Random walks on the mapping class group

Candidato: Federico Vigolo
Relatore: Dott. Roberto Frigerio
Introduction

This work is mainly concerned with discrete random walks on graphs and an interesting application of random walks in the specific setting of mapping class groups. As its name suggest, a random walk on a graph is a sequence of nodes obtained starting at a certain base point and then moving randomly around the graph. More precisely, at each step of a random walk one choose where to move next following a certain transition probability that depends only on the specific node you are sitting on (i.e. transition probabilities depend on space but not on time). The simplest example is when at each step one moves to any of the adjacent nodes with equally distributed probability: such a random walk is called the simple random walk on the graph. Random walks are a very interesting object. In fact, apart from the numerous applications they have in mathematics, they also represent a fundamental tool in many mathematical models for various subjects like physics, computer science, economics and biology.

When considering a random walk on a graph, it is clear that the structure of the graph greatly affects the random walk itself. Our main concern will be to understand how the geometry of the underlying graph affects the asymptotic behaviour of random walks. Any information in this sense is extremely useful because it can be used in both directions: to predict how a random walk will evolve knowing the geometry of the graph or, conversely, to understand the geometry of a graph knowing how evolve random walks on this graph.

The strongest results will be obtained when dealing with some extremely regular graphs, that are graphs coming from groups. Specifically, with every finitely generated group one can associate a graph called the Cayley graph of the group. Actually, Cayley graphs are not unique in that they depend on the choice of a generating set. Still, the asymptotic structure of Cayley graphs does not depend on any choice and this is what we are going to study.

After having developed some machinery, we will focus our study on a very specific setting, that is mapping class groups of surfaces. The mapping class group of a surface is the group of homeomorphisms of this surface considered up to isotopy. Apart from its intrinsic interest and the information that it provides about the surface, the mapping class group is also widely studied
as it is a fundamental tool to construct explicit examples of 3-manifolds. Indeed, much of the modern theory of 3-manifolds has been developed after Thurston described how it is possible to define geometric structures on a 3-manifold in terms of elements of the mapping class group of a surface.

It turns out that the elements of the mapping class group can be of three types. That is, they may have finite order (periodic elements), they may fix a set of disjoint curves (reducible elements), or they may be pseudo-Anosov. In many senses, the latter is definitely the most interesting and complicated case. It is quite difficult to construct explicit examples of pseudo-Anosov elements. Still, they are by far the most common type. Our final objective is to make apparent the predominance of pseudo-Anosov elements showing that a random walk in the mapping class group will almost surely end up walking among them.

We will now give a brief overview of the contents of each chapter of this thesis. In the first chapter we introduce Cayley graphs of groups and the fundamental tool used to describe their structure at infinity, that is the notion of quasi-isometry between metric spaces. Then, we review some well-known properties that are invariant under quasi-isometry, such as rates of growth of groups or amenability. The latter is a very important concept first defined by Von Neumann in response to the celebrated Banach-Tarski paradox. We will spend some time proving the equivalence between various definitions of amenability and we will try to draw parallels between growth and amenability when possible.

We conclude the chapter introducing δ-hyperbolicity. This is an interesting generalization of the notion of negative curvature to general metric spaces. We will mainly need this concept in order to define the boundary at infinity of an hyperbolic metric space.

In the second chapter we develop the techniques of surface theory that we need to study random walks on the mapping class group. Specifically, we carefully define all the objects we will deal with and then review known facts about topological and geometric structures of surfaces and their homeomorphisms. The first objective is to state the Nielsen-Thurston classification of elements of the mapping class group into periodic, reducible and pseudo-Anosov. Then we explore some related results, often focusing on properties of pseudo-Anosov elements. While outlining this highly developed part of surface theory, we will generally avoid giving proofs. We will try to give appropriate references when needed.

In contrast, all the results of Section 2.2 are proven in detail. In this section we introduce the curve complexes and prove that they are δ-hyperbolic. Further, we make apparent some strong relations between curve complexes and mapping class groups. One of the reasons why we decided to provide complete proofs is that there are some relatively new techniques to deal with
curves complexes and they are both interesting and elementary. Also in Section 2.4 many facts are proved thoroughly. Thus, we will take some more time to speak about orbifolds, finite subgroups of the mapping class group and their centralizers.

In the third chapter we will deal with random walks. We begin giving the definitions we are using throughout the chapter and we provide some examples. Then we restrict our attention to reversible random walks and show how closely these are related to amenability by proving Kesten’s criterion for amenability (the statement of this criterion is more or less that if a random walk on a group tends to walk away decidedly then the group cannot be amenable). Later, we develop the theory of boundaries for random walks and we show once more how this is related to asymptotic properties like growth and amenability of the underlying space.

Finally, in the last chapter we put at use the machinery developed before. We start by giving some simple examples to show how the boundaries of random walks can sometimes be identified with geometric boundaries. This serves as a motivation for the techniques used in the remainder of the chapter. All the remaining pages are devoted to the study of random walks on mapping class groups of surfaces. The main result we prove is that the $n$-th step of a random walk on the mapping class group of a surface will almost surely be a pseudo-Anosov element as $n$ goes to infinity. In order to do that, we follow the proof given by J. Maher in [Mah11].

Our attention in this work is mainly devoted to geometric, rather than probabilistic, arguments. In particular, the sections regarding the purely probabilistic aspects of random walks are much down-to-earth and, if possible, we try to avoid using sophisticated theorems there. By this reason, we will sometimes prefer to produce lengthy but elementary proofs of facts that could easily follow from more profound theories.
## Contents

**Introduction** iii

1 Geometric group theory 1

1.1 Basic concepts ................................................. 1
  1.1.1 Cayley graphs ........................................... 1
  1.1.2 Quasi-isometries ....................................... 3
  1.1.3 Growth of groups ....................................... 6

1.2 Amenability .................................................... 7
  1.2.1 Amenability for graphs ................................. 8
  1.2.2 Amenability for groups ................................. 11

1.3 Hyperbolicity of metric spaces ............................... 16
  1.3.1 Thin triangles ........................................... 16
  1.3.2 The Gromov boundary ................................... 20
  1.3.3 Horoball neighbourhoods ................................ 22
  1.3.4 A criterion for hyperbolicity .......................... 25

2 Surface theory 29

2.1 Basic facts and definitions ................................... 29
  2.1.1 Hyperbolic surfaces .................................... 29
  2.1.2 Homotopies and isotopies ............................... 34
  2.1.3 Curves and geodesics .................................... 35
  2.1.4 Mapping class groups ................................... 36

2.2 Curve complexes ................................................ 41
  2.2.1 Introduction to curve complexes ....................... 41
  2.2.2 Hyperbolicity of curves complexes ..................... 44
  2.2.3 The curve complex and the mapping class group ....... 53

2.3 Classification of homeomorphisms ............................. 56
  2.3.1 Foliations ................................................ 56
  2.3.2 Pseudo-Anosov homeomorphisms ......................... 60
  2.3.3 The Nielsen-Thurston classification ................... 63

2.4 Subgroups of the mapping class group ........................ 65
  2.4.1 Orbifolds ................................................ 65
  2.4.2 Two-dimensional orbifolds .............................. 68
Chapter 1

Geometric group theory

The objects of study of geometric group theory are the relations between algebraic and geometric properties of groups. Being groups algebraic structures, it is clear what we mean by algebraic properties. On the contrary from the definitions it is not at all clear what a geometric property of a group could be. Thus, the first objective of this chapter is to show how it is possible to put interesting metrics on finitely generated groups. Then we proceed studying possible consequences of the existence of actions of groups in metric spaces and we develop part of the theory of Gromov hyperbolic metric spaces.

1.1 Basic concepts

In this section we will introduce some basic notions of geometric group theory. Namely, we will firstly define Cayley graphs and quasi-isometries and then we will state the fundamental Milnor-Švarc Theorem and illustrate some of its important consequences. To conclude, we will define one of the simplest invariant under quasi-isometries of graphs, that is the rate of growth of metric balls.

All of this material is standard and can be found in any introductory text book. See for example [DK13].

1.1.1 Cayley graphs

A (unoriented) graph $G$ is given by a finite or countable set of vertices (or nodes) $V(G)$ and a set of (unoriented) pairs of nodes called (unoriented) edges. Notice that one can think of graphs as combinatorial simplicial complexes of dimension one letting the set of nodes be the 0-skeleton and the set of edges the 1-skeleton. In particular, graphs can naturally be seen as topological spaces taking the associate geometric simplicial complex, where the geometric simplicial complex is the topological space obtained from a combinatorial
simplicial complex identifying each \( n \)-simplex with the standard \( n \)-simplex of \( \mathbb{R}^{n+1} \) (the standard \( n \)-simplex of \( \mathbb{R}^{n+1} \) is the convex closure of the \( n + 1 \) vectors of the standard basis).

**Definition 1.1.1.** A graph is *locally finite* if every node is contained only in finitely many edges. The *degree* of a node of a locally finite graph is the number of edges that contain that node. A graph is of *bounded degree* if every node has degree bounded by the same constant \( K \).

**Remark 1.1.2.** One could wish to define graphs in such a way that multiple edges between the same pair of nodes are allowed. In that case the topological analogous is that of \( \Delta \)-complexes of dimension one (or CW-complexes, since in such a low dimension these topological complexes coincide).

Oriented graphs are very similar, the only difference is that the edges are oriented pairs of nodes \( e = (a, b) \) and the topological analogous is an (oriented) \( \Delta \)-complex. We will not be dealing with oriented graphs until Section 3.2.4 and even there their usage will be extremely basic.

We will always deal with connected graphs, *i.e.* graphs whose associated topological space is connected. That is because the sets of nodes of these graphs are naturally endowed with a distance function. Indeed, we define the distance between two nodes \( x, y \in V(G) \) as the length of the shortest path joining them

\[
d(x, y) := \min \left\{ n \mid \exists a_0, \ldots, a_n, \{a_i, a_{i+1}\} \in E(G) \forall i \right\}.
\]

Actually, since each edge of \( G \) is identified with the interval \([0, 1] \subset \mathbb{R}\), it is easy (but tedious) to define the distance between two generic points of the topological complex associated to \( G \) as the length of the shortest curve joining them.

Here it comes the main reason why we are interested in graphs. Let \( \Gamma \) be a finitely generated group and \( S \) a finite generating set. Then we can consider the graph \( C_S(\Gamma) \) whose vertices are the elements of \( \Gamma \) and two nodes \( h, g \in \Gamma \) form an edge if and only if \( h = gs^{\pm 1} \) with \( s \in S \).

**Definition 1.1.3.** The graph \( C_S(\Gamma) \) is the *Cayley graph* associated the group \( \Gamma \) with generating set \( S \).

We prefer to consider only finite generating sets because the graph obtained this way is of bounded degree with bound at most \( 2|S| \) and hence the geometry of \( C_S(\Gamma) \) will be much more significant (see *e.g.* Remark 1.1.9).

**Remark 1.1.4.** Notice that the distance between two elements \( g, h \in \gamma \) in the Cayley graph \( C_S(\Gamma) \) is given by

\[
d(g, h) = \min \{ n \mid h = gs_1^{\pm 1}s_2^{\pm 1}\cdots s_n^{\pm 1}, \ s_i \in S \}.
\]
That is, \( d(g, h) \) is the length of the shortest word \( w \) with letters in \( S \cup S^{-1} \) such that \( h = gw \). Later on, we will often consider the group \( \Gamma \) itself as a metric space and we will refer to this distance function as the word distance.

Recall that the group \( \Gamma \) acts on itself by left and right multiplication. That is, any element \( g \in \Gamma \) acts on \( \Gamma \) via the left translation \( L_g \) or the right translation \( R_g \) given by:

\[
L_g(h) = gh \\
R_g(h) = hg.
\]

From the above expression of the word distance, it is clear that for any choice of generating set \( S \) the action by left multiplication is an action by isometries on \( \Gamma \). Actually, the left multiplication gives an action by isometries of the whole Cayley graph \( C_S(\Gamma) \) because it sends edges on edges. As a corollary we obtain that the Cayley graph of a group is transitive. That is, for every pair of nodes \( g, g' \in C_S(\Gamma) \) there exists an isometry of \( C_S(\Gamma) \) sending \( g \) to \( g' \) (for example, \( L_{g'g^{-1}} \) will do).

**Remark 1.1.5.** On the contrary, in general the right multiplication does not induce an action on the Cayley graph and it is not an action by isometry on \( \Gamma \). Still, the right multiplication by \( g \) has the nice property that it sends every element \( h \in \Gamma \) to an element at distance \( d_w(g, e) \) from \( h \).

### 1.1.2 Quasi-isometries

The definition of the Cayley graph of a group \( \Gamma \) depends heavily on the choice of a generating set. Still, the asymptotic properties of the metric space \( C_S(\Gamma) \) are independent of any choice. The right concept to deal with the structure at infinity of a metric space is that of quasi-isometry. In what follows the term coarse will mean ‘up to constants’.

**Definition 1.1.6.** Given two metric spaces \((X, d_X)\) and \((Y, d_Y)\) and two positive constants \( L \) and \( A \), an \((L, A)\)-quasi-isometry from \( X \) to \( Y \) is a (not necessarily continuous) map \( f : X \to Y \) such that

(i) \( f \) is coarsely bi-Lipschitz. That is, for every choice of \( x \) and \( x' \) in \( X \) we have

\[
\frac{1}{L} d_X(x, x') - A \leq d_Y(f(x), f(x')) \leq L d_X(x, x') + A.
\]

(ii) \( f \) is coarsely surjective. That is, for every element \( y \in Y \) there exists an element \( x \in X \) such that \( d_Y(f(x), y) \leq A \).

Two metric spaces are quasi-isometric if there exists a quasi-isometry between them.
It is easy to verify that being quasi-isometric is an equivalence relation of metric spaces. An equivalent definition of quasi-isometry that will come handy later on is that a coarsely Lipschitz map \( f: X \to Y \) is a quasi-isometry if and only if there exists a coarsely Lipschitz map \( g: Y \to X \) such that \( f \circ g \) and \( g \circ f \) are coarsely equal to the identity. That is, there is a constant \( A \) such that for every \( x \in X \) and \( y \in Y 
abla
abla
abla
abla
abla
\)
\[ d_X(g \circ f(x), x) \leq A \quad \text{and} \quad d_Y(f \circ g(y), y) \leq A. \]

In this case we say that \( g \) is a quasi-inverse for \( f \) and vice versa.

**Remark 1.1.7.** Sometimes it will be convenient not to think of quasi-isometries as well-defined function, but only as coarsely well-defined. That is, to define a quasi-isometry \( f \) we might prefer not to specify a precise value for \( f(x) \), rather to define it only up to some uncertainty \( A \), i.e. for every point \( x \) we only know that its image lies in a certain set with diameter bounded by \( A \) (where \( A \) is clearly a constant independent on the specific point \( x \)).

Notice that if a set \( Y \subseteq X \) is an \( A \)-dense subset of a metric space \( X \) (that is, for every \( x \in X \) there exists \( y \in Y \) with \( d(x, y) \leq A \)) then the inclusion \( Y \hookrightarrow X \) is a \((1, A)\)-quasi-isometry. It follows that the set of vertices \( V(G) \) of a graph is quasi-isometric to the whole graph \( G \). In particular, a group \( \Gamma \) with the word metric induced by a generating set \( S \) is quasi-isometric to the Cayley graph \( C_S(\Gamma) \).

The following proposition is as simple as essential:

**Proposition 1.1.8.** Let \( \Gamma \) be a finitely generated group. If \( S \) and \( S' \) are two finite generating sets then the Cayley graphs \( C_S(\Gamma) \) and \( C_{S'}(\Gamma) \) are naturally quasi-isometric.

**Proof.** Let \( d_S \) and \( d_{S'} \) denote the word metrics induced by \( S \) and \( S' \). It is enough to prove that the identity map \( (\Gamma, d_S) \xrightarrow{id} (\Gamma, d_{S'}) \) is a quasi-isometry. Since the identity is clearly a coarse inverse for itself, we only need to show that it is coarsely Lipschitz (in both directions).

Since \( S \) is finite, there exists a constant \( L \) such that every element of \( S \) can be written as a word of elements of \( S' \) and \( (S')^{-1} \) of length at most \( L \). It clearly follows that for every \( g, h \in \Gamma \)
\[ d_S(g, h) \leq L d_S(g, h). \]

The same arguments prove also the converse inequality. \( \square \)

**Remark 1.1.9.** For the proof of Proposition 1.1.8 the finiteness of the generating sets is essential.

As a corollary we obtain that the Cayley graph of a finitely generated group is well-defined up to quasi-isometry. Moreover, if \( F: \Gamma \to \Gamma' \) is an isomorphism between finitely generated groups, then it is a quasi-isometry.
1.1. BASIC CONCEPTS

with respect to the word metrics (it is not important the specific word metric we are using because all of them are quasi-isometric).

Recall that a geodesic on a metric space $X$ is a map $\gamma: [a, b] \to X$ where $[a, b] \subset \mathbb{R}$ is a connected interval and for every $x$ and $y$ in $[a, b]$ the distance $d_X(\gamma(x), \gamma(y))$ is equal to $|x - y|$. A metric space $X$ is geodesic if every pair of points $x, y \in X$ is linked by a geodesic, i.e. there exists $\gamma$ with $\gamma(a) = x$ and $\gamma(b) = y$. Notice that connected graphs are geodesic metric spaces.

A metric space $X$ is proper if the closed balls $B_r(x)$ are compact for every point $x \in X$ and radius $r \geq 0$. An action $\Gamma \curvearrowright X$ is properly discontinuous if for every compact set $K \subseteq X$ there are only finitely many $g \in \Gamma$ such that $K \cap gK \neq \emptyset$.

**Theorem 1.1.10** (Milnor–Švarc). Let $X$ be a proper geodesic metric space and $\Gamma$ a group acting on $X$ by isometries. If the action $\Gamma \curvearrowright X$ is properly discontinuous and the quotient $X/\Gamma$ is compact, then $\Gamma$ is finitely generated and for every $x \in X$ the map sending an element $g \in \Gamma$ to $g(x)$ is a quasi-isometry between $\Gamma$ and $X$.

An immediate but interesting corollary is obtained applying Milnor–Švarc Theorem to the universal cover of Riemannian manifolds.

**Corollary 1.1.11.** If $M$ is a compact Riemannian manifold and $\tilde{M}$ is its universal Riemannian cover, then the fundamental group $\pi_1(M)$ is finitely generated and quasi-isometric to $\tilde{M}$.

The following is a useful corollary and we are going to need it later on.

**Corollary 1.1.12.** If $\Gamma$ is a finitely generated group and $H < \Gamma$ is a subgroup of finite index, then $H$ is finitely generated and it is quasi-isometric to $\Gamma$.

**Proof.** Notice that the Cayley graph $C_S(\Gamma)$ is a geodesic metric space and it is also proper because the generating set $S$ is finite. Being a subgroup of $\Gamma$, the left multiplication of $H$ gives an action by isometries on $C_S(\Gamma)$ and the quotient $H \setminus \Gamma$ is finite by hypothesis. It follows that $H \setminus C_S(\Gamma)$ is a finite graph and hence it is compact, thus Theorem 1.1.10 applies.

We conclude this subsection defining the natural analogous of embeddings in the setting of quasi-isometries.

**Definition 1.1.13.** Given two metric spaces $(X, d_X)$ and $(Y, d_Y)$, a map $f: X \to Y$ is a $(L, A)$-quasi-isometric embedding if it satisfies

$$\frac{1}{L}d_X(x, x') - A \leq d_Y(f(x), f(x')) \leq Ld_X(x, x') + A,$$

for every $x$ and $x'$ in $X$. 
Also in this case it may be convenient to think of quasi-isometric embeddings as coarsely well defined maps. Notice that $f: X \to Y$ is a quasi-isometric embedding if and only if it is a quasi-isometry between $X$ and its (coarse) image $f(X)$. In particular, a coarsely Lipschitz map $f: X \to Y$ is a quasi-isometric embedding if and only if there exists a map $g: Y \to X$ whose restriction to $f(X)$ is coarsely Lipschitz and both $g|_{f(X)} \circ f$ and $f \circ g|_{f(X)}$ are coarsely equivalent to the identity.

### 1.1.3 Growth of groups

One of the simplest invariants under quasi-isometry is the rate of growth of the volume of metric balls (when such a volume is defined). Let $\mathbb{R}_+ = [0, \infty)$ denote the set of positive real numbers. We can define a (partial) ordering between increasing functions $f, g: \mathbb{R}_+ \to \mathbb{R}_+$ setting $f \preceq g$ if there exist positive constants $\alpha, C_1, C_2$ such that

$$f(x) \leq C_1 g(\alpha x) + C_2$$

for every $x \in \mathbb{R}_+$. We say that two such functions have the same growth if $f \preceq g$ and $g \preceq f$. In this case we write $f \asymp g$.

**Remark 1.1.14.** Since we are dealing with increasing functions and we are interested only in their asymptotic behaviour, it makes perfectly sense to compare a function $f: \mathbb{R}_+ \to \mathbb{R}_+$ with a function defined only on the natural numbers $g: \mathbb{N} \to \mathbb{R}_+$.

Let $G$ be a connected graph and $x \in V(G)$ a base point. Then one can look at the growth of the function that to each radius $r$ assign the number of nodes in the ball of radius $r$ centred at $x$. Since it is more natural to look at graphs as discrete spaces, we will consider only integer radii and for convenience we use closed metric balls. Thus we define $G_{x,G}: \mathbb{N} \to \mathbb{N}$ as

$$G_{x,G}(n) = |B_n(x)| = \# \{ y \in V(G) \mid d(x, y) \leq n \}.$$

It is easy to see that if two connected graphs of bounded degree $G$ and $G'$ are quasi-isometric then the volume of their metric balls has the same growth (the choice of the base points is uninfluential):

$$G_{x,G} \asymp G_{x',G'}.$$

Notice that one can actually see that if $G$ is quasi-isometrically embedded in $G'$ then $G_{x,G} \preceq G_{x',G'}$.

As a corollary, we have that it is well-defined the *growth* of a finitely generated group $\Gamma$. That is, the growth of $\Gamma$ is the equivalence class of the growth function $G_\Gamma$ (in the notation $G_\Gamma$ we completely dropped the dependence on the base point because Cayley graphs are transitive).
1.2. AMENABILITY

**Definition 1.1.15.** A finitely generated group $\Gamma$ has *exponential growth* if $\mathfrak{G}_\Gamma \gtrsim e^x$ and it has *polynomial growth* if $\mathfrak{G}_\Gamma \lesssim x^n$ for some $n \in \mathbb{N}$. If its growth is neither exponential nor polynomial then $\Gamma$ has *intermediate growth*.

**Remark 1.1.16.** It is a non-trivial result that there exist finitely generated groups with intermediate growth (see [GP08] for an introduction on the subject).

**Example 1.1.17.** The free groups $F_n$ have exponential growth while the Euclidean lattices $\mathbb{Z}^n$ have polynomial growth. More in detail, the growth of $\mathbb{Z}^n$ is equivalent to that of $x^n$. Since the growth of $x^n$ and $x^m$ are not equivalent if $n \neq m$, we obtain that $\mathbb{Z}^n$ is not quasi-isometric to $\mathbb{Z}^m$.

The following is a well-known analytical lemma:

**Lemma 1.1.18 (Fekete).** If $a_n$ is a sequence of real numbers such that for every $n,m \in \mathbb{N}$ it satisfies $a_{n+m} \leq a_n + a_m$ (that is, $a_n$ is sub-additive), then there exists the limit

$$\lim_{n \to \infty} \frac{a_n}{n} = \inf_{m \in \mathbb{N}} \frac{a_m}{m}.$$

Since Cayley graphs of groups are transitive, it is easy to note that the growth function of a finitely generated group $\Gamma$ is *sub-multiplicative*:

$$\mathfrak{G}_\Gamma(n + m) \leq \mathfrak{G}_\Gamma(n)\mathfrak{G}_\Gamma(m).$$

By the Fekete Lemma, it follows that there exists the limit

$$\lim_{n \to \infty} \log \left( \frac{\mathfrak{G}_\Gamma(n)}{n} \right).$$

Then it is easy to prove the following:

**Proposition 1.1.19.** A finitely generated group $\Gamma$ has exponential growth if and only if the limit of $\log (\mathfrak{G}_\Gamma(n))/n$ is strictly positive. Moreover, $\Gamma$ has super-polynomial growth if and only if the limit of $\log (\mathfrak{G}_\Gamma(n))/\log(n)$ is $+\infty$.

### 1.2 Amenability

Here we will explore the notion of amenability for graphs and groups. Amenability is closely related to the growth, because in some sense it tells us how numerous are the different paths of escape from a node. This interpretation let us understand why amenability will play such an interesting role in Chapter 3.

In the following subsections we will give various criteria for amenability involving isoperimetric inequalities, paradoxical decompositions and invariant means. And we will try to use them to relate the property of amenability for a group with its growth and the amenability of its quotients and subgroups.

We will follow the exposition of [Pet13]. Still, many proofs are slightly revised or expanded.
1.2.1 Amenability for graphs

From now on all the graphs $G$ will be of bounded degree. Given a subset of vertices of a graph $X \subseteq V(G)$, we will denote by $\partial E X \subseteq E(G)$ the set of edges of $G$ having one endpoint lying on $X$ and the other lying on the complement. Sometimes it will be convenient to consider boundary vertices instead of boundary edges, we will therefore write $\partial^i X$ and $\partial^o X$ for the sets of endpoints of edges in $\partial E X$ lying inside and outside $X$ respectively.

Definition 1.2.1. A bounded degree graph $G$ is amenable if there exists a sequence of connected subsets of vertices $F_n \subseteq V(G)$ such that

$$\frac{|\partial E F_n|}{|F_n|} \to 0.$$ 

Such a sequence is called a Følner sequence. A Følner sequence is said to be a Følner exhaustion if we also have that $F_n \subseteq F_{n+1}$ and $F_n \nearrow V(G)$.

Remark 1.2.2. Since $G$ is a graph with bounded degree, the cardinalities of $\partial E X$, $\partial^i X$ and $\partial^o X$ linearly bound one another. In particular, in definition 1.2.1 we could have used both $\partial^i F_n$ and $\partial^o F_n$ instead of $\partial E F_n$ obtaining the same result. For simplicity, throughout this thesis we will often check if a graph is amenable using $\partial^i F_n$.

Example 1.2.3. The euclidean lattices $\mathbb{Z}^d$ are clearly amenable because the metric balls $B_n(0)$ form a Følner exhaustion. Conversely, it is easy to see that the $n$-regular trees $T_n$ with $n \geq 3$ are non-amenable.

Example 1.2.4. Let $G$ be the graph defined as follows: take the infinite graph of natural numbers $\mathbb{N}$ and at every node between $2^{2k}$ and $2^{2k+1}$ attach a copy of $T_3$. Such a $G$ is amenable because the sets $F_n = \{ i \mid 2^{2n+1} < i < 2^{2n+2} \}$ form a Følner sequence, but one can see that $G$ does not admit any Følner exhaustion (recall that the sets of a Følner exhaustion must be connected).

It is interesting to note that the amenability of a graph is strictly linked with the presence of isoperimetric inequalities. Indeed, it is easy to prove the following:

Proposition 1.2.5. A graph $G$ is non-amenable if and only if it satisfies a linear isoperimetric inequality, i.e. there exists a constant $\alpha > 0$ such that for every $X \subseteq V(G)$ we have $|\partial E X| \geq \alpha |X|$.

Proof. From the definition it is clear that non-amenability is equivalent to the existence of a constant $\alpha > 0$ such that $|\partial E X| \geq \alpha |X|$ for every connected set $X \subseteq V(G)$. Thus we only need to show that the same is true also if $X$ is disconnected.

One implication is trivial. For the other, let $X = X_1 \amalg \cdots \amalg X_n$ be the decomposition of $X$ in connected components. Notice that we also have
\[ \partial E X = \partial E X_1 \uplus \cdots \uplus \partial E X_n. \] If \( G \) is non-amenable then there exists \( \alpha > 0 \) such that \( |\partial E X_i| \geq \alpha |X_i| \) for every \( i = 1, \ldots, n \). Thus we have
\[ |\partial E X| = |\partial E X_1| + \cdots + |\partial E X_n| \geq \alpha (|X_1| + \cdots + |X_n|) = \alpha |X| \]
and this proves the proposition.

We are now going to give an equivalent condition for amenability of graphs. First of all we need a definition.

**Definition 1.2.6.** A paradoxical decomposition of a graph \( G \) is a pair of injective maps \( \alpha, \beta : V(G) \to V(G) \) with disjoint images such that both \( d(\alpha(x), x) \) and \( d(\beta(x), x) \) are bounded and \( V(G) = \alpha(V(G)) \uplus \beta(V(G)) \).

We will prove that a graph is non-amenable if and only if it admits a paradoxical decomposition. Paradoxical decomposition are very much related to the celebrated Banach-Tarski paradox and much of the initial interest on amenability originated from there. Actually, the first definitions of amenability where given only for groups and involved invariant means (see the next subsection). Følner’s condition for amenability came only later as a useful criterion for amenability. To prove the equivalence between Følner condition and the non-existence of paradoxical decomposition we will need a classical result of graph theory known as the Hall-Rado Marriage Theorem.

**Theorem 1.2.7.** Given two countable sets \( A, B \) and a function \( f : A \to \mathcal{P}(B) \) such that for every \( X \subseteq A \)
\[ |X| \leq \left| \bigcup_{a \in X} f(a) \right| \]
and for every \( Y \subseteq B \)
\[ |Y| \leq |Y_f| \]
where \( Y_f = \{ a \in A \mid f(a) \cap Y \neq \emptyset \} \), then there exists a bijection \( H : A \to B \) such that \( H(a) \in f(a) \) for every \( a \in A \).

Now we can easily relate amenability and paradoxical decompositions.

**Proposition 1.2.8.** A graph \( G \) of bounded degree \( D \) admits a paradoxical decomposition if and only if it is non-amenable.

**Proof.** If there exists a paradoxical decomposition \( \alpha, \beta : V(G) \to V(G) \), we will show that \( G \) satisfies a linear isoperimetric inequality. For simplicity instead of using the edge boundary \( \partial E X \) we will use the interior node boundary \( \partial^n V X \) (see Remark 1.2.2). Let \( k \) be the constant that bounds the misplacement of \( \alpha \) and \( \beta \). For every finite \( X \subseteq V(G) \) we have that both \( \alpha(X) \)
and $\beta(X)$ are contained in $N_k(X)$, where $N_k(X)$ denotes the neighbourhood of $X$ of radius $k$. Notice that $N_k(X) = N_k(\partial^m_V X) \cup X$, and hence we have:

$$2|X| = |\alpha(X) \cup \beta(X)| \leq |N_k(\partial^m_V X) \cup X| \leq |\partial^m_V X|D^k + |X|.$$ 

Therefore, we obtain

$$|\partial^m_V X| \geq D^{-k}|X|.$$ 

Conversely, notice that for every constant $k \in \mathbb{N}$ we have

$$|N_k(X)| = |X| + |\partial^m_V (N_1(X))| + |\partial^m_V (N_2(X))| + \cdots + |\partial^m_V (N_k(X))|$$

Thus, if $G$ satisfies a linear isoperimetric inequality with constant $\alpha$, then taking a constant $k$ greater or equal to $\alpha^{-1}$ yields

$$|N_k(X)| \geq |X| + \alpha[|N_1(X)| + |N_2(X)| + \cdots + |N_k(X)|] \geq 2|X|.$$ 

Now, if we take two copies of $V(G)$ and consider the function

$$F: V(G) \amalg V(G) \to \mathcal{P}(V(G))$$

obtained letting $F(x) = B_k(x)$ for any $x$ belonging to either of the $V(G)$’s, then we see that $F$ satisfies the hypotheses of Theorem 1.2.7. Indeed, we have that $F(A \amalg B) = N_k(A \cup B)$ and thus its cardinality is big enough. The other condition is clear.

We conclude that there exists a bijective function $H: V(G) \amalg V(G) \to V(G)$ with $H(x) \in F(x)$ for every $x$ in $V(G) \amalg V(G)$ and we observe that the components of $H$ form the required paradoxical decomposition. 

Now that we have quite an interesting concept for graphs we would like to extend it to groups. We are therefore eager to prove that amenability is invariant under quasi-isometries.

**Proposition 1.2.9.** If two graphs of bounded degree $H$ and $G$ are quasi-isometric, then $H$ is amenable if and only if so is $G$.

**Proof.** It is enough to show that if $H$ is amenable and $f: G \to H$ is an $(L,A)$-quasi-isometry, then also $G$ is amenable. Let $F_n$ be a Følner sequence for $H$, we claim that $T_n = f^{-1}(F_n)$ is a Følner sequence for $G$.

