a finitely generated group, the behavior of the Green function does not depend on the choice of the pair $x, y \in G$. In what follows we will assume the random walk starts at e, *i.e.* $X_0 = e$. In particular, we only consider $G(e, e|z)$ and its radius of convergence, given by

$$\frac{1}{\limsup_{n\to\infty} p^{(n)}(e, e)^{\frac{1}{n}}}$$

according to the Cauchy-Hadamard formula [11, chapter 4.1].

Definition 2.13. The number

$$\rho := \limsup_{n\to\infty} p^{(n)}(e, e)^{\frac{1}{n}}$$

is called the exponential decay rate of the random walk on G.

Note that, since $p^{(n)}(e,e) \leq 1$ for all $n \geq 1$, $G(e,e|z)$ converges absolutely for all $z \in \mathbb{C}$ with $|z| < 1$, so its radius of convergence is at least 1, and then $\rho \leq 1$.

When studying random walks, an important feature of their behaviour is whether they are recurrent or transient.

Definition 2.14. A state $x \in G$ is called recurrent if $G(x,x|1) = \infty$.

On the other hand, if a state $x \in G$ is such that $G(x,x|1)$ converges, then x is called *transient*.

This definition appeals several remarks, though.

Remark 2.15. (i) For irreducible Markov chains, it can be shown that all states are either transient or recurrent [4, corollary 13.4.5]. Since this applies to the random walk on a finitely generated group, we will say the walk is recurrent if $G(e,e|1)$ diverges.

(ii) The above definition is equivalent to requiring that almost surely the walk visits each state infinitely often. We refer to [4, proposition 13.4.2] for a proof of this fact.

With these remarks, we directly deduce the next result.

Corollary 2.16. If a random walk on G is recurrent, then $\rho = 1$.

Proof. Since the walk is recurrent, $G(e,e|1)$ diverges, so its radius of convergence is at most 1. It is also at least 1, so in fact the radius of convergence equals 1, and thus so does ρ. □

Let's illustrate the previous concepts with some explicit computations. Recall that for two sequences of real numbers $(u_k)_{k\geq 0}$ and $(v_k)_{k\geq 0}$, we denote $u_k \simeq v_k$ if

$$\lim_{k\to\infty} \frac{u_k}{v_k} = 1.$$

In words, this means $(u_k)_{k\geq 0}$ and $(v_k)_{k\geq 0}$ have the same asymptotic behaviour.

Example 2.17. (i) Let $G = \mathbb{Z}$ be the group of integers, generated by $S = \{-1, 1\}$. The simple random walk on G is then the process $(X_n)_{n\geq 0}$ defined by $X_0 = 0$ and $X_n = \xi_1 + \cdots + \xi_n$, $n \geq 1$, where each ξ_i has the Bernoulli distribution of parameter $\frac{1}{2}$. To determine $p^{(n)}(0,0)$ explicitly, first note that $p^{(2k+1)}(0,0) = 0$, as it is impossible to come back to the origin with an odd number of steps. So we are left to compute $p^{(2k)}(0,0)$, for every $k \geq 1$. Among the 2^{2k} possible walks, those which start and end to 0 have an equal amount of steps to the right and steps to the left. Such a walk is then the same thing as the choice of k objects among $2k$. Thus

$$p^{(2k)}(0,0) = \frac{1}{2^{2k}}\binom{2k}{k} = \frac{1}{2^{2k}}\frac{(2k)!}{(k!)^2}.$$

Using Stirling's formula, $k! \simeq \sqrt{2\pi k}k^k \mathrm{e}^{-k}$, we obtain

$$p^{(2k)}(0,0) \simeq \frac{1}{2^{2k}} \frac{\sqrt{2\pi 2k}(2k)^{2k}\mathrm{e}^{-2k}}{2\pi k k^{2k}\mathrm{e}^{-2k}} = \frac{1}{\sqrt{\pi k}}$$

and since $\sum_{k=1}^{\infty} \frac{1}{\sqrt{\pi k}} = \infty$, it follows that $\sum_{n=1}^{\infty} p^{(n)}(0,0) = \sum_{k=1}^{\infty} p^{(2k)}(0,0) = \infty$. Therefore, the chain is recurrent, and its exponential decay rate equals 1 by Corollary 2.16.

(ii) A similar reasoning can be done for the simple random walk on $\mathbb{Z}^2$. Among the 4^{2k} walks of length $2k$, those which come back to the origin in $2k$ steps must count an equal amount of steps to the left and to the right, and an equal amount of steps upwards and steps downwards. Such a walk is then the same as the choice of an integer $j \in \{0, \ldots, n\}$, the choice of j steps to the right, and $n - j$ steps upwards. It then follows that

$$\begin{aligned}
p^{(2k)}(0,0) &= \frac{1}{4^{2k}} \sum_{j=0}^{k} \binom{2k}{2j}\binom{2j}{j}\binom{2(k-j)}{k-j} \\
&= \frac{1}{4^{2k}} \sum_{j=0}^{k} \frac{(2k)!}{(2j)!(2k-2j)!} \frac{(2j)!}{(j!)^2} \frac{(2k-2j)!}{((k-j)!)^2} \\
&= \frac{1}{4^{2k}} \frac{(2k)!}{(k!)^2} \sum_{j=0}^{k} \binom{k}{j}\binom{k}{k-j} \\
&= \left(\frac{1}{2^{2k}}\binom{2k}{k}\right)^2
\end{aligned}$$

where the last equality relies on the combinatorial identity $\sum_{j=0}^{k} \binom{k}{j}\binom{k}{k-j} = \binom{2k}{k}$.

Using point (i) above, we deduce that

$$p^{(2k)}(0,0) \simeq \frac{1}{\pi k}.$$

Since $\sum_{k=1}^{\infty} \frac{1}{\pi k}$ diverges, so does $G(0,0|1)$, proving that the simple random walk on $\mathbb{Z}^2$ is recurrent. In particular, $\rho = 1$.

Similar techniques and combinatorial identities establish that, for $\mathbb{Z}^3$, the return probability $p^{(2k)}(0,0)$ behaves, up to a constant, as $\frac{1}{k^{3/2}}$. More generally, for the simple random walk on $\mathbb{Z}^d$, $p^{(2k)}(0,0)$ behaves as $\frac{1}{k^{d/2}}$, and therefore the walk is recurrent if and only if $d \in \{1, 2\}$. This is a celebrated result, known as Polya's theorem. See for instance [5, chapter I.B], and [9] for the original article of Polya.

This result also provides a counter-example to the converse of Corollary 2.16: the simple random walk on $\mathbb{Z}^3$ has exponential decay rate equals to 1, but is transient. Hence, for a random walk, the value of ρ does not characterise its recurrent or transient behaviour. Here are some reasons to explain this difference.

First of all, observe that the difference between the recurrence of the random walk on $\mathbb{Z}^2$ and the transience of the walk on $\mathbb{Z}^3$ is due to the difference between the nature of the two series

$$\sum_{n=1}^{\infty} \frac{1}{n}, \; \sum_{n=1}^{\infty} \frac{1}{n^{3/2}}.$$

and the difference between these two series is subtle. The second converges while the first one diverges, but this only relies on the fact that the sequence $\frac{1}{n^{3/2}}$ goes to 0 a little bit faster than $\frac{1}{n}$. Comparing the reasoning above for $\mathbb{Z}^2$ and the one for $\mathbb{Z}^3$ in [5], these two sequences are the consequences of the number of neighbours and dimensions in each model when computing probabilities. Broadly speaking, in dimension $d \geq 3$, the random walk has too much space to escape. This already suggests the recurrent/transient behaviour is dependent of the geometry of the lattice at a small scale.

On the other hand, the exponential decay rate ignores completely the difference that can occur between two sequences such as $\frac{1}{n}$ and $\frac{1}{n^{3/2}}$, and gives only information on the (non-)exponential behaviour of the return probabilities $p^{(n)}(e,e)$. Therefore, its values can only reflect a property of the large-scale geometry of the underlying lattice. It turns out amenability of a group is the good notion to consider for measuring its size, when regarding it from far away, and can thus traduce a (non-)exponential behaviour of $p^{(n)}(e,e)$. This is what Kesten proved in 1959, and what we will establish in the sequel.

Lastly, to avoid parity problems we encountered in the two examples above, we will now work mostly with $p^{(2n)}(e,e)$ rather than $p^{(n)}(e,e)$. It does not affect the exponential decay rate, since

$$\limsup_{n\to\infty} p^{(n)}(e,e)^{\frac{1}{n}} = \limsup_{n\to\infty} p^{(2n)}(e,e)^{\frac{1}{2n}}.$$

Indeed, we have $p^{(2n)}(e,e) \geq p^{(n)}(e,e)p^{(n)}(e,e) = (p^{(n)}(e,e))^2$, so taking the $2n-$th root and the limsup, we obtain $\limsup_{n\to\infty} p^{(n)}(e,e)^{\frac{1}{n}} \leq \limsup_{n\to\infty} p^{(2n)}(e,e)^{\frac{1}{2n}}$. The reverse inequality comes from the definition of the limsup.

2.3 The Markov operator

In this part, we introduce the key tool for proving Kesten's theorem. The operator we construct is the bridge between the amenability condition (C) we derived above, and the exponential decay rate of a random walk.

From now on, we only consider the simple random walk on a finitely generated group G, with a symmetric generating set S, such that $e \notin S$.

Definition 2.18. The operator $M\colon \ell^2(G) \longrightarrow \ell^2(G)$, defined as

$$(Mf)(g) := \frac{1}{|S|}\sum_{s\in S} f(s^{-1}g)$$

for all $f \in \ell^2(G)$ and $g \in G$, is called the Markov operator associated to the simple random walk.

In words, this definition says the value of Mf at g is obtained by averaging the values of f at the nearest neighbours of g, group elements of the form $s^{-1}g$, $s \in S$.

First, let us show that M indeed has good properties. It will allow us to invoque what we proved in Section 1.

Lemma 2.19. (i) M is linear.

(ii) M is a bounded operator, and $\|M\| \le 1$.

(iii) M is self-adjoint.

Proof. (i) Fix $f, g \in \ell^2(G)$, $\lambda \in \mathbb{C}$ and $h \in G$. Then

$$\begin{aligned} M(\lambda f + g)(h) &= \frac{1}{|S|}\sum_{s\in S}(\lambda f + g)(s^{-1}h) \\ &= \frac{\lambda}{|S|}\sum_{s\in S} f(s^{-1}h) + \frac{1}{|S|}\sum_{s\in S} g(s^{-1}h) \\ &= \lambda(Mf)(h) + (Mg)(h) \\ &= (\lambda Mf + Mg)(h). \end{aligned}$$

Henceforth $M(\lambda f + g) = \lambda Mf + Mg$, for every $f, g \in \ell^2(G)$, $\lambda \in \mathbb{C}$.

(ii) Let $f \in \ell^2(G), f \neq 0$. We compute that

$$\begin{aligned} \|Mf\|_2^2 &= \sum_{g\in G} |(Mf)(g)|^2 \\ &= \sum_{g\in G}\left|\frac{1}{|S|}\sum_{s\in S} f(s^{-1}g)\right|^2 \\ &= \sum_{g\in G}\frac{1}{|S|^2}\left|\sum_{s\in S} f(s^{-1}g)\right|^2 \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{g\in G} \frac{1}{|S|} \sum_{s\in S} f(s^{-1}g)^2 \\
&= \frac{1}{|S|} \sum_{s\in S} \sum_{g\in G} f(s^{-1}g)^2 \\
&= \|f\|_2^2
\end{aligned}
$$

by using Cauchy-Schwarz inequality first and thereafter Fubini's theorem, which applies since $\|f\|_2 < \infty$ and all terms are positive, to permute the order of summation. Hence $\|Mf\|_2 \leq \|f\|_2$ for all $f \in \ell^2(G), f \neq 0$, and we deduce $\|M\| \leq 1$.