First of all we want to show that the $T_n$ are big enough, so let us consider the subsets of ‘very internal’ vertices of $F_n$:

$$I_n = \{ x \in F_n \mid B_A(x) \subseteq F_n \}.$$ 

Since $f$ is $A$-quasi-surjective, we have that $\forall x \in I_n$ there exists $y \in G$ such that $d(f(y), x) \leq A$ and hence $y \in T_n$. We deduce that the image of $T_n$
contains at least \( |I_n|/D^A \) vertices, where \( D \) is the bound on the degree of vertices of \( H \). Thus we have:
\[
|T_n| \geq \frac{|I_n|}{D^A} \geq \frac{|F_n| - D^A|\partial^o F_n|}{D^A} \geq \lambda |F_n|
\]
with \( \lambda \) that approaches to \( 1/D^A \) when \( n \) increases.

Now we are done if we can bound \( |\partial^o F_n| \) linearly with \( |\partial^o F_n| \). Since \( y \in \partial^o T_n \) if and only if there is a neighbouring vertex lying outside \( T_n \), we have that \( f(y) \) is \((L + A)\)-close to \( \partial^o F_n \). Now, we can conclude the proof using the quasi-injectivity of \( f \) to obtain
\[
\frac{|\partial^o F_n|}{C^L(1+A)} \leq |\partial^o F_n|D^{L+A},
\]
where \( C \) is the bound on the degrees of \( G \).

**Corollary 1.2.10.** Let \( \Gamma \) be a finitely generated group and let \( S \) and \( S' \) be finite generating sets for \( \Gamma \). Then, if the Cayley graph \( C_S(\Gamma) \) is amenable so is \( C_{S'}(\Gamma) \).

### 1.2.2 Amenability for groups

We will now focus once more on groups. First of all, we say that a finitely generated group is **amenable** if so is any of its Cayley graphs (or equivalently, if all its Cayley graphs are amenable).

**Proposition 1.2.11.** If a Cayley graph \( C_S(\Gamma) \) is amenable then it admits a Følner exhaustion.

**Proof.** Let \( F_n \) be a Følner sequence. First of all we notice that for every \( k \in \mathbb{N} \), if \( n \) is large enough the set \( F_n \) must contain a ball of radius \( k \). In fact, if we have that for every \( g \in F_n \) there exists \( h \in \partial^o F_n \) with \( d(h,g) \leq k \), then we deduce that \( |\partial^o F_n|2^k \geq |F_n| \) and by hypothesis this cannot be the case if \( n \) is large. Therefore, up to translation, we can assume that \( F_n \) contains the ball of radius \( k \) centred at the identity \( e \in \Gamma \).

Now to get the desired exhaustion we only need to take an appropriate subsequence of \( F_n \). Let \( n(k) \) be the smallest number such that \( B_k(e) \subseteq F_n(k) \). Set \( n_1 = 1 \), then \( F_{n_1} \subseteq B_k(e) \) for \( k \) sufficiently large. Set \( n_2 = n(k) \) and repeat the same procedure. Eventually we end up with a subsequence \( F_{n_m} \) which is a Følner exhaustion.

In view of Proposition 1.2.5 we note that a non-amenable group must have exponential growth. The converse is not true, as it is shown in next example.
Example 1.2.12. The lamplighter group $\mathbb{Z}_2 \wr \mathbb{Z}$ is defined as the semi-direct product

$$\mathbb{Z}_2 \wr \mathbb{Z} := \left( \bigoplus_{\mathbb{Z}} \mathbb{Z}_2 \right) \rtimes_{\sigma} \mathbb{Z}$$

where $\sigma$ is the action given by the translation of bi-infinite sequences of $\mathbb{Z}_2$. (Recall that for any action $\sigma : H \curvearrowright G$ the semi-direct product $G \rtimes_{\sigma} H$ is the group whose set is $G \times H$ and the operation is given by $(g, h) \cdot (g', h') := (g\sigma_h(g'), hh')$.)

To get some intuition on the behaviour of the lamplighter group, imagine that the real line is a bi-infinite road and on any integer position there is a lamp that can be turned on or off. Then every element of $\mathbb{Z}_2 \wr \mathbb{Z}$ is given by a configuration $f : \mathbb{Z} \to \mathbb{Z}_2$ with only finitely many lamps turned on and a marker $m \in \mathbb{Z}$. The product of two elements $(f, m)$ and $(f', m')$ sums the configuration $f$ with the configuration $f'$ shifted by $m$ and then sums the positions of the markers.

A nice symmetric set of generators is given by $s = (1, 0)$, $R = (0, 1)$ and $L = (0, -1)$. For any element $(f, m)$ we have that multiplying to the right by $R$ or $L$ moves the marker $m$ of one position right or left. Multiplying on the right by $s$ switches the state of the lamp at position $m$.

It is easy to see that the lamplighter group has exponential growth. For example, one can notice that $|B_n(e)| \geq 2^{n/2}$ because using only the generators $s$ and $R$ it is possible to obtain with at most $n$ steps any configuration $f : \mathbb{Z} \to \mathbb{Z}_2$ where the only lamps turned on lie in positions between 0 and $n/2 - 1$.

Still, the lamplighter group is amenable. Let $F_n$ be the set of all the elements of $\mathbb{Z}_2 \wr \mathbb{Z}$ where all the lamps turned on and the marker lie between the position $-n$ and $n$. The cardinality of $F_n$ is $(2n + 1)2^{2n+1}$ and it is easy to see that the elements of $\partial_0^1 F_n$ are those where the marker stay in position $-n$ or $n$ and thus $\partial_0^1 F_n$ has cardinality $2 \cdot 2^{2n+1}$. Thus the sets $F_n$ form a Følner exhaustion for the lamplighter group.

We are now going to give another condition equivalent to the amenability due to Von Neumann. Recall that a finitely additive probability measure on a countable set $X$ is a function

$$\mu : \mathcal{P}(X) \to [0, 1]$$

such that $\mu(X) = 1$, $\mu(\emptyset) = 0$ and for any pair of disjoint sets $A, B \subseteq X$ we have $\mu(A \sqcup B) = \mu(A) + \mu(B)$.

Remark 1.2.13. If a finitely additive probability measure $\mu$ is also $\sigma$-additive (i.e. the measure of a union of countably many disjoint sets is equal to the countable summation of their measures), then it is a probability measure on the complete $\sigma$-field $\mathcal{P}(X)$ of $X$. 
We say that a finitely-additive probability measure on a countable group \( \Gamma \) is right-invariant if for any set \( X \subseteq \Gamma \) the measure \( \mu(X) \) is equal to the measure \( \mu(Xg) \) of its right translate \( Xg \) for every element \( g \in \Gamma \). Similarly, it is left-invariant if \( \mu(X) = \mu(gX) \) for every \( g \) and \( X \). Notice that since the group \( \Gamma \) is countable and transitive, a probability measure on \( \Gamma \) cannot be both \( \sigma \)-additive and left or right-invariant.

We can now state and prove Von Neumann’s condition for amenability.

**Theorem 1.2.14.** A finitely generated group \( \Gamma \) is amenable if and only if there exists a finitely additive right-invariant probability measure \( \mu \) defined on all subsets of \( \Gamma \).

**Proof of Theorem 1.2.14.** First let us show that if \( \Gamma \) admits such a measure \( \mu \) then it cannot be non-amenable. If it was, there would be a paradoxical decomposition \( \alpha, \beta : \Gamma \to \Gamma \). Notice that \( \alpha(x) \) can be written as \( xg \) for some \( g \in \Gamma \) and we have that \( d(x, \alpha(x)) \) is equal to the length of \( g \). Since \( \alpha \) is at bounded distance from the identity function, we deduce that only finitely many \( g_1, \ldots, g_n \in \Gamma \) are admissible and hence we obtain a finite partition \( \Gamma = X_1 \Pi \cdots \Pi X_n \) such that \( \alpha|_{X_i} = R_{g_i} \). That is, the restriction of \( \alpha \) to \( X_i \) is equal to the right translation by \( g_i \). We can do the same the same for \( \beta \) and we obtain some elements \( h_1, \ldots, h_m \in \Gamma \) and a partition \( \Gamma = Y_1 \Pi \cdots \Pi Y_m \) such that \( \beta|_{Y_i} = R_{h_i} \).

Now we have:

\[
\Gamma = \alpha(\Gamma) \Pi \beta(\Gamma) = (X_1 g_1 \Pi \cdots \Pi X_n g_n) \Pi (Y_1 h_1 \Pi \cdots \Pi Y_m h_m),
\]

whence we obtain a contradiction because

\[
1 = \mu(\Gamma) = \mu(X_1 g_1) + \cdots + \mu(X_n g_n) + \mu(Y_1 h_1) + \cdots + \mu(Y_m h_m) = 2.
\]

For the other implication, let \( F_n \) be a Følner sequence, we would like to define a left-invariant probability measure setting

\[
\mu(A) := \lim \frac{|A \cap F_n|}{|F_n|}.
\]

Still, such a limit could not exists. What we need then is a coherent way to choose a limit value for that sequence. The standard way to do so is via ultrafilters. The usage we make of ultrafilters is very basic: all we need to know is that when \( \mathcal{U} \) is a non-principal ultrafilter it allows us to choose a limit \( \mathcal{U} \text{-} \lim(a_n) \) also if the sequence \( a_n \) does not converge to any value. Moreover, we have that if the limit of \( a_n \) does exist then it coincides with \( \mathcal{U} \text{-} \lim(a_n) \).

So, we chose a non-principal ultrafilter \( \mathcal{U} \) and we set

\[
\mu(A) := \mathcal{U} \text{-} \lim \frac{|A \cap F_n|}{|F_n|},
\]
obtaining this way an additive probability measure on $\Gamma$. Therefore, now we only need to show that such a $\mu$ is invariant under right multiplication by elements of $\Gamma$.

Let $g$ be an element of $\Gamma$. Then

$$\left| |Ag^{-1} \cap F_n| - |A \cap F_n| \right| = |A \cap F_n| - |A \cap (F_n \triangle F)|$$

where $X \triangle Y$ represent the symmetric difference $(X \setminus Y) \cup (Y \setminus X)$. If the length of $g$ is $\ell(g) \leq k$, then for any subset $X \subseteq \Gamma$ the symmetric difference $X \triangle X_g$ is contained in $N_k(\partial^\ell F_n)$. Hence we have:

$$\left| |Ag^{-1} \cap F_n| - |A \cap F_n| \right| \leq |F_n \triangle F_n|$$

for every $A \subseteq \Gamma$ and $g \in \Gamma$.

A concept closely related to finitely additive probability measures is that of means. By definition, a mean on a countable set $X$ is a linear functional $m$ on $\ell^\infty(X) \to \mathbb{R}$ such that for every function $f \in \ell^\infty(X)$ its mean satisfies

$$\inf_{x \in X} f(x) \leq m(f) \leq \sup_{x \in X} f(x)$$

(recall that $\ell^\infty(X)$ is the space of bounded real valued functions of $X$).

Given a group $\Gamma$, we say that a mean $m$ on $\Gamma$ is right-invariant if it is invariant under pre-composition with right translations. That is, for every $f \in \ell^\infty(\Gamma)$ and $g \in \Gamma$ we have $m(f) = m(f \circ R_g)$. The definition of left-invariant means is the analogue with pre-composition with left translations. The following is a routine check:

**Proposition 1.2.15.** A finitely generated group $\Gamma$ admits a right-invariant finitely additive probability measure if and only if it has a right-invariant mean.

**Sketch of the proof.** If $\Gamma$ has a right-invariant mean $m$ then we can define a probability measure $\mu$ on $\Gamma$ letting $\mu(X) := m(1_X)$ where $1_X$ is the indicator function of $X$. Such a measure is finitely additive because $m$ is linear and it is right-invariant because

$$\mu(X_g) = m(1_{Xg}(x)) = m(1_X(xg^{-1})) = m(1_X \circ R_{g^{-1}}(x)) = m(1_X(x)).$$

Conversely, if $\mu$ is a right-invariant measure then we can define the mean of an indicator function as $m(1_X) := \mu(X)$ and by linearity we obtain a mean on the set of step functions. Being step functions a dense subset of $\ell^\infty(\Gamma)$, we can extend $m$ by continuity to the whole $\ell^\infty(\Gamma)$.

\square
1.2. AMENABILITY

We say that a mean is bi-invariant (or simply invariant) if it is both left and right-invariant. The following is a standard lemma:

**Lemma 1.2.16.** If a group \( \Gamma \) has a right-invariant mean \( m \), then it also admits a bi-invariant mean \( \tilde{m} \).

**Proof.** First of all notice that the mean \( m^{-1} \) defined as

\[
m^{-1}_x(f(x)) := m_x(f(x^{-1}))
\]

is left-invariant (we used the notation \( m_x \) to denote that the mean is taken with respect to the variable \( x \)).

Now, for any function \( f \in \ell^\infty(\Gamma) \) consider the function \( F_f : \Gamma \to \mathbb{R} \) given by \( F_f(g) := m(f \circ L_g) \). Clearly, \( F_f \) belongs to \( \ell^\infty(\Gamma) \), thus it is well-defined its mean. We define \( \tilde{m} \) as

\[
\tilde{m}(f) := m^{-1}(F_f).
\]

Such a \( \tilde{m} \) is clearly left-invariant. Moreover it is right-invariant because

\[
\tilde{m}(f \circ R_g) = m^{-1}_x(F_{f \circ R_g}(x)) = m^{-1}_x(m(f \circ R_g \circ L_x)) = m^{-1}_x(m(f \circ R_g \circ L_x \circ R_g))
\]

and the latter is equal to \( m^{-1}_x(F_f(x)) \) because \( m \) is right-invariant. \( \Box \)

**Corollary 1.2.17.** A finitely generated group \( \Gamma \) is amenable if and only if it has a bi-invariant mean.

**Remark 1.2.18.** Historically, the first definition of amenability regarded the existence of invariant means and it was introduced by Von Neumann during his studies on the Banach-Tarski paradox. Følner condition came only later.

**Remark 1.2.19.** With the same proof of Lemma 1.2.16 one can show that also the existence of left-invariant means implies the existence of bi-invariant means. The same holds also for invariant finitely additive probability measures.

Now it is easy to prove some other properties of amenable groups. For example, we have the following:

**Proposition 1.2.20.** If a finitely generated group \( \Gamma \) is amenable then all its finitely generated subgroups are amenable. Moreover, if \( 1 \to H \to \Gamma \to Q \to 1 \) is a short exact sequence, then \( \Gamma \) is amenable if and only if so are \( H \) and \( Q \).

**Proof.** If \( \Gamma \) is amenable with invariant probability measure \( \mu \) and \( H < \Gamma \) is a subgroup, we can define an invariant probability measure \( \nu \) on \( H \) as follows. For every coset \( H_i \in \Gamma/H \) choose a representative \( g_i \in \Gamma \) with \( H_i = g_iH \). Then for every \( A \subseteq H \) we set

\[
\nu(A) := \mu \left( \bigcup_{H_i \in \Gamma/H} g_iA \right)
\]
and we obtain an invariant finitely additive probability measure.

For the second part, if $1 \to H \to \Gamma \to Q \to 1$ is exact and $\Gamma$ is amenable, then $Q$ is clearly amenable because the push forward of $\mu$ gives an invariant probability measure on $Q$. Conversely, let $H$ and $Q$ be amenable with invariant means $m_H$ and $m_Q$ respectively. Given a function $f: \Gamma \to \mathbb{R}$ and a coset $gH \in \Gamma/H$ it is well-defined the mean of $f$ restricted to $gH$

$$m_{gH}(f) := m_H(f \circ L_g|_H)$$

and it does not depend on the choice of representative $g$ because $m_H$ is bi-invariant. Now that for every $f$ we have defined the function $m_{(\cdot)}(f): Q \to \mathbb{R}$, we can define a mean on $\Gamma$ taking its mean on $Q$:

$$m(f) := m_Q(m_{gH}(f))$$

and the result is an invariant mean.

1.3 Hyperbolicity of metric spaces

The object of study of this section is a notion of negatively curved metric space successfully used by M. Gromov to obtain a number of striking results. In particular, we will define $\delta$-hyperbolic spaces, the Gromov boundary of a $\delta$-hyperbolic space and recall that quasi-isometric hyperbolic spaces have homeomorphic boundaries. Then we will prove some technical results and a criterion for hyperbolicity that will be used in Chapters 2 and 4.

Our exposition follows mainly [BH99], but it is adapted with contents from [Vä05] and [DK13].

1.3.1 Thin triangles

If $X$ is a metric space, a geodesic triangle is a triangle formed by three geodesic segments $\alpha$, $\beta$ and $\gamma$ of $X$. We say that a geodesic triangle is $\delta$-thin if each segment is contained in the $\delta$-neighbourhood of the other two (Figure 1.1: A $\delta$-thin triangle.)
1.3. HYPERBOLICITY OF METRIC SPACES

1.1). Recall that a metric space is geodesics if every pair of points is linked by a geodesic. Now we define the central object of interest of this section.

**Definition 1.3.1.** A geodesic metric space $X$ is $\delta$-hyperbolic if all geodesic triangles in $X$ are $\delta$-thin. A geodesic metric space is hyperbolic if it is $\delta$-hyperbolic for some $\delta \geq 0$.

The notion of $\delta$-hyperbolicity is generally used to study the large-scale behaviour of metric spaces. Indeed, one usually let $\delta$ be as big as it is necessary in order to get hyperbolicity. As an example, notice that a geodesic space with finite diameter is trivially $\delta$-hyperbolic for every $\delta \geq \text{diam}(X)$.

**Example 1.3.2.** If a graph $G$ is a tree (that is, all the closed loops in $G$ are trivial), then it is 0-hyperbolic.

**Example 1.3.3.** The hyperbolic plane $\mathbb{H}^2$ is $\delta$-hyperbolic for some positive $\delta$. In fact, one can prove that the area of geodesic triangles is bounded and hence it is not possible for the edges to be too far apart (otherwise one could find a very large hemisphere contained in the geodesic triangle). Since every geodesic triangle in the $n$-th dimensional hyperbolic space $\mathbb{H}^n$ is contained in an embedded hyperbolic plane, we deduce that also $\mathbb{H}^n$ is $\delta$-hyperbolic with the same $\delta$.

It is easy to see that the thin-triangle condition on a hyperbolic space imply that also geodesic $n$-gons cannot be too large.

**Lemma 1.3.4.** Let $X$ be a $\delta$-hyperbolic space. If the geodesic segments $\alpha_1, \ldots, \alpha_n$ form an $n$-gon $n X$, then $\alpha_1$ is contained in the neighbourhood of $\alpha_2 \cup \cdots \cup \alpha_n$ of radius $\delta \log_2(n - 1)$.

**Proof.** The proof is immediate by induction. Indeed, taking two geodesics $\beta_0$ and $\beta_1$ from the endpoints of $\alpha_1$ to an appropriate endpoint of one of the $\alpha_i$, one can split a $(2^n + 1)$-gon in two $2^{n-1}$-gons (see Figure 1.2). Thus we conclude because $\alpha_1$ is close to the $\beta_j$'s and these are close to the other $\alpha_i$'s by induction.
If $X$ is a geodesic metric space, given two points $x$ and $y$ in $X$ we will usually denote by $[x,y]$ a geodesic of $X$ with endpoints $x$ and $y$ (we will continue to confuse geodesics with their images). By definition, such a geodesic exist for every pair of points, but it could well be non-unique. In contrast, if $X$ is $\delta$-hyperbolic one can apply the thin-triangle condition to the degenerate triangles where one of the edges is a single point. Thus we obtain that two geodesic whose endpoints coincide are one contained in the $\delta$-neighbourhood of the other.

Recall that for any pair of subsets $A, B$ of a metric space $X$ it is defined the Hausdorff distance between them as

$$d_H(A, B) = \inf \{ C \mid A \subseteq N_C(B) \text{ and } B \subseteq N_C(A) \}.$$ 

In particular, we have noticed that two geodesics with the same endpoints in a $\delta$-hyperbolic metric space have Hausdorff distance smaller than or equal to $\delta$.

The above fact generalizes greatly. Given a metric space $X$ we define a $(L, A)$-quasi-geodesic $\gamma$ as an $(L, A)$-quasi-isometric embedding $\gamma: [a, b] \to X$ of an interval of the real line in $X$ (see Subsection 1.1.2). The following fundamental lemma is sometimes referred to as the Morse Lemma:

Lemma 1.3.5 (Morse). Let $X$ be a $\delta$-hyperbolic space, $\gamma: [a, b] \to X$ an $(L, A)$-quasi-isometry and $\alpha = [\gamma(a), \gamma(b)]$ a geodesic with the same endpoints of $\gamma$. Then there exists a constant $C$ depending only on $L, A$ and $\delta$ that bounds the Hausdorff distance between $\gamma$ and $\alpha$

$$d_H(\gamma, \alpha) \leq C(L, A, \delta).$$

See [BH99] for a direct proof or [DK13] for a proof using ultrafilters and asymptotic cones. The following is a very important consequence of the Morse Lemma:

Proposition 1.3.6. Let $X$ and $Y$ be geodesic metric spaces and $f: X \to Y$ an $(L, A)$-quasi-isometric embedding. If $Y$ is $\delta'$-hyperbolic then $X$ is $\delta'$-hyperbolic with $\delta'$ depending only on $L, A$ and $\delta$.

Proof. Let $\alpha, \beta, \gamma$ be a geodesic triangle in $X$. Then the composition of these geodesics with the quasi-isometry $f$ gives rise to a triangle in $Y$ whose edges are the $(L, A)$-quasi-geodesics $f(\alpha), f(\beta), f(\gamma)$. By the Morse Lemma, there exist three geodesics $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$ whose Hausdorff distance from the quasi-geodesics is bounded by a constant $C(L, A, \delta)$.

Being the triangle $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$ $\delta$-thin, we have that $f(\alpha)$ is contained in the $[2C(L, A, \delta) + \delta]$-neighbourhood of $f(\beta) \cup f(\gamma)$. We conclude then that $\alpha$ itself is contained in a neighbourhood of $\beta \cup \gamma$ of radius

$$\delta'(L, A, \delta) := L(2C(L, A, \delta) + \delta) + LA$$

because $f$ is an $(L, A)$-quasi-isometric embedding. \qed
1.3. HYPERBOLICITY OF METRIC SPACES

Corollary 1.3.7. Hyperbolicity of metric spaces is invariant under quasi-isometries.

At this point it is natural to define hyperbolic groups as the finitely generated groups whose Cayley graphs are $\delta$-hyperbolic.

Example 1.3.8. The free group $F_n$ of rank $n$ is hyperbolic because its Cayley graph with respect to the standard generators is the infinite regular tree $T_{2n}$ that is 0-hyperbolic.

Example 1.3.9. The Euclidean lattice $\mathbb{Z}^n$ is not hyperbolic because its Cayley graph with respect to the standard generators is naturally quasi-isometric to $\mathbb{R}^n$ and this is not hyperbolic.

Example 1.3.10. In view of Theorem 1.1.10, if $M$ is a compact Riemannian manifold then its fundamental group $\pi_1(M)$ is hyperbolic if and only if so is the Riemannian fundamental cover $\tilde{M}$. (Notice that this also include the case of the Euclidean lattice. Indeed, $\mathbb{Z}^n$ is the fundamental group of the flat $n$-dimensional torus.)

The theory of hyperbolic groups is very interesting. Still, its study goes beyond the scopes of this work in that we will only need to deal with weakly hyperbolic groups (see Subsection 2.2.3).

Before concluding this subsection we give an equivalent condition for hyperbolicity. Let $X$ be a metric space with an origin $o \in X$, we define the Gromov product between two points $x, y \in X$ with respect to $o$ as

$$(x|y)_o := \frac{1}{2} [d(x, o) + d(y, o) - d(x, y)] .$$

To get an insight on the geometric meaning of the Gromov product, let $X$ be the Euclidean plane $\mathbb{R}^2$ and draw a triangle $x, y, z$ and the inscribed circle $C$. Then the Gromov product $(x|y)_z$ is equal to the distance from $z$ of the intersections of $C$ with the edges $[z, x]$ and $[z, y]$. (Equivalently, the quantities $(x|y)_z, (y|z)_x$ and $(z|x)_y$ are the unique lengths such that the sum of two of them is equal to the distance between the corresponding origins. See Figure 1.3.)

In view of the above geometric interpretation, it does not come as a surprise that when $X$ is a $\delta$-hyperbolic space the Gromov product $(x|y)_o$ is close to the distance between $o$ and the geodesics $[x, y]$. In particular, it seems reasonable that for every triple of points $x, y, z \in X$ the Gromov products satisfy

$$(x|y)_o \geq \min \{(x|z)_o, (z|y)_o\} - \delta'$$

for some constant $\delta'$ depending on $\delta$. Something more is actually true:

Proposition 1.3.11. A geodesic metric space $X$ is $\delta$-hyperbolic if and only if there exists a constant $\delta' \geq 0$ such that for every choice of four points $o, x, y, z \in X$ Inequality (1.1) holds.
CHAPTER 1. GEOMETRIC GROUP THEORY

Figure 1.3: Geometric interpretation of the Gromov product.

See [BH99] for a proof and for other equivalent conditions of hyperbolicity.

Remark 1.3.12. Some authors prefer to define the hyperbolicity using Inequality (1.1). Such a definition has the advantage that can be used for every metric space, not only for geodesic ones.

1.3.2 The Gromov boundary

Given a geodesic space $X$ with a base point $o \in X$, one would be tempted to define a boundary at infinity of $X$ as the set of possible exit directions from $o$. To be more precise, consider the geodesic rays exiting from $o$ i.e. the geodesics $\gamma: [0, \infty) \to X$ with $\gamma(0) = o$. We say that two geodesic rays are equivalent if they have bounded Hausdorff distance and we can define a boundary of exit directions from $o$ taking the quotient

$$\partial_o X := \{\text{geodesic rays}\} / \sim.$$ 

For our scopes such a definition is not very satisfactory because in general it is not well-behaved with respect to quasi-isometries. Thus one can try to adapt it defining quasi-geodesic rays as infinite quasi-geodesics $\gamma: [0, \infty) \to X$. Then we say that two quasi-geodesic rays are equivalent if they have finite Hausdorff distance and we define a boundary as

$$\partial_{q.g} X := \{\text{quasi-geodesic rays}\} / \sim.$$ 

Notice that there is a natural inclusion $\partial_o X \hookrightarrow \partial_{q.g} X$. Moreover, we observe that if $f: X \to Y$ is a quasi-isometric embedding then it is naturally defined a boundary map $\partial f: \partial_{q.g} X \to \partial_{q.g} Y$ sending a quasi-geodesic ray $\gamma$ to the composition $f \circ \gamma$. Such a construction is clearly functorial. That is, it respects the composition $(\partial(g \circ f) = \partial g \circ \partial f)$ and when $f: X \to X$ is coarsely equivalent to the identity map then $\partial f = \text{id}_{\partial_{q.g} X}$.

In general, quasi-geodesic rays can be quite wild so that such a boundary $\partial_{q.g} X$ is of difficult use. Still, if the space $X$ is also proper and $\delta$-hyperbolic
then one can prove that every quasi-geodesic ray stays close to a geodesic. Moreover, in this case it is also possible to define a topology on $X := X \cup \partial_{q.g.}X$ such that $X$ is compact and the following holds:

**Theorem 1.3.13.** Let $X$ be a proper $\delta$-hyperbolic space. Then for every fixed origin $o \in X$ the natural inclusion $\partial_o X \hookrightarrow \partial_{q.g.}X$ is a bijection. Moreover, if $X'$ is also a proper hyperbolic space and $f : X \to X'$ is a quasi-isometric embedding then the boundary map $\partial f : \partial_{q.g.}X \to \partial_{q.g.}X'$ is a continuous and injective.

**Example 1.3.14.** Once one has properly defined the topology on the boundary at infinity, it is easy to check that the boundary at infinity of the hyperbolic space $H^n$ is naturally homeomorphic to the sphere $S^{n-1}$ seen as the boundary of the Poincaré disc $D^n \subset \mathbb{R}^n$.

Unfortunately, later on we will need to deal with hyperbolic spaces that are not proper. In such a general settings Theorem 1.3.13 does not hold (its proof relies heavily on the Ascoli-Arzelà Theorem to construct geodesics).

To overcome this difficulty, we need to further generalize the definition of boundary.

Heuristically, notice that for any pair of points in the boundary of the hyperbolic space $\xi, \eta \in \partial_{q.g.}H^n = S^{n-1}$ there is a bi-infinite geodesic in $H^n$ whose extremities tend to $\xi$ and $\eta$. In particular, the distance of such a geodesic $[\xi, \eta]$ from the origin $0 \in D^n = \mathbb{H}^n$ is finite. It turns out that the same holds for every proper $\delta$-hyperbolic space $X$ and any origin $o \in X$. Taking for granted that the Gromov product $(x|y)_o$ is an approximation of the distance of $o$ from the geodesics $[x, y]$, we are not surprised from the fact that two sequences $(x_n)_{n \in \mathbb{N}}$ and $(y_m)_{m \in \mathbb{N}}$ turns out to converge to the same point in the boundary at infinity $\partial_{q.g.}X$ if and only if there exists the limit

$$\lim_{n,m \to \infty} (x_n|x_m)_o = +\infty.$$

From our heuristic we are induced to give the following:

**Definition 1.3.15.** Let $X$ be a $\delta$-hyperbolic space and $o \in X$ a fixed origin. We say that a sequence $x_n \in X$ **converges at infinity** if there exists the limit

$$\lim_{n,m \to \infty} (x_n|x_m)_o = +\infty.$$

Two sequences $x_n$ and $y_m$ converging at infinity are **asymptotic** if there exists the limit

$$\lim_{n \to \infty} (x_n|y_n)_o = +\infty.$$

Notice that the definition of convergence at infinity and asymptotic sequences does not depend on the choice of the origin $o$. Moreover, using the hyperbolicity condition given by Proposition 1.3.11 it is easy to see that being asymptotic is an equivalence relation (this is false for generic metric spaces).
CHAPTER 1. GEOMETRIC GROUP THEORY

**Definition 1.3.16.** The boundary at infinity of a \( \delta \)-hyperbolic space is defined as the quotient of the set of sequences converging at infinity where two sequences are identified if and only if they are asymptotic

\[
\partial_\infty X := \{ (x_n)_{n \in \mathbb{N}} \mid x_n \text{ converging at infinity} \} / \sim.
\]

First of all, we remark that this definition of the boundary at infinity is coherent with the previous definition (see [BH99]):

**Proposition 1.3.17.** If \( X \) is a proper \( \delta \)-hyperbolic metric space then the two boundaries \( \partial_{q,g} X \) and \( \partial_\infty X \) coincide naturally.

As before, one can define a topology on \( \overline{X} = X \cup \partial_\infty X \). This time the space \( \overline{X} \) need not be compact tough (actually, one can define a metametric on \( X \) and \( \overline{X} \) is equal to the metametric-completion. See [Vä05]). The topology on \( \overline{X} \) is defined in such a way that a sequence \( x_n \in X \) converges at the infinity (with respect to the topology) if and only if it is converging at infinity (as in Definition 1.3.15). Moreover such a sequence does actually converge to the point \( [x_n] \in \partial_\infty X \) and two sequences converging at infinity are asymptotic if and only if they converge to the same point of \( \partial_\infty X \).

Now, let \( f: X \to Y \) be a quasi-isometric embedding. Also in this case we have an obvious candidate for defining a boundary map \( \partial f: \partial_\infty X \to \partial_\infty Y \) because we can send a sequence \( (x_n)_{n \in \mathbb{N}} \) of \( X \) to its image \( (f(x_n))_{n \in \mathbb{N}} \).

**Theorem 1.3.18.** Let \( X \) and \( X' \) be \( \delta \)-hyperbolic spaces and \( f: X \to Y \) a quasi-isometric embedding. Then sending a sequence \( (x_n)_{n \in \mathbb{N}} \) of \( X \) to its image \( (f(x_n))_{n \in \mathbb{N}} \) yields a well-defined map \( \partial f: \partial_\infty X \to \partial_\infty Y \) that is continuous and injective.

See [Vä05] for a proof. Notice that the fact that such \( \partial f \) is well-defined is a non trivial result because a priori the Gromov product is not well behaved under quasi-isometries.

**Remark 1.3.19.** The proof of Theorem 1.3.18 relies on the fact that \( X \) is geodesic. Actually, something less than geodesicity is also sufficient, but in full generality the theorem is false (see [BH99] for a counterexample).

### 1.3.3 Horoball neighbourhoods

In this subsection we discuss briefly a lemma that is needed in Subsection 4.2.4. If \( X \) is a \( \delta \)-hyperbolic space and \( A \) is a subset of \( X \), one can naturally define the boundary at infinity of \( A \) as the set of equivalence classes of sequences \( x_n \) contained in \( A \) and converging at infinity in \( X \). That is,

\[
\partial_\infty A := \{ [x_n] \mid x_n \in A, \ (x_n|x_m)_o \to \infty \} \subseteq \partial_\infty X
\]

where \( o \in X \) is any fixed origin.
Remark 1.3.20. If one properly defines the topology on $\overline{X} = X \cup \partial_{\infty}X$, then the boundary at infinity of a set $A$ is $\partial_{\infty}A = \overline{A} \cap \partial_{\infty}X$.