(iii) We must show that $\langle Mf, g\rangle = \langle f, Mg\rangle$ for every pair $f, g \in \ell^2(G)$. First, suppose that f, g are both $\mathbb{R}-$valued and positive. Expanding the definition of M, one has

$$
\begin{aligned}
\langle Mf, g\rangle &= \sum_{h\in G} (Mf)(h)g(h) \\
&= \sum_{h\in G} \frac{1}{|S|} \sum_{s\in S} f(s^{-1}h)g(h) \\
&= \sum_{s\in S} \frac{1}{|S|} \sum_{h\in G} f(s^{-1}h)g(h) \\
&= \sum_{s\in S} \frac{1}{|S|} \sum_{t\in G} f(t)g(st) \\
&= \sum_{t\in G} \frac{1}{|S|} \sum_{s\in S} f(t)g(s^{-1}t) \\
&= \sum_{t\in G} f(t)(Mg)(t) \\
&= \langle f, Mg\rangle
\end{aligned}
$$

and each permutation of sums is justified by Fubini's theorem, which we may use since all terms are positive. The fifth equality relies on the fact that S is symmetric, so we can safely change the argument of g without altering the value of the sum.

Thus $\langle Mf, g\rangle = \langle f, Mg\rangle$ holds for positive $\mathbb{R}-$valued functions. From there, we get the equality for arbitrary $\mathbb{R}-$valued functions, by splitting $f = f^+ - f^-$ and $g = g^+ - g^-$ into their positive and negative parts. Lastly, this also implies the equality for $\mathbb{C}-$valued functions, by writing every $f \in \ell^2(G)$ as $f = \mathrm{Re}(f) + i\mathrm{Im}(f)$, and using linearity of the inner product in the first variable, and anti-linearity in the second. □

Remark 2.20. Point (iii) above and an induction on $n \geq 1$ implies that M^n is self-adjoint for all $n \geq 1$.

A crucial property of M is that one can recover the transition probabilities of the model from it.

Proposition 2.21. For every $g, h \in G$, $\langle M\delta_h, \delta_g\rangle = p(g,h)$.
More generally, $\langle M^n\delta_h, \delta_g\rangle = p^{(n)}(g,h)$ for every $g, h \in G$, $n \geq 1$.

Proof. By the definition of M, we have

$$\langle M\delta_h, \delta_g\rangle = \sum_{t\in G}(M\delta_h)(t)\delta_g(t) = \sum_{t\in G}\frac{1}{|S|}\sum_{s\in S}\delta_h(s^{-1}t)\delta_g(t).$$

If g and h are not related by a generator, there is no $s \in S$ such that $\delta_h(s^{-1}t)\delta_g(t) \neq 0$, so $\langle M\delta_h, \delta_g\rangle = 0$, which agrees with $p(g,h)$.

On the other hand, if g and h are nearest-neighbour, there is exactly one $s \in S$ such that $g = sh$, and $\langle M\delta_h, \delta_g\rangle$ reduces to $\frac{1}{|S|} = p(g,h)$, which shows the first identity. For the second, we do an induction on $n \geq 1$. The case $n = 1$ is handled. Suppose then $n > 1$, and that the identity holds up to the $n-$th power of M. Using self-adjointness of M and the induction hypothesis, one has

$$\begin{aligned}
\langle M^{n+1}\delta_h, \delta_g\rangle &= \langle M^n\delta_h, M\delta_g\rangle \\
&= \langle M^n\delta_h, \frac{1}{|S|}\sum_{s\in S}s\delta_g\rangle \\
&= \frac{1}{|S|}\sum_{s\in S}\langle M^n\delta_h, \delta_{sg}\rangle \\
&= \frac{1}{|S|}\sum_{s\in S}p^{(n)}(sg,h) \\
&= \sum_{s\in S}p(g,sg)p^{(n)}(sg,h) \\
&= p^{(n+1)}(g,h)
\end{aligned}$$

and the last equality follows from the total probability formula. This concludes the inductive step, and also our proof. □

From this, we derive an important consequence, namely that $(p^{(2n)}(e,e))_{n\geq 1}$ has the behaviour we expect. This also justify the terminology "decay rate".

Corollary 2.22. The sequence $(p^{(2n)}(e,e))_{n\geq 0}$ is decreasing.

Proof. Let $n \geq 0$. By the previous proposition, and the self-adjointness of M^{n+2} we can write $p^{(2n+2)}(e,e) = \langle M^{2n+2}\delta_e, \delta_e\rangle = \langle M^n\delta_e, M^{n+2}\delta_e\rangle$. The Cauchy-Schwarz inequality then implies

$$p^{(2n+2)}(e,e) \leq \|M^n\delta_e\|\|M^{n+2}\delta_e\| \leq \|M^n\delta_e\|^2 = p^{(2n)}(e,e)$$

using that $\|M^2\| \leq 1$. □

In fact, M contains much more information about the model than just the transition probabilities. To prove the next proposition, we appeal the next basic fact from real analysis.

Lemma 2.23. Let $(u_n)_{n\geq 0}$ be a sequence of positive real numbers.
It holds that

$$\liminf_{n\to\infty} \frac{u_{n+1}}{u_n} \leq \liminf_{n\to\infty} u_n^{\frac{1}{n}} \leq \limsup_{n\to\infty} u_n^{\frac{1}{n}} \leq \limsup_{n\to\infty} \frac{u_{n+1}}{u_n}.$$

Proof. Let $\ell := \limsup_{n\to\infty} \frac{u_{n+1}}{u_n}$, and $\varepsilon > 0$. There exists $N \in \mathbb{N}$ such that $\frac{u_{n+1}}{u_n} < \ell + \varepsilon$ if $n \geq N$. Hence, for $n \geq N$, one has

$$\frac{u_n}{u_N} = \frac{u_n}{u_{n-1}} \frac{u_{n-1}}{u_{n-2}} \cdots \frac{u_{N+1}}{u_N} < (\ell + \varepsilon)^{n-N}$$

which implies $u_n^{\frac{1}{n}} < u_N^{\frac{1}{n}} (\ell+\varepsilon)^{1-\frac{N}{n}} = (\ell+\varepsilon)\left(\frac{u_N}{(\ell+\varepsilon)^N}\right)^{\frac{1}{n}}$. Taking the limsup of both sides, it follows that

$$\limsup_{n\to\infty} u_n^{\frac{1}{n}} < \ell + \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, we get the third inequality. The first one is shown similarly, and the second is obvious. □

Additionally, recall that a function $f \in \ell^2(G)$ is *finitely supported* if $|\mathrm{supp}(f)| < \infty$, where $\mathrm{supp}(f) := \{x \in G \mid f(x) \neq 0\}$.

Here is the link between the Markov operator and the simple random walk on G.

Proposition 2.24. One has $\|M\| = \rho$.

Proof. To begin, note that for all $n \geq 1$, $p^{(2n)}(e,e) = \langle M^{2n}\delta_e, \delta_e\rangle$ by Proposition 2.21. This implies $p^{(2n)}(e,e) \leq \|M^{2n}\|$ by Proposition 1.12, which applies since M^{2n} is bounded and self-adjoint. Thus $p^{(2n)}(e,e)^{\frac{1}{2n}} \leq \|M^{2n}\|^{\frac{1}{2n}}$ for all $n \geq 1$, and taking the limsup yields

$$\rho = \limsup_{n\to\infty} p^{(2n)}(e,e)^{\frac{1}{2n}} \leq \limsup_{n\to\infty} \|M^{2n}\|^{\frac{1}{2n}} = \lim_{n\to\infty} \|M^{2n}\|^{\frac{1}{2n}} = \|M\|$$

where the last equality comes from Proposition 1.15. Hence $\rho \leq \|M\|$.

For the reverse inequality, fix $f \in \ell^2(G)$ finitely supported. Note that by Cauchy-Schwarz, and the fact that M is self-adjoint, we have

$$\|M^{n+1}f\|^2 = \langle M^{n+1}f, M^{n+1}f\rangle = \langle M^n f, M^{n+2}f\rangle \leq \|M^n f\|\|M^{n+2}f\|$$

so the sequence $\left(\frac{\|M^{n+1}f\|}{\|M^n f\|}\right)_{n\geq 0}$ is increasing. It is also bounded from above by 1, so its limit exists, and

$$\frac{\|Mf\|}{\|f\|} \leq \lim_{n\to\infty} \frac{\|M^{n+1}f\|}{\|M^n f\|}.$$

Now $\lim_{n\to\infty} \frac{\|M^{n+1}f\|}{\|M^n f\|} = \lim_{n\to\infty} \|M^n f\|^{\frac{1}{n}}$ by Lemma 2.23, so we are left to show this last limit is less than or equal to ρ. To do this, using self-adjointness of M we write

$$\|M^n f\|^2 = \langle M^{2n} f, f\rangle = \sum_{g\in G} (M^{2n} f)(g)\overline{f(g)} = \sum_{g\in G}\sum_{h\in G} p^{(2n)}(g,h) f(h)\overline{f(g)}$$

Since f has finite support, both sums run over a finite set of elements. Let $\varepsilon > 0$. By Lemma 2.12, the radius of convergence of $G(g,h|z)$ does not depend on g,h, so for each pair $g,h \in G$ giving a contribution to the sum, there is $N_{g,h} \in \mathbb{N}$ such that $p^{(2n)}(g,h)^{\frac{1}{2n}} \leq \rho + \varepsilon$ for all $n \geq N_{g,h}$. Letting $N := \max_{g,h} N_{g,h}$, we get

$$\|M^n f\|^2 \leq (\rho+\varepsilon)^{2n} C(f)$$

where $C(f) < \infty$ is a constant depending only on f. Taking the $2n-$th root and the limsup, it follows that

$$\frac{\|Mf\|}{\|f\|} \leq \lim_{n\to\infty} \|M^n f\|^{\frac{1}{n}} \leq \rho + \varepsilon.$$

As $\varepsilon > 0$ is arbitrary, this yields $\frac{\|Mf\|}{\|f\|} \leq \rho$ for every $f \in \ell^2(G)$ finitely supported. Since the latter subspace is dense in $\ell^2(G)$ (cf. [4, theorem 4.3.1]), we can use Remark 2.25 to conclude that $\|M\| \leq \rho$, as claimed. □

Remark 2.25. To complete the above proof, it remains to see that

$$\|M\| = \sup\left\{\frac{\|Mf\|}{\|f\|} \mid f \neq 0, |\mathrm{supp}(f)| < \infty\right\}.$$

Denote $r := \sup_{f\neq 0} \frac{\|Mf\|}{\|f\|}$ and r' the same supremum but running over finitely supported non-zero functions. Clearly, one has $r' \leq r$. For the reverse inequality, fix $f \neq 0$ and $\varepsilon > 0$. By density, we may pick $f' \in \ell^2(G)$ finitely supported such that $\|f - f'\|_2 < \frac{\varepsilon}{r+r'}$. It implies

$$\|f'\|_2 = \|f' - f + f\|_2 \leq \|f' - f\|_2 + \|f\|_2 < \frac{\varepsilon}{r+r'} + \|f\|_2$$

and it follows that

$$\begin{aligned}\|Mf\|_2 &= \|Mf - Mf' + Mf'\|_2 \\ &\leq \|Mf - Mf'\|_2 + \|Mf'\|_2 \\ &\leq r\|f - f'\|_2 + r'\|f'\|_2\end{aligned}$$

$$
\begin{aligned}
&< r\varepsilon + r'\left(\frac{\varepsilon}{r+r'} + \|f\|_2\right) \\
&= \varepsilon + r'\|f\|_2
\end{aligned}
$$

using the definitions of r and r' for the second inequality. Letting $\varepsilon \to 0$, we obtain $\frac{\|Mf\|_2}{\|f\|_2} \leq r'$ for all $f \neq 0$, and thus $r \leq r'$. This shows $r = r'$, as claimed.

Since $\rho = \|M\| = \lim_{n\to\infty} \|M^n\|^{\frac{1}{n}}$, ρ is often also called the spectral radius of the random walk.

2.4 Kesten's theorem

Here is the statement and the proof of Kesten's theorem.

Theorem 2.26. A group G is amenable if and only if $\rho = 1$.

Proof. Suppose first that G is amenable. By Corollary 2.7, G has the property (C), which exactly says the Markov operator has $\|M\| \geq 1$. On the other hand, $\|M\| \leq 1$, so $\|M\| = 1$, and by Proposition 2.24, we get $\rho = \|M\| = 1$.

Conversely, let us assume G is not amenable. Again, by Corollary 2.7, G does not have (C), and this implies that $\|M\| < 1$. As $\|M\| = \rho$, we get $\rho < 1$, and this finishes the proof. □

In words, as already hinted above, Kesten's theorem tells us the group is non-amenable if and only if the return probabilities at the origin $p^{(n)}(e,e)$ decay exponentially fast. This is quite intuitive: a non-amenable group has an expansive geometry, and its Cayley graph escapes very fast to infinity. In that situation, it is highly unexpected to come back to the origin after a large amount of steps, so $p^{(n)}(e,e)$ must decrease fast enough.

A direct implication of this characterization is a huge family of transient random walks, namely all those on non-amenable finitely generated groups.

Corollary 2.27. If G is not amenable, then the simple random walk on G is transient.

Proof. Since G is non-amenable, $\rho < 1$, and the contrapositive of Corollary 2.16 gives the claim. □

Propositions 2.21 and 2.24 in the previous subsection shows the Markov operator contains a lot of relevant informations of the model of simple random walks on finitely generated groups. This allows one to widely use tools from functional analysis and Hilbert spaces theory to tackle questions in probability theory. Corollary 2.27 above illustrates well this idea.

2.5 Simple random walks on free groups

In general, computing the exact value of the spectral radius is difficult, if not out of reach, and figuring out whether an interesting group is amenable or not via Kesten's criterion is in general hard. There is however one class of groups for which an explicit formula for the spectral radius is known, namely non-abelian free groups. This is also a result due to Kesten [8], that we establish in this part. The first step towards the proof is to introduce two other sequences of probabilities.

For $x, y \in G$, let $f^{(n)}(x, y)$ be the probability to reach y from x in n steps for the first time, and let $u^{(n)}(x, y)$ be the probability to reach y from x for the first time in n steps, with at least one non-trivial step. More precisely, we set $u^{(0)}(x, x) = 0$. Note that on the other hand $f^{(0)}(x, x) = 1$.

Note also that $f^{(n)}(x, x) = 0$ for all $n \geq 1$, and $f^{(n)}(x, y) = u^{(n)}(x, y)$ for all $n \geq 0$ if $x \neq y$. For $z \in \mathbb{C}$, we write

$$F(x, y|z) := \sum_{n=0}^{\infty} f^{(n)}(x, y)z^n, \; U(x, y|z) := \sum_{n=0}^{\infty} u^{(n)}(x, y)z^n$$

for the corresponding generating functions. They satisfy non-trivial relations with the Green function.

Lemma 2.28. Let $x, y \in G$, $z \in \mathbb{C}$. Then the following holds.

(i) $G(x, y|z) = F(x, y|z)G(y, y|z)$.

(ii) $G(x, x|z) = \dfrac{1}{1 - U(x, x|z)}$.

(iii) $U(x, x|z) = \sum_{y \in G} p(x, y)zF(y, x|z)$.

(iv) $F(x, y|z) = \sum_{w \in G} p(x, w)zF(w, y|z)$ if $x \neq y$.

Proof. (i) We start by proving that

$$p^{(n)}(x,y) = \sum_{k=0}^{n} f^{(k)}(x,y)p^{(n-k)}(y,y). \tag{1}$$

If $x = y$, this clearly holds, since $f^{(k)}(x,x) = 0$ for $k \geq 1$. We then consider the case $x \neq y$, and we prove the equality by induction on $n \geq 0$.

Suppose $n = 0$. Then both the left and the right hand sides equal 0, as $f^{(0)}(x,y) = p^{(0)}(x,y) = 0$. Now let $n \geq 1$, and suppose the equality holds up to $n-1$. By conditioning on the first step, and using the induction hypothesis, we get

$$\begin{aligned}
p^{(n)}(x,y) &= \sum_{z\in G} p(x,z)p^{(n-1)}(z,y) \\
&= \sum_{z\in G} p(x,z)\left(\sum_{k=0}^{n-1} f^{(k)}(z,y)p^{(n-1-k)}(y,y)\right) \\
&= \sum_{k=0}^{n-1} \underbrace{\left(\sum_{z\in G} p(x,z)f^{(k)}(z,y)\right)}_{=f^{(k+1)}(x,y)} p^{(n-1-k)}(y,y) \\
&= \sum_{k=1}^{n} f^{(k)}(x,y)p^{(n-k)}(y,y) \\
&= \sum_{k=0}^{n} f^{(k)}(x,y)p^{(n-k)}(y,y)
\end{aligned}$$

since $f^{(0)}(x,y) = 0$. The third equality relies on Fubini's theorem, to permute the order of summation. Hence we have (1) for every $n \geq 0$. By the definition of multiplication of generating functions, this equality exactly means

$$G(x,y|z) = F(x,y|z)G(y,y|z).$$

(ii) Observe that, again conditioning on the first step, one has

$$u^{(k)}(x,x) = \sum_{y\in G} p(x,y)u^{(k-1)}(y,x) = \sum_{y\in G} p(x,y)f^{(k-1)}(y,x) \tag{2}$$

and this holds for all $k \geq 0$ if we set $f^{(-1)}(x,y) := 0$ for all $x,y \in G$. Now for all $n \geq 1$, we compute that

$$\begin{aligned}
\sum_{k=0}^{n} u^{(k)}(x,x)p^{(n-k)}(x,x) &= \sum_{k=0}^{n}\left(\sum_{y\in G} p(x,y)f^{(k-1)}(y,x)\right)p^{(n-k)}(x,x) \\
&= \sum_{y\in G} p(x,y)\sum_{k=1}^{n} f^{(k-1)}(y,x)p^{(n-k)}(x,x)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{y \in G} p(x,y) \sum_{k=0}^{n-1} f^{(k)}(y,x) p^{(n-1-k)}(x,x) \\
&= \sum_{y \in G} p(x,y) p^{(n-1)}(y,x) \\
&= p^{(n)}(x,x)
\end{aligned}
$$

using (2) for the first equality, Fubini's theorem for the second to permute sums, (1) for the fourth and the total probability formula for the last one. This means the n−th coefficient of $G(x,x|z)$ coincides with the n−th coefficient of $U(x,x|z)G(x,x|z)$ for every $n \geq 1$, while for $n = 0$ we have $p^{(0)}(x,x) = 1$ and $u^{(0)}(x,x)p^{(0)}(x,x) = 0$. Hence

$$
G(x,x|z) - U(x,x|z)G(x,x|z) = 1
$$

as announced.

(iii) This follows from (2).

(iv) We use the same strategy as before. Let $x \neq y$. Then one has

$$
\begin{aligned}
\sum_{w \in G} p(x,w) z F(w,y|z) &= \sum_{w \in G} p(x,w) z \left(\sum_{k=0}^{\infty} f^{(k)}(w,y) z^k \right) \\
&= \sum_{k=0}^{\infty} \underbrace{\left(\sum_{w \in G} p(x,w) f^{(k)}(w,y) \right)}_{=f^{(k+1)}(x,y)} z^{k+1} \\
&= \sum_{k=1}^{\infty} f^{(k)}(x,y) z^k \\
&= F(x,y|z)
\end{aligned}
$$

using that $f^{(0)}(x,y) = 0$ if $x \neq y$, and Fubini's theorem for permuting the two sums. This shows (iv) and finishes the proof. □

The above results hold for any finitely generated group. However, if we restrict to non-abelian free groups, whose Cayley graphs have a tree structure, the generating function $F(x,y|z)$ has an additional property of "transitivity", leading to an explicit formula for the Green function $G(x,y|z)$, and thus also for its radius of convergence.

Hence, from now on, let $G = F_k$ be the non-abelian free group of rank k, and

$$
S = \{a_1, \ldots, a_k, a_1^{-1}, \ldots, a_k^{-1}\}
$$

be the standard generating set for G.

Lemma 2.29. The graph $\mathrm{Cay}(G,S)$ is an infinite $2k$−regular tree.

Proof. It is clear $\mathrm{Cay}(G,S)$ is infinite and $2k$–regular. Moreover, S generates G, so $\mathrm{Cay}(G,S)$ is connected. Towards a contradiction, suppose there is a cycle of length $n \geq 3$. We then have a sequence of edges

$$e_1 = (u, s_1),\ e_2 = (us_1, s_2),\ \ldots,\ e_n = (us_1 \ldots s_{n-1}, s_n)$$

and $u = us_1 \ldots s_n$. Henceforth, $s_1 \ldots s_n = e$ is a non-trivial relation in $G = F_k$, which is not free. This is the desired contradiction, and we have the claim. □

In a tree, for any two vertices x, y there exists a *unique* path $c(x,y)$ between x and y. The length of the path $c(x,y)$ is the number of edges it contains, and we denote it $\mathrm{d}(x,y)$.

Here is the transitivity property announced above.

Lemma 2.30. Let $x, y \in V(\mathrm{Cay}(G,S))$.
If $w \in c(x,y)$, then $F(x,y|z) = F(x,w|z)F(w,y|z)$.

Proof. Since the path $c(x,y)$ between x and y is unique, the random walk must pass to w when going from x to y. Conditioning with respect to the first visit to w, this gives

$$f^{(n)} = \sum_{k=0}^{n} f^{(k)}(x,w)f^{(n-k)}(w,y)$$

for all $n \geq 0$, and thus $F(x,y|z) = F(x,w|z)F(w,y|z)$. □

All we need is in place to determine the spectral radius of the simple random walk on a free group.

Theorem 2.31. For the simple random walk on $G = F_k$, one has $\rho = \frac{\sqrt{2k-1}}{k}$.