We write the following simple facts as a lemma:

Lemma 1.3.21. Let $X$ be a $\delta$-hyperbolic space and $o \in X$ a fixed origin.

(i) If $(x_n)_{n \in \mathbb{N}}$ is a sequence converging at infinity, then every infinite subsequence $x_{n_k}$ converges at infinity and is asymptotic to $x_n$.

(ii) If $(x_n)_{n \in \mathbb{N}}$ is a sequence converging at infinity and $(y_n)_{n \in \mathbb{N}}$ is another sequence such that $(x_n|y_n)_o$ tends to infinity, then also $y_n$ converges at infinity and it is asymptotic to $x_n$.

Proof. The first assertion is obvious. To prove (ii) we only need to show that $y_n$ converges at infinity. To do so, it is enough to apply twice the condition for hyperbolicity of Proposition 1.3.11:

$$(y_n|y_m)_o \geq \min \{ (y_n|x_n)_o, (x_n|y_m)_o \} - \delta'$$

$$\geq \min \{ (y_n|x_n)_o, (x_n|x_m)_o, (x_m|y_m)_o \} - 2\delta'$$

thus $(y_n|y_m)_o$ tends to infinity with $n$ and $m$. \qed

For any subset $A \subset X$ it is convenient to define the nearest point projections of a point $x \in X$ on $A$. A nearest point projection is a point of $A$ that realizes the distance $d(x, A)$. Such a point needs not be unique. Actually, if $A$ is not closed such a projection could not exist at all. In this case a nearest point projection will be a point that approximates the distance up to a small error that will be henceforth ignored.

If a metric space $(X, d)$ has a fixed origin $o \in X$, we define the L-horoball neighbourhood of a subset $A \subset X$ as

$$\Theta_L(A) := \bigcup_{x \in A} B_{\|x\|+L}(x),$$

where $\|x\|$ denotes the distance of $x$ from the origin $o$ and $B_{\|x\|+L}(x)$ is the ball of radius $d(o, x) + L$ centred at $x$. We will need the following:

Lemma 1.3.22. Let $X$ be a $\delta$-hyperbolic space and $o \in X$ a fixed origin. Then for every constant $L$ and every subset $A \subset X$, the Gromov boundary of $A$ coincides with that of its L-horoball neighbourhood

$$\partial_{\infty}A = \partial_{\infty}(\Theta_L(A)).$$

Proof. It is clear that $\partial_{\infty}A \subset \partial_{\infty}\Theta_L(A)$ for every $L$, so we only need to show that taking the horoball neighbourhood does not add any point at infinity. As an intermediate step it is convenient to enlarge the set $A$ in such a way
that it contains many geodesics. Let \( A' \) be the set obtained taking the union of the images of all the geodesics with endpoints in \( A \):

\[
A' := \bigcup \{ \gamma([a,b]) \mid \gamma : [a,b] \to X \text{ geodesic s.t. } \gamma(a), \gamma(b) \in A \}.
\]

We claim that \( A' \) has the same boundary at infinity of \( A \).

Indeed, let \( z_n \) be a sequence in \( A' \). For every \( n \), there exist \( x_n \) and \( y_n \) in \( A \) such that \( z_n \) lies in a geodesic \([x_n, y_n]\). If the Gromov products \((z_n|x_n)_o\) and \((x_n|y_n)_o\) are bounded then we have

\[
K \geq (z_n|x_n)_o + (x_n|y_n)_o
\]

\[
= \frac{1}{2} \left[ \|x_n\| + \|y_n\| - d(x_n, z_n) - d(z_n, y_n) \right] + \|z_n\|
\]

\[
= (x_n|y_n)_o + \|z_n\|
\]

thus \( \|z_n\| \) is bounded and \( z_n \) cannot converge at infinity. It follows that if \( z_n \) does converge at infinity then at least one between \((z_n|x_n)_o\) and \((x_n|y_n)_o\) is unbounded. Taking a subsequence if necessary, we can assume that there exists the limit

\[
\lim_{n \to \infty} (z_n|x_n)_o = +\infty
\]

and hence by Lemma 1.3.21 \( z_n \) is asymptotic to a sequence of elements of \( A \) and hence \( \partial_\infty A' \subseteq \partial_\infty A \).

Now all we have to prove is that the boundary \( \partial_\infty \Theta_L(A) \) is contained in \( \partial_\infty A' \). Let \( y_n \) be a sequence of points in \( \Theta_L(A) \) converging to a point at infinity and let \( z_n \) be a nearest point projection of \( y_n \) in \( A' \). We will show that the Gromov product \((y_n|z_n)_o\) tends to infinity and hence \( y_n \) and \( z_n \) converge to the same limit in the Gromov boundary.

By definition there is a sequence \( x_n \in A \) such that \( d(y_n, x_n) \leq \|x_n\| + L \). For every \( n \), let \([x_n, y_n], [y_n, z_n], [z_n, x_n]\) be a geodesic triangle with vertices \( x_n, y_n, z_n \) (Figure 1.4). Notice that by construction the geodesic \([z_n, x_n]\) is contained in \( A' \). Let \( p_n \) be the point in the arc \([z_n, y_n]\) at distance \( 2\delta \) from \( z_n \). Since \( z_n \) realizes the minimal distance between \( y_n \) and \( A \), \( p \) cannot be \( \delta \)-close to \([z_n, x_n]\). Hence there exists \( q_n \) in \([x_n, y_n]\) with \( d(p_n, q_n) \leq \delta \).

By triangle inequality we have

\[
d(z_n, x_n) \leq d(x_n, q_n) + 3\delta,
\]

\[
d(z_n, y_n) \leq d(y_n, q_n) + 3\delta.
\]

Summing yields

\[
d(z_n, x_n) \leq d(x_n, q_n) + d(y_n, q_n) + 6\delta - d(z_n, y_n)
\]

\[
= d(x_n, y_n) + 6\delta - d(z_n, y_n)
\]

\[
\leq \|x_n\| + L + 6\delta - d(z_n, y_n).
\]
Then we can conclude:

\[(y_n|z_n)_o = \frac{1}{2} \left[ \|y_n\| + \|z_n\| - d(z_n, y_n) \right] \]

\[\geq \frac{1}{2} \left[ \|y_n\| + \|x_n\| - d(z_n, x_n) - d(z_n, y_n) \right] \]

\[\geq \frac{1}{2} \left[ \|y_n\| - L - 6\delta \right] \]

and the latter tends to infinity by hypothesis.

\[\square\]

### 1.3.4 A criterion for hyperbolicity

This section is devoted to the proof of a criterion for hyperbolicity that will come handy in Subsection 2.2.2. We will continue to use the notation \([x, y]\) to denote a geodesic (or any geodesic, depending on the context) with endpoints on \(x\) and \(y\). First of all, here it is a simple analytical lemma:

**Lemma 1.3.23.** Let \(f : \mathbb{R}_+ \rightarrow \mathbb{R}_+\) be a positive increasing function. If there exists a constant \(C \in \mathbb{R}\) and a real number \(t_0 > 0\) such that \(f(t) \leq f(f(t)) + C\) and \(f(t) < t - C\) for every \(t > t_0\), then \(f\) is bounded.

**Proof.** We claim that \(f\) is bounded by \(f(t_0)\). If \(t \leq t_0\) the claim is true because \(f\) is increasing. Let \(t > t_0\). Assuming \(f(t) > t_0\) yields \(f(f(t)) < f(t) - C\) by hypothesis. Hence we have

\[f(t) \leq f(f(t)) + C < f(t)\]

contradiction.

\[\square\]

Now we can state and prove the criterion for hyperbolicity.
Proposition 1.3.24. If $X$ is a geodesic metric space and for every pair of points $x, y \in X$ is given a set $A(x, y) \subseteq X$ such that

(i) $A(x, y) \subseteq N_K(A(x, z) \cup A(z, y))$ for every $z \in X$,

(ii) $\text{diam}(A(x, y)) \leq K$ whenever $d(x, y) \leq 1$,

for a fixed constant $K$, then $X$ is $\delta$-hyperbolic.

Proof. The idea is to show that the sets $A(x, y)$ are not too far from the actual geodesics between $x$ and $y$. For any positive real number $t > 0$, let $\Delta(t)$ denote the supremum of the distances reached by the sets $A(x, y)$ from the geodesics $[x, y]$ as $x$ and $y$ vary among the pairs of elements at distance at most $t$ one from the other. Equivalently, $\Delta(t)$ is the smallest radius such that $A(x, y)$ is contained in a neighbourhood of $[x, y]$:

$$\Delta(t) = \inf \{ r \mid A(x, y) \subseteq N_r([x, y]) \text{ for every geodesic } [x, y], \text{d}(x, y) \leq t \}.$$ 

We begin by showing that $\Delta(t)$ has at most logarithmic growth. By the hypotheses it is clear that $\Delta(t) \leq K$ whenever $t \leq 1$. If $t > 1$, let $[x, y]$ be a geodesic between two points at distance $t$ and let $x'$ be the mid-point of this geodesic. Then $A(x, y)$ is contained in a $K$-neighbourhood of $A(x, x') \cup A(x', y)$. Let $[x, x']$ and $[x', y]$ be the initial and final segments of $[x, y]$. If $d(x, x') = d(x, y)/2 \leq 1$ we conclude that $A(x, y)$ is contained in a $2K$-neighbourhood of $[x, x'] \cup [x', y] = [x, y]$, otherwise we take the midpoints of $[x, x']$ and $[x', y]$ and repeat the process. By induction we conclude that for $t > 1$

$$\Delta(t) \leq K(\log_2(t) + 2). \quad (1.2)$$

Now we will show that $\Delta$ cannot have logarithmic growth without being bounded. Fix a distance $t > 0$, two points $x, y$ with $0 < d(x, y) \leq t$ and a geodesic $[x, y]$ between them. By definition, for any point $z \in A(x, y)$ the ball of radius $\Delta(t)$ centred at $z$ intersects the geodesic $[x, y]$ and clearly so does the ball of radius $\Delta(t) + 2K$. Let $x'$ be the first point of $[x, y]$ meeting the ball $B_{\Delta+2K}(z)$ and $y'$ the last doing so (Figure 1.5). Notice that by the triangle inequality $d(x', y') \leq 2\Delta(t) + 4K$.

We observe that $x'$ and $y'$ split the geodesic in three geodesic segments $[x, x'], [x', y']$ and $[y', y]$ ($x$ and $x'$ or $y$ and $y'$ could coincide but this is not an issue). By the thin triangle property we have that $A(x, y)$ is contained in the $2K$-neighbourhood of $A(x, x') \cup A(x', y') \cup A(y', y)$ then so does the point $z$. Still, we have that $z$ cannot be close to the sets $A(x, x')$ or $A(y', y)$ because they are contained in the $\Delta(t)$-neighbourhoods of the segments $[x, x']$ and $[y', y]$ and these are far from $z$ by construction. Thus we have that $d(z, A(x', y')) \leq 2K$. Since $d(x, x') \leq 2\Delta(t) + 4K$, we obtain

$$d(z, [x, y]) = d(z, [x', y']) \leq d(z, A(x', y')) + \Delta(2\Delta(t) + 4K)$$

$$\leq \Delta(2\Delta(t) + 4K) + 2K.$$
1.3. HYPERBOLICITY OF METRIC SPACES

Figure 1.5: The ball of radius \( \Delta(t) + 2K \) centred in \( z \in A(x, y) \) meets the geodesic \([x, y]\) in two points \( x' \) and \( y' \).

Since the above inequality holds for every choice of \( x, y \) and \( z \in A(x, y) \), taking the supremum we obtain that the function \( \Delta \) satisfies

\[
\Delta(t) \leq \Delta(2\Delta(t) + 4K) + 2K.
\]

Let \( f(t) := 2\Delta(t) + 4K \). Then we have

\[
f(t) \leq f(f(t)) + 4K.
\]

By Equation (1.2) we also have that \( f(t) \leq 2K \log_2(t) + 8K \). Thus \( f \) satisfy the hypotheses of Lemma 1.3.23 and hence it is bounded. In particular, also \( \Delta(t) \) is bounded by some constant \( D \) independent of \( t \).

So far we have proven that for every geodesic \([x, y]\) the set \( A(x, y) \) is contained in the \( D \)-neighbourhood of \([x, y]\). Conversely, we will now show that \([x, y]\) is contained in a \( 2D \)-neighbourhood of \( A(x, y) \). By contradiction, assume there exists a point \( z \in [x, y] \) such that \( d(z, A(x, y)) > 2D \). Since \( A(x, y) \) is connected we conclude that there must exist an element \( w \in A(x, y) \) that is close both to the initial segment \([x, z]\) and the ending one \([z, y]\), i.e. there are two points \( w', w'' \) in \([x, y]\) such that \( d(w, w') \) and \( d(w, w'') \) are both smaller than or equal to \( D \) and \( z \) lies in the segment \([w', w'']\) \( \subset [x, y] \) (Figure 1.6).

Since we have \( d(w', w'') \leq d(w', w) + d(w, w'') \leq 2D \), at least one between \( w' \) and \( w'' \) is at distance smaller than or equal to \( D \) from \( z \) and hence \( d(z, w) \leq 2D \), contradiction.

It is now easy to conclude that \( X \) is hyperbolic. Indeed, for every geodesic
Figure 1.6: In $A(x, y)$ there is a point $w$ that is $D$-close to two points $w'$ and $w''$ preceding and succeeding $z$ respectively.

triangle $[x, y], [y, z], [z, x]$ we have:

$$[x, y] \subseteq N_{2D}(A(x, y)) \subseteq N_{2D}(N_K(A(x, y)) \cup N_K(A(y, z))) \subseteq N_{2D}(N_K(N_D([x, y])) \cup N_K(N_D([y, z]))) = N_{3D+K}([x, y] + [y, z])$$

and hence $X$ is $3D + K$ hyperbolic. $\square$
Chapter 2

Surface theory

The theory of surfaces is a highly developed subject. The aim of this chapter is to introduce various fundamental concepts such as mapping class groups, curves complexes and classifications of homeomorphism in order to build a solid background that in Chapter 4 will allow us to study random walks on mapping class groups of surfaces.

We tried to be as self-contained as possible. And when it was not possible to provide proper arguments we tried to give adequate references.

2.1 Basic facts and definitions

In this section we define many of the objects that we will be using for the remainder of the chapter. We also state many known results, generally without providing any argument. A sort of exception Subsection 2.1.4 where we try to give the ideas of the proofs of some of the theorems there stated. For, we think that those sketchy arguments may help to familiarise with the concepts so far defined.

All the results stated in this section are proved in [FM11].

2.1.1 Hyperbolic surfaces

A surface is a topological manifold of dimension 2. It is a know fact that in dimension 2 every topological manifold admits a unique differential structure up to diffeomorphism. In particular, surfaces are also differentiable manifolds and two surfaces are homeomorphic if and only if they are diffeomorphic.

We will consider only orientable surfaces of finite type, that is, surfaces which are obtained from orientable compact surfaces with (possibly empty) boundary removing a finite number of points (punctures). In order to be able to properly address to punctures, it will sometimes be useful to think of them as marked points of $S$. Given three natural numbers $g, b, p \geq 0$, we will denote by $S_{g,b}^p$ the surface of genus $g$ with $b$ boundary components and
CHAPTER 2. SURFACE THEORY

Figure 2.1: A surface of genus 2, with 4 punctures and 3 boundary components.

$p$ punctures.

It is a classical result that for every connected orientable surface $S$ of finite type there exist unique $g, b,$ and $p$ such that $S$ is homeomorphic to the surface $S_{g,b}^p$. From now on we will always assume surfaces to be orientable, connected and of finite type. (The only exception are the possibly disconnected surfaces obtained cutting other surfaces along curves. See below.)

A curve on a surface $S$ is a continuous map $\alpha: I \to S$ where $I$ is the unit interval $[0, 1] \subset \mathbb{R}$. We will generally assume curves to be smooth or at least piecewise smooth without explicit mention. Moreover we will often ignore the actual parametrization of curves focusing only on their images. A curve is closed if its endpoints coincide (i.e. it is given by a map $\alpha: S^1 \to \mathbb{R}$ where $S^1$ is the unit circle). It is simple if the map is injective. We will often refer to closed (simple) curves simply as curves. If $S$ has non-empty boundary $\partial S$, a proper arc is a map $\gamma: I \to S$ such that $\{0, 1\} = \gamma^{-1}(\partial S)$. Also in this case when we write arc we often mean that the arc is proper, (piecewise) smooth and simple.

The Euler characteristic of a topological space is defined as the alternating sum of the Betti numbers. An easy calculation shows that the Euler characteristic of a surface $S_{g,b}^p$ is

$$\chi(S_{g,b}^p) = 2 - 2g - b - p.$$  

If $\alpha$ is a closed simple curve on a surface $S$, a tubular neighbourhood of $\alpha$ is a neighbourhood $N(\alpha) \subset S$ that is homeomorphic to an annulus $S^1 \times I$. We say that the surface $S'$ is obtained cutting $S$ along $\alpha$ if it is homeomorphic to the surface obtained removing the interior of a tubular neighbourhood of $\alpha$ from $S$ (Figure 2.2). Such an operation is topologically well-defined.

If $S' = S \setminus \overset{\circ}{N}(\alpha)$ is the surface obtained cutting $S$ along $\gamma$, then it has the same Euler characteristic as $S$. (Notice that if a surface is disconnected, its Euler characteristic is equal to the sum of the characteristics of its components.) Similarly, if $\gamma$ is a proper arc it is well-defined the cutting
operation, and if \( S' \) is the surface obtained cutting \( S \) along \( \gamma \), then

\[
\chi(S') = \chi(S) + 1.
\]

As a consequence of the classification of surfaces and the invariance of the Euler characteristic under cuts along simple curves we readily obtain that on a surface there are only finitely many types of simple closed curves up to diffeomorphism. For example, let \( \alpha \) and \( \beta \) be non-separating smooth simple closed curves (i.e. the curves \( \alpha \) and \( \beta \) do not disconnect \( S \)) and let \( S' \) and \( S'' \) be the surfaces obtained cutting along \( \alpha \) and \( \beta \) respectively. Then \( \chi(S') = \chi(S'') = \chi(S) \) and both \( S' \) and \( S'' \) have the same number of punctures of \( S \) and two more boundary components. Then by the classification of surfaces \( S' \) and \( S'' \) must be diffeomorphic. Gluing back the boundary components one gets a diffeomorphism of \( S \) that sends \( \alpha \) to \( \beta \). In general, even if a curve \( \alpha \) is separating there are only finitely many possibilities for the surface obtained cutting \( S \) along \( \alpha \) and the claim follows as above.

Recall that a simply connected complete Riemannian manifold with constant curvature \( K = 1, 0 \) or \(-1\) is isometric to the unit sphere \( S^n \), the Euclidean space \( \mathbb{R}^n \) or the Hyperbolic space \( \mathbb{H}^n \) respectively. A surface \( S \) endowed with a Riemannian metric \( g \) is hyperbolic if it has totally geodesic boundary and the metric \( g \) is complete, of constant curvature \(-1\) and of finite area. If a hyperbolic surface has empty boundary then its universal Riemannian cover is \( \mathbb{H}^2 \). If it has non-empty geodesic boundary then the universal cover is a convex subset of \( \mathbb{H}^2 \).

Recall that the hyperbolic plane is \( \delta \)-hyperbolic and hence it has a well-defined boundary at infinity \( \partial_{\infty} \mathbb{H}^2 \). The orientation-preserving isometries of \( \mathbb{H}^2 \) are of three mutually exclusive types: \textit{elliptic} if they fix a point of \( \mathbb{H}^2 \); \textit{parabolic} if they fix a single point in \( \partial_{\infty} \mathbb{H}^2 \); \textit{hyperbolic} if they fix exactly two points in \( \partial_{\infty} \mathbb{H}^2 \) (Figure 2.3). If \( \varphi \) is a hyperbolic isometry of \( \mathbb{H}^2 \), the unique geodesic \( \gamma \) in \( \mathbb{H}^2 \) whose endpoints coincide with the fixed points of \( \varphi \) is the axis of the isometry. One can see that \( \varphi \) acts as a translation on \( \gamma \) and that the points of \( \gamma \) realize the minimum on \( \mathbb{H}^2 \) of the translation lengths of this isometry.

The group \( \text{Isom}^+(\mathbb{H}^2) \) of orientation-preserving isometries of the hyperbolic plane is canonically isomorphic to \( \text{PSL}(2, \mathbb{R}) \) and the induced topology
coincides with the topology of uniform convergence on compact subsets. Recall that an action of a group \( \Gamma \) on a paracompact topological space \( X \) is properly discontinuous if for every compact set \( K \subseteq X \) there exist only finitely many \( g \in \Gamma \) such that \( gK \cap K \neq \emptyset \). The action is free if every \( g \in \Gamma \setminus \{e\} \) has no fixed points in \( X \). It turns out that a group of isometries of \( \mathbb{H}^2 \) acts properly discontinuously on \( \mathbb{H}^2 \) if and only if it is a discrete subgroup of \( \text{Isom}^+(\mathbb{H}^2) \). It follows that every hyperbolic surface without boundary is the quotient of \( \mathbb{H}^2 \) by a discrete subgroup of isometries \( \Gamma \subset \text{Isom}^+(\mathbb{H}^2) \) that contains no elliptic elements. Moreover, if a surface is compact (i.e. without punctures) then all the elements of \( \Gamma \) are hyperbolic isometries.

**Remark 2.1.1.** It is easy to see that a discrete group \( \Gamma < \text{Isom}^+(\mathbb{H}^2) \) does not contain elliptic elements if and only if it is torsion free. When \( \Gamma \) is discrete but has elliptic elements the quotient is not a manifold, but an orbifold. We will deal with orbifolds in Subsection 2.4.1

The following holds:

**Proposition 2.1.2.** Given three real numbers \( a, b, c \geq 0 \), there exists a geodesic hexagon with right angles in \( \mathbb{H}^2 \) whose odd edges have length \( a, b \) and \( c \) respectively (Figure 2.4). Moreover, such a hexagon is unique up to isometry.

![Figure 2.3: Different types of isometries of \( \mathbb{H}^2 \).](image)

**Figure 2.4: An hexagon whose odd edges have length \( a, b \) and \( c \).**
2.1. BASIC FACTS AND DEFINITIONS

Let $D, D'$ be two geodesic hexagons with right angles and odd edges long $a, b$ and $c$. Gluing them along the even edges yields a hyperbolic pair of pants. That is, a disk with two holes with a hyperbolic metric and geodesic boundary. The lengths of the boundary components are $2a, 2b$ and $2c$.

**Remark 2.1.3.** If one or more of the lengths $a, b, c$ are set to zero, the corresponding pair of pants is degenerate. That is, some boundary component is shrunk to a puncture.

**Remark 2.1.4.** It is easy to verify that the area of a pair of pants is bounded by a constant independent of $a, b$ and $c$. For example, it is enough to triangulate it and bound the area of a geodesic triangle. Actually, using Gauss-Bonnet Theorem (see below) one directly finds out that the area of every hyperbolic pair of pants is precisely $2\pi$.

We say that a closed simple curve $\alpha$ on a surface $S$ is essential if it is not null-homotopic nor boundary parallel, i.e. $\alpha$ is not a separating curve that bounds a disk with a single puncture (or hole). We say that a proper arc $\gamma$ is essential if it does not bound a disk at the boundary (see Figure 2.5).

Let $S$ be a surface with negative Euler characteristic and boundary $\partial S = \beta_1 \sqcup \cdots \sqcup \beta_k$. One can show that taking a maximal set of disjoint and pairwise non-homotopic essential curves $\alpha_1, \ldots, \alpha_n$ one gets a pants decomposition for the surface (Figure 2.6). In particular, for every choice of lengths $b_1, \ldots, b_k$ and $a_1, \ldots, a_n$ one can put a hyperbolic structure of finite area on $S$ such that the lengths of the curves $\alpha_i$ and the geodesic boundary components $\beta_j$ are $a_i$ and $b_j$ respectively. That is done simply by gluing together appropriate pairs of pants. (Punctures are obtained gluing degenerate pairs of pants.)
The Gauss-Bonnet Theorem is a classical result that says that for every Riemannian surface $S$, if $R$ is a compact region with geodesic boundary, then the integral of the curvature is proportional to the Euler characteristic

$$2\pi \chi(R) = \int_R K(x) dA.$$ 

In particular, if the metric has constant curvature $K$, then we have $2\pi \chi(R) = K \text{Area}(R)$. It follows that a compact surface has a hyperbolic structure only if its Euler characteristic is strictly negative.

It is possible to generalize the Gauss-Bonnet Theorem to the non-compact case:

**Theorem 2.1.5.** If $S$ is a surface of finite type with a complete Riemannian metric of finite area such that the curvature $K$ is absolutely integrable, then

$$2\pi \chi(S) = \int_S K(x) dA.$$ 

See [Ros82] for a proof. It follows that a surface of finite type admits a hyperbolic structure (i.e. a hyperbolic metric of finite area) if and only if its Euler characteristic is strictly negative.

### 2.1.2 Homotopies and isotopies

An **isotopy** of a surface $S$ is a continuous map $F: S \times I \to S$ where $I = [0,1]$ is an interval of real numbers, such that for every time $t \in I$ the map $F_t = F|_{S \times \{t\}}: S \to S$ is a homeomorphism. Two homeomorphisms $f, g$ of $S$ are **isotopic** if there exists an isotopy $F$ with $F_0 = f$ and $F_1 = g$. It is a classical result that if two homeomorphism $\varphi, \psi: S \to S$ are homotopic then they are also isotopic.

Another well-known result is that if $\alpha$ and $\beta$ are two simple closed curves in $S$, then they are **freely homotopic** (i.e. exists a homotopy sending one to the other, possibly without fixing any base point) if and only if they are isotopic (i.e. there exists an isotopy sending one to the other. The **isotopy class** of a simple closed curve $\alpha$ is the set simple closed curves freely homotopic to $\alpha$. (Here curves are identified with their images. In particular, they are not oriented.)

Let $\alpha$ and $\beta$ be two simple closed curves in $S$. The **geometric intersection number** between $\alpha$ and $\beta$ is the minimal number of intersections among curves in the isotopy class of $\alpha$ and $\beta$ that intersect transversely

$$i(\alpha, \beta) := \min \{ \#(\alpha' \cap \beta') \mid \alpha' \sim \alpha, \beta' \sim \beta, \alpha' \text{ transverse to } \beta' \}.$$ 

Two curves $\alpha$ and $\beta$ are in **minimal position** if they realize the minimal number of intersections.
2.1. BASIC FACTS AND DEFINITIONS

It is possible to prove that two curves $\alpha$ and $\beta$ are in minimal position if and only if there are no bigons among them. That is, there are no embedded disks with boundary consisting of an arc of $\alpha$ and an arc of $\beta$. In a similar fashion, one can define the intersection number between proper arcs and it turns out that two arcs are in minimal position if there are no bigons nor bigons on the boundary (see Figure 2.7).

There is a strong generalization of the fact that freely homotopic curves are in fact isotopic. A mult curve $\mu$ in a surface $S$ is a properly embedded compact 1-manifold whose components are essential (i.e. a collection of disjoint essential arcs and curves). Two multicurves $\mu, \nu$ are in minimal position if all of their components are in minimal position. The following holds:

**Proposition 2.1.6.** Let $\mu, \nu$ and $\mu', \nu'$ be two pairs of multicurves of $S$ in minimal position. If $\mu$ is homotopic to $\mu'$ and $\nu$ is homotopic to $\nu'$, then there exists an isotopy of $S$ sending $\mu$ onto $\mu'$ and $\nu$ onto $\nu'$.

(Recall that according to our definition homotopies and isotopies are not required to fix the boundary $\partial S$ pointwise. Still, they are required to fix the boundary components as sets. That is, the endpoints of an arc under an homotopy must lie on $\partial S$ at every time.)

2.1.3 Curves and geodesics

In this section we will restrict our attention to curves that are not homotopically trivial. In particular the surface $S$ will never be the sphere $S^2$ nor the disk $D^2$.

It is a classical result that if $S$ is a closed hyperbolic surface then for every non-trivial simple closed curve $\alpha$ there exists a unique closed geodesic $\hat{\alpha}$ that is freely homotopic to $\alpha$. Moreover, one can see that for any pair of non-trivial simple closed curves $\alpha$ and $\beta$, the corresponding geodesics $\hat{\alpha}$ and $\hat{\beta}$ are in minimal position.
The same facts hold even when $S$ is equal to the torus $T^2$ with a flat metric (but different techniques are employed). Some extra care is needed when the surface $S$ is not closed, even if it is hyperbolic. Indeed, a proper arc will generally be freely homotopic to infinitely many geodesic arcs and a curve around a puncture will be homotopic to none.

Recall that a curve $\alpha$ is essential if it is not null-homotopic nor boundary parallel and that a proper arc $\gamma$ is essential if it does not bound a disk at the boundary. Then it is true that for every essential curve $\alpha$ or arc $\gamma$ on a hyperbolic surface $S$ there exists a unique geodesic $\hat{\alpha}$ and a unique geodesic arc $\hat{\gamma}$ of minimal length that are freely homotopic to $\alpha$ and $\gamma$. Again, geodesics and geodesic arcs of minimal length are in minimal position.

As a corollary, we obtain that any family of curves can be simultaneously put in minimal position. Specifically, let $S$ be a surface with negative Euler characteristic and $\{[\alpha_i] \mid i \in I\}$ a set of homotopy classes of essential arcs or curves in $S$. Choose a hyperbolic structure for $S$. Then, to obtain a family of arcs and curves representing the classes $[\alpha_i]$ that are in minimal position with respect to this geometric structure.

Again, the same holds for surfaces with zero Euler characteristic (i.e. the torus $T^2$, the annulus $\mathbb{S}^1 \times I$, the once punctured disk and the twice punctured sphere) using some ad hoc techniques.

### 2.1.4 Mapping class groups

We write $\text{Homeo}^+(S, \partial S)$ to denote the group of orientation-preserving homeomorphisms of $S$ that restrict to the identity on the boundary $\partial S$ and $\text{Homeo}_0(S, \partial S) < \text{Homeo}^+(S, \partial S)$ to denote the subgroup of those that are isotopic to the identity through an isotopy that restricts to the identity on $\partial S$ at every time.

**Definition 2.1.7.** The **mapping class group** of a surface $S$ is the group of orientation-preserving homeomorphisms of $S$ that restrict to the identity on $\partial S$ up to isotopies that restrict to the identity on $\partial S$:

$$\text{Mod}(S) := \text{Homeo}^+(S, \partial S)/\text{Homeo}_0(S, \partial S).$$

**Remark 2.1.8.** In literature the definition of mapping class groups tends to vary, especially for what concerns the behaviour of homeomorphisms and isotopies on the components of the boundary $\partial S$. We have defined the mapping class group as in [FM11] and this is sometimes referred to as the **mapping class group relative to the boundary**.

Notice that by our definition, every orientation-preserving homeomorphism of a surface $S^p_{g,0}$ without boundary identifies an element of the mapping class group. In particular, it may permute the punctures. On the contrary, if
the boundary has more components we also require that the homeomorphism does not exchange them.

We will now give the fundamental example of an element of the mapping class group. Let $\gamma$ be is a simple closed curve and $N$ a tubular neighbourhood of $\gamma$. By definition $N$ is homeomorphic to an annulus $S^1 \times I$. Let $h : I \rightarrow \mathbb{R}$ be a smooth function such that $h(0) = 0$ and $h(1) = 2\pi$, then the map

$$f : S^1 \times I \rightarrow S^1 \times I$$

$$(\theta, t) \rightarrow (\theta + h(t), t)$$

is a homeomorphism of the annulus that restricts to the identity at the boundary (Figure 2.8). This implies that $f$ can be extended to a homeomorphism $T_\gamma$ of the whole surface letting $T_\gamma$ act as the identity on the exterior of $N$. The homeomorphism $T_\gamma$ is unique up to isotopy and is called the Dehn-twist about $\gamma$ (Figure 2.9). In particular, Dehn-twists are well-defined elements of the mapping class group $\text{Mod}(\mathcal{S})$.

One can prove that a Dehn-twist about an essential curve $\gamma$ is always an element of infinite order of the mapping class group, for example by showing that there exists an essential curve $\alpha$ intersecting $\gamma$ and then proving that $\alpha$ and $T_\gamma^n \alpha$ have always positive intersection number (this concludes because homeomorphisms preserve intersection numbers and a curve has zero intersection number with itself).

When a curve is not essential extra care is needed. Indeed, let $\gamma$ be a boundary-parallel simple closed curve. Then the Dehn-twists about $\gamma$ is trivial if $\gamma$ is the boundary of a punctured disc, it is of infinite order if $\gamma$ is
the boundary of a disc with a hole. Together with the fact that permutations of boundary components are not allowed, this is the only difference between punctures and boundaries.