Proof. First we note that $f^{(k)}(x,y) = f^{(k)}(z,w)$ if $x \sim y$ and $v \sim w$ are two pairs of neighbours. Hence $F(x,y|z) = F(v,w|z)$, and for brievety we denote this series by $F(z)$. Now for every $x, y \in G$, there is a unique path $c(x,y)$ of length $\mathrm{d}(x,y)$ in $\mathrm{Cay}(G,S)$ connecting them, so Lemma 2.30 gives

$$F(x,y|z) = F(z)^{\mathrm{d}(x,y)}.$$

Suppose $x, y \in G$ are neighbours. Then, by Lemma 2.28(iv), we obtain

$$F(z) = F(x,y|z) = \sum_{w \sim x} \frac{1}{2k} zF(w,y|z) = \sum_{w \sim x} \frac{1}{2k} zF(z)^{\mathrm{d}(w,y)}.$$

In this sum, one term corresponds to $w = y$, and in this case $\mathrm{d}(w, y) = 0$. For the others $2k-1$ neighbours of x, we have $\mathrm{d}(w, y) = 2$, and thus

$$F(z) = \frac{1}{2k}z + \frac{2k-1}{2k}zF(z)^2.$$

Solving this quadratic equation, we end out with

$$F(z) = \frac{k \pm \sqrt{k^2 - (2k-1)z^2}}{(2k-1)z}$$

and since $F(0) = 0$ by definition, 0 must be an apparent singularity, so this imposes

$$F(z) = \frac{k - \sqrt{k^2 - (2k-1)z^2}}{(2k-1)z}.$$

From there, part (iii) of Lemma 2.28 implies $U(x, x|z) = zF(z) = \frac{k-\sqrt{k^2-(2k-1)z^2}}{2k-1}$, and part (ii) then yields

$$G(x, x|z) = \frac{1}{1 - U(x, x|z)} = \frac{2k-1}{(k-1) + \sqrt{k^2 - (2k-1)z^2}}.$$

Finally, by point (i) of Lemma 2.28, we have $G(x, y|z) = F(z)^{\mathrm{d}(x,y)}G(y, y|z)$ and it follows that

$$G(x, y|z) = \frac{2k-1}{(k-1) + \sqrt{k^2 - (2k-1)z^2}}\left(\frac{k - \sqrt{k^2 - (2k-1)z^2}}{(2k-1)z}\right)^{\mathrm{d}(x,y)}.$$

for all $x, y \in G$. To obtain the radius of convergence R of $G(x, y|z)$, we use Pringsheim's theorem [11, chapter 8.1], which assures R equals the smallest positive singularity of $G(x, y|z)$. From the expression above, this singularity is $\frac{k}{\sqrt{2k-1}}$, and hence

$$\rho = \frac{1}{R} = \frac{\sqrt{2k-1}}{k}$$

as announced. This concludes the proof. □

From this result, we derive two immediate consequences.

Corollary 2.32. (i) F_k is not amenable, for all $k \geq 2$.
(ii) The simple random walk on F_k is transient, for all $k \geq 2$.

Proof. (i) If $k \geq 2$, then $\rho(F_k) = \frac{\sqrt{2k-1}}{k} < 1$, so F_k is not amenable by Kesten's criterion.

(ii) This follows from (i) and Corollary 2.27. □

Remark 2.33. Here, we considered only simple random walks on free groups, whose Cayley graphs are $2k$–regular trees. Similar results as the ones above can be shown in fact for k–regular trees. See [13, chapter 1.1] for further details.

3. Boundaries for random walks on groups

The purpose of this section is to establish a second criterion relating the amenability of a group to the behaviour of random walks it carries. Kesten's theorem showed us amenability of a finitely generated group is equivalent to a slow decrease of the return probabilities at the origin. On the other hand, we are now going to be interested in the behaviour of a random walk at infinity, and we will sketch the theoretic foundations of what is called the *boundary theory* for random walks on groups.

The result we will prove, at least partially, states that a countable group is amenable if and only it carries a probability measure such that the associated Poisson boundary is trivial.

3.1 Harmonic functions on groups

Let G be a countable discrete group, and μ be a probability measure on G.

Definition 3.1. A function $f : G \longrightarrow \mathbb{R}$ is $\mu-$harmonic if

$$f(g) = \sum_{h \in G} f(gh)\mu(h)$$

for all $g \in G$.

In words, a $\mu-$harmonic function has a sort of mean value property: the value of f at $g \in G$ is the average, according to the distribution μ, of the values of f around g.

This definition appeals several remarks.

Remark 3.2. (i) For a function f on G, being $\mu-$harmonic depends on μ.

(ii) The sum appearing in the right hand side of Definition 3.1 is absolutely convergent in at least two cases: if μ is finitely supported, and if f is bounded. In the sequel, we will assume nothing about the support of μ, but we will restrict our purposes to the case $f \in \ell^\infty(G)$, and all sums involved exist then.

Without delay, let us provide easy examples of harmonic functions.

Example 3.3. (i) For any pair (G, μ), constant functions are always $\mu-$harmonic. Indeed, if $f(x) = a \in \mathbb{R}$ for all $g \in G$, we have

$$\sum_{h \in G} f(gh)\mu(h) = \sum_{h \in G} a\mu(h) = a \sum_{h \in G} \mu(h) = a = f(g)$$

for all $g \in G$, whence f is $\mu-$harmonic.

(ii) Let $\mu = \delta_e$. Then, if f is a function on G and $g \in G$, we obtain directly

$$\sum_{h\in G} f(gh)\delta_e(h) = f(g)$$

so any function is $\mu-$harmonic.

(iii) Let $G = \mathbb{Z}$, and $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. We claim that $f : \mathbb{Z} \longrightarrow \mathbb{R}$ is $\mu-$harmonic if and only if $f(n) = an + b$ for some $a, b \in \mathbb{R}$, and all $n \in \mathbb{Z}$. First of all, if f has this form, it is easily seen to be $\mu-$harmonic, since then

$$\frac{1}{2}f(n-1) + \frac{1}{2}f(n+1) = \frac{1}{2}(a(n-1)+b) + \frac{1}{2}(a(n+1)+b) = an+b = f(n)$$

for all $n \in \mathbb{Z}$. Conversely, suppose f is $\mu-$harmonic. Let $a := f(1) - f(0)$, and $b := f(0)$. We show that $f(n) = an + b$ by induction on $n \geq 0$. For $n = 0, 1$, the equality holds by definition of a and b. If furthermore it holds up to $n \geq 2$, then f being $\mu-$harmonic implies

$$f(n+1) = 2\left(f(n) - \frac{1}{2}f(n-1)\right) = 2(an+b) - (a(n-1)+b) = a(n+1)+b$$

showing that $f(n) = an + b$ for all $n \geq 0$. The same can be done for $n < 0$, using this time $f(n-1) = 2(f(n) - \frac{1}{2}f(n+1))$. This proves the claim.

As promised, we will now focus on *bounded* $\mu-$harmonic functions. We then introduce

$$\ell^\infty_\mu(G) := \{f \in \ell^\infty(G) \mid f \text{ is } \mu\text{-harmonic}\}.$$

This subset of $\ell^\infty(G)$ has the following properties.

Lemma 3.4. (i) $\ell^\infty_\mu(G)$ is a convex subspace of $\ell^\infty(G)$.

(ii) $\ell^\infty_\mu(G)$ is closed with respect to the topology induced by $\|\cdot\|_\infty$.

Proof. (i) To start, note that as observed above constant functions are in $\ell^\infty_\mu(G)$, which is therefore not empty. In particular, the zero function is $\mu-$harmonic. Next, suppose $f_1, f_2 \in \ell^\infty_\mu(G)$, and $\alpha, \beta \in \mathbb{R}$. Fix $g \in G$. Then we have

$$\begin{aligned}(\alpha f_1 + \beta f_2)(g) &= \alpha f_1(g) + \beta f_2(g)\\ &= \alpha\sum_{h\in G} f_1(gh)\mu(h) + \beta\sum_{h\in G} f_2(gh)\mu(h)\\ &= \sum_{h\in G}(\alpha f_1 + \beta f_2)(gh)\mu(h)\end{aligned}$$

using that f_1, f_2 are μ−harmonic. This proves that $\alpha f_1+\beta f_2$ is μ−harmonic, and $\ell^\infty_\mu(G)$ is a subspace of $\ell^\infty(G)$. The particular case where $\alpha = t, \beta = 1-t$ with $t \in [0, 1]$ proves it is convex.

(ii) Fix a sequence $(f_n)_{n\geq 0}$ in $\ell^\infty_\mu(G)$ converging to f in $\ell^\infty_\mu(G)$. Fix $\varepsilon > 0$. The hypothesis implies there is $N \in \mathbb{N}$ such that $\|f - f_N\|_\infty < \varepsilon$. Then for $g \in G$ one has

$$\begin{aligned}\left|f(g) - \sum_{h\in G} f(gh)\mu(h)\right| &= \left|f(g) - f_N(g) + f_N(g) - \sum_{h\in G} f(gh)\mu(h)\right| \\ &= \left|f(g) - f_N(g) + \sum_{h\in G} f_N(gh)\mu(h) - \sum_{h\in G} f(gh)\mu(h)\right| \\ &< \underbrace{|f(g) - f_N(g)|}_{\leq \|f-f_N\|_\infty} + \sum_{h\in G} \underbrace{|f_N(gh) - f(gh)|}_{\leq \|f-f_N\|_\infty} \mu(h) \\ &\leq 2\|f - f_N\|_\infty \\ &< 2\varepsilon.\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, this yields to $f(g) = \sum_{h\in G} f(gh)\mu(h)$ for all $g \in G$, so $f \in \ell^\infty_\mu(G)$, which is therefore closed in $\ell^\infty_\mu(G)$. □

In particular, the last point of the lemma implies that $\ell^\infty_\mu(G)$ is a Banach space, as a closed subset of a complete space is complete.

That subspace being defined, we can now formulate a *Liouville property* for G.

Definition 3.5. A pair (G, μ) is Liouville if $\ell^\infty_\mu(G) = \mathbb{R} \cdot \mathbf{1}_G$.

By Example 3.3(iii) a μ−harmonic function on $\mathbb{Z}$, for $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, is affine. In particular, if it is bounded, then it must be constant. This shows $(\mathbb{Z}, \mu)$ is Liouville.

Remark 3.6. As observed before, the class of harmonic functions on G for a given measure depends strongly on the measure, and thus so does the space $\ell^\infty_\mu(G)$. In particular, it is not true that if a pair (G, μ) is Liouville, then (G, μ') is Liouville for any other probability measure μ' on G. Consider for instance $G = \mathbb{Z}$, $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ and $\mu' = \delta_0$.

In fact, there is a deep relation between the class of μ−harmonic functions on a group G and its (non-)amenability. This relation, established by Kaimanovich and Vershik [6], can be stated as follows. It will be one of our main ingredients to understand functions on the Poisson boundary, in the next section.

Theorem 3.7. Let G be a countable discrete group.
G is amenable if and only if there exists a probability measure μ on G such that (G, μ) is Liouville.

3.2 The Poisson boundary

Let G be a countable discrete group, and let μ be a probability measure on G. Consider the random walk on G driven by μ. Out of the infinite product space $G^{\mathbb{N}^*}$ together with the product measure $\mu^{\mathbb{N}^*}$, consider the map

$$\begin{aligned} P\colon G^{\mathbb{N}^*} &\longrightarrow G^{\mathbb{N}} \\ (h_n)_{n\geq 1} &\longmapsto (w_n)_{n\geq 0} \end{aligned}$$

where $w_0 := e$ and $w_n := h_1 \ldots h_n$ for all $n \geq 1$.

On the target space, let $\mathcal{F}$ denote the product $\sigma-$algebra, generated by cylinder sets, and $\mathbb{P}$ be the probability measure given by the push-forward of $\mu^{\mathbb{N}}$ under P. Equivalently, $\mathbb{P}$ is the product measure $\bigotimes_{n=0}^{\infty} \mu^{*n}$, with the convention that $\mu^{*0} = \delta_e$ is the Dirac mass at the neutral element $e \in G$.

We usually call $(G^{\mathbb{N}}, \mathbb{P})$ the *path space*, while the initial product $(G^{\mathbb{N}^*}, \mu^{\mathbb{N}^*})$ is the *step space*, or the *space of increments*.

On the path space, we also define the *time shift*

$$\begin{aligned} T\colon G^{\mathbb{N}} &\longrightarrow G^{\mathbb{N}} \\ (w_n)_{n\geq 0} &\longmapsto (w_{n+1})_{n\geq 0}. \end{aligned}$$

This is a $\mathcal{F}-$measurable map, since for $g \in G$ and $k \geq 1$ fixed, $T^{-1}(C_g^k) = C_g^{k-1} \in \mathcal{F}$ while $T^{-1}(C_g^0) = \emptyset$. When looking at the behaviour of the random walk at infinity, the first thing to do is to identify those that coincide after a certain time. For two trajectories $x = (x_n)_{n\geq 0}, y = (y_n)_{n\geq 0}$, we set

$$x \sim y \iff \exists n, m \geq 0,\ T^n x = T^m y.$$

The first thing to check is the following.

Lemma 3.8. $\sim$ is an equivalence relation.

Proof. Reflexivity is clear by choosing for instance $n = m = 0$, and symmetry is given by the definition itself. For the transitivity, suppose $x \sim y$ and $y \sim z$ for three trajectories x, y, z. We then have integers $n, m, k, p \geq 0$ such that

$$T^n x = T^m y,\ T^k y = T^p z$$

and, without loss of generality, we may suppose $m \geq k$. It follows that

$$T^n x = T^m y = T^{m-k}(T^k y) = T^{m-k}(T^p z) = T^{m+p-k} z$$

whence $x \sim z$, concluding the proof. □

In the sequel, we will be interested in the subspace of $L^\infty(G^{\mathbb{N}}, \mathbb{P})$ consisting of T−invariant functions, namely functions $f \in L^\infty(G^{\mathbb{N}}, \mathbb{P})$ such that $f \circ T = f$. This subspace will be denoted $L^\infty(G^{\mathbb{N}}, \mathbb{P})^T$.

Let us now introduce stationary measures. Fix (B, ν) a measure space, on which G acts measurably. Recall that $\mu * \nu$ is the push-forward of $\mu \otimes \nu$ by the map $G \times B \longrightarrow B$, $(g, b) \longmapsto gb$.

Definition 3.9. The measure ν is μ−stationary if $\mu * \nu = \nu$.

Among all G−spaces carrying a μ−stationary measure, one has a universal property, distinguishing it from the others. This is the so called *Poisson boundary*, whose existence and properties will be admitted. Further comments on this construction, which relies on measurable partitions and the Rokhlin's correspondence can be found in [7], or in [2, appendix I].

Theorem 3.10. There exists a measure G−space (B_{PF}, ν_{PF}), with ν_{PF} being μ−stationary, and a measurable, G−equivariant, T−invariant map

$$\mathbf{bnd} \colon (G^{\mathbb{N}}, \mathbb{P}) \longrightarrow (B_{PF}, \nu_{PF})$$

such that $\nu_{PF} = \mathbf{bnd}_*(\mathbb{P})$, and satisfying the following universal property: for every G−space (B, λ) with a μ−stationary measure λ and a G−equivariant, T−invariant map $\varphi \colon (G^{\mathbb{N}}, \mathbb{P}) \longrightarrow (B, \lambda)$ such that $\lambda = \varphi_*(\mathbb{P})$, there exists a G−equivariant map $\psi \colon (B_{PF}, \nu_{PF}) \longrightarrow (B, \lambda)$ such that $\psi \circ \mathbf{bnd} = \varphi$.
Moreover, the induced map

$$L^\infty(B_{PF}, \nu_{PF}) \longrightarrow L^\infty(G^{\mathbb{N}}, \mathbb{P})^T$$

given by precomposition with **bnd** is an isomorphism.

3.3 The Poisson transform

In this part, we define a correspondence between the μ−harmonic functions on G and the measurable bounded functions on a G−space equipped with a μ−stationary measure. We call it the *Poisson transform*.

Definition 3.11. Let ν be a μ−stationary probability measure on a G−space B. The Poisson transform is the map

$$\begin{aligned} \hat{P}\colon L^\infty(B,\nu) &\longrightarrow \ell^\infty_\mu(G) \\ f &\longmapsto \hat{P}f \end{aligned}$$

where $(\hat{P}f)(g) := \displaystyle\int_B f(gx)\,\mathrm{d}\nu(x)$, for all $g \in G$.

This is a well-defined map. To see this, first observe that for all $f \in L^\infty(B,\nu)$ and $g \in G$, one has

$$\int_B f(gx)\,\mathrm{d}(\mu * \nu)(x) = \sum_{h\in G} \mu(h) \int_B f(ghx)\,\mathrm{d}\nu(x).$$

If $f = \mathbf{1}_A$, for a measurable $A \subset B$, the equality holds since

$$\begin{aligned} \int_B \mathbf{1}_A(gx)\,\mathrm{d}(\mu * \nu)(x) &= \int_B \mathbf{1}_{g^{-1}A}(x)\,\mathrm{d}(\mu * \nu)(x) \\ &= (\mu * \nu)(g^{-1}A) \\ &= \sum_{h\in G} \mu(h)\nu(h^{-1}g^{-1}A) \\ &= \sum_{h\in G} \mu(h)\nu((gh)^{-1}A) \\ &= \sum_{h\in G} \mu(h) \int_B \mathbf{1}_{(gh)^{-1}A}(x)\,\mathrm{d}\nu(x) \\ &= \sum_{h\in G} \mu(h) \int_B \mathbf{1}_A(ghx)\,\mathrm{d}\nu(x) \end{aligned}$$

and by linearity, it holds for every step functions (finite linear combinations of indicator functions). Since they form a dense subset of $L^\infty(B,\nu)$, the equality holds for all $f \in L^\infty(B,\nu)$.

We can thus prove the next lemma.

Lemma 3.12. The map $\hat{P}\colon L^\infty(B,\nu) \longrightarrow \ell^\infty_\mu(G)$ is well-defined, and linear.

Proof. Let $f \in L^\infty(B,\nu)$. Then, for a fixed $g \in G$, we compute that

$$\sum_{h\in G} (\hat{P}f)(gh)\mu(h) = \sum_{h\in G} \mu(h) \int_B f(ghx)\,\mathrm{d}\nu(x)$$

$$
\begin{aligned}
&= \int_B f(gx)\ \mathrm{d}(\mu * \nu)(x) \\
&= \int_B f(gx)\ \mathrm{d}\nu(x) \\
&= \hat{P}f(g)
\end{aligned}
$$

since ν is $\mu-$stationary. This proves that $\hat{P}f$ is $\mu-$harmonic. It is also bounded because f is bounded. Hence $\hat{P}$ is well-defined. Its linearity follows from the pointwise definition of a linear combination of functions and the fact that the integral is linear. □

To invert the Poisson transform $\hat{P}$, the key observation is that a bounded harmonic function on G gives rise to a bounded martingale, built as follows. Let $(\mathcal{F}_n)_{n\geq 0}$ be the canonical filtration[2] associated to the random walk. Fix $f \in \ell^\infty_\mu(G)$, and define

$$
\begin{aligned}
f_n \colon G^{\mathbb{N}} &\longrightarrow \mathbb{R} \\
\mathbf{w} = (w_n)_{n\geq 0} &\longmapsto f(w_n)
\end{aligned}
$$

for all $n \geq 0$. The sequence of random variables $(f_n)_{n\geq 0}$ forms a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$, since for a fixed trajectory $\mathbf{w} = (w_n)_{n\geq 0}$, one has

$$
\begin{aligned}
\mathbb{E}[f_{n+1}(\mathbf{w})|\mathcal{F}_n] &= \mathbb{E}[f(w_{n+1})|w_1, \ldots, w_n] \\
&= \sum_{g\in G} f(w_n g)\mu(g) \\
&= f(w_n) \\
&= f_n(\mathbf{w})
\end{aligned}
$$

using that f is $\mu-$harmonic for the third equality. Additionally, f_n is bounded for all $n \geq 0$ since f is bounded, and this bound is uniform in n. By the convergence theorem for bounded martingales (cf. Appendix B), the limit

$$\lim_{n\to\infty} f_n(\mathbf{w})$$

exists for $\mathbb{P}-$almost every path $\mathbf{w} \in G^{\mathbb{N}}$, and we thus set

$$
\begin{aligned}
F \colon G^{\mathbb{N}} &\longrightarrow \mathbb{R} \\
\mathbf{w} = (w_n)_{n\geq 0} &\longmapsto \begin{cases} \lim\limits_{n\to\infty} f_n(\mathbf{w}) & \text{if the limit exists} \\ 0 & \text{otherwise} \end{cases}.
\end{aligned}
$$

Note that F is $T-$invariant, because

$$F(T(\mathbf{w})) = F((w_{n+1})_{n\geq 0}) = \lim_{n\to\infty} f(w_{n+1}) = \lim_{n\to\infty} f(w_n) = F(\mathbf{w})$$

[2]as defined in Appendix B.

for all $\mathbf{w} \in (G^{\mathbb{N}}, \mathbb{P})$. Hence, by Theorem 3.10, there exists $\tilde{F} \in L^\infty(B_{PF}, \nu_{PF})$ such that $F = \tilde{F} \circ \mathbf{bnd}$. This way, we define a map

$$\overline{P}\colon \ell^\infty_\mu(G) \longrightarrow L^\infty(B_{PF}, \nu_{PF})$$

sending any $\mu-$harmonic function on G to a bounded measurable function on the Poisson boundary. We now claim that $\hat{P}$ and $\overline{P}$ are mutual inverses.

In order to show the last assertion, we will appeal the next basic result from integration theory.

Lemma 3.13. Let $(E, \mathcal{A}), (F, \mathcal{B})$ be two measure spaces, and let μ be a measure on E. Let $f : E \longrightarrow F$, $g\colon F \longrightarrow \mathbb{R}$ be measurable. Then it holds that

$$\int_E g \circ f(x) \, \mathrm{d}\mu(x) = \int_F g(y) \, \mathrm{d}f_*\mu(y).$$

Proof. This is another density argument. Suppose first $g = \mathbf{1}_B$ for some $B \in \mathcal{B}$. Then one gets directly that

$$\int_F \mathbf{1}_B(y) \, \mathrm{d}f_*\mu(y) = f_*\mu(B) = \mu(f^{-1}(B)) = \int_E \mathbf{1}_{f^{-1}(B)}(x) \, \mathrm{d}\mu(x) = \int_E \mathbf{1}_B(f(x)) \, \mathrm{d}\mu(x)$$

and from here the result also holds for finite linear combinations of indicator functions. Since any positive measurable function is a pointwise increasing limit of a sequence of step functions, the monotone convergence theorem gives the equality for every measurable $g\colon F \longrightarrow [0, \infty)$. This implies the claim for all functions, by decomposing $g = g^+ - g^-$ into its positive and negative parts. □

Proposition 3.14. We have $\hat{P} \circ \overline{P} = \mathrm{Id}_{\ell^\infty_\mu(G)}$ and $\overline{P} \circ \hat{P} = \mathrm{Id}_{L^\infty(B_{PF}, \nu_{PF})}$. Consequently, $\ell^\infty_\mu(G) \cong L^\infty(B_{PF}, \nu_{PF})$.

Proof. Let us prove the first identity. The second will be taken for granted, and a proof can be found in [1, section 2.4].