Recall that a curve $\alpha$ is non-separating if $S \setminus \alpha$ is connected. The following fundamental theorem holds:

**Theorem 2.1.9** (Dehn-Lickorish). If $S$ is a closed surface without punctures and boundary components, then its mapping class group is generated by Dehn-twists about finitely many non-separating simple closed curves.

**Remark 2.1.10.** If $S$ is the closed surface of genus $g$, its mapping class group is generated by the Dehn-twists about the $2g + 1$ curves of figure 2.10. Moreover, it is possible to prove that the mapping class group is actually finitely presented.

Let $S$ be a closed surface and $P = \{p_1, \ldots, p_n\}$ a set of marked points in $S$. If we denote by $S^*$ the surface obtained puncturing $S$ at the points $p_1, \ldots, p_n$, then every homeomorphism of $S$ that fixes the set $P$ (possibly permuting its elements) gives naturally rise to a homeomorphism of $S^*$. Conversely, one can show that every homeomorphism of $S^*$ extends to a homeomorphism of $S$ that clearly fixes $P$. Moreover, every isotopy on $S^*$ induces an isotopy on $S$, thus we have a map $\mathcal{F}: \text{Mod}(S^*) \to \text{Mod}(S)$ that is called the **forgetful map**. Still, there is no inverse for such a map because it is not injective. Indeed, it may well be that a homeomorphism fixing $P$ is isotopic to the identity but not through an isotopy fixing the set $P$ at all times.

The kernel of the forgetful map is quite well understood. Let $S$ be a surface with $\chi(S) < 0$ (possibly with punctures and non-empty boundary) and let $x \in S$ be a marked point. Denote by $\text{Mod}(S, x)$ the group of orientation-preserving homeomorphisms of $S$ that restrict to the identity on $\partial S$ and fix the marked point $x$ considered up to isotopies relative to the boundary that also fix $x$ at every time. Notice that $\text{Mod}(S, x)$ is actually the subgroup of $\text{Mod}(S \setminus \{x\})$ composed by the homeomorphisms that fix the puncture $x$ (this is a proper subgroup whenever $S$ already has punctures).

Then one can define the **push map** $\text{Push}: \pi_1(S, x) \to \text{Mod}(S, x)$ assigning to each loop $\gamma \in \pi_1(S, x)$ the homeomorphism obtained “pointing your finger
at \( x \) and then dragging a small neighbourhood of \( x \) along the path \( \gamma^{-} \) (Figure 2.11).

To define the push map more formally, if \( \gamma \) is a simple loop we can define \( \text{Push}(\gamma) \) as we did for Dehn-twists. Let \( N \) be a tubular neighbourhood such that \( \gamma \) is given by its heart \( \gamma = S^1 \times \{1/2\} \subset N \) and let \( h: I \to \mathbb{R} \) be such that \( h(0) = h(1) = 0 \) and \( h(1/2) = 2\pi \). The map

\[
f: S^1 \times I \to S^1 \times I
\]

restricts to the identity on \( \partial N \) and hence extends to the whole surface. We define \( \text{Push}(\gamma) \) as the isotopy class of the obtained homeomorphism.

**Remark 2.1.11.** Notice that \( \text{Push}(\gamma) \) is isotopic to the composition \( T^{-} \gamma^{-1} T^{-1} \gamma \) where \( \gamma^{-} \) and \( \gamma^{+} \) are the curves parallel to \( \gamma \) given by \( S^1 \times \{0\} \) and \( S^1 \times \{1\} \).

To complete the definition of the push map it is enough to notice that \( \pi_1(S,x) \) is generated by simple loops, so that one can define the push map for complicate paths composing push maps along simple loops. One then can verify that the obtained map is well-defined up to isotopy.

**Remark 2.1.12.** Actually, some extra care is needed. Indeed, functions are composed from the right to the left while we usually compose paths from the left to the right. Thus, to make the push map an homomorphism we should drag \( x \) along \( \gamma^{-1} \) instead of \( \gamma \). Still, we will not pay attention to this subtlety since we are not going to use the push map quantitatively.

Having defined the push map, we can finally state the Theorem of the Birman exact sequence.

**Theorem 2.1.13.** If \( S \) is a surface with \( \chi(S) < 0 \) and \( x \in S \) is a marked point, then the following sequence is exact.

\[
0 \to \pi_1(S,x) \xrightarrow{\text{Push}} \text{Mod}(S,x) \xrightarrow{\varphi} \text{Mod}(S) \to 0.
\]

As above, let \( P = \{x_1, \ldots, x_n\} \) be a set of marked point of a surface \( S \), set \( S^* = S \setminus P \) and denote by \( \text{Mod}(S,P) \) the subgroup of \( \text{Mod}(S^*) \) consisting
of homeomorphisms that fix each of the $x_i$'s. Applying multiple times the Birman exact sequence, one can quite explicitly describe $\operatorname{Mod}(S,P)$ in terms of $\operatorname{Mod}(S)$ and Dehn-twists about finitely many simple loops on $S^*$.

Since $\operatorname{Mod}(S,P)$ is the subgroup of $\operatorname{Mod}(S^*)$ of elements that restrict to the identity on $P$, we have a natural exact sequence

$$0 \rightarrow \operatorname{Mod}(S,P) \rightarrow \operatorname{Mod}(S^*) \rightarrow S_n \rightarrow 0$$

where $S_n$ is the group of permutations of $n$ elements. In particular, if $\operatorname{Mod}(S,P)$ is generated by some elements $g_1, \ldots, g_m$ and $f_1, \ldots, f_k$ are elements of $\operatorname{Mod}(S^*)$ whose action on $P$ generates $S_n$, then $\operatorname{Mod}(S^*)$ is generated by $\{g_1, \ldots, g_m, f_1, \ldots, f_k\}$.

In order to find such $f_i$'s one can define half-twists as follows. Let $\gamma$ be a separating curve surrounding two punctures, i.e. such that $S \setminus \gamma$ is disconnected and one component is a twice punctured disk. Then one can perform a half of a Dehn-twist about $\gamma$ and this can be extended to a homeomorphism of the whole surface rotating of $\pi$ the punctured disk. Such a map gives a transposition between punctures and it is easy to see that the symmetric group $S_n$ is generated by transpositions.

So far we have seen how to reconstruct the mapping class group of a surface with punctures from the mapping class group of the surface obtained filling the punctures. A similar arguments allow us to understand also how change the mapping class groups when adding boundary components.

Indeed, the surface with boundary $S_{g,b}^p$ contains an embedded copy of $S_{g,0}^{p+b}$ obtained ‘forgetting’ the boundary components. This induces a map between their mapping class groups $\operatorname{Mod}(S_{g,b}^p) \rightarrow \operatorname{Mod}(S_{g,0}^{p+b})$ whose kernel is the subgroup generated by Dehn-twists about curves parallel to the boundary components. Moreover, if $P = \{x_1, \ldots, x_b\}$ is the set of newly added punctures, then the image is $\operatorname{Mod}(S_{g,0}^{p+b}, P)$ thus we have an exact sequence

$$0 \rightarrow \mathbb{R}^b \rightarrow \operatorname{Mod}(S_{g,b}^p) \rightarrow \operatorname{Mod}(S_{g,0}^{p+b}, P) \rightarrow 0.$$
2.2. Curve complexes

In this section we will define various simplicial complexes that can be associated with a surface and we will show that when they are connected they naturally are $\delta$-hyperbolic metric spaces. Moreover, we will also show that it is possible to define a relative metric on the mapping class group of a surface such that the resulting metric space is quasi isometric to one of this complex; obtaining as a result that mapping class groups are relatively hyperbolic groups.

In contrast with Section 2.1, here full details are provided. The definitions are given following [FM11], while the proof of the $\delta$-hyperbolicity is on the lines of [HPW13]. Finally, the last subsection follows closely arguments from [MM99]. See [MM99] and [MM00] for a profound investigation on relations among curve complexes, mapping class groups and Teichmüller spaces.

2.2.1 Introduction to curve complexes

The curve complex of a surface $S$ is the simplicial complex $\mathcal{C}(S)$ defined as follows:

- the 0-skeleton is the set of homotopy classes of (unoriented) essential closed curves
  \[ \mathcal{C}^{(0)}(S) = \{ [\alpha] \mid \alpha \text{ (unoriented) essential curve in } S \}; \]

- $n+1$ vertices $a_0, \ldots, a_n \in \mathcal{C}^{(0)}(S)$ span an $n$-simplex if and only if they have disjoint representatives. In other words, the simplex $\sigma(a_0, \ldots, a_n)$ belongs to the complex $\mathcal{C}(S)$ if and only if there exist simple closed curves $\alpha_0, \ldots, \alpha_n$ such that $a_i = [\alpha_i]$ and $\alpha_i \cap \alpha_j = \emptyset$ for every $i \neq j$. 

Figure 2.13: A generating set of Dehn-twists and half Dehn-Twists for a generic surface.

Theorem 2.1.14. Let $S_{g,b}^p$ be a hyperbolic surface, possibly with punctures or boundary. Then the mapping class group $\text{Mod}(S_{g,b}^p)$ is generated by finitely many Dehn-twists and half-twists (Figure 2.13).
CHAPTER 2. SURFACE THEORY

Remark 2.2.1. Since it is always possible to simultaneously put a family of essential curves in minimal position, a set of classes of essential curves $a_0, \ldots, a_n$ spans an $n$-simplex of $\mathcal{C}(S)$ if and only if all the intersection numbers $i(a_i, a_j)$ are zero.

Notice that from the point of view of homotopy classes of simple closed curves boundary components and punctures are the same thing. In particular, if $S'$ is a surface obtained by $S$ deleting a boundary component (and hence obtaining a puncture) then the curve complexes $\mathcal{C}(S)$ and $\mathcal{C}(S')$ are the same. Keeping this in mind, for the remainder of the chapter we will prefer to work only with compact surfaces (i.e. with boundary but without punctures).

Remark 2.2.2. For every surface $S$, the dimension of the simplices of the curve complex $\mathcal{C}(S)$ is bounded by a constant depending only on the surface $S$. Indeed, if $S_{g,b}$ has genus $g$ and $b$ boundary components and $\alpha_1, \ldots, \alpha_n$ are non-homotopic disjoint essential curves, then $n \leq 3g - 3 + b$ (this is the number of curves necessary to obtain a pants decomposition of $S_{g,b}$). Still, apart from few cases the curve complexes are definitely not locally finite: if the surface is complicated enough it is easy to see that every essential curve admits infinitely many disjoint essential curves.

The main reason why curve complexes are important is that mapping class groups act naturally on them. If $\varphi$ is a homeomorphism of $S$, then it induces a map of $\mathcal{C}^{(0)}(S)$ onto itself sending the class $[\alpha]$ to the class $[\varphi(\alpha)]$. This map can be extended to the whole complex of curves. For, it is clear that whenever $\sigma([\alpha_0], \ldots, [\alpha_n])$ belongs to $\mathcal{C}(S)$ so does $\sigma([\varphi(\alpha_0)], \ldots, [\varphi(\alpha_n)])$ because images of disjoint curves remain disjoint. Hence it is possible to linearly extend the map in the interior of each simplex. Moreover, it is clear that two isotopic homeomorphisms induce the same map between the vertices of $\mathcal{C}(S)$ and hence they induce the same map on $\mathcal{C}(S)$ at all.

We have thus defined an action $\text{Mod}(S) \curvearrowright \mathcal{C}(S)$. This action is a fundamental tool in the study of mapping class groups because the action of a homeomorphism on the essential curves (almost) completely determines the homeomorphism up to isotopy.

We now want to define a structure of metric space on the curve complex. To do so, we need the complex to be connected.

Definition 2.2.3. A surface $S$ is sporadic if it is a sphere with at most $4$ punctures (or holes) or a torus with at most one puncture (or hole).

The reason for the above definition is that it is easy to note that any two essential curves on a sporadic surface must intersect, and hence the curve complex is disconnected (recall that a simplicial complex is connected if and only so is its 1-skeleton). On the contrary, one can prove that the curve complex of a non-sporadic surface is connected.
Remark 2.2.4. In the next section we will obtain a proof of the connectedness of the curve complex of a non-sporadic surface as a by-product of the proof of hyperbolicity.

Remark 2.2.5. Usually, when one needs to deal with curve complexes of sporadic surfaces one changes the definition in such a way that two classes of curves are linked by an edge if their intersection number is the smallest possible. The resulting graph is the same for both $S^4_{0,0}$ and $S^4_{1,0}$ and it is known as the Farey graph.

Let $S$ be a non-sporadic surface. Then its complex of curves admits a natural structure of metric space. For, every $n$-simplex can be endowed with a metric by means of the natural identification with the standard $n$-simplex of $\mathbb{R}^{n+1}$. Then one can define the distance between two points in $C(S)$ as the length of the shortest path in $C(S)$ joining them. The obtained metric space is clearly geodesic and the mapping class group acts by isometries on it.

In what follows it will be convenient to work only with the 1-skeleton of $C(S)$ instead of the whole curve complex. For this reason we define the curve graph $C_g(S)$ as the graph determined by the 1-skeleton of the curve complex $C_g(S) := C_1(S)$. Again, this graph has a natural metric and the mapping class group acts by isometries on it. For our purposes the curve graph and the curve complex are equivalent because of the following.

Lemma 2.2.6. For any $n \geq 1$, let $C$ be a connected simplicial complex of dimension $n+1$ and let $C^{(n)}$ be its n-skeleton. If both complex are endowed with their natural metrics $d_C$ and $d_{C^{(n)}}$, then the inclusion $C^{(n)} \hookrightarrow C$ is a quasi-isometry.

Proof. Both the spaces are quasi-isometric to their set of vertices $C^{(0)}$, hence it is enough to prove that the two metrics induced on $C^{(0)}(S)$ restricting $d_C$ and $d_{C^{(n)}}$ are quasi-isometric. It is clear that $d_C \leq d_{C^{(n)}}$, the tricky part is to show the converse inequality. We need to show that every geodesic $\gamma$ connecting two vertices in $C$ can be well approximated by a path contained in $C^{(n)}$.

Since $\gamma$ is a geodesic, its restriction to each simplex is a straight line. When $\gamma$ travel within the $n$-skeleton we can keep it as it is; when it crosses an $(n+1)$-simplex $\sigma$ we have to change that piece with a path contained in the boundary $\partial \sigma$. The lemma follows if we manage to prove that there exists a constant $K > 0$ depending only on the dimension $n$, such that for every pair of points $x, y \in \partial \sigma$ there exists a path in $\partial \sigma$ whose length is smaller than or equal to $K \|x - y\|$ (where $\|x - y\|$ is the distance between $x$ and $y$ in $\sigma$ realized as the standard Euclidean $(n+1)$-simplex).

This is an exercise of Euclidean geometry: let $x$ and $y$ be two points on $\partial \sigma$. We can assume they lie in two different $n$-faces $\tau, \tau'$ of $\sigma$. Let $z$ be the point in $\tau \cap \tau'$ that minimizes $\|x - z\| + \|z - y\|$. Then the angle $\theta$ formed by $\overline{xz}$ and $\overline{zy}$ is greater or equal to the dihedral angle $\theta_n$ between $\tau$ and $\tau'$.
Such an angle $\theta_n$ is strictly greater than 0 and depends only on the dimension $n$ (it is easy to compute it by induction). The cosine law imply

$$\|x - y\|^2 = \|x - z\|^2 + \|z - y\|^2 - 2\|x - z\|\|z - y\| \cos(\theta) \geq \|x - z\|^2 + \|z - y\|^2 - 2\|x - z\|\|z - y\| \cos(\theta_n).$$

Recalling that we need to bound the ratio

$$\frac{\|x - y\|^2}{(\|x - z\| + \|z - y\|)^2},$$

an easy calculation shows that this expression attains a minimum when $\|x - z\| = \|z - y\|$ and this minimum is equal to

$$K^2 := \frac{1 - \cos(\theta_n)}{2}.$$

Thus $K = \sqrt{1/2 - \cos(\theta_n)/2}$ is the constant depending only on the dimension that we were looking for.

Applying finitely many times Lemma 2.2.6, we find that all the $n$-skeleta of $\mathcal{C}(S)$ are quasi-isometric to $\mathcal{C}(S)$ itself. In particular we have the following:

**Corollary 2.2.7.** For every non-sporadic surface, the natural inclusion $\mathcal{C}_g(S) \hookrightarrow \mathcal{C}(S)$ of the curve graph into the curve complex is a quasi-isometry.

**Remark 2.2.8.** Since any graph $G$ is quasi-isometric to its set of vertices $V(G)$, it will often be convenient to work only with vertices of graphs instead of using the whole graph. This could lead to some confusion because we will always implicitly treat metric spaces as if they were geodesic (that is, the distance between two pairs of point is realized by a path joining them). Still, strictly speaking the space $V(G)$ is highly non-geodesic (it is totally disconnected). In this case, it is understood that when we speak about geodesics or paths realizing distances we are meaning that they are discrete paths. That is, a path is no longer a continuous map from an interval to a metric space, rather a finite number of steps of length one in $V(G)$. (Actually, in Subsection 2.2.3 we will need also paths with steps of length $1/2$. Their meaning will be clear.)

### 2.2.2 Hyperbolicity of curves complexes

In this section we will give a simple proof of the hyperbolicity of the curve complex following [HPW13]. When a surface has punctures or boundary components, it is often awkward to work with the curve complex because it is more difficult to check whether a closed curve is essential or not. For this reason it will be convenient to work with a different complex of curves,
that is the arc complex $A(S)$. The arc complex is defined analogously to the curve complex, except that the vertices are isotopy classes of essential arcs instead of isotopy classes of essential curves.

$\bullet$ $A^{(0)}(S) := \{ [\gamma] \mid \gamma \text{ essential arc of } S \}$;

$\bullet$ $\sigma([\gamma_0], \ldots, [\gamma_n]) \in A(S)$ if and only if $i([\gamma_i], [\gamma_j]) = 0$ for every $i$ and $j$.

As for the curve complex, the arc complex has finite dimension and hence it is quasi-isometric to the arc graph $A_g(S) := A^{(1)}(S)$. For convenience we will state and prove various results for the arc graph, but it is understood that the same holds for the arc complex.

Remark 2.2.9. For the whole subsection we will work with surfaces without punctures. That is because we want to have as many essential arcs as possible. This assumption will not be restrictive because our object of interest is the curve complex and from its point of view punctures are the same as boundary components. Alternatively, we could consider punctures as marked points and we could allow arcs to have endpoints on marked points.

Now, let $\alpha$ and $\beta$ be two arcs in minimal position and assume they have a preferred endpoint $\alpha_0$ and $\beta_0$ respectively. We say that an arc with endpoints $\alpha_0$ and $\beta_0$ is a unicorn arc between $\alpha$ and $\beta$ if it is composed of a sub-arc of $\alpha$ and a sub-arc of $\beta$ (Figure 2.14). Note that unicorn arcs are essential because $\alpha$ and $\beta$ are in minimal position. We define the unicorn path between $\alpha$ and $\beta$ as the set of unicorn arcs.

The unicorn path between two arcs heavily depends on the choice of the preferred endpoints, henceforth for the whole section it will be convenient to assume that arcs always come with a fixed preferred endpoint. The actual choice of preferred endpoints is uninfluent in what follows.

We now fix some notation. If $p$ and $q$ are two points on an arc $\gamma$, we write $\gamma^p_q$ for the sub-arc of $\gamma$ going from $p$ to $q$. Moreover, if $\gamma$ has a preferred endpoint $\gamma_0$, we write $\gamma^p_0$ for the sub-arc going from $\gamma_0$ to $p$. Every unicorn arc between $\alpha$ and $\beta$ is hence of the form $u^p = \alpha^p_0 \cup \beta^p_0$ with $p$ in the intersection $\alpha \cap \beta$. Notice that not every point $p$ of the intersection defines an unicorn arc.
because $\alpha_0^p \cup \beta_0^p$ could auto-intersects. We denote the unicorn path between $\alpha$ and $\beta$ by

$$U(\alpha, \beta) := \{ u^p \mid u^p \text{ unicorn arc between } \alpha \text{ and } \beta \}. $$

There is a natural ordering among the unicorn arcs between $\alpha$ and $\beta$. Indeed, one can set $u^p \prec u^q$ if $\alpha_0^p \subset \alpha_0^q$. Hence the biggest unicorn arc is $\alpha$ itself while the smallest is $\beta$.

So far, for every fixed pair of arcs in minimal position we have defined the ordered set

$$U(\alpha, \beta) = \alpha \prec \gamma_1^p \prec \cdots \prec \gamma_k^p \prec \beta. $$

Let $\alpha'$ and $\beta'$ be another pair of arcs in minimal position such that $\alpha' \sim \alpha$ and $\beta' \sim \beta$ are homotopic (and the homotopies send the preferred endpoints onto the preferred endpoints). Then by Proposition 2.1.6 we have that the whole unicorn path $U(\alpha', \beta')$ is isotopic to the unicorn path $U(\alpha, \beta)$ (recall that both homotopies and isotopies are not required to fix the boundary components pointwise). That is,

$$U(\alpha', \beta') = \alpha \prec (\gamma')_1^p \prec \cdots \prec (\gamma')_k^p \prec \beta$$

and $(\gamma')_i^p \sim \gamma_i^p$ for every $i = 1, \ldots, k$. Thus, for every pair of elements $[\alpha], [\beta]$ of the arc complex $\mathcal{A}(S)$ it is well-defined the unicorn path between them $U([\alpha], [\beta]) \subset \mathcal{A}(S)$. Since our focus is on complexes of arcs and curves, from now on we will consider unicorn paths only up to homotopy. In particular, to represent a unicorn path between two homotopy classes of arcs we will generally choose some convenient representatives in minimal position and work with them.

We can now prove the following:

**Proposition 2.2.10.** The arc graph of any compact surface $S$ is connected and $\delta$-hyperbolic.

**Proof.** To prove connectedness it is enough to prove that the unicorn path between two classes of essential arcs $a$ and $b$ is actually a connected path in $\mathcal{A}(S)$.

Let $\alpha$ and $\beta$ be arcs in minimal position with $a = [\alpha]$ and $b = [\beta]$. If they are disjoint, then the classes $a$ and $b$ have at most distance one (and $U(\alpha, \beta) = \{ \alpha, \beta \}$ is trivially a path joining them). In the general case, it is enough to show that if $u^q$ is the unicorn arc successive to $u^p$ under the ordering $\prec$, then $[u^p]$ and $[u^q]$ have distance at most one in the arc complex.

Notice that $q$ must lie in $\beta_0^p$ because otherwise the segment $\beta_0^q \subset u^q$ would meet $\alpha_0^p$ in $p$ and hence $u^q$ would auto intersect. We have

$$u^p = \alpha_0^p \cup \beta_0^p \cup \beta_0^q$$

$$u^q = \alpha_0^q \cup \beta_0^q \cup \beta_0^q.$$
If we show that the interior of $\alpha_p^q$ and $\beta_p^q$ do not intersect we are done because it is enough to translate slightly $u^q$ on one side to make it disjoint from $u^p$ (Figure 2.15). If $\alpha_p^q \cap \beta_p^q$ is not empty, then letting $q' \neq q$ be the first point of $\alpha$ meeting $\beta_p^q$ we obtain a unicorn path $u^p \prec u^{q'} \prec u^q$, but $u^q$ should be the successor of $u^p$, which is a contradiction.

To prove $\delta$-hyperbolicity it is enough to show that the family of unicorn paths $U(a,b)$ with $a$ and $b$ vertices of $\mathcal{A}_g(S)$ satisfy the hypotheses of Proposition 1.3.24. We have already noticed that if $d(a,b) = 1$ then $\text{diam}(U(a,b)) = 1$. It only remain to prove that for every choice of $a,b,c$ in $\mathcal{A}_g(S)$ the triangle $U(a,b), U(a,c), U(c,b)$ is ‘thin’. We will actually prove that it is 1-thin.

Choose some representatives $\alpha, \beta, \gamma$ in minimal position and consider an unicorn arc $u^p \in U(\alpha, \beta)$. If $u^p$ and $\gamma$ are disjoint, then $u^p \in N_1(\gamma)$ and we are done. Otherwise, walk along $\gamma$ until you first meet a point $q \in u^p$ (Figure 2.16). Suppose that $q$ lies in $\alpha$, then the arc

$$v^q = \gamma_0^q \cup \alpha_0^q$$

belongs to $U(\alpha, \gamma)$. Translating slightly aside $v^q$ we can make it disjoint from $u^p$, and hence $u^p \in N_1(U(\alpha, \gamma))$.

\[\square\]

Remark 2.2.11. Actually, our usage of Proposition 1.3.24 was slightly improper because we defined the paths $U(a,b)$ only for the vertices of the curve graph. This could be both formalized working directly on the set of edges considering discrete geodesics or defining similar paths also for points in the edges in the obvious way.

Remark 2.2.12. Using a more refined version of Proposition 1.3.24 one can prove that the arc complex is actually 7-hyperbolic.
CHAPTER 2. SURFACE THEORY

Figure 2.16: Triangles of unicorn paths are 1-thin.

Figure 2.17: Obtaining an essential arc at distance 1 from an essential curve $\alpha$.

For convenience, we also define the arc and curve complex $\mathcal{AC}(S)$ of a surface as the simplicial complex whose vertices are essential arcs and curves up to free homotopy and a collection of vertices span a simplex if they can be realized disjointly. As for the other complexes, also $\mathcal{AC}(S)$ is finite dimensional and hence quasi-isometric to the arc and curve graph $\mathcal{AC}_g(S)$.

Let $S$ be a surface with at least one boundary component. Since $\mathcal{A}(S)$ is a subcomplex of $\mathcal{AC}(S)$, we have an obvious 1-Lipschitz injection

$$\iota: \mathcal{A}(S) \hookrightarrow \mathcal{AC}(S).$$

Notice that this map is 1-coarsely-surjective because if $\alpha$ is an essential curve of $S$, then one can find an arc $\gamma$ joining $\alpha$ with $\partial S$. Taking a boundary component of a tubular neighbourhood of $\alpha \cup \gamma$ we get an essential arc at distance one from $[\alpha]$ (Figure 2.17). As a consequence we have that $\mathcal{AC}(S)$ is connected.

Actually, $\iota$ is something more than coarsely-surjective. Indeed if $\alpha$ and $\beta$ are two disjoint curves then one can find an arc $\gamma$ joining one of them to $\partial S$ without intersecting the other curve. Then the essential arc obtained with the above procedure is disjoint from both $\alpha$ and $\beta$. We have thus shown...
that when \( d([\alpha], [\beta]) = 1 \) we can find an arc in \( A(S) \) at distance at most one from both \([\alpha]\) and \([\beta]\).

As a last note, observe that if two arcs \([\alpha], [\beta] \in A(S)\) are at distance one from a curve \([\gamma] \in AC(S)\), then we can simultaneously put the three of them in minimal position obtaining that \( \alpha \cup \beta \) is disjoint from \( \gamma \). This implies that a unicorn path \( U([\alpha], [\beta]) \) is bounded to stay at distance one from \( \gamma \). This fact will be essential for the following:

**Lemma 2.2.13.** The arc and curve graph \( AC_g(S) \) of a compact surface with at least one boundary component is \( \delta \)-hyperbolic.

**Proof.** We want to use again the criterion given by Proposition 1.3.24 and the family of sets we would like to use are neighbourhoods of unicorn paths.

For any vertex \( a \in AC_g(S) \), choose an essential arc \( \hat{a} \) at distance one from \( a \) (if \( a \) itself is an arc let \( \hat{a} = a \)) and fix a preferred endpoint for every arc in \( A_g(S) \). We claim that the family of sets

\[
A(a, b) = N_1(U(\hat{a}, \hat{b})) \subset AC_g(S)
\]

satisfies the hypotheses of the criterion of hyperbolicity.

By definition \( a \) and \( b \) belongs to \( A(a, b) \). Moreover it is clear that for every other \( c \)

\[
A(a, b) \subset N_1(A(a, c) \cup A(c, b))
\]

because the same holds for the unicorn paths in \( A_g(S) \) and the inclusion is 1-Lipschitz.

All it remains to prove is that given two vertices \( a, b \in AC_g(S) \) at distance one the diameter of \( A(a, b) \) is bounded. If they are both arcs then \( U(\hat{a}, \hat{b}) = \{\hat{a}, \hat{b}\} \) and hence \( A(a, b) \) has diameter less then 3. If only \( a \) is an arc, we have that both \( \hat{a} \) and \( \hat{b} \) are at distance one from \( b \) and hence the whole unicorn path \( U(\hat{a}, \hat{b}) \) is bounded to stay at distance one from \( b \). Again the diameter of \( A(a, b) \) is less then 3.

If both \( a \) and \( b \) are closed curves, we have already noticed that there exists an arc \( c = \hat{c} \) that is at distance one from both \( a \) and \( b \). As before, since both \( \hat{a} \) and \( \hat{c} \) are at distance one from \( a \), so is \( U(\hat{a}, \hat{c}) \). The same is true for \( U(\hat{c}, \hat{b}) \). Since we already know that triangles of unicorn paths are 1-thin, we conclude that

\[
\text{diam } A(a, b) \leq \text{diam } N_1(A(a, c) \cup A(c, b)) \leq 3.
\]

Now we want to relate \( AC(S) \) with the curve complex \( C(S) \). Clearly also in this case we have an 1-Lipschitz inclusion

\[
\iota: C(S) \rightarrow AC(S).
\]
More interestingly, if $S$ is not the three-holed sphere we can also define a map $r$ from the vertices of $\mathcal{AC}(S)$ to $\mathcal{C}(S)$: if $\alpha$ is an essential closed curve then $r([\alpha]) = [\alpha]$ will do; if $\gamma$ is an essential arc then the definition is a little trickier.

Let $\alpha$ and $\beta$ be the components of $\partial S$ where lie the endpoints of $\gamma$ ($\alpha$ and $\beta$ could coincide) and denote by $N$ the regular neighbourhood of $\gamma \cup \alpha \cup \beta$. We claim that at least one of the components of $\partial N$ is essential. If $\alpha$ and $\beta$ coincide then $\partial N$ has two components which are given by (translates of) $\gamma$ and a sub-arc of $\alpha$. Such components cannot bound a disk since $\gamma$ is essential and if one of the components bounds a once-holed disk then the other is essential since $S$ is not the three-holed sphere.

If $\alpha$ and $\beta$ are different components, then $\partial N$ has only one component which is composed of $\alpha, \beta$ and two copies of $\gamma$. Such a curve is separating: on one side there is $N$, on the other $S \setminus N$. Now, $N$ is a two-holed disk and $S \setminus N$ cannot be a once-holed disk because $S$ is not a three-holed sphere (nor a disk since $\chi(S) < 0$).

Having proved our claim, we can set $r([\gamma])$ to be one of the essential components of $N$. Notice that $r$ is coarsely well-defined, because if $\partial N$ is composed of two essential curves then they have distance one.

By definition $r$ is an inverse for $\iota$. Vice versa, $\iota$ is a coarse-inverse for $r$ because every class $[\gamma]$ can be realized disjointly from $r([\gamma])$. Hence, if we could prove that $r$ is also coarsely-Lipschitz we would conclude that $\mathcal{AC}(S)$ and $\mathcal{C}(S)$ are quasi-isometric. When $S$ is non-sporadic, this is actually true and it is exactly what we are going to prove in the next lemma. Notice also that “en passant” we will prove that the latter is connected.

**Lemma 2.2.14.** If $S$ is a non-sporadic compact surface with at least one boundary component, then the retraction $r : \mathcal{AC}_g(S) \to \mathcal{C}_g(S)$ is 2-Lipschitz.

**Proof.** We only need to show that if two vertices $a$ and $b$ of $\mathcal{AC}_g(S)$ have
2.2. CURVE COMPLEXES

Figure 2.19: Qualitatively different arcs on a twice-holed torus.

Figure 2.20: The heart of an annulus gives rise to a non-separating curve in the once-holed torus.

distance one in $AC_g(S)$ then they images $r(a)$ and $r(b)$ have at most distance 2 in the graph $C_g(S)$.

If at least one between $a$ and $b$ is a closed curve then the claim is obvious, hence assume they are the isotopy class of two disjoint essential arcs $\alpha$ and $\beta$. We need to show that there exists an essential curve $\gamma$ disjoint from both $r(\alpha)$ and $r(\beta)$.

Cut $S$ along $\alpha$ and $\beta$, the resulting surface $S'$ has characteristic $\chi(S') = \chi(S) + 2$. In particular, if $\chi(S) < -2$ then at least one component $B$ of $S'$ cannot be a once-holed disk. In this case, we can find an essential closed curve $\gamma$ disjoint from both $\alpha$ and $\beta$ and hence from $r(\alpha)$ and $r(\beta)$. Indeed, if $B$ has non-zero genus we can choose a non-separating $\gamma$, otherwise we can choose a curve $\gamma$ surrounding exactly two holes (such a $\gamma$ is essential because $S$ is not the three-holed sphere).