Let $f \in \ell^\infty_\mu(G)$, and $g \in G$. For notational convenience, we denote by F the lift of $\overline{P}f$ to the path space. We start by computing that

$$\begin{aligned}
\hat{P}(\overline{P}f)(g) &= \int_{B_{PF}} \overline{P}f(gx) \, \mathrm{d}\nu_{PF}(x) \\
&= \int_{B_{PF}} (g^{-1}\overline{P}f)(x) \, \mathrm{d}\nu_{PF}(x)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{G^{\mathbb{N}}} ((g^{-1}\overline{P}f) \circ \mathbf{bnd})(\mathbf{w}) \,\mathrm{d}\mathbb{P}(\mathbf{w}) \\
&= \int_{G^{\mathbb{N}}} g^{-1}(\overline{P}f \circ \mathbf{bnd})(\mathbf{w}) \,\mathrm{d}\mathbb{P}(\mathbf{w}) \\
&= \int_{G^{\mathbb{N}}} F(g\mathbf{w}) \,\mathrm{d}\mathbb{P}(\mathbf{w}) \\
&= \int_{G^{\mathbb{N}}} \lim_{n\to\infty} f_n(g\mathbf{w}) \,\mathrm{d}\mathbb{P}(\mathbf{w}) \\
&= \lim_{n\to\infty} \int_{G^{\mathbb{N}}} f_n(g\mathbf{w}) \,\mathrm{d}\mathbb{P}(\mathbf{w})
\end{aligned}
$$

using the definition of $\hat{P}$ for the first equality, and the definition of the action of G on a space functions, provided it acts on the domain of those functions, for the second and the fifth equality. The third equality follows from Lemma 3.13, recalling that $\nu_{PF} = \mathbf{bnd}_*\mathbb{P}$, and the fourth one from the $G-$equivariance of $\mathbf{bnd}$. The sixth is the definition of F, and the last one is due to the dominated convergence theorem, which we may apply since f_n is bounded for all $n \geq 0$ and since constant functions on a probability space are integrable. Writing a trajectory $\mathbf{w} \in G^{\mathbb{N}}$ as $(e, h_1, h_1h_2, \dots)$ with its sequence of increments $\mathbf{h} = (h_i)_{i\geq 1}$, the last integral equals

$$\int_{G^{\mathbb{N}^*}} f(gh_1\cdots h_n) \,\mathrm{d}\mu^{\mathbb{N}}(\mathbf{h}).$$

Since the integrand depends only on the first n coordinates, this reduces to integrate over the first n copies of G, each endowed with the measure μ, so

$$
\begin{aligned}
\int_{G^{\mathbb{N}^*}} f(gh_1\cdots h_n) \,\mathrm{d}\mu^{\mathbb{N}^*}(\mathbf{h}) &= \int_G \cdots \int_G f(gh_1\cdots h_n) \,\mathrm{d}\mu(h_1)\dots\mathrm{d}\mu(h_n) \\
&= \sum_{h_1\in G} \cdots \sum_{h_n\in G} f(gh_1\cdots h_n)\mu(h_n)\mu(h_{n-1})\cdots\mu(h_1).
\end{aligned}
$$

Now, using n times the fact that f is $\mu-$harmonic, we end it up with $f(g)$. This establishes that

$$\hat{P}(\overline{P}f)(g) = \lim_{n\to\infty} \int_{G^{\mathbb{N}}} f_n(g\mathbf{w}) \,\mathrm{d}\mathbb{P}(\mathbf{w}) = \lim_{n\to\infty} f(g) = f(g)$$

for all $g \in G$, so $\hat{P} \circ \overline{P}(f) = f$, and the first equality is shown. Note that this implies linearity of $\overline{P}$, as the inverse of a bijective linear map is linear as well. Thus we have an isomorphism $\ell^\infty_\mu(G) \cong L^\infty(B_{PF}, \nu_{PF})$. □

We have now all we need to state and prove the main result of this section.

Theorem 3.15. Let G be a countable discrete group.
G is amenable if and only if there exists a probability measure on G such that the associated Poisson boundary (B_{PF}, ν_{PF}) is trivial.

Proof. Suppose G is amenable. By Theorem 3.7, there exists a probability measure μ on G such that $\ell_\mu^\infty(G)$ reduces to constant functions. By Proposition 3.14, it follows that $L^\infty(B_{PF}, \nu_{PF})$ also reduces to constant functions, which implies that (B_{PF}, ν_{PF}) is trivial.

Conversely, suppose there exists μ such that the associated Poisson boundary is trivial. Again, it means $L^\infty(B_{PF}, \nu_{PF})$ consists of constant functions, and thus so does $\ell_\mu^\infty(G)$. Hence (G, μ) is Liouville, and G is amenable by Theorem 3.7. $\square$

A. Riesz representation theorem

In this first appendix, we provide a proof of the Riesz representation theorem, used to construct adjoint operators on Hilbert spaces.

We begin by introducing orthogonal complements in pre-Hilbert spaces.

Definition A.1. Let $\mathcal{H}$ be a pre-Hilbert space. Let $S \subset \mathcal{H}$ be non-empty. The set

$$S^\perp = \{x \in \mathcal{H} \mid \forall y \in S,\ \langle x, y\rangle = 0\}$$

is called the orthogonal complement of S.

Note that S need not to be a vector subspace of $\mathcal{H}$. However, its orthogonal complement is always a subspace : if $x, y \in S^\perp$ and $\lambda \in \mathbb{C}$, then

$$\langle x + \lambda y, z\rangle = \langle x, z\rangle + \lambda\langle y, z\rangle = 0$$

for all $z \in S$, so $x + \lambda y \in S^\perp$. It is furthermore closed in $\mathcal{H}$, because if $(x_n)_{n\geq 0}$ in $S^\perp$ converges to $x \in \mathcal{H}$, it implies

$$\langle x, y\rangle = \langle \lim_{n\to\infty} x_n, y\rangle = \lim_{n\to\infty} \langle x_n, y\rangle = 0$$

for all $y \in S$, and thus $x \in S^\perp$.

Additionally, note that $S \cap S^\perp \subset \{0\}$. Indeed, if $x \in S \cap S^\perp$, then $\|x\|^2 = \langle x, x\rangle = 0$, so $x = 0$.

Orthogonal complements are useful because they provide a splitting of $\mathcal{H}$ as a direct sum, under a completeness assumption.

Theorem A.2. Let $\mathcal{H}$ be a complex Hilbert space, and $G \subset \mathcal{H}$ a closed subspace. Then every $z \in \mathcal{H}$ can be written uniquely as $z = x + y$ with $x \in G$ and $y \in G^\perp$. Moreover, in this decomposition, one has

$$\|z - x\| = \inf_{w\in G} \|z - w\|,\ \ \|z - y\| = \inf_{w\in G^\perp} \|z - w\|.$$

In this case, we usually write $\mathcal{H} = G \oplus G^\perp$. In addition, the second part of the theorem says that the elements x, y appearing in the decomposition of $z \in \mathcal{H}$ minimize the distance of z to G and $G^\perp$.

Proof. For the uniqueness part, suppose $z \in \mathcal{H}$ has two decompositions $z = x_1 + y_1 = x_2 + y_2$, $x_1, x_2 \in G$, with $y_1, y_2 \in G^\perp$. Then one gets

$$x_1 - x_2 = y_1 - y_2 \in G \cap G^\perp = \{0\}$$

so $x_1 = x_2$ and $y_1 = y_2$. Let's focus now on the existence part. For brievety, denote $\delta := \inf_{w \in G} \|z - w\|$. Let $(a_n)_{n\geq 0}$ be a sequence in G such that $\lim_{n\to\infty} \|z - a_n\| = \delta$. By the parallelogram law, for all $n, m \geq 0$, we compute

$$2\|z - a_n\|^2 + 2\|z - a_m\|^2 - \|a_n - a_m\|^2 = \|2z - a_n - a_m\|^2 = 4\left\|z - \frac{a_n + a_m}{2}\right\|^2 \geq 4\delta^2$$

and the last inequality holds by definition of δ, since $\frac{a_n+a_m}{2} \in G$. Hence we get

$$\|a_n - a_m\|^2 \leq 2\|z - a_n\|^2 + 2\|z - a_m\|^2 - 4\delta^2 \xrightarrow[n,m\to\infty]{} 0$$

by the choice of the sequence $(a_n)_{n\geq 0}$. This means $(a_n)_{n\geq 0}$ is Cauchy, and since $\mathcal{H}$ is complete, it then converges to $x \in \mathcal{H}$, which must be in G since it is a closed subspace, and $x \in G$ satisfies

$$\|z - x\| = \inf_{w\in G} \|z - w\|.$$

This proves the first claim. Now we show that $z - x \in G^\perp$, and $z = x + (z - x)$ will be the desired decomposition. Note that if $w \in G$ and $\lambda \in \mathbb{C}$, then $x + \lambda w \in G$, so

$$\|z - x\|^2 \leq \|z - x - \lambda w\|^2 = \|z - x\|^2 + |\lambda|^2\|w\|^2 - 2\mathrm{Re}\lambda\langle w, z - x\rangle$$

and this leads to

$$2\mathrm{Re}\lambda\langle w, z - x\rangle \leq |\lambda|^2\|w\|^2. \tag{3}$$

If $\lambda > 0$, then dividing by λ and taking the limit $\lambda \to 0$ provides $\mathrm{Re}\langle w, z - x\rangle \leq 0$. On the other hand, replacing λ by $-i\lambda$ in (3), taking $\lambda > 0$, dividing by it and letting $\lambda \to 0$ provides $\mathrm{Im}\langle w, z-x\rangle \leq 0$. G being a subspace, these two inequalities also holds for $-w$ instead of w. Finally, we conclude that

$$\mathrm{Re}\langle w, z - x\rangle = \mathrm{Im}\langle w, z - x\rangle = 0$$

and thus $\langle w, z - x\rangle = 0$ for all $w \in G$. This proves $z - x \in G^\perp$, as claimed.
It remains to see that $\|x\| = \|z - (z - x)\| = \inf_{w\in G^\perp} \|z - w\|$. If $w \in G^\perp$, then Pythagore's theorem implies

$$\|z - w\|^2 = \|z - x + x - w\|^2 = \|z - x - w\|^2 + \|x\|^2 \geq \|x\|^2$$

so $\|z - w\|^2 \geq \|x\|^2$ for all $w \in G^\perp$, and this is an equality if and only if $w = z - x \in G^\perp$, which proves $\|z - (z - x)\| = \inf_{w\in G^\perp} \|z - w\|$. Hence we are done. □

Riesz representation theorem is now a direct consequence of the above.

Theorem A.3. Let $\mathcal{H}$ be a complex Hilbert space, and $f \in \mathcal{H}^*$.
Then there exists a unique $y \in \mathcal{H}$ such that

$$f(x) = \langle x, y\rangle$$

for all $x \in \mathcal{H}$. Moreover, $\|f\| = \|y\|$.