The only non-sporadic compact surface $S$ with at least one hole and $\chi(S) = -2$ is the twice-holed torus. If the endpoints of $\alpha$ lie in different components of $\partial S$ then cutting $S$ along $\alpha$ yields a once-holed torus and then cutting along $\beta$ yields an annulus (the only surface with two boundary components and zero characteristic). Taking $\gamma$ to be the heart of the annulus determines a curve of $S$ disjoint from $\alpha$ and $\beta$. Such a curve is essential because it is non-separating (it follows from the fact that there exists a section of the annulus that intersects $\gamma$ in a single point and that gives rise to a closed curve in $S$ when gluing back the cuts. Figure 2.20).

If both the endpoints of $\alpha$ lie in a component of $\partial S$ and those of $\beta$ lie in the other, then $r(\alpha)$ and $r(\beta)$ are clearly disjoint. Hence it only
remains to work out the case where both \( \alpha \) and \( \beta \) have the endpoints in a single component of the boundary of a twice holed torus. If cutting along \( \alpha \) separates, then one of the resulting component must be an annulus while the other is a once-holed torus and \( \beta \) must be in the latter. As before, cutting this torus along \( \beta \) yields an annulus whose heart \( \gamma \) is a non-separating curve.

Therefore, we can assume both \( \alpha \) and \( \beta \) are non-separating. In particular, for each of them the retraction \( r \) could have been defined in two ways. This implies that if the endpoints of \( \alpha \) and \( \beta \) are not alternating then either \( r(\alpha) \) and \( r(\beta) \) are disjoint or we can find a curve disjoint from both taking the other possible definition of \( r(\alpha) \). Finally, if the endpoints of \( \alpha \) and \( \beta \) are alternating then \( r(\alpha) \) and \( r(\beta) \) intersect in exactly one point. This implies that the complement of \( r(\alpha) \cup r(\beta) \) is connected and hence must be a twice-holed disk. Then the closed curve surrounding those two holes is essential.

\[ \square \]

**Corollary 2.2.15.** If \( S \) is a non-sporadic compact surface with at least one boundary component then its curve complex \( \mathcal{C}(S) \) is connected and is quasi-isometric to the arc and curve complex \( AC(S) \). In particular, \( \mathcal{C}(S) \) is \( \delta \)-hyperbolic.

The last effort of this section is to extend the hyperbolicity of the curve complex to the case of closed surfaces.

**Theorem 2.2.16.** The complex of curves of a non-sporadic surface \( S \) is connected and \( \delta \)-hyperbolic.

**Proof.** Notice that this theorem deals only with complexes of curves and not arcs, hence boundary components and puncture are the same thing. This is the reason why we dropped the compactness hypothesis.

We have already proven the thesis if \( S \) has at least one boundary component (or puncture), hence it only remains to deal with the case of closed surfaces. This can easily be done with the following trick.

Choose a hyperbolic metric on \( S \). Then every element of \( \mathcal{C}(S) \) can be uniquely realized as a geodesic of \( S \). Clearly, the complement of a closed geodesic on a surface is a dense open subset of \( S \). Since \( \mathcal{C}(S) \) is countable, the Baire Theorem implies that the intersection of these complements is dense in \( S \), hence we can choose point that lies outside every closed geodesic. Then one can consider the punctured surface \( S^* \) obtained removing on such a point and this yields an embedding

\[
\mathcal{C}(S) \hookrightarrow \mathcal{C}(S^*).
\]

This embedding is clearly 1-Lipschitz. Moreover, ‘forgetting’ the puncture of \( S^* \) yields a 1-Lipschitz retraction \( \mathcal{C}(S^*) \to \mathcal{C}(S) \) (essential curves in \( S^* \) remain essential also in \( S \)).
This implies that the embedding is a quasi-isometric embedding, and hence \( C(S) \) is hyperbolic because so is \( C(S^*) \).

**Remark 2.2.17.** In general the embedding \( C(S) \hookrightarrow C(S^*) \) of the proof of Theorem 2.2.16 is not a quasi-isometry because it is not coarsely surjective.

### 2.2.3 The curve complex and the mapping class group

Given a geodesic metric space \((X, d)\) and a family \( \{Y_i\} \) of subsets of \( X \), one can define a new metric on \( X \) imposing that the \( Y_i \)'s have diameter at most one. Namely, for every subset \( Y_i \) add a new point \( \zeta_i \) to \( X \) and link it to every point of \( Y_i \) with a path of length \( 1/2 \). Considering the path metric in this enlarged space \( \hat{X} \) yields a geodesic metric space \((\hat{X}, d_e)\) called the *electric space*. The metric on \( X \) induced by the inclusion \( X \hookrightarrow \hat{X} \) is called the *electric or relative distance* of \( X \) (here 'relative' stands for 'relative to the sets \( Y_i \)').

Let \( \Gamma \) be a finitely generated group with word distance \( d_w \) and let \( H_1, \ldots, H_n \) be subgroups of \( \Gamma \). Then one can consider the electric metric \( d_e \) induced on \( \Gamma \) by the family \( \{gH_i\} \) of left cosets of the \( H_i \)'s.

**Definition 2.2.18.** A finitely generated group \( \Gamma \) is *weakly relatively hyperbolic* with respect to the subgroups \( H_1, \ldots, H_n \) if the electric space \((\hat{\Gamma}, d_e)\) is \( \delta \)-hyperbolic.

**Remark 2.2.19.** If \( d_w \) is the word metric of \( \Gamma \) given by a finite generating set \( S \), then the electric distance on \( \Gamma \) is the word metric with respect to the (possibly infinite) set \( S \cup H_1 \cup \cdots \cup H_n \).

\[
d(g, h) = \min \{ n \mid h = gs_1^{-1}s_2^{-1}\cdots s_n^{-1}, s_i \in S \cup H_1 \cup \cdots \cup H_n \}.\]

In particular, it is still true that \( \Gamma \) acts by isometries on \((\Gamma, d_e)\) via left multiplication.

We will now prove that the mapping class group \( \text{Mod}(S) \) of a non-sporadic surface \( S \) is weakly relatively hyperbolic with respect to the family of subgroups that we are now going to define. Recall that in \( S \) there exist only finitely many types of essential curves up to diffeomorphism. Thus there also are only finitely many isotopy classes of curves up to the action of the mapping class group. Choose a representative \( a_1, \ldots, a_n \in C^{(0)}(S) \) for each orbit of this action (Figure 2.21).

Now, let \( H_i = \text{Fix}(a_i) \) be the subgroup of \( \text{Mod}(S) \) that fixes the \( i \)-th representative and consider the electric space \((\widehat{\text{Mod}}(S), d_e)\) defined by these subgroups. For every isotopy class \( b \in C^{(0)}(S) \), if \( a_i \) is the unique representative of its class up to the action of \( \text{Mod}(S) \), then the elements \( g \in \text{Mod}(S) \) such that \( g(a_i) = b \) form a left coset \( gH_i \). Conversely, every element of a left coset \( gH_i \) maps \( a_i \) to the same isotopy class \( b = g(a_i) \). Hence we have a
one-to-one correspondence between isotopy classes of curves and cosets of the $H_i$’s and by definition the latter correspond exactly to the extra points added to $\text{Mod}(S)$ to form the electric space $\hat{\text{Mod}}(S)$. Denoting by $\zeta_b$ the left coset associated to $b$, we have thus defined a natural map $\Phi: C^{(0)}(S) \to \hat{\text{Mod}}(S)$ sending $b$ to $\zeta_b$.

**Proposition 2.2.20.** If $S$ is a non-sporadic surface, the map $\Phi: C^{(0)}(S) \to \hat{\text{Mod}}(S)$ is a quasi-isometry between the electric space $(\hat{\text{Mod}}(S), d_e)$ and the curve graph $C_g(S)$ (and hence the curve complex $C(S)$).

**Proof.** It is clear that $\Phi$ is coarsely-surjective because every $g \in \text{Mod}(S)$ belongs to the left coset $\zeta_b$ where $b = g(a_i)$ and hence $d(g, \Phi(b)) = d(g, \zeta_b) = 1/2$. We will now prove that $\Phi$ is a Lipschitz map.

The same argument showing that the action of $\text{Mod}(S)$ has a finite number of orbits in $C^{(0)}(S)$ shows also that in $S$ there are only finitely many pairs of isotopy classes of disjoint non-isotopic curves up to diffeomorphism. For any such orbit, choose a representative $(b_j, b'_j)$ with $j = 1, \ldots, k$. Then for any $j$ there are two indices $i(j), i'(j)$ and two elements $h_j$ and $h'_j$ of $\text{Mod}(S)$ such that $h_j(a_{i(j)}) = b_j$ and $h'_j(a_{i'(j)}) = b'_j$. Since there are only finitely many representatives, there exists a constant $C$ that bounds the electric lengths $\|h_j\|_e$ and $\|h'_j\|_e$ for every $j$ (where $\|\cdot\|_e := d_e(\cdot, e)$ is the electric distance from the identity of $\text{Mod}(S)$).

Let $c$ and $c'$ be two curves at distance one in $C(S)$. Then there exists a representative $(b_j, b'_j)$ and a diffeomorphism $g \in \text{Mod}(S)$ such that $g(b_j) = c$ and $g(b'_j) = c'$. By definition, $g \circ h_j$ and $g \circ h'_j$ send respectively $a_{i(j)}$ and $a_{i'(j)}$ to $c$ and $c'$ and hence they belong to the cosets identified by $\zeta_c$ and $\zeta_{c'}$ respectively. Hence we have

$$d_e(g \circ h_j, \zeta_c) = d_e(g \circ h'_j, \zeta_{c'}) = \frac{1}{2}$$

and we conclude that

$$d_e(\zeta_c, \zeta_{c'}) \leq d_e(g \circ h_j, g \circ h'_j) + 1 = d_e(h_j, h'_j) + 1 \leq 2C + 1,$$

thus $\Phi$ is $(2C + 1)$-Lipschitz.
2.2. CURVE COMPLEXES

To obtain the other inequality we define a quasi-inverse $\Psi: \hat{\text{Mod}}(S) \to \mathcal{C}^{(0)}(S)$. For every element $\zeta_b$ it is natural to define $\Psi(\zeta_b) := b$. Still, for an element $g \in \text{Mod}(S)$ we do not have a preferred image, hence we let $\Psi(g)$ be any of the $g(a_i)$ for $i = 1, \ldots, n$. This map is coarsely well-defined in that the uncertainty for $\Psi(g)$ is bounded by the diameter $D$ of the set of representatives $\{a_i \mid i = 1, \ldots, n\}$ because $g$ acts as an isometry on the curve complex (for example, taking the representatives as in Figure 2.21 $D$ is equal to 1. It is clear that $\Phi$ and $\Psi$ are coarse-inverse one to each other, so we only need to prove that $\Psi$ is Lipschitz.

Since $\hat{\text{Mod}}(S)$ is a geodesic space where the steps of a geodesic are always long 1 or $1/2$, it is enough to show that $d(\Psi(g), \Psi(g'))$ is bounded whenever $g$ and $g'$ have distance 1 or $1/2$. If $d_e(g, \zeta_b) = 1/2$, by definition $g$ belongs to the coset $\zeta_b$ hence there exists $a_i$ such that $g(a_i) = b$. It follows that the distance $d(\Psi(g), \Psi(\zeta_b))$ is bounded by the uncertainty $D$. It only remains to bound the distance $d(\Psi(g), \Psi(g'))$ when $g' = gh$ for an element $h$ of the original generating set $S$. Up to the uncertainty $D$, we may suppose that both $\Psi(g)$ and $\Psi(g')$ are obtained applying $g$ and $g'$ to the same representative $a_i$. Whence

$$d(\Psi(g), \Psi(g')) \leq d(g(a_i), g(h(a_i))) + D = d(a_i, h(a_i)) + D$$

and the latter is bounded by

$$\max_{h \in S} d(a_i, h(a_i)) + D$$

which is clearly bounded because $S$ is finite.

Remark 2.2.21. Notice that for every choice of an essential curve $\tilde{c}$ in $S$ the map $\Psi_{\tilde{c}}: (\text{Mod}(S), d_e) \to \mathcal{C}(S)$ sending $g$ to $g(\tilde{c})$ is a quasi-inverse for $\Phi$ because $\Psi_{\tilde{c}}$ coincides with the function $\Psi$ defined in the proof of Proposition 2.2.20 up to the constant

$$K = \max_{i=1,\ldots,n} d(\tilde{c}, a_i).$$

In view of Theorem 2.2.16, we promptly obtain the following.

Corollary 2.2.22. The mapping class group of a non-sporadic surface is weakly relatively hyperbolic with respect to the subgroups $H_i$.

Remark 2.2.23. Notice that the induced map between the Gromov boundaries

$$\partial_\infty \Phi: \partial_\infty \mathcal{C}(S) \to \partial_\infty (\hat{\text{Mod}}(S))$$

is natural. Indeed, the definition of $\Phi$ depends only on the choice of some classes of essential curves $a_1, \ldots, a_n$ (see above). Once the representatives are
fixed, by Remark 2.2.21, for every essential curve $\tilde{c} \in \mathcal{C}(S)$ the quasi-isometry $\Psi_{\tilde{c}}$ is a quasi-inverse for $\Phi$ and hence the boundary map $\partial_{\infty}\Psi_{\tilde{c}}$ does not depend on the choice of $\tilde{c}$. Conversely, fixing a curve $\tilde{c}$ we deduce that also $\partial_{\infty}\Phi$ does not depend on the choice of the representatives.

## 2.3 Classification of homeomorphisms

The aim of this section is to introduce various results regarding a topological classification of homeomorphisms. To prove in detail what we are going to need in the next sections, various refined tools are required. Thus, it was not possible to develop all the necessary theories. We will introduce foliations of surfaces and we will focus on the fact that they provide a topological description of some asymptotic characteristic of homeomorphisms and complexes of curves. Then we will state the Nielsen-Thurston classification for homeomorphisms and some related results that will be used later on. We will not introduce Teichmüller spaces.

Our definitions are somewhat in the middle between those of [FM11] and [FLP79]. For a complete treatment on foliations and the classification of homeomorphisms see [FM11] and [FLP79]. Otherwise, see [CB88] (here they use geodesic laminations instead of foliations, but the theory is almost the same). Various results stated at the end of Subsection 2.3.3 are proven in [Iva92]. Notice that in what follows we focus on surfaces with puncture but with empty boundary while various authors prefer to work with surfaces with boundary but without punctures. Still the theory in those cases is very similar (recall that we discussed relations between homeomorphisms of surfaces with punctures or boundary in 2.1.4).

### 2.3.1 Foliations

Let $S$ be a closed surface. A singular foliation $\mathcal{F}$ of $S$ is a partition of $S$ in connected subsets called leaves such that there is an atlas of smooth charts where the leaves correspond to the horizontal lines or to the levels of a $k$-pronged saddle with $k \geq 3$ (Figure 2.22).

The points corresponding to the centres of the saddles are the singular points, all the other ones are regular points. For simplicity, we think of singular points as degenerate leaves so that all the other leaves are smooth curves. Notice that by compactness there are only finitely many singular points.

If $S$ is a surface with some marked points $p_1, \ldots, p_n$ (punctures), we define the singular foliations of $S$ as the singular foliations of the underlying surface, with the difference that the marked points can also correspond to the centres of saddles with only 1 prong (Figure 2.23).

It is possible to define foliations for surfaces with boundary and everything we are going to say holds with due adjustments in that case too. We prefer
2.3. CLASSIFICATION OF HOMEOMORPHISMS

2.3. CLASSIFICATION OF HOMEOMORPHISMS

Figure 2.22: A regular point of a foliation and a singular point corresponding to a 3-pronged saddle.

Figure 2.23: Punctures corresponding to the centres of 1 and 2-pronged saddles.

not to speak about such foliations because we are not going to need them. For the remainder of this chapter we will always assume $S$ to be a surface with punctures but without boundary components.

If $\alpha : I \to S$ is a smooth arc, we say that it is \textit{transverse} to a foliation $F$ if it misses all the singular points and its interior is transverse to the leaves of $F$ at every point. Two foliations $F, G$ are \textit{transverse} if each leaf of one (apart from singular points) is transverse to the other. In particular, $F$ and $G$ have the same set of singular points (Figure 2.24a).

Given two arcs $\alpha$ and $\beta$ transverse to a foliation $F$, a \textit{leaf-preserving isotopy} from $\alpha$ to $\beta$ is a map $H : I \times I \to S$ such that

(i) $H|_{I \times \{0\}} = \alpha$ and $H|_{I \times \{1\}} = \beta$;

(ii) $H|_{I \times \{t\}}$ is a smooth arc transverse to $F$ at every time $t$;

(iii) both $H|_{\{0\} \times I}$ and $H|_{\{1\} \times I}$ are contained in a single leaf.

A \textit{transverse measure} on a foliation $F$ is a function $\mu$ that assigns to every smooth arc $\alpha$ transverse to $F$ a positive measure $\mu(\alpha)$ that is $\sigma$-additive with respect to the composition of arcs and is invariant under transverse isotopies. Moreover, we require $\mu$ to be \textit{regular}, i.e. for every regular point of the foliation there must exist a local chart $U \to \mathbb{R}_{x,y}$ such that the leaves
CHAPTER 2. SURFACE THEORY

(a) A common singular point of two transverse foliations \( \mathcal{F} \) and \( \mathcal{G} \)

(b) Leaf-preserving isotopy

Figure 2.24: Arcs transverse to foliations.

correspond to horizontal lines and \( \mu \) is induced by \( |dy| \). A measured foliation is a foliation \( \mathcal{F} \) equipped with a transverse measure \( \mu \).

By definition, if \( \mathcal{F} \) is a foliation of a surface \( S \), there exists a smooth atlas on \( S \setminus \{ \text{singular points} \} \) such that leaves correspond to horizontal lines. In particular, the transition maps between intersecting charts must be of the form

\[
\psi_{ij}(x, y) = (h(x, y), g(y)).
\]

If \( \mathcal{F} \) also has a transverse measure \( \mu \), we can find a similar atlas whose transition maps are simply given by

\[
\psi_{ij}(x, y) = (h(x, y), c_{ij} \pm y)
\]

and the measure \( \mu(\alpha) \) is locally obtained as the integral

\[
\mu(\alpha) = \int_\alpha |dy|.
\]

Such charts are called natural.

Notice that the form \( |dy| \) given by an atlas of natural charts is globally well-defined on \( S \) except at the singular points (that are negligible). In particular, for any not necessarily transverse arc \( \alpha: I \to S \) the integral \( \int_\alpha |dy| \) is defined (we impose that the set \( \alpha^{-1}\{\text{singular points} \} \) does not contribute to the total). If \( \alpha \) is also transverse to \( \mathcal{F} \) then this integral is equal to \( \mu(\alpha) \).

Remark 2.3.1. With this definition the integral \( \int_\alpha |dy| \) along a path that moves up and then comes down to its starting leaf is not zero (Figure 2.25). Notice that such an arc \( \alpha \) cannot be transverse to the foliation.

Let \( (\mathcal{F}, \mu) \) be a measured foliation. Then for every closed curve \( \alpha: S^1 \to S \) we can define

\[
I(\mathcal{F}, [\alpha]) := \inf_{\gamma \in [\alpha]} \int_\gamma |dy|
\]
2.3. **CLASSIFICATION OF HOMEOMORPHISMS**

![Figure 2.25: An inessential arc where the integral of the transverse measure is non-zero despite the fact that its endpoints are on the same level.](image)

where the infimum is taken over the curves \( \gamma \) freely homotopic to \( \alpha \). In particular, this gives us a well-defined functional on the vertices of the curve complex

\[
I(F, \cdot) : C^0(S) \to \mathbb{R}.
\]

We will refer to this functional as the *intersection form*.

Two foliations \( F \) and \( G \) of a surface \( S \) are equivalent if their intersection form coincide, *i.e.* for every closed curve \( \alpha \)

\[
I(F, [\alpha]) = I(G, [\alpha]).
\]

We define the *space of measured foliations* \( \mathcal{MF}(S) \) as the quotient of the set of measured foliations of \( S \) by this equivalence relation.

For simplicity of notation, later on we will write \( S \) instead of \( C^0(S) \) for the set of essential curves. Then, by definition, the intersection form gives an embedding

\[
\mathcal{MF}(S) \hookrightarrow \mathbb{R}^S
\]

and this embedding induces a topology on \( \mathcal{MF}(S) \) taking the product topology of \( \mathbb{R}^S \) (this is equal to the topology of pointwise convergence). It turns out that the image of this embedding does not contain the zero element, thus it is well-defined the projection

\[
\mathcal{MF}(S) \hookrightarrow \mathbb{R}^S \setminus \{0\} \to \mathbb{P} \mathbb{R}^S.
\]

The image of this map is called the *space of projective measured foliations* and is denoted by \( \mathcal{PMF}(S) \).

It is clear that if two measured foliations \( F, G \) of a surface \( S \) differ by an isotopy then they are equivalent in \( \mathcal{MF}(S) \). Moreover, it is also easy to see that they are equivalent if it is possible to go from one to the other contracting a compact singular leaf. (Figure 2.26)

Such a transformation is called a *Whitehead move*. It turns out that two foliations are equivalent if and only if they differ by finitely many isotopies and Whitehead moves (see [FLP79]).
2.3.2 Pseudo-Anosov homeomorphisms

If $\varphi$ is a diffeomorphism of a (punctured) surface $S$ and $(F, \mu)$ is a measured foliation, then it is clearly well-defined the image foliation $\varphi \cdot (F, \mu) = (\varphi \cdot F, \varphi_* \mu)$ whose leaves are the images of the leaves of $F$ and the transverse measure $\varphi_* \mu$ is the push-forward of the measure $\mu$. The same holds if $\varphi$ is a homeomorphism that restricts to a diffeomorphism away from the singular points of $F$. Now we can give the following:

**Definition 2.3.2.** An homeomorphism $\varphi: S \to S$ is pseudo-Anosov if there are two transverse measured foliations $(F^s, \mu^s)$ and $(F^u, \mu^u)$ such that $\varphi$ restricts to a diffeomorphism of $S \setminus \{\text{singular points of } F^u\}$ and

$$\varphi \cdot (F^s, \mu^s) = (F^s, \frac{1}{\lambda} \mu^s)$$

$$\varphi \cdot (F^u, \mu^u) = (F^u, \lambda \mu^u)$$

where $\lambda > 1$ is a real number called the dilatation factor of $\varphi$. The foliations $(F^s, \mu^s)$ and $(F^u, \mu^u)$ are called stable and unstable foliation respectively.

**Remark 2.3.3.** Notice that since $\lambda$ is strictly greater than 1, such a homeomorphism $\varphi$ cannot be differentiable at the singular points. A standard way to produce homeomorphisms of this kind is to consider Teichmüller maps of the surface. (See [FM11].)

If $\varphi$ is a diffeomorphism, it is clear that the class of $\varphi \cdot (F, \mu)$ in $\mathcal{MF}(S)$ only depends on the isotopy class of $\varphi$. Since every homeomorphism of $S$ is isotopic to a diffeomorphism, the action of $\text{Diffeo}(S)$ on measured foliations induces a well defined action of the mapping class group on space of measured foliations

$$\text{Mod}(S) \curvearrowright \mathcal{MF}(S).$$

If $f \in \text{Mod}(S)$ is a mapping class, we say that it is pseudo-Anosov if it has a pseudo-Anosov representative. It turns out that the classes of $(F^s, \mu^s)$ and $(F^u, \mu^u)$ are the unique fixed points in $\mathcal{PMF}(S)$ under the action of $f$. 

![Figure 2.26: A Whitehead move between singular foliations.](image)
and are thus they are well-defined up to rescaling the transverse measure. Moreover, one can see that also the dilatation factor \( \lambda \) is well-defined.

A standard method to construct explicit examples of pseudo-Anosov homeomorphisms is via liftings of pseudo-Anosov homeomorphisms of the torus (these are much easier to describe). Indeed, we have the following:

**Lemma 2.3.4.** Let \( p: S' \to S \) be a finite smooth covering map between surfaces and let \( \varphi \in \text{Homeo}(S) \) is a homeomorphism of \( S \) that admits a lifting \( \varphi' \in \text{Homeo}(S) \). Then, if \( \varphi \) is pseudo-Anosov so is \( \varphi' \).

**Proof.** If \( \mathcal{F} \) is a foliation of \( S \), it is clear that \( p^{-1}(\mathcal{F}) \) gives a foliation of \( S' \) whose singular points are the pre-images of the singular points of \( \mathcal{F} \). Moreover, if \( \mu \) is a transverse measure for \( \mathcal{F} \), it locally gives a transverse measure for \( p^{-1}(\mathcal{F}) \). By additivity we can extend this local transverse measure to a transverse measure of \( p^{-1}(\mathcal{F}) \) and this is well-defined because leaf-preserving isotopies of \( p^{-1}(\mathcal{F}) \) descend to leaf preserving isotopies of \( \mathcal{F} \) (and vice versa).

Now it is enough to observe that if \( (\mathcal{F}^s, \mu^s) \) and \( (\mathcal{F}^u, \mu^u) \) are the stable and unstable foliations of \( \varphi \), then their pre-images under \( p \) are stable and unstable foliations for \( \varphi' \). \( \square \)

**Remark 2.3.5.** Actually, to obtain pseudo-Anosov homeomorphisms as liftings of pseudo-Anosov homeomorphism of the torus one need to extend Lemma 2.3.4 to branched covers. The same proof works in this case too.

We are interested in the action of pseudo-Anosov element on the complex of curves \( \mathcal{C}(S) \) of a non-sporadic surface \( S \). In particular, we want to make apparent some similarities between pseudo-Anosov homeomorphisms and hyperbolic isometries of \( \mathbb{H}^2 \).

Recall that in Subsection 2.2.2 we proved that the curve complex \( \mathcal{C}(S) \) is \( \delta \)-hyperbolic and hence it is well-defined its boundary at infinity (Subsection 1.3.2). In [Kla99] it is proven that the Gromov boundary of the curve complex \( \mathcal{C}(S) \) can be naturally embedded in the space of topological foliations (that is the quotient space of foliations obtained from \( \mathcal{M}\mathcal{F} \) forgetting the measures). Then one can show that if \( g \) is a pseudo-Anosov element of \( \text{Mod}(S) \) then its stable and unstable foliations belong to the image of the boundary of \( \mathcal{C}(S) \). Moreover they are the only points in the Gromov completion \( \overline{\mathcal{C}(S)} = \mathcal{C}(S) \cup \partial_{\infty}\mathcal{C}(S) \) fixed by the action of \( g \) (the action on the curve complex is coherent with the action on the space of foliations). This is the same behaviour of a hyperbolic isometry of the hyperbolic space \( \mathbb{H}^n \). Indeed the only fixed points under such an isometry are the two endpoints of its axis in the Gromov boundary \( \partial_{\infty}\mathbb{H}^n \).

To push further the analogy, recall that we described an embedding

\[
\mathcal{M}\mathcal{F}(S) \hookrightarrow \mathbb{R}^S
\]
induced by the intersection form \( I \). Also the set of essential curve can be embedded in \( \mathbb{R}^S \) sending a curve \( a \in S \) to the functional identified by the intersection numbers:

\[
a \mapsto i(\cdot, a) \in \mathbb{R}^S.
\]

In this context it makes sense the convergence of a sequence of essential curves to a measured foliation. Since it is clear that the image of an essential curve is not the zero functional, we can also consider the projectivization

\[
S \hookrightarrow \mathbb{R}^S \to \mathbb{PR}^S
\]

and the following holds:

**Theorem 2.3.6.** Let \( S \) be a closed surface of genus at least 2. If \( g \) is a pseudo-Anosov element of \( \text{Mod}(S) \) then for every essential curve \( a \in S \) the iteration of \( g \) satisfies

\[
\lim_{n \to +\infty} [g^n(a)] = [(F^u, \mu^u)]
\]

and

\[
\lim_{n \to +\infty} [g^{-n}(a)] = [(F^s, \mu^s)]
\]

where \((F^u, \mu^u)\) and \((F^s, \mu^s)\) are the stable and unstable foliations of \( g \).

Thus the action of a pseudo-Anosov element on the complex of curves reminds very closely the action of a hyperbolic isometry of \( \mathbb{H}^n \). For a proof of Theorem 2.3.6 see [FM11, Corollary 14.24].

Recall now that for any choice of an essential curve \( a \in S \) the map \( \Psi_a : \text{Mod}(S) \to \mathcal{C}(S) \) sending \( f \) to \( f(a) \) is a quasi-isometry between \( \mathcal{C}(S) \) and the mapping class group with the relative metric (see Subsection 2.2.3). In particular, also the Gromov boundary \( \partial_\infty \text{Mod}(S) \) is identified with a subset of topological foliations. We will need the following:

**Proposition 2.3.7.** If \( S \) is a non-sporadic surface and \( g \in \text{Mod}(S) \) is pseudo-Anosov, then the boundary at infinity of the set \( \{g^k \mid k \in \mathbb{Z}\} \subset \text{Mod}(S) \) consists uniquely of the two points \([F^u, \mu^u])\) and \([F^s, \mu^s)]\).

**Sketch of the proof.** Since \( \Psi_a \) is a quasi-isometry, the boundary at infinity of \( \{g^k \mid k \in \mathbb{Z}\} \subset \text{Mod}(S) \) coincides with that of \( \{\Psi_a(g^k) \mid k \in \mathbb{Z}\} \subset \mathcal{C}(S) \).

Notice that for every mapping class \( f \in \text{Mod}(S) \) we have \( \Psi_a(f^n) = f^n(a) \). Thus, the thesis (at least for closed surfaces) would follows from Theorem 2.3.6 if we knew that the pointwise convergence in \( \overline{\mathcal{C}(S)} \) seen as a subset of \( \mathbb{PR}^S \) implies convergence also in the topology of the Gromov completion \( \overline{\mathcal{C}(S)} = \mathcal{C}(S) \cup \partial_\infty \mathcal{C}(S) \).

To prove the proposition one would have to extend Theorem 2.3.6 to surfaces with punctures. Then it would be enough to study the maps defined in [Kla99] and check their relations with the topology of the boundary as it is defined in [Vä05].
2.3. CLASSIFICATION OF HOMEOMORPHISMS

2.3.3 The Nielsen-Thurston classification

In what follows \( S \) will always be a surface with punctures (without boundary).

**Definition 2.3.8.** A mapping class \( f \in \text{Mod}(S) \) is *periodic* if it has finite order.

By definition, if \( k \) is the order of a periodic mapping class \( f \), then for every homeomorphism \( \varphi \in f \) the \( k \)-th power of \( \varphi \) is isotopic to the identity. It is a non-trivial result that there exists a representative \( \varphi \) for \( f \) such that \( \varphi^k \) is the identity. If \( \chi(S) < 0 \), one way to prove that \( f \) has a representative of order \( k \) is trying to find a hyperbolic metric where \( f \) can be realized as an isometry and then conclude because the only isometry isotopic to the identity is the identity itself. The reason why it should be easier to find such an isometry is that the mapping class group acts naturally on the space of hyperbolic structures and then everything one has to do is to prove that such an action has fixed points. (See [FM11, Theorem 7.1].)

Actually, a much stronger result holds. Indeed, any finite subgroup of the mapping class group of a surface \( S \) with negative Euler characteristic can be realized as a subgroup of isometries for some hyperbolic metric on \( S \).

**Theorem 2.3.9** (Nielsen Realization Theorem). Let \( S \) be a surface, possibly with punctures but without boundary components. If \( \chi(S) < 0 \) and \( F \) is a finite subgroup of the mapping class group \( \text{Mod}(S) \), then there exists a hyperbolic metric of finite area on \( S \) such that \( F \) can be realized as a subgroup of \( \text{Isom}^+(S) \).

The Nielsen Realization Theorem is a very profound result and was first proven by Kerckhoff in the '80 using earthquakes on Teichmüller spaces (see [Ker83]). In the case of surfaces with punctures a combinatorial proof has been recently found by Hensel, Osajda and Przytycki (see [HOP12]).

**Definition 2.3.10.** A mapping class \( f \in \text{Mod}(S) \) is *reducible* if there exists a set of isotopy classes of simple closed curves \( S = \{c_1, \ldots, c_n\} \) such that \( f(S) = S \) and the curves \( c_i \) can be realized disjointly. Such a set of curves \( S \) is called a *reduction system*.

The reason for the name is easily explained. Let \( f \) be a reducible mapping class. Up to take an appropriate power of \( f \), we can suppose that \( f(c_i) = c_i \) for every class \( c_i \) in \( S \). Choose some disjoint representatives \( \gamma_1, \ldots, \gamma_n \) for the \( c_i \)'s and let \( \varphi \) be a representative of \( f \). Then the images of the \( \gamma_i \)'s under \( f \) form a multicurve and by hypothesis the multicurves \( \{\gamma_1, \ldots, \gamma_n\} \) and \( \{f(\gamma_1), \ldots, f(\gamma_n)\} \) are homotopic and hence isotopic (see Subsection 2.1.2). Thus there exists a representative \( \varphi' \) for \( f \) that fixes all of the \( \gamma_i \)'s. In particular, \( \varphi' \) induces a homeomorphism of \( S' \) where \( S' \) is the surface obtained cutting \( S \) along the curves \( \gamma_i \). It follows that one can study reducible homeomorphisms by looking at homeomorphisms of simpler surfaces.
CHAPTER 2. SURFACE THEORY

Remark 2.3.11. It is clear that a Dehn-twist about a curve is a reducible element that is not periodic and it is easy to see that there are periodic elements that are also reducible. It is actually more complicated to show that there exists periodic elements that are not reducible. An example for such a homeomorphism can be obtained realizing a surface of genus $g$ gluing the edges of a $(4g + 2)$-gon and then considering the homeomorphism given by the rotation of one click clockwise (or counter-clockwise). (Figure 2.27).

The following theorem is a corner stone of the theory of homeomorphisms of surfaces (see [FM11, Theorem 13.2] for a proof using Teichmüller mappings. Otherwise, see [FLP79] or [CB88] for a proof closer to the original work of Thurston).

Theorem 2.3.12 (Nielsen-Thurston). Let $S$ be a surface with genus $g \geq 0$, possibly with punctures. Then every mapping class $f \in \text{Mod}(S)$ is either periodic, reducible or pseudo-Anosov. Further, if it is pseudo-Anosov then it is neither periodic nor reducible.

In analogy with the three types of homeomorphism, we can also define similar categories for subgroups if the mapping class group. Namely, a subgroup $F < \text{Mod}(S)$ is reducible if it has a reduction system. That is, there exists a set of disjoint curves fixed by all the elements of $F$. Otherwise it is irreducible.

We conclude this subsection stating some results that we will need later on.

Theorem 2.3.13 (Introduction of [Iva92], Theorem 2). Let $S$ be a surface with $\chi(S) < 0$. If $F < \text{Mod}(S)$ is an infinite irreducible subgroup, then it satisfies one of the following:

(i) $F$ has a cyclic subgroup of finite index generated by a pseudo-Anosov element;
in $F$ there are two pseudo-Anosov elements that fix different foliations and generate a free group of rank 2.

**Theorem 2.3.14.** Let $S$ be a surface with $\chi(S) < 0$. If a subgroup of the mapping class group $F < \text{Mod}(S)$ is infinite then it contains an element of infinite order.

**Sketch of the proof.** If $F$ is irreducible then by Theorem 2.3.13 it contains a pseudo-Anosov element (that has infinite order). If it is reducible there exists a reduction system $S = \{c_1, \ldots, c_n\}$. Up to taking a finite index subgroup, we can suppose that $F$ fixes each of the $c_i$'s. Moreover, taking some representatives $\alpha_i \in c_i$ we can suppose that $F$ also fixes each component of $S$ cut along the $\alpha_i$'s and thus $F$ restricts to subgroups of the mapping class groups of those components. Reducing again if necessary, one either finds a component where $F$ is irreducible (and hence contains a pseudo-Anosov element) or $F$ must contain a Dehn-twist about one of the reducing curves.

**Theorem 2.3.15** ([Iva92], Lemma 8.13). Let $S$ be a surface with $\chi(S) < 0$. If $g$ is a pseudo-Anosov element of the mapping class group $\text{Mod}(S)$, then the infinite cyclic group generated by $g$ has finite index in the centralizer $C(g) < \text{Mod}(S)$.

## 2.4 Subgroups of the mapping class group

The main objects of study of this section will be finite subgroups of the mapping class group. To deal with them we need to introduce the theory of orbifolds. Also in this case, the proofs of various fundamental results are omitted. Still, we preferred to provide details in Subsections 2.4.3 and 2.4.4. One reason being that various facts we need are only stated for closed surfaces from the authors that we are following. In Subsection 2.4.5 are proven many of the technical tools we will use in Chapter 4.

The theory of orbifold is here developed following [Sco83] and [Thu80], while [MH75] plays an important role in Subsection 2.4.4. Subsection 2.4.5 follows the lines suggested from [Mah11].

### 2.4.1 Orbifolds

We need to generalize the idea of manifolds in order to work with quotients of non-necessarily free properly discontinuous actions. To do this we need to consider spaces that are locally homeomorphic to quotients of $\mathbb{R}^n$.

**Definition 2.4.1.** An $n$-dimensional orbifold is a paracompact Hausdorff topological space equipped with a covering by open sets $U_i$ and homeomorphisms $\phi_i: \tilde{U}_i / \Gamma_i \to U_i$ where $\tilde{U}_i$ is an open set of $\mathbb{R}^n$ and $\Gamma_i$ is a finite group acting by homeomorphisms on $\tilde{U}_i$. Moreover, we require that:
(i) the covering \( \{ U_i \} \) is closed under finite intersection;

(ii) whenever \( U_j \subset U_i \) there exist an inclusion \( \varepsilon_{ji} : \Gamma_j \hookrightarrow \Gamma_i \) and an \( \varepsilon_{ji} \)-equivariant embedding \( \psi_{ji} : \tilde{U}_j \hookrightarrow \tilde{U}_i \) such that the following diagram commutes.

\[
\begin{array}{ccc}
\tilde{U}_j & \xrightarrow{\psi_{ji}} & \tilde{U}_i \\
\downarrow \phi_j & & \downarrow \phi_i \\
\tilde{U}_j / \Gamma_j & \xrightarrow{\psi_{ji}/\varepsilon_{ji}} & \tilde{U}_i / \varepsilon_{ji}(\Gamma_j)
\end{array}
\]

In analogy with the theory of manifolds, morphisms between orbifolds are those continuous map which respect the quotient structures. An orbifold is smooth if the \( \Gamma_i \)'s act by diffeomorphisms and all the above maps are smooth.

Actually, the embeddings \( \psi_{ij} \) and the injection \( \varepsilon_{ji} \) should be thought of as defined only up to conjugation by elements of \( \Gamma_j \) because in general it is not true that given \( U_k \subset U_j \subset U_i \) then \( \psi_{ki} = \psi_{ji} \circ \psi_{kj} \).

Given an orbifold \( O \), let \( X_O \) denote the underlying topological space. For any point \( x \in O \) contained in a local coordinate system \( U \cong \tilde{U} / \Gamma \) one defines the isotropy group \( \Gamma_x < \Gamma \) as the stabilizer of a pre-image of \( x \) in \( \tilde{U} \). The isotropy group is well-defined up to isomorphism. We say that a point is regular if its isotropy group is trivial. The set of non-regular points is the singular locus of \( O \). It is easy to see that the singular locus is a closed subset of \( X_O \) and it turns out that it is nowhere dense.

If a group \( \Gamma \) acts properly discontinuously on a manifold \( M \), then the quotient admits a natural structure of orbifold. Indeed, for every point \( x \in M / \Gamma \) let \( x_0 \in \pi^{-1}(x) \) and \( U \) a small open set of \( M \) homeomorphic to \( \mathbb{R}^n \) such that \( \pi^{-1}(x) \cap U = \{ x_0 \} \). Since the action is properly discontinuous, we can assume that for every \( f \in \Gamma \) with \( f(x_0) \neq x_0 \) the image \( f(U) \) is disjoint from \( U \). Considering the finite intersection

\[
V = \bigcap_{f \in \Gamma_{x_0}} f(U)
\]

we obtain an open set of \( \mathbb{R}^n \) where \( \Gamma_{x_0} < \Gamma \) acts by homeomorphisms and whose quotient is a neighbourhood of \( x \) in \( M / \Gamma \). Then one shows that it is also possible to construct a cover closed by intersection. Moreover, one can
see that the singular locus of $M/\Gamma$ is the image of the set of points that are
fixed by some elements of $\Gamma$.

Recall that a map $p: X \to Y$ between topological spaces is a cover if for
every point $y \in Y$ there exists a connected open neighbourhood $y \in U$ such
that $p^{-1}(U) = \bigsqcup V_i$ and $p|_{V_i}: V_i \to U$ is a homeomorphism for every such $V_i$.
To define orbifold covers one will then allow $p|_{V_i}: V_i \to U$ to be a quotient.
More precisely, a map between orbifolds $p: \mathcal{O}' \to \mathcal{O}$ is an orbifold cover if for
every point $y \in Y$ there exists a connected neighbourhood $U \cong \tilde{U}/\Gamma$ such
that the inverse image $p^{-1}(U)$ is union of open sets $V_i \cong \tilde{V}/\Gamma_i$ and there are
inclusions $\varepsilon_i: \Gamma_i \hookrightarrow \Gamma$ such that $V \cong \tilde{U}/\varepsilon_i(\Gamma_i)$ and the restriction of $p$ to the
components $V_i$ is given by the composition of quotients

\[
\begin{array}{ccc}
\tilde{V}_i & \longrightarrow & \tilde{U}_i \\
\downarrow & \downarrow & \downarrow \\
\tilde{V}_i/\Gamma_i & \cong & \tilde{U}/\varepsilon_i(\Gamma_i) \\
\cong & & \cong \\
V_i & \overset{p}{\longrightarrow} & U_i
\end{array}
\]

Notice that a function could be an orbifold cover without being a cover
between the underlying topological spaces. Conversely, a function $p$ could be
a topological cover between the underlying topological spaces of two orbifolds
without being an orbifold cover.

**Example 2.4.2.** If $\Gamma$ acts properly discontinuously on a manifold $M$ and
$\Gamma' < \Gamma$ is a subgroup, then it is easy to see that the quotient $M/\Gamma' \to M/\Gamma$
is an orbifold cover.

One can define the universal orbifold cover of an orbifold $\mathcal{O}$ as an orbifold
cover $\tilde{p}: \tilde{\mathcal{O}} \to \mathcal{O}$ which dominates any other cover $p: \mathcal{O}' \to \mathcal{O}$. That is, once
a regular point $x \in \mathcal{O}$ and pre-images $\tilde{x} \in \tilde{\mathcal{O}}$ and $x' \in \mathcal{O}'$ are fixed, there
exists an orbifold covering map $p'$ sending $\tilde{x}$ to $x'$ such that the following
diagram commutes.

\[
\begin{array}{ccc}
\tilde{\mathcal{O}} & \overset{p'}{\longrightarrow} & \mathcal{O}' \\
\downarrow^{\tilde{p}} & & \downarrow^{p} \\
\mathcal{O} & & 
\end{array}
\]

**Theorem 2.4.3.** Every orbifold admits a universal orbifold cover. Moreover,
such cover is unique up to isomorphism.
Let $O$ be an orbifold. Since universal orbifold covers are regular, in analogy with the theory of covers of topological spaces one can define the orbifold fundamental group $\pi_1(O)$ as the group of deck transformations of the universal cover $\tilde{p}: \tilde{O} \to O$.

**Remark 2.4.4.** If $M$ is a manifold and $p: O \to M$ is an orbifold cover, then also $O$ must be a manifold and the map $p$ is actually a cover of topological spaces in the usual sense. Thus, if an orbifold $O$ is covered by a manifold $M$, then its orbifold universal cover is the topological universal cover $\tilde{M}$. It follows that if $O \cong M/\Gamma$ where $M$ is a simply connected manifold and $\Gamma$ acts properly discontinuously on $M$, then the orbifold fundamental group $\pi_1(O)$ is isomorphic to $\Gamma$.

### 2.4.2 Two-dimensional orbifolds

We now want to restrict our attention to orbifolds obtained quotienting surfaces. Henceforth all orbifolds $O$ will be smooth 2-dimensional orbifolds.

For any point $x \in O$, let $U = \tilde{U}/\Gamma$ be a local coordinate system and $\tilde{x} \in \tilde{U}$ a pre-image of $x$. Then we can find a $\Gamma$-invariant Riemannian metric $\tilde{g}$ on $\tilde{U}$ by choosing a Riemannian metric $g$ and taking its mean under the action of the group $\tilde{g} = \frac{1}{|\Gamma|} \sum_{g \in \Gamma} g^*(g)$.

Considering the differentials $d_{\tilde{x}}g$ we obtain an action of the isotropy group $\Gamma_x$ on the tangent space $T_{\tilde{x}}\tilde{U}$. Since $\tilde{g}$ is $\Gamma$-invariant, the exponential map $\exp_{\tilde{x}}: T_{\tilde{x}}\tilde{U} \to \tilde{U}$ is equivariant under the action of $\Gamma_x$. That is, for any $g \in \Gamma$ fixing $\tilde{x}$ and $v \in T_{\tilde{x}}\tilde{U}$ contained in the domain of definition of the exponential map we have

$$\exp_{\tilde{x}}(d_{\tilde{x}}g(v)) = g(\exp_{\tilde{x}}(v)).$$

If the radius $r$ of a ball centred at the origin in $T_{\tilde{x}}\tilde{U}$ is small enough, the exponential map is a diffeomorphism with the image and we can assume that

$$g[\exp_{\tilde{x}}(B_r(0_{\tilde{x}}))] \cap \exp_{\tilde{x}}(B_r(0_{\tilde{x}})) \neq 0$$

only when $g$ is an element of $\Gamma_x$. Hence, we can construct a new local coordinate system around $x$ considering

$$\begin{array}{cccc}
B_r(0_{\tilde{x}}) & \xrightarrow[\exp_{\tilde{x}}]{\exp_{\tilde{x}}(B_r(0_{\tilde{x}}))} & \tilde{U} \\
\downarrow & & \downarrow \\
B_r(0_{\tilde{x}})/\Gamma_x \xrightarrow[\exp_{\tilde{x}}/\Gamma_x]{\exp_{\tilde{x}}(B_r(0_{\tilde{x}}))/\Gamma_x} & \tilde{U}/\Gamma_x \cong U.
\end{array}$$

In this new local system of coordinates, the open subset of $\mathbb{R}^2$ is actually a round ball and the group acts by Euclidean isometries, i.e. $\Gamma_x$ is a finite
subgroup of $O(2, \mathbb{R})$. Hence $\Gamma_x$ can only be a rotation group, a reflection group or a dihedral group. In particular the singular locus is composed of cone points, reflection lines and corner reflectors.

This implies that the underlying space of a two-dimensional orbifold is a surface (possibly with boundary). Exploiting the classification of surfaces, one can then classify two-dimensional orbifolds via the combinatorial data given by the type of the underlying surface and the number of cone points and corner reflectors and their orders (which means the order of their local groups). Note that it is important not to confuse the manifold structure of $X_\mathcal{O}$ with the orbifold structure of $\mathcal{O}$.

One can ask whether an orbifold admits a geometric structure of constant curvature, i.e. it can be realized as a quotient of a homogeneous space by a discrete group of isometries. Actually the first question at all is if every orbifold is covered by a manifold (or equivalently, if the orbifold universal cover is a manifold). The answer is negative, but not dramatically. Indeed ‘most’ orbifolds are covered by a manifold.

**Theorem 2.4.5.** The only two-dimensional orbifolds that are not covered by a surface are the following (Figure 2.29):

(i) the sphere $S^2$ with one cone point of order $p$;
(ii) the sphere $S^2$ with two cone points of order $p$ and $q$, with $p \neq q$;
(iii) the disk $D^2$ with one cone corner reflector of order $p$;
(iv) the disk $D^2$ with two cone corner reflectors of order $p$ and $q$, with $p \neq q$.

We say that an orbifold is **good** if it is covered by a manifold, otherwise it is **bad**. If an orbifold admits a geometric structure then it is good because it is covered by a homogeneous space. Quite surprisingly, the converse is also true.

**Theorem 2.4.6.** Every good two-dimensional orbifold is isomorphic to the quotient of one among $S^2$, $\mathbb{R}^2$ or $\mathbb{H}^2$ by a discrete group of isometries.

The last result gives also information about surfaces. For example, if $S$ is a smooth surface and $\Gamma$ acts properly discontinuously on $S$, then the
CHAPTER 2. SURFACE THEORY

quotient $O = S/\Gamma$ is an orbifold. Since $O$ admits a geometric structure, we
deduce that $S$ admits a Riemannian metric of constant curvature such that
$\Gamma$ acts by isometries. In particular, every finite group $\Gamma < \text{Diffeo}(S)$ acts by
isometries with respect to some metric of constant curvature on $S$.

Remark 2.4.7. It is interesting to contrast this result with the Nielsen Re-
alization Problem (Theorem 2.3.9). The great difficulty in that problem is
that it is not at all obvious that finite groups of the mapping class group
can be realized as finite groups of diffeomorphisms.

The next step is to ask whether every good orbifold is finitely covered by
a surface. Also in this case the answer is often positive.

Theorem 2.4.8. If $O$ is a two-dimensional good orbifold whose fundamental
group is finitely generated, then it is finitely covered by a surface.

To see why this theorem should be true, let $O = \mathbb{H}^2/\Gamma$ where $\Gamma$ is a
finitely generated discrete subgroup of $\text{Isom}^+(\mathbb{H}^2) \cong \text{PSL}(2, \mathbb{R})$. The Selberg
Lemma implies that there exists a finite index subgroup $H < \Gamma$ which is
torsion free. In particular, this implies that $H$ does not contain any elliptic
element (see Remark 2.1.1). As a consequence we have that $S := \mathbb{H}^2/H$ is a
hyperbolic surface and the quotient $S \to O$ is a finite cover.

We have already noticed that if an orbifold $O$ is isomorphic to the quotient
of a simply connected manifold by a discrete group $\Gamma$, then its fundamental
group is $\Gamma$ itself. Recall that by Theorem 2.4.6 we know that any good
orbifold is the quotient of a simply connected homogeneous space. Then, at
least when $O$ is hyperbolic and orientable, we can proceed as above to prove
Theorem 2.4.8. The analysis of spherical and Euclidean orbifolds is generally
simpler than the hyperbolic case and one can work out the non-orientable
case taking two-folds orientation covers.

2.4.3 Orbifold characteristic

We now want to define an analogue of the Euler characteristic for orbifolds.
First of all we require that such an orbifold characteristic $\chi(O)$ coincides
with the standard Euler characteristic whenever the orbifold $O$ is a smooth
surface. Moreover, a peculiar property of the Euler characteristic is that it is multiplicative under finite topological covers. That is, if \( S \to S' \) is a \( d \)-sheeted cover between surfaces (the degree \( d \) of a cover is the number of pre-images of any point of the base space) then \( \chi(S) = d\chi(S') \). Mimicking this feature, we would like to define the Euler characteristic for orbifolds in such a way that if \( p: O' \to O \) is a \( d \)-sheeted orbifold cover then \( \chi(O') = d\chi(O) \)
(similarly, the degree of an orbifold cover is the cardinality of the pre-image of a regular point).

In view of Theorem 2.4.8, these requirements uniquely define the characteristic for every good orbifold with finitely generated fundamental group. For, any such orbifold \( O \) is of the form \( S/\Gamma \) for some finite group \( \Gamma \) and hence \( \chi(O)/|\Gamma| \) must be equal to \( \chi(S) \). For our purposes, it will be convenient to find a more explicit formula.

We say that a two-dimensional orbifold \( O \) is of finite type if its singular locus consists only of finitely many cone points \( \{x_1, \ldots, x_k\} \) and the underlying surface \( X_O \) is of finite type.

Remark 2.4.9. Notice that according to our definitions an orbifold of finite type has no reflection lines nor corner reflectors and the surface \( X_O \) is orientable. Notice also that \( X_O \) turns out to have empty boundary because boundary components on the underlying surface are generated by reflection lines and corner reflections. This assumptions are not essential and one can develop the theory also in the non-orientable setting. We preferred not to do so because we will never need to work in non-orientable contexts.

Example 2.4.10. If an (orientable) orbifold \( O \) is compact and its singular points are all cone points, then it is of finite type because \( X_O \) is of finite type (it is compact) and cone points form a discrete subset of \( X_O \).

Example 2.4.11. If \( S \) is a surface of finite type (with empty boundary) and \( \Gamma < \text{Diffeo}^+(S) \) is a finite subgroup of orientation-preserving diffeomorphisms then \( O = S/\Gamma \) is of finite type. For, all the singular points are cone points because reflection lines and corner reflectors can only be obtained through orientation-reversing diffeomorphism. Then one can prove that also \( X_O \) is of finite type and that there are only finitely many cone points. (One way to do it is to choose an hyperbolic metric on \( S \) such that \( \Gamma \) acts by isometries. This induces a (non-complete) hyperbolic metric of finite area on \( X_O \setminus \{\text{cone points}\} \) and one can use the Gauss-Bonnet formula (adapted for non-geodesic boundary) to show that there are only finitely many cone points.)

Given an orbifold of finite type \( O \), choose for any cone point \( x_i \) an embedded disk \( D_i \subset X_O \) containing \( x_i \) and take these disks small enough so that they are all disjoint. We denote by \( \tilde{X}_O \) the surface obtained from \( X_O \)
removing the interior of these disks
\[ \hat{X}_\mathcal{O} = X_\mathcal{O} \setminus \bigcap_{i=1}^k \hat{D}_i. \]

Inspired by the fact that cutting a surface along a closed simple curve does not change its Euler characteristic, we are tempted to define
\[ \chi(\mathcal{O}) := \chi(\hat{X}_\mathcal{O}) + \sum_{i=1}^k \chi(D_i) \]
where \( \chi(\hat{X}_\mathcal{O}) \) is the Euler characteristic and \( \chi(D_i) \) should be defined keeping in mind that \( D_i \) contains a cone point. In particular, if the cone point \( x_i \in D_i \) has order \( d_i \), then \( D_i \) should be thought of as the quotient of a disc \( D^2 \) by a rotation of order \( d_i \). Thus we have a \( d_i \)-fold branched cover \( D^2 \to D_i \) and we are hence induced to define \( \chi(D_i) = \chi(D^2)/d_i = 1/d_i \). We give the following:

**Definition 2.4.12.** If \( \mathcal{O} \) is an orbifold of finite type with \( k \) cone points \( x_1, \ldots, x_k \) of orders \( d_1, \ldots, d_k \), then its orbifold characteristic is
\[ \chi(\mathcal{O}) := \chi(\hat{X}_\mathcal{O}) + \sum_{i=1}^k \frac{1}{d_i}. \]

By the definition, it is clear that if an orbifold of finite type \( \mathcal{O} \) is a surface then its orbifold characteristic coincides with its Euler characteristic. To show that the orbifold characteristic has the properties we required at the beginning, all we have to prove is that the characteristic of an orbifold cover is the right multiple of that of the base space.

**Proposition 2.4.13.** If \( p: \mathcal{O}' \to \mathcal{O} \) is an \( n \)-sheeted orbifold cover between orbifolds of finite type, then
\[ \chi(\mathcal{O}') = n \chi(\mathcal{O}). \]

**Proof.** As above, let \( x_1, \ldots, x_k \) be the singular points of \( \mathcal{O} \) of order \( d_1, \ldots, d_k \) and let \( D_1, \ldots, D_k \) be disjoint disks containing them. Write \( \hat{X}_\mathcal{O} \) for the finite type surface obtained from \( X_\mathcal{O} \) removing the interior of the disks \( D_i \). Now, assume the disks \( D_i \) are small enough to be trivializing sets for the cover \( p \) and let \( \hat{X}_{\mathcal{O}'} = p^{-1}(\hat{X}_\mathcal{O}) \). Then \( p \) restrict to an \( n \)-sheeted topological cover between surfaces of finite type
\[ p|_{\hat{X}_{\mathcal{O}'}^\ast}: \hat{X}_{\mathcal{O}'} \to \hat{X}_\mathcal{O}. \]

The inverse image of any disk \( D_i \) is given by a disjoint union of disks \( D'_{ij} \) with \( j = 1, \ldots, k_i \) and each of these disks can possibly contain a cone
point $y_{ij}$ of order $d_{ij}$. Notice that the order $d_i$ of the cone point $x_i$ must be a multiple of each of the $d_{ij}$. That is, for every $j = 1, \ldots, k_i$ there exists $a_{ij}$ such that $d_i = a_{ij}d_{ij}$.

The key point is that $a_{ij}$ is equal to the number of sheets of the branched cover $D'_{ij} \to D_i$ and this is given by the degree of the restriction to the boundary of the cover map $p$

$$p|_{\partial D'_{ij}} : \partial D'_{ij} \to \partial D_i.$$ 

Thus we have

$$\sum_{j=1}^{k_i} a_{ij} = \sum_{j=1}^{k_i} \deg(p|_{\partial D'_{ij}})$$

and the latter is equal to $n$ because $p|_{\tilde{X}_{O'}}$ is an $n$-fold cover.

Summing up we conclude:

$$\chi(O') = \chi(\tilde{X}_{O'}) + \sum_{i,j} \chi(D'_{ij})$$

$$= \chi(\tilde{X}_{O'}) + \sum_{i=1}^{k} \sum_{j=1}^{k_i} \frac{1}{d_{ij}}$$

$$= n\chi(\tilde{X}_{O}) + \sum_{i=1}^{k} \sum_{j=1}^{k_i} \frac{a_{ij}}{d_i}$$

$$= n\chi(\tilde{X}_{O}) + \sum_{i=1}^{k} \frac{n}{d_i} = n\chi(O).$$

Now that we have a well-defined characteristic, we can find an analogue to the Gauss-Bonnet theorem. Indeed, if $S$ is a surface with a Riemannian metric of finite area and $\Gamma < \text{Isom}^+(S)$ is a finite group, then the orbifold $O = S/\Gamma$ is naturally a metric space (it has a Riemannian metric well-defined away from the singular points) and its area is equal to $\text{Area}(S)/|\Gamma|$.

By the Gauss-Bonnet formula, we know that if a surface $S$ admits a complete metric of constant curvature $K$ and finite area, then $2\pi \chi(S) = K \text{Area}(S)$ (see Subsection 2.1.1). Since both the area and the characteristic are scaled by the same constant under finite covers, we deduce that for every finite subgroup $\Gamma < \text{Isom}^+(S)$ we have

$$2\pi \chi(S/\Gamma) = K \text{Area}(S/\Gamma).$$

Recall that by Theorem 2.4.6 every good orbifold admits a geometric structure. If $O$ is an orbifold of finite type, then one can prove that its orbifold fundamental group is finitely generated. In particular, 2.4.8 applies
to orbifolds of finite type and we deduce that such an orbifold $O$ is obtained quotienting a Riemannian surface of constant curvature $S$ by a finite group of isometries. Thus, if the area of $O$ is finite, then it satisfy $K \text{Area}(O) = 2\pi \chi(O)$.

The last observation immediately implies that an orbifold of finite type admits only one kind of geometric structure of finite area and this is determined by the sign of its orbifold characteristic. If an orbifold has negative characteristic, it is easy to show that it admits a hyperbolic structure of finite area using the same techniques we used to build hyperbolic structures for surfaces (see Subsection 2.1.1). We say that an orbifold is \textit{hyperbolic} if it of finite type and it is endowed with a hyperbolic structure of finite area.

An unexpected application of the Gauss-Bonnet formula for orbifold is given by the following proposition.

\textbf{Proposition 2.4.14.} If $O$ is a hyperbolic orbifold, then its area is at least $\pi/21$.

\textit{Sketch of the Proof.} We already observed that when an orbifold of finite type $O$ has a hyperbolic structure of finite type then the area is given by \[ \text{Area}(O) = -2\pi \chi(O), \] where $\chi(O)$ is of the form 

\[ 0 > \chi(O) = \chi(\hat{X}_O) + \sum \frac{1}{d_i} = \chi(X_O) + \sum \left( 1 - \frac{1}{d_i} \right). \]

Hence it is enough to prove that the biggest strictly negative number that is obtained in such a way is $-1/42$. A simple case check leads to the result. \qed

\textbf{Corollary 2.4.15.} If $S$ is a hyperbolic surface and $\Gamma < \text{Isom}^+(S)$ is a finite subgroup, then $|\Gamma| \leq 42|\chi(S)|$.

\textit{Proof.} The quotient $S/\Gamma$ is naturally a hyperbolic orbifold and hence must have area greater or equal to $\pi/21$. Since $S \to S/\Gamma$ is a cover with $|\Gamma|$ sheets, we have 

\[ \frac{\pi}{21} \leq \text{Area}(S/\Gamma) = \frac{\text{Area}(S)}{|\Gamma|}. \]

By the Gauss-Bonnet formula we also have that $\text{Area}(S) = -2\pi \chi(S)$. Thus we obtain 

\[ |\Gamma| \leq -42\chi(S) \]

as required. \qed

In view of the Nielsen Realization Theorem (Theorem 2.3.9) we also obtain this remarkable result:

\textbf{Theorem 2.4.16.} Every finite subgroup of the mapping class group of a surface $S$ with $\chi(S) < 0$ has cardinality bounded by $42|\chi(S)|$. 

2.4. SUBGROUPS OF THE MAPPING CLASS GROUP

2.4.4 Orbifold mapping class group

For the remainder of the chapter, $S$ will always be a surface with negative Euler characteristic and without boundary components. Given a finite subgroup $F < \text{Mod}(S)$, by the Nielsen Realization Theorem there exists a hyperbolic metric on $S$ such that $F$ can be realized as a subgroup of isometries $F' < \text{Isom}(S)$. Thus the quotient $O = S/F'$ is a hyperbolic orbifold and the quotient map $p: S \to O$ is an orbifold cover (recall that an orbifold is hyperbolic if it is of finite type and it has a hyperbolic structure of finite area).

Remark 2.4.17. Sometimes we will directly write $S/F$ with $F < \text{Mod}(S)$. If that is the case, it is understood that the quotient is meant with respect to a realization of $F$ in $\text{Isom}(S)$. A priori the resulting orbifold may depend on the specific realization. Still, this will not be an issue because we will only use the fact that $S \to S/F$ is an orbifold cover without being particularly interested in the actual orbifold that we are covering.

Recall that a map between orbifolds $\varphi: O \to O'$ is a morphism if it is coherent with the local quotient structure. Since the singular loci of hyperbolic orbifolds consist only of cone points, it is easy to see that the isomorphisms of an orbifold $\varphi: O \to O'$ are actually homeomorphisms of the underlying surface $\varphi: X_O \to X_O$ that send cone points of a certain order to cone points of the same order. Vice versa, such homeomorphisms are isomorphisms with respect to the orbifold structure and we will often call them orbifold homeomorphisms.

The orbifold mapping class group $\text{Mod}(O)$ of the orbifold $O$ is defined as the set of orbifold homeomorphisms of $O$ up to isotopies which preserve punctures and cone points. That is, two orbifold homeomorphisms are isotopic if they are isotopic as homeomorphisms of the underlying surface $\varphi: X_O \to X_O$ via an isotopy that also fixes each cone point at every time. Notice that when $O$ is a surface its orbifold mapping class group coincides with the usual mapping class group, hence the notation $\text{Mod}(O)$ is justified.

We now need to explore some relations between $\text{Mod}(S)$ and $\text{Mod}(O)$. Denote by $O^*$ the punctured surface obtained by $O$ replacing each cone point with a puncture (strictly speaking, writing $X_O^*$ instead of $O^*$ would be more coherent with the notation used so far. Still, we prefer to use the latter for sake of readability). Notice that $\text{Mod}(O)$ is naturally identified with the subset of $\text{Mod}(O^*)$ consisting of homeomorphisms that fix the sets of punctures coming from cone points with the same order. In particular, we have that $\text{Mod}(O)$ is a subgroup of finite order of $\text{Mod}(O^*)$. Moreover, letting $S^* = p^{-1}(O^*)$ we obtain a normal cover $p: S^* \to O^*$ whose group of deck transformations is $F'$. 
Notice that if an orbifold homeomorphism $\varphi$ of $\mathcal{O}$ lifts to a homeomorphism $\tilde{\varphi}$ of $S$ (i.e. there exists $\tilde{\varphi}$ which commutes with the projection, $p \circ \tilde{\varphi} = \varphi \circ p$), then $\tilde{\varphi}$ must lie in the normalizer $N(F') < \text{Homeo}(S)$. Indeed, $\tilde{\varphi}$ restricted to $S^*$ is a lifting of $\varphi$ restricted to $\mathcal{O}^*$.

Given $\psi_1 \in F'$, then $\tilde{\varphi} \circ \psi_1$ is another lifting of $\varphi$. Since $p : S^* \to \mathcal{O}^*$ is a normal cover, this implies that there exists $\psi_2 \in F'$ such that $\psi_2 \circ \tilde{\varphi} = \tilde{\varphi} \circ \psi_1$ on $S^*$. By continuity we deduce that the equality holds also in $S$.

Conversely, it is easy to see that every element $\tilde{\varphi} \in N(F')$ induces an orbifold homeomorphism $\varphi$ setting $\varphi(x) = p(\tilde{\varphi}(\tilde{x}))$ where $\tilde{x}$ is a pre-image of $x$ under $p$. Hence we have a homomorphism $\Psi : N(F') \to \text{Homeo}(\mathcal{O})$ with $\ker(\Psi) = F'$ whose image consists of the orbifold homeomorphisms which admit liftings to $S$.