Proof. If $f \equiv 0$, we choose $y = 0$ and we are done. Assume now that f is a nontrivial functional. Since f is continuous, $\mathrm{Ker}(f)$ is a closed proper subspace of $\mathcal{H}$, and $\mathrm{Ker}(f)^{\perp}$ is not empty. Furthermore, it has dimension at least 1, since otherwise Theorem A.2 would imply $\mathcal{H} = \mathrm{Ker}(f)$, contradicting the fact that f is nontrivial. Let $y_0 \in \mathrm{Ker}(f)^{\perp}$, with $\|y\| = 1$. Then every $x \in \mathcal{H}$ can be written

$$x = x - \langle x, y_0\rangle y_0 + \langle x, y_0\rangle y_0$$

and since $\langle x, y_0\rangle y_0 \in \mathrm{Ker}(f)^{\perp}$, this forces $x - \langle x, y_0\rangle y_0 \in \mathrm{Ker}(f)$. Thus $f(x - \langle x, y_0\rangle y_0) = 0$ for all $x \in \mathcal{H}$, and it follows that

$$f(x) = f(\langle x, y_0\rangle y_0) = \langle x, y_0\rangle f(y_0) = \langle x, \overline{f(y_0)} y_0\rangle$$

for all $x \in \mathcal{H}$. We set then $y := \overline{f(y_0)} y_0$, and the first claim holds. The second follows, since

$$|f(x)| = |\langle x, y\rangle| \leq \|x\|\|y\|$$

for every $x \in \mathcal{H}$ by Cauchy-Schwarz, giving $\|f\| \leq \|y\|$. For $x = y$, one has

$$|f(y)| = |\langle y, y\rangle| = \|y\|^2 = \|y\|\|y\|$$

and thus $\|f\| = \|y\|$. This achieves the proof. □

B. Convergence of bounded martingales

In this appendix, we briefly introduce the theory of martingales. To do so, we assume known the basic properties of conditional expectation of random variables. These properties, and eventually their proofs, can be found for instance in [4, chapter 12].

More precisely, we show the almost sure convergence of bounded martingales, result used above to define the inverse Poisson transform.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(X_n)_{n\geq 0}$ a stochastic process, *i.e.* a sequence of $\mathbb{R}-$valued random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition B.1. A filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence $(\mathcal{F}_n)_{n\geq 0}$ of sub$-\sigma-$algebras of $\mathcal{F}$ such that

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n \subset \ldots$$

In this case, we call $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ a filtered probability space.

For instance, given $(X_n)_{n\geq 0}$ a stochastic process, the sequence $(\mathcal{F}_n)_{n\geq 0}$ defined by $\mathcal{F}_n := \sigma(X_0, \ldots, X_n)$ is a filtration on Ω, called the *canonical* filtration.

Definition B.2. A sequence $(X_n)_{n\geq 0}$ is adapted to the filtration $(\mathcal{F}_n)_{n\geq 0}$ if X_n is $\mathcal{F}_n-$measurable for all $n \geq 0$.

In the sequel, we fix an abstract filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$. Here is the main definition.

Definition B.3. A sequence $(M_n)_{n\geq 0}$ of random variables is a martingale if it satisfies the following properties :

(i) $(M_n)_{n\geq 0}$ is adapted.

(ii) $\mathbb{E}[|M_n|] < \infty$ for all $n \geq 0$.

(iii) $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ for all $n \geq 0$.

Moreover, $(M_n)_{n\geq 0}$ is a submartingale if $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geq M_n$ for all $n \geq 0$, and a supermartingale if $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$, for all $n \geq 0$.

In words, a process is a martingale if, in average, its value tomorrow, knowing the past up to now, is the same as its value today.

Example B.4. (i) Let $(X_k)_{k\geq 0}$ be a sequence of independent and identically distributed random variables, such that $\mathbb{E}[|X_0|] < \infty$. Set $S_n := \sum_{k=0}^{n} X_k$ and $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$. If $\mathbb{E}[X_0] = 0$, then $(S_n)_{n\geq 0}$ is a martingale. Indeed, as above, $(S_n)_{n\geq 0}$ is adapted to $(\mathcal{F}_n)_{n\geq 0}$. The fact that $\mathbb{E}[|X_0|] < \infty$ and linearity of expectation implies directly $\mathbb{E}[|S_n|] < \infty$ for all $n \geq 0$, and we also have

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_n + X_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_n|\mathcal{F}_n] + \mathbb{E}[X_{n+1}|\mathcal{F}_n] = S_n + \mathbb{E}[X_{n+1}] = S_n$$

using linearity of conditional expectation, the fact that S_n is $\mathcal{F}_n$−measurable, and that X_{n+1} is independent of $X_0, \dots, X_n$. The same proof shows that $(S_n)_{n\geq 0}$ is a submartingale if $\mathbb{E}[X_0] \geq 0$, and a supermatingale if $\mathbb{E}[X_0] \leq 0$.

(ii) If $(\mathcal{F}_n)_{n\geq 0}$ is a filtration and $X \in L^1$, the sequence $(M_n)_{n\geq 0}$ defined as $M_n := \mathbb{E}[X|\mathcal{F}_n]$ is a martingale. By definition, M_n is $\mathcal{F}_n$−measurable, and integrable since $\mathbb{E}[|M_n|] = \mathbb{E}[|\mathbb{E}[X|\mathcal{F}_n]|] \leq \mathbb{E}[|X|] < \infty$ and $X \in L^1$. For the third condition, one computes that

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_n] = M_n$$

for all $n \geq 0$, using the tower property for conditional expectations, valid since $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

(iii) If $(M_n)_{n\geq 0}$ and $(N_n)_{n\geq 0}$ are martingales, then $(M_n + N_n)_{n\geq 0}$ is a martingale. The first and the second properties are straightforward, and for the third one it suffices to write

$$\mathbb{E}[M_{n+1} + N_{n+1}|\mathcal{F}_n] = \mathbb{E}[M_{n+1}|\mathcal{F}_n] + \mathbb{E}[N_{n+1}|\mathcal{F}_n] = M_n + N_n.$$

Note that if $(M_n)_{n\geq 0}$ is a martingale, then $\mathbb{E}[M_n] = \mathbb{E}[\mathbb{E}[M_n|\mathcal{F}_0]] = \mathbb{E}[M_0]$ for all $n \geq 0$, *i.e.* the sequence $(\mathbb{E}[M_n])_{n\geq 0}$ is constant. Likewise, $(\mathbb{E}[M_n])_{n\geq 0}$ is bounded below (resp. above) by $\mathbb{E}[M_0]$ if $(M_n)_{n\geq 0}$ is a submartingale (resp. supermartingale).

In view of point (iii) above, we now describe how to get submartingales from martingales. This is an application of Jensen's inequality for conditional expectations.

Proposition B.5. Let $\varphi\colon \mathbb{R} \longrightarrow [0, \infty)$ be convex.
Let $(M_n)_{n\geq 0}$ be a martingale, and suppose $\mathbb{E}[|\varphi(M_n)|] < \infty$ for all $n \geq 0$.
Then $(\varphi(M_n))_{n\geq 0}$ is a submartingale.

Proof. First note that $\varphi(M_n)$ is $\mathcal{F}_n$−measurable, because M_n is and φ is measurable. By Jensen's inequality, we have

$$\mathbb{E}[\varphi(M_{n+1})|\mathcal{F}_n] \geq \varphi(\mathbb{E}[M_{n+1}|\mathcal{F}_n]) = \varphi(M_n)$$

for all $n \geq 0$, using that $(M_n)_{n\geq 0}$ itself is a martingale. This proves the claim. □

Here is another powerful mean for constructing martingales from arbitrary adapted processes. First, we need the following definition.

Definition B.6. A collection $(H_n)_{n\geq 1}$ of $\mathbb{R}-$valued random variables is a predictable process if H_n is $\mathcal{F}_{n-1}-$measurable and bounded, for all $n \geq 1$.

Note that, because of the inclusion $\mathcal{F}_{n-1} \subset \mathcal{F}_n$, a predictable process is in particular adapted.

Lemma B.7. Let $(M_n)_{n\geq 0}$ be an adapted process, and $(H_n)_{n\geq 1}$ be a predictable process.
Let $H \cdot M$ be the process defined as $(H \cdot M)_0 := 0$ and

$$(H \cdot M)_n := \sum_{k=1}^{n} H_k(M_k - M_{k-1})$$

for all $n \geq 1$. Then the following holds.

(i) If $(M_n)_{n\geq 0}$ is a martingale, then $((H \cdot M)_n)_{n\geq 0}$ is a martingale.

(ii) If $(M_n)_{n\geq 0}$ is a submartingale (resp. supermartingale), and if $H_n \geq 0$ for all $n \geq 1$, then $((H \cdot M)_n)_{n\geq 1}$ is a submartingale (resp. supermartingale).

Proof. (i) Firstly, for all $n \geq 0$, $(H \cdot M)_n$ is $\mathcal{F}_n-$measurable as a sum of $\mathcal{F}_n-$measurable random variables, so $((H \cdot M)_n)_{n\geq 0}$ is adapted. Also, M_n is integrable and H_n is bounded for all $n \geq 0$, so $(H \cdot M)_n \in L^1$ for all $n \geq 0$. For the third condition, we have

$$\begin{aligned}
\mathbb{E}[(H \cdot M)_{n+1}|\mathcal{F}_n] &= \mathbb{E}\left[\sum_{k=1}^{n+1} H_k(M_k - M_{k-1})\middle|\mathcal{F}_n\right] \\
&= \mathbb{E}[(H \cdot M)_n|\mathcal{F}_n] + \mathbb{E}[H_{n+1}(M_{n+1} - M_n)|\mathcal{F}_n] \\
&= (H \cdot M)_n + H_{n+1}(\mathbb{E}[M_{n+1}|\mathcal{F}_n] - \mathbb{E}[M_n|\mathcal{F}_n]) \\
&= (H \cdot M)_n
\end{aligned}$$

using the $\mathcal{F}_n-$measurability of $(H \cdot M)_n$ and H_{n+1} for the third equality, and the fact that $(M_n)_{n\geq 0}$ is a martingale for the fourth one. This proves (i).

(ii) We can do the same computation as above, and replace the last equality by $\geq$ (resp. $\leq$) if (M_n) is a submartingale (resp. supermartingale) and $H_n \geq 0$ for all $n \geq 0$. □

The last notion we introduce is the one of stopping times.

Definition B.8. A random variable $T : \Omega \longrightarrow \mathbb{N} \cup \{\infty\}$ is a stopping time if $\{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$.

We observe that this is equivalent to require that $\{T \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$. Indeed, if the latter is true, then $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\} \in \mathcal{F}_n$ since also $\{T \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$. Conversely, if T is as in Definition B.8, then we can write

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{T = k\}$$

and use that $\{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ for all $k = 0, \dots, n$ to obtain $\{T \leq n\} \in \mathcal{F}_n$, as wanted.

We have now all necessary tools to establish almost sure convergence of bounded martingales. This relies on the following idea.

First, fix a sequence $x = (x_n)_{n\geq 0}$ of real numbers. For all pairs of real numbers $a < b$, we define two time sequences $(S_k(x))_{k\geq 1}$, $(T_k(x))_{k\geq 1}$ as $S_1(x) := \inf\{n \geq 0 \mid x_n \leq a\}$, $T_1(x) := \inf\{n \geq S_1(x) \mid x_n \geq b\}$ and by induction

$$S_{k+1}(x) := \inf\{n \geq T_k(x) \mid x_n \leq a\},\ T_{k+1}(x) := \inf\{n \geq S_{k+1}(x) \mid x_n \geq b\}$$

with the convention $\inf \emptyset = \infty$. Then, for all $n \geq 0$, we set

$$N_{a,b}^{(n)}(x) := \sum_{k=1}^{\infty} \mathbf{1}_{\{x \in \mathbb{R}^{\mathbb{N}} \mid T_k(x) \leq n\}},\ N_{a,b}^{(\infty)}(x) := \sum_{k=1}^{\infty} \mathbf{1}_{\{x \in \mathbb{R}^{\mathbb{N}} \mid T_k(x) < \infty\}}.$$

$N_{a,b}^{(n)}(x)$ represents the number of climbs beyond the level b by the sequence $(x_n)_{n\geq 0}$ before the rank n, and $N_{a,b}^{(\infty)}(x)$ is the total number of such climbs.

The next lemma from analysis is the key tool we need.