Actually, an analogous result holds for the mapping class groups. Let $\tilde{\text{Mod}}(\mathcal{O})$ denote the subgroup of the orbifold mapping class group $\text{Mod}(\mathcal{O})$ whose elements admit liftings in $\text{Mod}(S)$ (notice that if a homeomorphism $\varphi$ lifts than so does every homeomorphism isotopic to $\varphi$). The relative mapping class group $\text{Mod}(S, \mathcal{O}) < \text{Mod}(S)$ is defined as the subgroup of isotopy classes of liftings of homeomorphisms of $\mathcal{O}$. The following holds.

**Theorem 2.4.18.** If $S$ is a surface with $\chi(S) < 0$ and $F < \text{Mod}(S)$ is a finite subgroup and $\mathcal{O}$ is the quotient orbifold, then the relative mapping class group $\text{Mod}(S, \mathcal{O})$ is equal to the normalizer $N(F) < \text{Mod}(S)$. Moreover, the above map $\Psi$ is well-defined up to isotopies and induces a map

$$\Phi : N(F) \to \tilde{\text{Mod}}(\mathcal{O})$$

with kernel $F$ and image $\tilde{\text{Mod}}(\mathcal{O})$.

A proof can be found in [MH75]. We are now going to show that the images of the maps $\Phi$ and $\Psi$ are finite index subgroups. We begin with a simple algebraic lemma:
Lemma 2.4.19. For every natural number \( n \), a finitely generated group \( \Gamma \) has only finitely many subgroups of index \( n \).

Proof. For any subgroup \( H < \Gamma \) of index \( n \) the group \( \Gamma \) acts by left multiplication on the set \( \Gamma / H \) of left cosets \( gH \subset \Gamma \). The set \( \Gamma / H \) is a finite set with \( n \) elements and \( H \) is equal to the stabilizer of the coset \( H \in \Gamma / H \). Thus the number of subgroups of index \( n \) in \( \Gamma \) is bounded by the number of actions of \( \Gamma \) on a set of \( n \) elements.

To conclude we have to show that there are only finitely many homomorphisms from \( \Gamma \) to the symmetric group \( S_n \), but this is trivially true because it is enough to specify the image of a finite generating set of \( \Gamma \). \( \square \)

Lemma 2.4.20. If \( p: E \to X \) is a finite regular cover and \( \pi_1(E) \) is finitely generated, then the subgroup of homeomorphisms of \( X \) that admit liftings to \( E \) is of finite index in \( \text{Homeo}(X) \).

Proof. Recall that a map \( \varphi: Y \to X \) lifts to a mapping \( \tilde{\varphi}: Y \to E \) sending the point \( y_0 \in Y \) to a point \( e_0 \in p^{-1}(\varphi(y_0)) \) if and only if \( \varphi_* \left( \pi_1(Y,y_0) \right) \) is contained in \( p_* \left( \pi_1(E,e_0) \right) \).

Since the cover is regular, the image \( p_* \left( \pi_1(E,e_0) \right) \) is a normal subgroup and hence it does not depend on the choice of any base point. For this reason we can forget about the base point \( p(e_0) \) and see \( \pi_1(E) \) as a subgroup of \( \pi_1(X) \).

Now, let \( \varphi: X \to X \) be a homeomorphism. By the above remark we have that \( \varphi \) lifts to a homeomorphism of \( E \) if and only if the isomorphism \( \varphi_*: \pi_1(X) \to \pi_1(X) \) sends \( \pi_1(E) \) onto \( \pi_1(E) \). Let \( d \) be the degree of the covering map \( p \). Then \( \varphi_* \left( \pi_1(E) \right) \) must be a normal subgroup of index \( d \) in \( \pi_1(X) \). By Lemma 2.4.19 we know that there are only finitely many subgroups of \( \pi_1(X) \) of order \( d \), thus the subgroup of homeomorphisms that send \( \pi_1(E) \) onto itself has finite index in \( \text{Homeo}(X) \). \( \square \)

Corollary 2.4.21. The images of the maps \( \Psi \) and \( \Phi \) of Theorem 2.4.18 have finite index in \( \text{Homeo}(O) \) and \( \text{Mod}(O) \) respectively.

Proof. Notice that an orbifold homeomorphisms of \( O \) lifts to \( S \) if and only if its restriction to \( O^* \) lifts to a homeomorphism of \( S^* \). (One implication is clear, the other follows easily extending by continuity a lifting in \( \text{Homeo}(S^*) \) to a lifting in \( \text{Homeo}(S) \).) Thus it is clear that the image of \( \Psi \) has finite index in \( \text{Homeo}(O) \) because by Lemma 2.4.20 it is equal to the intersection of \( \text{Homeo}(O) \) with a finite index subgroup of \( \text{Homeo}(O^*) \).

It is also clear that the image of \( \Phi \) has finite index in \( \text{Mod}(O) \) because it is equal to the image of \( \text{Im}(\Psi) \) under the surjective map \( \text{Homeo}(O) \to \text{Mod}(O) \). \( \square \)

The following proposition follows easily from Corollary 2.4.21.
Proposition 2.4.22. If \( S \) is a surface with \( \chi(S) < 0 \), then in the mapping class group \( \text{Mod}(S) \) there are only finitely many conjugacy classes of finite groups.

Proof. Since the cardinality of a finite subgroup \( F < \text{Mod}(S) \) is bounded by Theorem 2.4.16, it is enough to prove that there are only finitely many conjugacy classes of groups with fixed cardinality. Let \( F \) and \( F' \) be finite subgroups of \( \text{Mod}(S) \) with \( |F| = |F'| \). The quotient orbifolds \( O = S/F \) and \( O' = S/F' \) have both orbifold characteristic equal to \( \chi(S)/|F| \).

By the combinatorial description of orbifolds in terms of underlying surfaces and cone points it is easy to see that there are only finitely many orbifolds with a fixed orbifold characteristic. Hence we can assume that the orbifolds \( O \) and \( O' \) are homeomorphic.

Let \( \varphi: O \to O' \) be such a homeomorphism. If \( \varphi \) lifts to an homeomorphism \( \tilde{\varphi}: S \to S \) then the conjugation by \( \tilde{\varphi} \) sends the group of deck transformations of \( O \) to that of \( O' \), thus we have \( F' = [\tilde{\varphi}] F [\tilde{\varphi}]^{-1} \). We can hence conclude because by definition we have that such a lifting \( \tilde{\varphi} \) exists if and only if \( \varphi \) lies in the image of the above map \( \Psi \) and by Corollary 2.4.21 this is a finite index normal subgroup of \( \text{Homeo}(O) \).

The last result we need is some sort of control over the type of homeomorphisms. Since we have an inclusion

\[
\text{Mod}(O) \subset \text{Mod}(O) \hookrightarrow \text{Mod}(O^*)
\]

it is natural to ask whether \( \Phi: N(F) \to \text{Mod}(O^*) \) is type-preserving.

Proposition 2.4.23. For every element \( f \in \text{Mod}(O) \) and for every lifting \( \tilde{f} \in \Phi^{-1}(f) \) the type of \( f \) as an element of \( O^* \) is the same as that of \( \tilde{f} \) in \( \text{Mod}(S) \).

Proof. Since \( \Phi \) has finite kernel, it is clear that \( f \) is periodic if and only if so is \( \tilde{f} \). Moreover, if \( f \) fixes a finite family of disjoint essential curves \( a_1, \ldots, a_n \) then \( \tilde{f} \) fixes the finite family \( p^{-1}(a_1), \ldots, p^{-1}(a_n) \) and these curves are disjoint and essential (see Lemma 2.4.25 below). The pseudo-Anosov case is slightly more delicate.

Let \( \varphi \) be an orbifold homeomorphism of \( O \) that lifts to a homeomorphism \( \tilde{\varphi} \) of \( S \). We have to show that if \( \varphi \) is a pseudo-Anosov homeomorphism of \( O^* \), then \( \tilde{\varphi} \) is a pseudo-Anosov homeomorphism of \( S \).

If \( (F^s, \mu^s) \) and \( (F^u, \mu^u) \) are the stable and unstable foliations of \( \varphi \) on \( O^* \), then by Lemma 2.3.4 we have that the restriction of \( \tilde{\varphi} \) to \( S^* \) is a pseudo-Anosov homeomorphism with stable and unstable foliations \( p^{-1}(F^s, \mu^s) \) and \( p^{-1}(F^u, \mu^u) \). To prove the lemma it is hence enough to show that these foliations are actually foliations of \( S \) and not only of \( S^* \).

The only reason why a foliation of \( S^* \) could not be a foliation of \( S \) is that one of the punctures of \( S^* \) that are regular points in \( S \) corresponds to
the centre of a 1-pronged saddle. Still, the extra punctures of $S^*$ are exactly
the pre-images of the cone points. In particular, if $x$ is a cone point of $O$
of order $d$ and $F$ is a foliation of $O^*$ where $x$ correspond to the centre of a
$k$-pronged saddle, then the pre-images $p^{-1}(x)$ are centres of saddles with $dk$
prongs. In particular, these cannot be 1-pronged saddles. 

Remark 2.4.24. Notice that we have not proven that if $\tilde{f}$ is reducible then so
is $f$. Thus, even if $\tilde{f}$ is reducible, a priori we only know that $f$ cannot be
pseudo-Anosov but it could well be periodic and irreducible.

2.4.5 Centralizers of finite subgroups

We define the curve complex $C(O)$ of an orbifold $O$ as the curve complex of the
punctured surface $O^*$. As before, let $F$ be a finite subgroup of the mapping
class group of a surface $S$ with $\chi(S) < 0$, $F' < \text{Isom}(S)$ a realization for
$F$ as a group of isometries and $O = S/F'$ the quotient orbifold. If $\alpha$ is a
simple closed curve in $O^*$, then its pre-image under the cover map $p^{-1}(\alpha)$ is
a union of disjoint curves $\beta_1, \ldots, \beta_n$ of $S$. We have the following:

Lemma 2.4.25. The curve $\alpha$ is an essential curve in $O^*$ if and only if all
of the $\beta_i$’s in $p^{-1}(\alpha)$ are essential in $S$.

Proof. We begin with the only if part. If $\beta_i$ is a separating curve and $N$
is a component of $S$ cut along $\beta_i$, then also $\alpha$ must be separating and
$p|_N: N \to p(N)$ must be an orbifold cover of a component of $O$ cut along $\alpha$.
Notice that it is possible to define orbifolds with boundary in the same way
one does for manifolds and the theory remains unchanged. In particular, it
is still defined the orbifold characteristic and is equal to

$$\chi(O) = \chi(X_O) - \sum_{\text{cone points}} (1 - \frac{1}{d_i})$$

where the underlying surface $X_O$ may have non-empty boundary and the
d_i’s are the orders of the cone points. By hypothesis, $p(N)$ cannot be a once
punctured disc nor a disk with a single cone point. If it is a disk with two
cone points of order two then it has characteristic $\chi(p(N)) = 0$. Since there
are no punctures in $p(N)$, also $N$ cannot have punctures and hence should be a
disk. Still, a disk cannot cover $p(N)$ because it has strictly positive Euler
characteristic. In all the other cases $p(N)$ has negative orbifold characteristic
and hence also $N$ must have negative characteristic. In particular $N$ cannot
be a once punctured disk.

For the converse implication we will show that the pre-image of an
inessential curve of $O^*$ is inessential in $S^*$ and a hence also on $S$. Assume
that $\alpha$ is a separating curve on $O^*$ and a component $M$ of $O^*$ cut along $\alpha$ is a
disk, possibly with a puncture. Then every component $N$ of $p^{-1}(M) \subset S^*$ has
non-empty boundary. If $M$ is a disk without punctures then by characteristic arguments also $N$ is a disk, hence its boundary (that is one of the $\beta_i$'s) is not essential. If $M$ is a once punctured disk then $\chi(M) = 0$ and hence $N$ can only be a once punctured disk or an annulus. Still, $N$ cannot be an annulus because $M$ is not compact. Thus $N$ is a once punctured disk and again we find that one of the $\beta_i$'s is non essential.

In a similar way it is also possible to show that the curves $\beta_i$ are pairwise non-isotopic. Since isotopies of $\mathcal{O}^*$ lift to isotopies of $S$, in view of Lemma 2.4.25 it is well-defined the relation that to each class of curves $c \in \mathcal{C}(\mathcal{O})$ assigns its pre-image $p^{-1}(c) \subseteq \mathcal{C}(\mathcal{S})$. Notice that the set $p^{-1}(c)$ has diameter at most one because the pre-images of a curve are disjoint. It follows that we have a coarsely well-defined injective map $p^* : \mathcal{C}(\mathcal{O}) \to \mathcal{C}(\mathcal{S})$.

Actually, the sets $p^{-1}(c)$ are $F$-invariant. For, by definition $F$ is realized as a group of isometries $F'$ which is the group of deck transformation of $p$ and this clearly fixes the sets $p^{-1}(\alpha)$ for every curve $\alpha$. Being the sets $p^{-1}(c)$ both invariant and of diameter one, we deduce that the image of the map $p^* : \mathcal{C}(\mathcal{O}) \to \mathcal{C}(\mathcal{S})$ is contained in $\text{Fix}_1(F)$ where $\text{Fix}_k(F) \subseteq \mathcal{C}(\mathcal{S})$ denotes the $k$-coarsely fix set:

$$\text{Fix}_k(F) := \{c \in \mathcal{C}(\mathcal{S}) \mid d(c, f(c)) \leq k \text{ for every } f \in F^*\}.$$ 

Moreover, the map $p^* : \mathcal{C}(\mathcal{O}) \to \text{Fix}_1(F)$ is coarsely surjective. Indeed, if a curve class $c \in \mathcal{C}(\mathcal{S})$ belongs to the coarse fix set $\text{Fix}_1(F)$, then for every $f, g \in F$ we have $d(f(c), g(c)) = d(c, f^{-1}g(c)) \leq 1$. Let $F'$ be a realization of $F$ by isometries and let $\alpha$ be the unique geodesic representing $c$. Then the curves $\varphi_i(\alpha)$ with $\varphi_i \in F'$ either coincide or are disjoint. Since $p^{-1}(p(\alpha)) = \{\varphi_i(\alpha) \mid \varphi_i \in F'\}$, we conclude that $p(\alpha)$ is a simple closed curve in $\mathcal{O}$. Moreover, by Lemma 2.4.25 $p(\alpha)$ is essential and hence $\alpha$ belongs to the coarse image of $p^*(\mathcal{C}(\mathcal{O}))$. Thus the map $p^*$ is a coarse bijection between $\mathcal{C}(\mathcal{O})$ and $\text{Fix}_1(F)$.

**Remark 2.4.26.** Since in the above arguments we made no requirements on the quotient $\mathcal{O}$, we have just proved that for every finite subgroup $F < \text{Mod}(\mathcal{S})$ if $\mathcal{O} = S/F$ is a triangle orbifold (that is, $\mathcal{O}^*$ is a three-punctured sphere), then $\text{Fix}_1(F)$ is empty in $\mathcal{C}(\mathcal{S})$.

The action of the normalizer $N(F) \subseteq \text{Mod}(\mathcal{S})$ restricts to an action on the $k$-coarsely fixed set. Indeed, let $c \in \text{Fix}_k(F)$ and $g \in N(F)$, then for every $f \in F$ the conjugate $f' = g^{-1}fg$ is also in $F$ and

$$d(g(c), fg(c)) = d(g(c), gf'(c)) = d(c, f'(c)) \leq k.$$ 

In particular, also the action of the centralizer $C(F) < N(F)$ restricts to $\text{Fix}_k(F)$. We say that an action on a metric space is *coarsely transitive* if there exists a constant $K$ such that for every pair of points $x$ and $y$ there exists an element $g$ of the group such that $d(g \cdot x, y) \leq K$. The following holds:
Proposition 2.4.27. If $F$ is a finite subgroup of the mapping class group, then its centralizer $C(F)$ acts coarsely transitively on $\text{Fix}_1(F) \subseteq C(S)$.

Proof. Since the group $F$ is finite, its centralizer is a finite index subgroup of $N(F)$. Hence it is enough to prove that the latter acts coarsely transitively on $\text{Fix}_1(F)$. Recall that by Theorem 2.4.18 the normalizer $N(F)$ is equal to the relative mapping class group $\text{Mod}(S, O)$ and there is a map $\Phi: \text{Mod}(S, O) \to \text{Mod}(O)$. Now the idea is to compare the action of $N(F)$ with a coarse action of $\text{Mod}(O)$.

First of all we notice that the statement is vacuously true if $O$ is a triangle orbifold because $\text{Mod}(O)$ (and thus $C(F)$) is finite and $\text{Fix}_1(F)$ is empty (see Remark 2.4.26).

In the other cases recall that we have a coarsely well-defined map $p^*: C(O) \to \text{Fix}_1(F)$ that is a coarse bijection. It follows that the action $\text{Mod}(O^*) \actson C(O)$ induces a coarse action $\text{Mod}(O^*) \actson \text{Fix}_1(F)$. Since the map $N(F) \to \text{Mod}(O^*)$ is defined taking the natural maps induced by the quotient, the action $N(F) \actson \text{Fix}_1(F)$ is coarsely coherent with the coarse action $\text{Mod}(O^*) \actson \text{Fix}_1(F)$. That is, for every $g \in N(F)$ and $c \in \text{Fix}_1(F)$, the class $g(c)$ is contained in the coarse image of $\Phi(g)(p(c))$ under $p^*$.

Summing up, to prove that the action $N(F) \actson \text{Fix}_1(F)$ is coarsely transitive it is enough to prove the same property for the coarse action induced by

$$\Phi: N(F) \to \text{Mod}(O^*) \actson C(O).$$

Recalling that the image of $\Phi$ is a finite index subgroup of $\text{Mod}(O)$ (see Corollary 2.4.21) and $\text{Mod}(O)$ has finite index in $O^*$, it is enough to show that the coarse action $\text{Mod}(O^*) \actson \text{Fix}_1(F)$ is coarsely transitive. If $O^*$ is the once-punctured torus or the four-punctured sphere then it is clear because the action $\text{Mod}(O^*) \actson C(O)$ is transitive. In the other cases $C(O)$ is connected and $p^*$ is 1-Lipschitz, hence the thesis follows because the action $\text{Mod}(O^*) \actson C(O)$ is coarsely surjective. (From the fact that there are only finitely many essential curves up to diffeomorphism follows easily that $\text{Mod}(O^*)$ acts coarsely transitively on $C(O)$.)

Recall that by Subsection 2.2.3 the mapping class group $\text{Mod}(S)$ of a non-sporadic surface $S$ is weakly relatively hyperbolic. In fact, we proved that $\text{Mod}(S)$ has a relative metric $d_v$ such that for every fixed base point $\bar{c} \in C(S)$ the map $\psi_{\bar{c}}: (\text{Mod}(S), d_v) \to C(S)$ obtained sending $g \in \text{Mod}(S)$ to $g(\bar{c}) \in C(S)$ is a quasi-isometry (see Remark 2.2.21). Thus the Gromov boundaries $\partial_\infty \text{Mod}(S)$ and $\partial_\infty C(S)$ are naturally identified. (This identification is natural in that it does not depend on the choice of $\bar{c}$. See Remark 2.2.23.)

Proposition 2.4.27 gives us the following:

Corollary 2.4.28. Let $S$ be a non-sporadic surface and $F < \text{Mod}(S)$ a finite subgroup. Then, under the natural identification of $\partial_\infty \text{Mod}(S)$ with
\( \partial_\infty C(S) \), the boundary at infinity of the centralizer \( \partial_\infty C(F) \) in the relative metric is naturally identified with \( \partial_\infty (\text{Fix}_1(F)) \subset \partial_\infty C(S) \).

**Proof.** Choose a base point \( \tilde{c} \in C(S) \) to obtain a quasi-isometry \( \Psi_{\tilde{c}} \) and let \( D \) denote the distance of \( \tilde{c} \) from the coarse fixed set \( \text{Fix}_1(F) \). We have already noticed that the action of \( C(F) \) fixes \( \text{Fix}_1(F) \), hence we have that for every \( g \in C(F) \) its image \( \Psi_{\tilde{c}}(g) = g(\tilde{c}) \) is \( D \)-close to \( \text{Fix}_1(F) \). It follows that \( \partial_\infty C(F) \subseteq \partial_\infty (\text{Fix}_1(F)) \).

The other inclusion follows easily from the fact that the action \( C(F) \rtimes \text{Fix}_1(F) \) is coarsely transitive.

Notice that \( \partial_\infty (\text{Fix}_1(F)) \) is obviously contained in \( \text{Fix}_0(F) \), where the latter is the set of points in \( \partial_\infty C(S) \) that are fixed by the action of \( F \). Moreover, the action of \( N(F) \) stabilizes both \( \partial_\infty (\text{Fix}_1(F)) \) and \( \text{Fix}_0(F) \) (that is, it fixes the sets without necessarily fixing their elements). Thus \( N(F) \) is contained in the stabilizers of those sets (by definition, the stabilizer of a set is the maximal subgroup that stabilizes it). It turns out that these inclusions are equalities:

**Proposition 2.4.29.** Let \( S \) be a non-sporadic surface and \( F < \text{Mod}(S) \) a finite subgroup such that \( S/F \) is not a triangle orbifold. Then the set \( \partial_\infty (\text{Fix}_1(F)) \) is equal to \( \text{Fix}_0(F) \) and the normalizer \( N(F) \) is equal to the stabilizer of \( \text{Fix}_0(F) \).

We will not prove this proposition because it requires the definition of Teichmüller space and we preferred not to introduce this theory. For a proof, see [Mah11, Section 2]. By the same reason we will not give the proof of the following:

**Proposition 2.4.30.** If \( F, F' \) are two finite subgroups of the mapping class group of \( \text{Mod}(S) \) of a non-sporadic surface, then \( \partial_\infty C(F) \cap \partial_\infty C(F') = \partial_\infty C(F'') \) where \( F'' \) is the subgroup generated by \( F \cup F' \).

**Remark 2.4.31.** The idea for both Proposition 2.4.29 and 2.4.30 is to work in the Teichmüller space \( T(S) \) and use the fact that the set of fixed points under the action of a finite subgroup \( F < \text{Mod}(S) \) is the image of a totally geodesic embedding of the Teichmüller space of the quotient orbifold.

We will also have to deal with centralizers of infinite groups. In particular, we will need to use the fact that their boundary at infinity is very small.

**Proposition 2.4.32.** Let \( S \) be a non-sporadic surface and \( g \) an element of infinite order in the mapping class group \( \text{Mod}(S) \).

(i) If \( g \) is pseudo-Anosov then \( \partial_\infty C(g) \) consists of two distinct points.

(ii) If \( g \) is reducible then \( \partial_\infty C(g) \) is empty.
2.4. SUBGROUPS OF THE MAPPING CLASS GROUP

Proof. The pseudo-Anosov case follows from Theorem 2.3.15. Indeed, since the group generated by \( g \) has finite index in \( C(g) \), then

\[
\partial_\infty C(g) = \partial_\infty(\{g^k \mid k \in \mathbb{Z}\})
\]

and by Proposition 2.3.7 the latter is equal to the classes of the stable and unstable foliations of \( g \).

Let \( g \) be an infinite order reducible element. Up to taking a power of \( g \), we can assume it is \emph{pure}, i.e. there are finitely many disjoint curves \( \alpha_i \subset S \) with \( i = 1, \ldots, n \) that are fixed by a representative of \( g \) and such that this representative also fixes all the components of \( S \setminus \{\alpha_i \mid i = 1, \ldots, n\} \) and its restriction to each of these component is either the identity or a pseudo-Anosov homeomorphism (see [Iva92]). It is not restrictive to take a power of \( g \) since \( C(g) \subset C(g^n) \).

Since \( g \) has infinite order, at least one of the curves \( \alpha_i \) bounds a component where \( g \) acts as a pseudo-Anosov or it is a Dehn-twist about \( \alpha_i \). In both cases, if a curve \( \beta \) intersect \( \alpha \), then it is easy to see that the classes \( g^n([\beta]) \) are all distinct. In particular, \( [\beta] \) cannot belong to the fixed set \( \text{Fix}(g) \).

As above, choose a base point \( \tilde{c} \in C(S) \) to define a quasi-isometry \( \Psi_{\tilde{c}} \) from the (electrified) mapping class group to the curve complex. Then the fixed set \( \text{Fix}(g) \) must be at bounded distance from \( \tilde{c} \) because it is composed of classes of curves disjoint from \( \alpha_i \) and hence it is contained in a neighbourhood of \( \tilde{c} \) of radius

\[
d(\tilde{c}, \{[\alpha_i] \mid i = 1, \ldots, n\}) + 1.
\]

Since the action of \( C(g) \) fixes the fixed set \( \text{Fix}(g) \) and \( \text{Fix}(g) \) has diameter 1, for every \( h \in C(g) \) we have

\[
d(\tilde{c}, h(\tilde{c})) \leq d(\tilde{c}, \text{Fix}(g)) + d(\text{Fix}(g), h(\tilde{c})) + 1 = 2d(\tilde{c}, \text{Fix}(g)) + 1.
\]

Hence the image of \( C(g) \) under the quasi-isometry \( \Psi_{\tilde{c}} \) is bounded and hence it has empty boundary at the infinity. \( \square \)

Corollary 2.4.33. If \( S \) is a non-sporadic surface and \( F \) is an infinite subgroup of \( \text{Mod}(S) \), then the boundary of its centralizer \( \partial_\infty C(F) \) has at most two points.

Proof. By Theorem 2.3.14, since \( F \) is infinite there exists an element \( f \in F \) of infinite order. Since \( C(F) \) is a subgroup of \( C(f) \), the thesis follows from Proposition 2.4.32. \( \square \)

Now that we have developed much of the theory regarding centralizers in the mapping class group, we can finally use it to prove the facts that we will need in Section 4.2. First of all, we define the class of groups we will be interested in.
Definition 2.4.34. A subgroup $H < \text{Mod}(S)$ is non-elementary if it contains two pseudo-Anosov elements with distinct fixed points in $\mathcal{PMF}(S)$.

Remark 2.4.35. By Theorem 2.3.13 we deduce that $H$ is non-elementary if and only if it contains a pseudo-Anosov element and it is not virtually cyclic.

Remark 2.4.36. Using some basic theory of the dynamics of the action of pseudo-Anosov elements on the space of measured foliations, one can easily prove that the boundary at infinity $\partial_\infty H$ of a non-elementary subgroup $H$ must be infinite. For example, if $f$ and $g$ are pseudo-Anosov classes and $[\mathcal{MF}^s]$ and $[\mathcal{MF}^u]$ are the stable and unstable foliation for $f$, then the conjugate $g^k fg^{-k}$ is pseudo-Anosov with stable and unstable foliations $g^k [\mathcal{MF}^s]$ and $g^k [\mathcal{MF}^u]$. In particular, $g^k [\mathcal{MF}^s]$ and $g^k [\mathcal{MF}^u]$ belong to $\partial_\infty H$. Thus one only has to show that these foliations are distinct to prove our claim.

Proposition 2.4.37. Let $S$ be a non-sporadic surface and $H$ a non-elementary subgroup of $\text{Mod}(S)$. If $\partial_\infty H$ is contained in $\partial_\infty \mathcal{C}(F)$ for some group $F < \text{Mod}(S)$, then there exists a group $F' < \text{Mod}(S)$ containing $F$ and such that $H$ is contained in the normalizer $N(F')$.

Proof. Notice that by Corollary 2.4.33 the group $F$ must be finite because the boundary of a non-elementary subgroup of the mapping class group is infinite. Take $F'$ to be a maximal subgroup such that $\partial_\infty H \subseteq \partial_\infty \mathcal{C}(F')$ (such a maximal group exists because the cardinality of a finite subgroup of $\text{Mod}(S)$ is bounded. See Theorem 2.4.16).

Recall that the left multiplication of $\text{Mod}(S)$ acts by isometries and hence it extends continuously to the boundary. For every element $h \in H$ we clearly have $h(\partial_\infty H) = \partial_\infty H$, hence $\partial_\infty H$ is also contained in $h(\partial_\infty \mathcal{C}(F')) \cap \partial_\infty \mathcal{C}(F')$. Notice that $h(\mathcal{C}(F')) = \mathcal{C}(hF'h^{-1})$ and hence also their boundaries at infinity coincide $h(\partial_\infty \mathcal{C}(F')) = \partial_\infty \mathcal{C}(hF'h^{-1})$. Thus we have

$$\partial_\infty H \subseteq \partial_\infty \mathcal{C}(F') \cap \partial_\infty \mathcal{C}(hF'h^{-1}).$$

By Proposition 2.4.30 we obtain that $\partial_\infty H$ is contained in the boundary at infinity of the centralizer of the group $F'' = \langle F', hF'h^{-1} \rangle$:

$$\partial_\infty H \subseteq \partial_\infty \mathcal{C}(F'').$$

As before, $F''$ must be finite because $\partial_\infty H$ is infinite. By the maximality of $F'$ we conclude that $F''$ is actually equal to $F'$. Thus $hF'h^{-1} < F'$ and hence $h \in N(F')$.

Corollary 2.4.38. Let $S$ be a non-sporadic surface and $H < \text{Mod}(S)$ a non-elementary subgroup. Then for every finite group $F < \text{Mod}(S)$ either $\partial_\infty \mathcal{C}(F)$ has infinitely many images under $H$ or there exists a finite group $F' \supseteq F$ such that $H$ is contained in the normalizer $N(F')$. \hfill $\square$
Proof. Assume $\partial\infty C(F)$ has only finitely many images under the action of $H$. Then taking the stabilizer of $\partial\infty C(F)$ we obtain a subgroup $H'$ that has finite index in $H$. By Theorem 2.3.13 it follows easily that also $H'$ is non-elementary.

By Corollary 2.4.28 and Proposition 2.4.29 we have that $\partial\infty C(F) = \text{Fix}_\beta(F)$ (Proposition 2.4.29 applies because the quotient $S/F$ is not a triangle orbifold since $\partial\infty C(F)$ is not empty). Hence $H'$ stabilizes $\text{Fix}_\beta(F)$ by construction. Applying Proposition 2.4.29 once more we obtain that $H'$ is contained in $N(F)$. Hence we conclude applying Proposition 2.4.37 to $H$.

Remark 2.4.39. Notice that when the group $F$ of Proposition 2.4.37 is contained in $H$ we can assume also $F'$ to be contained in $H$. This is because we can take $F'$ to be the maximal group contained in $H$ such that $\partial\infty H \subseteq \partial\infty C(F')$ and the above argument works the same because also the group $F'' = \langle F', hF'h^{-1} \rangle$ is contained in $H$. 


Chapter 3

Random Walks

In this chapter we will introduce the theory of random walks (or, discrete Markov chains). Our objective is to relate asymptotic properties of random walks on a graph with its large-scale geometry. In particular, we will find out that random walks are much more likely to spread out if the graph satisfies some isometric inequality. As a corollary, we will obtain new conditions for amenability. After that, we will study asymptotic properties of random walks in terms of harmonic functions and Poisson boundaries. In a certain sense, the Poisson boundary of a random walks is the probability space of possible asymptotic behaviours of sample paths. The existence of such a boundary is very closely related to the probabilistic tools we will use in Chapter 4.

Most of the time, when studying analytical and probabilistic properties of random walks we will not need the graph structure. Thus we will generally consider random walks on a countable set of possible states without specifying any link between states. If the reader likes to, he can think of the set of states as a graph where two states are linked if and only if there is a non-trivial probability to pass from one to the other. Still, such a graph will not be locally finite in general.

3.1 First examples and definitions

A random walk on a graph is exactly what its name suggests. Namely, one starts its walk on a node of the graph and each time moves to an adjacent node chosen at random. The mathematical concept that formalize random walks is that of Markov chain.

Definition 3.1.1. Given a countable state space $S$ and a probability space $\Omega$, a Markov chain (or random walk) is a sequence of random variables $\{X_n: \Omega \to S\}$ such that for every $n \in \mathbb{N}$ and $x \in S$ with $P\{X_n = x\} \neq 0$ the conditional probability satisfies

$$P[X_{n+1} = y \mid X_n = x] = p(x, y).$$

87
The values \( p(x, y) \) are called transition probabilities. We will sometimes refer to the set of transition probabilities \( P \) as the transition matrix or transition operator.

Given a state space \( S \) and a transition matrix \( P \), for every fixed probability distribution \( \theta \) on \( S \) we can find a Markov chain with starting distribution \( P_{X_0} = \theta \) (if \( X : \Omega \to S \) is a random variable, we denote by \( P_X \) the probability distribution induced by \( X \) on \( S \)). To see it, let \( \Omega := S^\mathbb{N} \) with the product \( \sigma \)-field and denote by \( Z_n : \Omega \to S \) the \( n \)-th projection. Then by Kolmogorov extension Theorem there is a probability distribution \( P_\theta \) on \( \Omega \) such that

\[
P_\theta[Z_n = x_n, Z_{n-1} = x_{n-1}, \ldots, Z_0 = x_0] = \theta(x_0)p(x_0, x_1) \cdots p(x_{n-1}, x_n).
\]

Letting \( X_n = Z_n \) we obtain the desired random walk.