Lemma B.9. Let $(x_n)_{n\geq 0} \in \mathbb{R}^{\mathbb{N}}$.
If $N_{a,b}^{(\infty)}(x) < \infty$ for all $a, b \in \mathbb{Q}$, then $(x_n)_{n\geq 0}$ converges in $\mathbb{R} \cup \{\pm\infty\}$.

Proof. We will prove the contrapositive. Suppose $(x_n)_{n\geq 0}$ does not converge in $\mathbb{R} \cup \{\pm\infty\}$. We distinguish two cases. First, assume $(x_n)_{n\geq 0}$ is bounded. Then it has an accumulation by Bolzano-Weierstrass, and since it does not converge, it has at least two distinct accumulation points, say $\ell \neq \ell'$. Without restriction, we may assume $\ell < \ell'$. Set then $\varepsilon := \frac{\ell'-\ell}{3}$. By density of $\mathbb{Q}$ in $\mathbb{R}$, we can choose a rational $a \in (\ell, \ell+\varepsilon)$ and a rational $b \in (\ell'-\varepsilon, \ell')$. Then ℓ and ℓ' being accumulation points of the sequence $(x_n)_{n\geq 0}$ implies that $N_{a,b}^{(\infty)}(x) = \infty$, proving the claim in this case.
Suppose now that $(x_n)_{n\geq 0}$ is not bounded. Here again, we distinguish cases : either the sequence is bounded below and not above, or bounded above and not below, or neither bounded below nor above. Let us suppose $(x_n)_{n\geq 0}$ is bounded below and not bounded above. Let then

$$C := \sup\{m \in \mathbb{R} \mid \forall n \in \mathbb{N},\ x_n > m\}$$

be the biggest real number that bounds $(x_n)_{n\geq 0}$ from below. Up to shifting the sequence by a constant, we can assume that $C = 0$. Now, for any fixed rational number $q > 0$, there are infinitely many $n \in \mathbb{N}$ with $0 < x_n < q$. Indeed, if there were only finitely many terms $x_{n_1}, \dots, x_{n_k}$ of the sequence such that $0 < x_{n_i} < q$, $i = 1, \dots, k$, we could set

$$s := \min_{1 \le i \le k} x_{n_k}$$

and the whole sequence would be bounded below by $s > 0 = C$, contradicting the choice of C. Thus we can choose an arbitrary $q \in \mathbb{Q}, q > 0$ and there is infinitely many terms of the sequence lying in $(0, q)$. Since $(x_n)_{n\ge0}$ is not bounded above, there is also infinitely many terms of the sequence above $q' := q + 1 \in \mathbb{Q}$. Thus $N^{(\infty)}_{q,q'}(x) = \infty$.
The case where $(x_n)_{n\ge0}$ is bounded above and not below is handled in a similar manner. Lastly, if the sequence is neither bounded below nor above, for all $a < b \in \mathbb{Q}$, there are infinitely many terms of the sequence below a and infinitely many terms above b. Hence $N^{(\infty)}_{a,b}(x) = \infty$ as well, and we are done. □

Let us clarify the notations we will use below. For $(M_n)_{n\ge0}$ a martingale and an integer $k \ge 0$, we will denote $N^{(k)}_{a,b}$ the random variable defined by

$$N^{(k)}_{a,b}(\omega) := N^{(k)}_{a,b}((M_n(\omega))_{n\ge0})$$

for all $\omega \in \Omega$. Also, $N^{(\infty)}_{a,b}$ is defined as $N^{(\infty)}_{a,b}(\omega) := N^{(\infty)}_{a,b}((M_n(\omega))_{n\ge0})$ for all $\omega \in \Omega$.

Likewise, for each $k \ge 1$, we define two random variables $S_k, T_k \colon \Omega \longrightarrow \mathbb{N} \cup \{\pm\infty\}$ by

$$S_k(\omega) := S_k((M_n(\omega))_{n\ge0}),\ T_k(\omega) := T_k((M_n(\omega))_{n\ge0})$$

for any $\omega \in \Omega$. It turns out S_k and T_k are both stopping times. Indeed, for instance one can write

$$\{T_k \le n\} = \bigcup_{0 \le m_1 < n_1 < \cdots < m_k < n_k \le n} \{X_{m_1} \le a,\ X_{n_1} \ge b, \dots,\ X_{m_k} \le a,\ X_{n_k} \ge b\}$$

which implies that $\{T_k \le n\} \in \mathcal{F}_n$ for all $n \ge 0$.

The convergence theorem will essentially follow from the next proposition. Recall that for two integers $n, m \in \mathbb{N}$ we denote $m \wedge n := \min(n, m)$, and for a real number $x \in \mathbb{R}$, we denote $(x)_+$ its positive part defined as $(x)_+ := x$ if $x > 0$, and $(x)_+ = 0$ otherwise.

Proposition B.10. Let $(M_n)_{n\ge0}$ be a martingale.
For all $a, b \in \mathbb{R}$, for all $n \ge 0$, we have

$$(b - a)\mathbb{E}[N^{(n)}_{a,b}] \le \mathbb{E}[(M_n - a)_+] - \mathbb{E}[(M_0 - a)_+].$$

Proof. Let $H_n := \sum_{k=1}^{\infty} \mathbf{1}_{\{S_k < n \leq T_k\}}$. Observe that

$$\{S_k < n\} = \{S_k \leq n-1\} \in \mathcal{F}_{n-1}, \ \{n \leq T_k\} = \{T_k \leq n-1\}^c \in \mathcal{F}_{n-1}.$$

and it follows that $\mathbf{1}_{\{S_k < n \leq T_k\}}$ is $\mathcal{F}_{n-1}$–measurable, and thus so is H_n. Hence $(H_n)_{n\geq 1}$ is a predictable process. Finally, let $\tilde{M}_n := (M_n - a)_+$, which is a submartingale by Proposition B.5. For $n \geq 1$, we compute

$$\begin{aligned}
(H \cdot \tilde{M})_n &= \sum_{k=1}^{n} H_k(\tilde{M}_k - \tilde{M}_{k-1}) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{\infty} \mathbf{1}_{\{S_k < i \leq T_k\}}(\tilde{M}_i - \tilde{M}_{i-1}) \\
&= \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} \mathbf{1}_{\{S_k \wedge n < i \leq T_k \wedge n\}}(\tilde{M}_i - \tilde{M}_{i-1}) \\
&= \sum_{k=1}^{\infty} \sum_{i=(S_k\wedge n)+1}^{T_k \wedge n} (\tilde{M}_i - \tilde{M}_{i-1}) \\
&= \sum_{k=1}^{N_{a,b}^{(n)}} \tilde{M}_{T_k \wedge n} - \tilde{M}_{S_k \wedge n} \\
&\geq (b-a) N_{a,b}^{(n)}.
\end{aligned}$$

We also have $\tilde{M}_n - \tilde{M}_0 = (1 \cdot \tilde{M})_n = (H \cdot \tilde{M})_n + ((1-H) \cdot \tilde{M})_n$. Since $1 - H \geq 0$ and since $(\tilde{M}_n)_{n\geq 0}$ is a submartingale, Lemma B.7 assures $((1-H) \cdot \tilde{M})_{n\geq 0}$ still is a submartingale, and in particular

$$\mathbb{E}[((1-H) \cdot \tilde{M})_n] \geq \mathbb{E}[((1-H) \cdot \tilde{M})_0] = 0$$

for all $n \geq 0$. We then conclude

$$\mathbb{E}[\tilde{M}_n - \tilde{M}_0] = \mathbb{E}[(H \cdot \tilde{M})_n] + \mathbb{E}[((1-H) \cdot \tilde{M})_n] \geq (b-a)\mathbb{E}[N_{a,b}^{(n)}]$$

for all $n \geq 0$, as announced. □

This proposition implies that bounded martingales have almost surely a finite number of climbs in a closed bounded interval.

Corollary B.11. Let $(M_n)_{n\geq 0}$ be a martingale. Suppose there is $C > 0$ such that $\mathbb{E}[|M_n|] < C$, for all $n \geq 0$. Then, for all $a, b \in \mathbb{R}$, $\mathbb{E}[N_{a,b}^{(\infty)}] < \infty$.
In particular, $N_{a,b}^{(\infty)} < \infty$ almost surely.

Proof. Under the hypothesis of boundedness, one gets

$$\mathbb{E}[(M_n - a)_+] \leq \mathbb{E}[|M_n - a|] \leq |a| + C$$

and it follows from Proposition B.10 that

$$\mathbb{E}[N_{a,b}^{(n)}] \leq \frac{\mathbb{E}[(M_n - a)_+] - \mathbb{E}[(M_0 - a)_+]}{b - a} \leq \frac{|a| + C}{b - a}$$

for all $n \in \mathbb{N}$, $a, b \in \mathbb{R}$. The dominated convergence theorem then implies

$$\mathbb{E}[N_{a,b}^{(\infty)}] = \lim_{n \to \infty} \mathbb{E}[N_{a,b}^{(n)}] \leq \frac{|a| + C}{b - a}$$

and in particular $N_{a,b}^{(\infty)}$ is finite almost surely. This concludes the proof. □

Here is then the convergence theorem.

Corollary B.12. Let $(M_n)_{n \geq 0}$ be a martingale. Suppose there is $C > 0$ such that $\mathbb{E}[|M_n|] < C$ for all $n \geq 0$. Then $(M_n)_{n \geq 0}$ converges almost surely.

Proof. By Corollary B.11, $N_{a,b}^{(\infty)} < \infty$ on a full measure set $\Omega_{a,b}$, for all $a, b \in \mathbb{Q}$. Then, by Lemma B.9, $(M_n)_{n \geq 0}$ converges on $\bigcap_{a,b \in \mathbb{Q}} \Omega_{a,b}$, which also has full measure. Thus $(M_n)_{n \geq 0}$ converges almost surely. □

References

[1] Yuri Bader and Yehuda Shalom, *Factor and normal subgroup theorems for lattices in products of groups*, Invent. Math. 163 (2006), no. 2, 415–454.

[2] I.P. Cornfeld, S.V. Fomin, and Ya.G. Sinai, *Ergodic theory*, Springer, New York, 1982.

[3] Erling Følner, *On groups with full Banach mean value*, Mathematica Scandinavica, 3, 243–254, 1955.

[4] Jean-François Le Gall, *Intégration, probabilités et processus aléatoires*, Ecole normale supérieure de Paris, 2006.

[5] Pierre de la Harpe, *Topics in geometric group theory*, Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2000.

[6] Vadim Kaimanovich and Anatoly Vershik, *Random walks on discrete groups: boundary and entropy*, The annals of probability (1983), 457–490.

[7] Vadim Kaimanovich, *Poisson boundaries for discrete groups*, Unpublished survey.

[8] Harry Kesten, *Symmetric random walks on groups*, Transactions of the American Mathematical Society, Vol. 92, No. 2. (Aug., 1959), pp. 336-354.

[9] George Polya, *Uber eine aufgabe betreffend die irrfahrt im strassennetz*, Math. Ann., 84:149–160, 1921.

[10] Hans Reiter, *Classical harmonic analysis and locally compact groups*, Clarendon Press, Oxford, 1968.

[11] Reinhold Remmert, *Theory of complex functions*, Springer Science and Business Media, 1991.

[12] Jean-Pierre Serre, *Arbres, amalgames, SL_2*, Astérisque, tome 46, Société mathématique de France, 1977.

[13] Wolfgang Woess, *Random walks on infinite graphs and groups*, Cambridge Tracts in Math., vol. 138, Cambridge University Press, Cambridge, 2000.