The space \( \Omega = S^\mathbb{N} \) is called the space of sample paths and its elements \( x = (x_n)_{n \in \mathbb{N}} \) are the sample paths. Given \( x \in S \), we will write \( P_x \) for the probability measure induced on \( \Omega \) by the initial distribution \( \theta = \delta_x \) (i.e. \( P_x \) represent the probability of undertaking a certain path when performing the random walk that started at the point \( x \) and moves accordingly to \( P \)).

Throughout the thesis we will often prove results on random walks that depend only on the state space and the transition operator. In those cases we will refer to the pair \((S, P)\) as the Markov process, keeping in mind that to be precise we have a Markov chain only when a starting distribution is specified.

**Example 3.1.2.** If \( G \) is a graph, the simple random walk on \( G \) starting at a node \( o \in V(G) \) is the Markov chain with state space \( V(G) \) and starting distribution \( \delta_o \) whose transition matrix is given by

\[
p(x, y) = \frac{1}{\sharp\{ \text{neighbours of } x \}}
\]

for every pair of adjacent nodes \( x, y \) in \( V(G) \). Notice that this is the most natural example, where we start our walk form the node \( o \) and at every step we move with equal probability to one of the neighbouring nodes.

In general, a random walk on a graph is called nearest neighbour if \( p(x, y) = 0 \) whenever \( x \) and \( y \) are not linked by an edge, i.e. in nearest neighbour random walks we are forced to walk along the edges of the graph.

We will be mainly interested on random walks on groups. Namely, given a probability measure \( \mu \) on a countable group \( \Gamma \), we can consider the Markov process with state space \( \Gamma \) and transition matrix \( p(g, h) := \mu(g^{-1}h) \). For any starting probability distribution \( \theta \) on \( \Gamma \), this defines a Markov chain \( \{X_i\} \) on \( \Gamma \). We will call \( \mu \) the generating measure.

Notice that:

\[
P_{X_1}(g) = \sum_{h \in \Gamma} \theta(h)\mu(h^{-1}g) = \theta * \mu(g),
\]
where $\theta * \mu$ is the convolution. In general, the induced probability $P_{X_n}$ is the $n$-fold convolution of $\theta$ with $\mu$, $P_{X_n} = \theta * \mu * \cdots * \mu$.

In a group the identity element $e \in \Gamma$ is a natural starting point. Therefore, if we do not explicitly define a starting distribution we are assuming $\theta = \delta_e$.

**Example 3.1.3.** If a probability measure $\mu$ has support contained in a finite set of generators $S \subset \Gamma$, then the induced random walk on the group gives rise to a nearest neighbour random walk on the Cayley graph $C_S(\Gamma)$.

If $S$ is a symmetric set of generators (i.e. $g \in S$ if and only if $g^{-1} \in S$), then we can obtain the simple random walk on the Cayley graph $C_S(\Gamma)$ by setting $\mu(g) = 1/|S|$ for any $g \in S$.

**Definition 3.1.4.** A random walk on a state space $S$ is called irreducible if for every pair of states $x, y \in S$ there exists $n$ such that $p_n(x, y) > 0$. That is, from every state $x$ it is possible to reach any other state $y$.

Notice that the simple random walk on a graph is irreducible if and only if the graph is connected and a random walk on a group $\Gamma$ is irreducible if and only if the support of the generating measure $\mu$ generates $\Gamma$ as a semi-group.

A Markov process $(S, P)$ acts naturally on measures and functions assigning to each state the mean value of its adjacent states. Namely, given a function $f : S \to \mathbb{R}$, we set $Pf : S \to \mathbb{R} \cup \{\pm \infty\}$ as

$$Pf(x) := \sum_{y \in S} p(x, y) f(y) = E_x[f(X_1)]$$

when the latter is defined.

It is clear that $Pf$ is well-defined for a bounded $f$. Still, we will have to deal with more general domains which we will specify in the next section. When considering $P$ as an operator, we will refer to it as the Markov operator.

## 3.2 Reversible random walks

Roughly speaking, a reversible random walk is a random walk such that the probability to pass from a state $x$ to a state $y$ is the same as the probability to come back from $y$ to $x$. Actually, this is not the true interpretation because the states will be weighted. (Notice that it makes sense to weight the states. For example, if there are many little town all linked to a big city, when walking randomly it is much more likely to go from a town to the city rather than from the city to a specific town.) It turns out that when a random walk is reversible it is possible to use analytic techniques to easily deduce various interesting results on the random walk.

The main object of study will be the Markov operator. In particular, it will be of great interest its norm because this is closely related to the
presence of linear isoperimetric inequalities. Our exposition mainly follows that of [Pet13]. In some circumstances we found convenient to integrate with material from [Woe00]. Sometimes, it is interesting to compare the theory we are going to develop with some classical results of graph theory. (See e.g. [Bol79].)

3.2.1 Reversible measures

**Definition 3.2.1.** A Markov chain on a state space $\mathcal{S}$ is **reversible** if there exists a measure $\lambda$ on $\mathcal{S}$ (different from the constant-zero measure) such that for every pair of states $x, y \in \mathcal{S}$

$$\lambda(x) p(x, y) = \lambda(y) p(y, x).$$

Such a $\lambda$ is called a **reversible measure**.

**Example 3.2.2.** We say that a random walk on a countable group induced by a probability measure $\mu$ is **symmetric** if $\mu(g) = \mu(g^{-1})$ for every $g \in \Gamma$. It is trivial that any symmetric random walk is reversible with constant reversible measure.

**Remark 3.2.3.** It is clear that if $\lambda$ is a reversible measure for a process $P$ then for every positive constant $a \in \mathbb{R}_+$ the scaled measure $a \lambda$ is also a reversible measure for $P$. Conversely, it is easy to see that the reversible measure of a Markov process is unique up to a scale factor.

An easy way to define reversible random walks is by means of **electric networks**. Namely, for every pair of states $x, y \in \mathcal{S}$ choose a conductance $c(x, y) = c(y, x) \geq 0$ in such a way that $C_x := \sum_{y \in \mathcal{S}} c(x, y)$ is finite for every state $x$. We can now define a Markov process setting

$$p(x, y) := \frac{c(x, y)}{C_x}.$$ 

The resulting Markov chain is reversible with reversible measure $\lambda(x) = C_x$.

Conversely, with every reversible Markov chain with a fixed reversible measure $\lambda$ one can associate an electric network setting $c(x, y) := \lambda(x) p(x, y)$. Hence we have a one-to-one correspondence

$$\{\text{reversible Markov processes with fixed measure}\} \leftrightarrow \{\text{electric networks}\}.$$ 

In view of this correspondence, we will often use the notation $C_x$ to denote the reversible measure $\lambda$ of a reversible Markov chain and we will always assume $c(x, y) = \lambda(x) p(x, y)$.

**Example 3.2.4.** The simple random walk on a graph $G$ is easily described as an electric network imposing a constant conductance on any edge of $G$. To obtain more general nearest neighbour reversible random walks it is enough to arbitrarily choose conductances for the edges.
3.2. REVERSIBLE RANDOM WALKS

Example 3.2.5. Consider a finite cyclic group $\mathbb{Z}/k\mathbb{Z}$ and for any parameter $p \in [0, 1]$ consider the transition probability given by

$$p(n, m) = \begin{cases} p & \text{if } m \equiv (n + 1) \text{ mod}(k) \\ 1 - p & \text{if } m \equiv (n - 1) \text{ mod}(k) \\ 0 & \text{otherwise} \end{cases}.$$  

It is easy to see that the induced Markov chain is reversible if and only if the parameter $p$ is equal to $1/2$.

Remark 3.2.6. We will usually assume $C_x > 0$ for every state $x$. For if a state has measure zero it is unlinked with all the other states and hence it is influential in the study of random walks.

When dealing with a reversible Markov process $(\mathcal{S}, P)$, it is often useful to use the additional structure of measure space of the state space $\mathcal{S}$. In particular, once a reversible measure $\lambda$ is fixed, we can consider the Hilbert space $\ell^2(\mathcal{S})$ given by

$$\ell^2(\mathcal{S}) := \{ f : \mathcal{S} \rightarrow \mathbb{R} \mid \langle f, f \rangle < \infty \},$$

where

$$\langle f, g \rangle_{\mathcal{S}} = \int_{\mathcal{S}} f(x)g(x)d\lambda(x) = \sum_{x \in \mathcal{S}} f(x)g(x)C_x.$$

Clearly $\langle \cdot, \cdot \rangle_{\mathcal{S}}$ is a positive definite scalar product on $\ell^2(\mathcal{S})$. Moreover, $\ell^2(\mathcal{S})$ is complete with respect to the natural norm $\|f\|_2 := \sqrt{\langle f, f \rangle_{\mathcal{S}}}$. From now on, when we are dealing with reversible Markov processes we imply that we have fixed a reversible measure. The actual choice of the reversible measure is of no importance because such measure is unique up to a scale factor. We will usually denote the norm $\|\cdot\|_2$ simply by $\|\cdot\|$ and the scalar product by $\langle \cdot, \cdot \rangle$.

It will sometimes be useful to restrict ourselves to other spaces such as the space of bounded functions $\ell^\infty$ or the space of functions with compact (finite) support $\ell_c$.

Remark 3.2.7. Notice that $\ell_c$ is a dense subset of $\ell^2$, while in general $\ell^\infty$ does not contain $\ell^2$ nor it is contained in there.

3.2.2 The Markov operator

In the next few pages we will show that the Markov operator $P$ as defined in (3.1) has various nice properties when restricted to $\ell^2(\mathcal{S})$. For the remainder of this section we will always assume that Markov processes are reversible.

Proposition 3.2.8. If a function $f : \mathcal{S} \rightarrow \mathbb{R}$ is in $\ell^2(\mathcal{S})$, then $Pf$ takes only finite values. Moreover, $Pf$ belongs to $\ell^2(\mathcal{S})$ and $\|Pf\| \leq \|f\|$.
CHAPTER 3. RANDOM WALKS

Proof. For any state $x \in S$ let $h_x(y) := \frac{p(x, y)}{C_y}$. Then $h_x$ is square integrable:

$$\|h_x\|^2 = \sum_{y \in S} \left( \frac{p(x, y)}{C_y} \right)^2 C_y$$

$$= \sum_{y \in S} p(x, y) \frac{p(x, y)}{C_y}$$

$$= \sum_{y \in S} p(x, y) \frac{p(y, x)}{C_x}$$

$$= \frac{1}{C_x} \sum_{y \in S} p(x, y) p(y, x) \leq \frac{p_2(x, x)}{C_x}$$

and the latter is clearly finite. Now, we have:

$$Pf(x) = \sum_{y \in S} p(x, y) f(y)$$

$$= \sum_{y \in S} \frac{p(x, y)}{C_y} f(y) C_y$$

$$= (\frac{p(x, y)}{C_y}, f(y))$$

$$\leq \|h_x\||f||,$$

where we used Cauchy-Schwartz for the last inequality. Hence $Pf(x)$ is finite for every $x \in S$.

It remains to prove that $\|Pf\| \leq \|f\|$. Consider the operator $\|Pf\|^2 = \sum_{x \in S} \left( \sum_{y \in S} f(y) p(x, y) \right)^2 C_x$

$$\leq \sum_{x \in S} \sum_{y \in S} f^2(y) p(x, y) C_x$$

$$= \sum_{y \in S} f^2(y) \sum_{x \in S} p(x, y) C_x$$

$$= \sum_{y \in S} f^2(y) \sum_{x \in S} p(y, x) C_y$$

$$= \sum_{y \in S} f^2(y) C_y = \|f\|^2,$$

where in the second passage we used Jensen inequality.

Recall that, given a linear operator between normed spaces

$L: (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y),$
one define the norm of the linear operator as

$$\|L\| := \sup_{v \in X, v \neq 0} \frac{\|L(v)\|_Y}{\|v\|_X}.$$ 

An operator $L$ is continuous if and only if it has bounded norm. A continuous operator is sometime called bounded operator.

In particular, Proposition 3.2.8 can be rephrased saying that the Markov operator, when restricted to $\ell^2(S)$, gives rise to a continuous operator $P: \ell^2(S) \to \ell^2(S)$ with $\|P\| \leq 1$.

**Remark 3.2.9.** Since the set of functions with finite support $\ell_c$ is a dense subset of $\ell^2$, and $P$ is continuous, we have that

$$\|P\| = \sup_{f \in \ell^2(S)} \frac{\|Pf\|}{\|f\|} = \sup_{f \in \ell_c(S)} \frac{\|Pf\|}{\|f\|}.$$ 

Recall that if $L: H \to H$ is a bounded linear operator of a Hilbert space on itself, then there exists a unique adjoint operator $L^*: H \to H$ such that

$$\langle Lv, w \rangle = \langle v, L^*w \rangle$$ 

for every $v, w$ in $H$. A bounded linear operator $L$ is self-adjoint if $L = L^*$.

**Proposition 3.2.10.** The linear operator $P: \ell^2(S) \to \ell^2(S)$ is self-adjoint.

**Proof.** Given $f, g \in \ell^2(S)$, we have:

$$\langle Pf, g \rangle = \sum_{x \in S} \left( \sum_{y \in S} f(y)p(x, y) \right) g(x)C_x$$

$$= \sum_{y \in S} f(y) \left( \sum_{x \in S} g(x)p(x, y)C_x \right)$$

$$= \sum_{y \in S} f(y)C_y \left( \sum_{x \in S} g(x)p(y, x) \right) = \langle f, Pg \rangle,$$

where we could exchange the order of summation because $|f(x)g(y)|p(x, y)C_x$ is summable:

$$\sum_{x, y \in S} |f(x)g(y)|p(x, y)C_x = \langle P|f||g \rangle \leq \|P||f||g\| \leq \|P\||f||g\|.$$ 

Alternatively, we could have shown that $\langle Pf, g \rangle = \langle f, Pg \rangle$ for every $f, g \in \ell_c$, concluding then by continuity. \qed
Remark 3.2.11. Notice that, denoting with $P_n$ the $n$-th composition of $P$
\[ P^n := P \circ P \circ \cdots \circ P, \]
we have
\[ P^n f(x) = \sum_{y \in S} p_n(x, y) f(y). \]
In particular, if we write $1_X$ for the indicator function of the subset $X \subseteq S$, then
\[ (P^n 1_x, 1_y) = C_y p_n(x, y). \] \hfill (3.2)

Remark 3.2.12. It is clear from the definition that, given two linear operators $X \overset{L}{\to} Y \overset{L'}{\to} Z$, then $\|L' \circ L\| \leq \|L\| \|L'\|$. In particular, $\|P^n\| \leq \|P\|^n$.

3.2.3 Norm of the Markov operator

We will now try to find an explicit formula for $\|P\|$. We begin by stating a simple analytic lemma:

**Lemma 3.2.13.** Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of positive real numbers. Then for any integer $k$
\[ \limsup_{n \to \infty} \sqrt[n+k]{a_n} = \limsup_{n \to \infty} \sqrt[n]{a_n}. \]

Noticing that for a reversible Markov process $0 \neq p(x, y) = c(x, y)/C_x$ if and only if $p(y, x) \neq 0$, we promptly obtain the following:

**Corollary 3.2.14.** Given four possible states $x, y, z, w \in S$ of a reversible random walk, such that there exist $k, k'$ with both $p_k(x, z)$ and $p_{k'}(y, w)$ strictly positive, then
\[ \limsup_{n \to \infty} (p_n(x, y))^{1/n} = \limsup_{n \to \infty} (p_n(z, w))^{1/n} \]

**Proof.** Since we have that
\[ p_{n-k-k'}(z, w)p_k(x, z)p_{k'}(w, y) \leq p_n(x, y) \leq \frac{p_{n+k+k'}(z, w)}{p_k(z, x)p_{k'}(y, w)}, \]
the thesis follows applying Lemma 3.2.13 to the sequence $p_n(z, w)$. \qed

Recall that a random walk is *irreducible* if for every $x, y \in S$ there exists $n$ with $p_n(x, y) > 0$. We can then give the following key concept:

*CHAPTER 3. RANDOM WALKS*
3.2. REVERSIBLE RANDOM WALKS

Definition 3.2.15. The spectral radius of an irreducible reversible Markov chain is

$$\rho(P) := \limsup_{n \to \infty} (p_n(o, o))^{1/n}$$

where $o \in S$ is a fixed origin.

It follows from Corollary 3.2.14 that the spectral radius is well-defined, i.e. it does not depend on the choice of $o$.

Theorem 3.2.16. Given an irreducible reversible Markov process $(S, P)$, then $\|P\| = \rho(P)$.

Proof. By (3.2) we have:

$$C_o p_n(o, o) = (P^n 1_o, 1_o) \leq \|P^n\| \|1_o\| \leq \|P\| \|1_o\|^2 = C_o \|P\|^n.$$

where for the first inequality we applied Cauchy-Schwartz. This clearly implies

$$\rho(P) = \limsup_{n \to \infty} \sqrt[n]{p_n(o, o)} \leq \|P\|.$$

To prove the converse inequality, we begin noticing that for a fixed $f \in \ell^2(S)$ the rate of decrease of $\|P^n f\|$ is monotonically decreasing. In fact, since $P$ is self-adjoint we have:

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

whence

$$\frac{\|P^{n+1} f\|}{\|P^n f\|} \leq \frac{\|P^{n+2} f\|}{\|P^{n+1} f\|}.$$ (3.3)

It is a general fact that for any sequence of positive real numbers $(a_n)_{n \in \mathbb{N}}$

$$\liminf_{n \to \infty} \frac{a_{n+1}}{a_n} \leq \liminf_{n \to \infty} a_n^{1/n} \leq \limsup_{n \to \infty} a_n^{1/n} \leq \limsup_{n \to \infty} \frac{a_{n+1}}{a_n}.$$

Since $\|P^{n+1} f\|/\|P^n f\|$ is monotone, we have that

$$\liminf_{n \to \infty} \frac{\|P^{n+1} f\|}{\|P^n f\|} = \limsup_{n \to \infty} \frac{\|P^{n+1} f\|}{\|P^n f\|}.$$ 

Therefore, if we could prove that for every $f \in \ell_c(S)$,

$$\limsup_{n \to \infty} \|P^n f\|^{1/n} \leq \rho(P).$$
then the theorem would follows because by equation (3.3) we would deduce that
\[ \frac{\| Pf \|}{\| f \|} \leq \rho(P) \]
and then we would conclude thanks to Remark 3.2.9.

So, let \( f \in \ell_c(S) \) be a function with finite support, then

\[
\| P^n f \|^2 = \langle P^n f, P^n f \rangle \\
= \langle p^{2n} f, f \rangle \\
= \sum_{x \in S} (p^{2n} f)(x)f(x)C_x \\
= \sum_{x,y \in S} p^{2n}(x,y)f(y)f(x)C_x.
\]

By Corollary 3.2.14, for every \( x, y \in S \) and for every \( \varepsilon > 0 \) we have that \( p_n(x,y) \leq (\rho(P) + \varepsilon)^n \) for every \( n \) large enough. Now, since the support of \( f \) is finite, we have that there exists a sufficiently large \( n_0 \) such that for every \( n > n_0 \):

\[
\| P^n f \|^2 \leq \sum_{x,y \in \text{Supp}(f)} (\rho(P) + \varepsilon)^{2n} |f(y)f(x)|C_x.
\]

for a suitable constant \( K \) independent of \( n \). We conclude that for any \( f \in \ell_c \)

\[
\limsup_{n \to \infty} \| P^n f \|^{1/n} \leq \rho(P) + \varepsilon
\]

and letting \( \varepsilon \) go to zero we complete the proof of the theorem.

We conclude this subsection providing a formula for \( P \) that will be useful in Section 3.2.5.

**Lemma 3.2.17.** The norm of a continuous self-adjoint operator \( L : \ell^2(S) \to \ell^2(S) \) can be computed as

\[
\| L \| = \sup_{f \in \ell_c(S)} \frac{|\langle Lf, f \rangle|}{\| f \|^2}.
\]

**Proof.** This can be proved in general using spectral theory on Hilbert spaces (see [Rud91, Theorem 12.25]). For a direct proof, notice that using Cauchy-Schwartz we obtain

\[
\sup_{f \in \ell_c(S)} \frac{|\langle Lf, f \rangle|}{\| f \|^2} \leq \sup_{f \in \ell_c(S)} \frac{\| Lf \|}{\| f \|} = \| L \|.
\]
Conversely, notice that
\[
\sup_{f \in \ell^c(S)} \frac{|\langle Lf, f \rangle|}{\|f\|} = \sup_{f \in \ell^2(S)} \frac{|\langle Lf, f \rangle|}{\|f\|}
\]
by continuity and let \( C \) be such a supremum. For any \( f, g \in \ell^2(S) \) we have:
\[
|\langle Lf, g \rangle| = \left| \frac{\langle L(f + g), f + g \rangle - \langle L(f - g), f - g \rangle}{4} \right|
\leq \left| \frac{\langle L(f + g), f + g \rangle + \langle L(f - g), f - g \rangle}{4} \right|
\leq C \frac{\langle f + g, f + g \rangle + \langle f - g, f - g \rangle}{4} = C \frac{\langle f, f \rangle + \langle g, g \rangle}{2}.
\]
Taking
\[ g = Lf \|f\| \|Lf\| \]
yields \( \|Lf\| \leq C\|f\| \), whence the thesis.

Notice that by the definition of the Markov operator it is clear that
\[ |\langle Pf, f \rangle| \leq \langle |P|f\rangle, |f\rangle \] for every function \( f \). Lemma 3.2.17 hence implies the following:

**Corollary 3.2.18.** The norm of the Markov operator is
\[
\|P\| = \sup_{f \in \ell^c(S)} \frac{\langle Pf, f \rangle}{\|f\|^2}.
\]

### 3.2.4 Energy and differentials

Given a Markov process \((S, P)\), it will be convenient to consider the state space \( S \) as the set of nodes of an oriented graph, so that the transition probabilities can be read as the probability of passing through a determinate edge in a given direction.

Given an oriented edge \( e \), its starting and ending points will be denoted by \( e^- \) and \( e^+ \) respectively. We will write \( \overset{\leftarrow}{E} \) for the set of all oriented edges which are walked trough with non-trivial probability
\[
\overset{\leftarrow}{E} := \{ e \mid p(e^-, e^+) > 0 \}.
\]

When dealing with functions \( \theta : \overset{\leftarrow}{E} \to \mathbb{R} \), to simplify the notation we will often write \( \theta(x\overset{\leftarrow}{y}) \) even if \( x\overset{\leftarrow}{y} \) is not an element of \( \overset{\leftarrow}{E} \). In such cases it is understood that \( \theta(x\overset{\leftarrow}{y}) \) equals zero.

We now want to mimic concepts of real analysis such as differentials and harmonicity for reversible Markov processes (equivalently, for electric networks).
Chapter 3. Random Walks

Notice that if \((S, P)\) is a reversible Markov process, then for every edge \(e \in \overrightarrow{E}\) the opposite edge \(\hat{e} := e^+e^-\) is itself in \(\overrightarrow{E}\). We will call a function \(\theta: \overrightarrow{E} \to \mathbb{R}\) symmetric if \(\theta(e) = -\theta(\hat{e})\) for every edge in \(\overrightarrow{E}\). We will sometime call symmetric functions flows on \(S\).

We define the differential of a function \(f: S \to \mathbb{R}\) as the symmetric function \(\nabla f: \overrightarrow{E} \to \mathbb{R}\) given by

\[
\nabla f(e) := (f(e^+) - f(e^-)) c(e)
\]

where \(c(e) = c(e^-, e^+\) is the conductance of the edge \(e\).

Vice-versa, given a symmetric function \(\theta: \overrightarrow{E} \to \mathbb{R}\), we can define the analogue of the divergence \(\nabla^* \theta: S \to \mathbb{R}\) as

\[
\nabla^* \theta(x) := \frac{1}{C_x} \sum_{y \in S} \theta(yx) c(\overrightarrow{xy}).
\]

Moreover, we define the scalar product of two flows \(\theta\) and \(\eta\) as

\[
\langle \theta, \eta \rangle_{\overrightarrow{E}} := \frac{1}{2} \sum_{e \in \overrightarrow{E}} \frac{\theta(e) \eta(e)}{c(e^-, e^+)}.
\]

Notice that this scalar product is positive definite. It is easy to check that considering only flows with finite norm we obtain an Hilbert space \(\ell^2(\overrightarrow{E})\).

Remark 3.2.19. All of these definitions have a physical interpretation. Imagine \(S\) to be a real electric network where the nodes \(x\) and \(y\) are connected by a wire with resistance \(1/c(x, y)\). Now, if we think of a function \(f: S \to \mathbb{R}\) as an electric potential, then the flow \(\nabla f\) represents exactly the current that is flowing through the network.

Moreover if \(\theta\) is a flow (i.e. a current), then \(\langle \theta, \theta \rangle_{\overrightarrow{E}}\) is the power that is being dissipated by the Joules effect.

Notice that the constant \(C_x\) represents the total conductance between the node \(x\) and the remaining network. In particular, if we consider \(f(x) = V\) and \(f(y) = 0\) for any other node \(y\), then the dissipated power is exactly \(V^2 C_x\). We can hence think to \(V^2 f(x)\) as the amount of electric power that one has to give to the node \(x\) to keep it at the voltage \(V\) if all the other nodes of the network are grounded. This means that, when considering a generic potential \(f \in \ell^2(S)\), the square norm \(\langle f, f \rangle_S\) is the power needed to keep the potential \(f\) if all the wires where broken and set to the ground.

Similarly, given a current \(\theta\), then \(\nabla^* \theta(x)\) is the potential that the current induces on the node \(x\) if we assume all the other nodes grounded.

Lemma 3.2.20. If a function \(f\) is in \(\ell^2(S)\), then \(\nabla f\) belongs to \(\ell^2(\overrightarrow{E})\).
3.2. REVERSIBLE RANDOM WALKS

Proof. Since \(c(x, y) = c(y, x)\), we have:

\[
\langle \nabla f, \nabla f \rangle_{\tilde{E}} = \frac{1}{2} \sum_{e \in \tilde{E}} ((f(e^+) - f(e^-))c(e^-, e^+))^2 c(e^-, e^+) \\
= \frac{1}{2} \sum_{x, y \in S} (f(y) - f(x))^2 c(x, y) \\
\leq \sum_{x, y \in S} (f^2(y) + f^2(x)) c(x, y) = 2 \sum_{x \in S} f^2(x) C_x = 2 \langle f, f \rangle_S,
\]

where for the last inequality we used that \((a + b)^2 \leq 2(a^2 + b^2)\). \(\square\)

Lemma 3.2.21. If a flow \(\theta\) is in \(\ell^2(\tilde{E})\), then \(\nabla^\ast \theta\) belongs to \(\ell^2(S)\). In particular, \(\nabla^\ast \theta(x)\) is finite for every state \(x\).

Proof. We have:

\[
\langle \nabla^\ast \theta, \nabla^\ast \theta \rangle_S = \sum_{x \in S} \left( \frac{1}{C_x} \sum_{y \in S} \theta(y \xrightarrow{\tilde{E}} x) \right)^2 C_x \\
= \sum_{x \in S} \left( \sum_{y \in S} \theta(y \xrightarrow{\tilde{E}} x) \right)^2 \frac{1}{C_x} \\
\leq \sum_{x \in S} \sum_{y, z \in S} |\theta(y \xrightarrow{\tilde{E}} x)||\theta(z \xrightarrow{\tilde{E}} x)| \frac{1}{C_x} \\
= \sum_{x, y, z \in S} \frac{|\theta(y \xrightarrow{\tilde{E}} x)|}{\sqrt{c(y, x)}} \sqrt{\frac{c(z, x)}{C_x}} \frac{|\theta(z \xrightarrow{\tilde{E}} x)|}{\sqrt{c(z, x)}} \sqrt{\frac{c(y, x)}{C_x}} \\
\leq \sum_{x, y, z \in S} \frac{(\theta(y \xrightarrow{\tilde{E}} x))^2 c(z, x)}{C_x} \frac{c(y, x)}{C_x},
\]

where we have just used Cauchy-Schwartz and then exchanged \(y\) and \(z\) in one of the factors. Now we can conclude since

\[
\sum_{x, y, z \in S} \frac{(\theta(y \xrightarrow{\tilde{E}} x))^2 c(z, x)}{C_x} \frac{c(y, x)}{C_x} = \sum_{x, y \in S} \frac{(\theta(y \xrightarrow{\tilde{E}} x))^2}{c(y, x)} = 2 \langle \theta, \theta \rangle_{\tilde{E}}.
\]

\(\square\)

Lemma 3.2.22. Given \(f \in \ell^2(S)\) and \(\theta \in \ell^2(\tilde{E})\), then

\[
\langle \nabla f, \theta \rangle_{\tilde{E}} = \langle f, \nabla^\ast \theta \rangle_S,
\]

i.e. \(\nabla\) and \(\nabla^\ast\) are adjoint operators.
Proof. By definition we have:

\[
\langle \nabla f, \theta \rangle \frac{1}{2} = \sum_{x,y \in S} \frac{(f(y) - f(x))c(x, y)\theta(x, y)}{c(x, y)}
\]

\[
= \frac{1}{2} \sum_{x,y \in S} f(y)\theta(x, y) - \frac{1}{2} \sum_{x,y \in S} f(x)\theta(x, y)
\]

\[
= \frac{1}{2} \sum_{x,y \in S} f(y)\theta(x, y) - \frac{1}{2} \sum_{x,y \in S} f(y)\theta(y, x) = \sum_{x,y \in S} f(y)\theta(x, y).
\]

On the other hand, the other expression is

\[
\langle f, \nabla^* \theta \rangle_S = \sum_{x \in S} f(x) \left( \frac{1}{C_x} \sum_{y \in S} \theta(y, x) \right) C_x
\]

\[
= \sum_{x,y \in S} f(x)\theta(y, x)
\]

and hence they coincide.

So far we have defined the analogue of gradient and divergence. It is hence natural to define the Laplacian of a function \( f : S \to \mathbb{R} \) as

\[
\triangle f := \nabla^* \nabla f.
\]

Notice that by Lemmas 3.2.20 and 3.2.21, if \( f \) is in \( \ell^2(S) \) then \( \triangle f \) is well-defined and it is itself in \( \ell^2(S) \).

Notice that \( \triangle = \text{id} - P \):

\[
\triangle f(x) = \sum_{y \in S} \frac{1}{C_x} (f(x) - f(y))c(y, x) = f(x) - Pf(x). \quad (3.4)
\]

We will therefore say that a function \( f \) is harmonic in \( x \) if \( f(x) = Pf(x) \).

**Remark 3.2.23.** The physical interpretation of harmonicity is quite simple. In fact if \( f \) is a fixed potential, then \( \nabla^* \nabla f(x) \) is the amount of current that flows into the node \( x \) rescaled by \( C_x \). If we imagine to prevent accumulation of charges on nodes by allowing the nodes to exchange charges with the ground, we have that a potential is harmonic on a node \( x \) if there is no current flowing in or out the network in that node. Equivalently, I am not spending energy on that particular node to keep the potential \( f \).

Lastly, we give the following:

**Definition 3.2.24.** The Dirichlet energy of a function \( f : S \to \mathbb{R} \) is

\[
\mathcal{E}(f) := \langle \nabla f, \nabla f \rangle_E.
\]

Given a constant \( \kappa > 0 \), a function \( f \) is said to satisfy the Dirichlet inequality of constant \( \kappa \) if

\[
\mathcal{E}(f) \geq \kappa \| f \|^2.
\]
3.2. REVERSIBLE RANDOM WALKS

Notice that the Dirichlet inequality does not depend on the choice of the reversible measure because both $\mathcal{E}(f)$ and $\|f\|^2$ scale linearly with the measure.

Dirichlet inequalities are closely related to the spectral radius of the associated Markov process. This can be easily seen in the following proposition.

**Proposition 3.2.25.** Given a reversible Markov process $(S, P)$, then $\|P\| \leq 1 - \kappa$ with $\kappa > 0$ if and only if every function $f \in \ell_c(S)$ satisfies the Dirichlet inequality of constant $\kappa$.

**Proof.** Notice that by Equation (3.4) we have:

$$\mathcal{E}(f) = \langle f, \nabla f \rangle_S = \|f\|^2 - \langle f, Pf \rangle_S.$$  

Hence the thesis follows easily from Corollary 3.2.18. \qed

### 3.2.5 Isoperimetric inequalities and amenability

We want to extend concepts like isoperimetric inequalities from graphs to reversible Markov processes.

We have already seen that with every reversible Markov process $(S, P)$ we can associate the (oriented) graph $(S, \overrightarrow{E})$. By definition, we can chose a reversible measure $\lambda$ on $S$ (or equivalently, we can represent $P$ as an electric network), so that for every $X \subseteq S$ we have

$$\lambda(X) = \sum_{x \in X} \lambda(x) = \sum_{x \in X} C_x.$$ 

All that remains to do is to define a measure for the boundary of sets $X \subseteq S$. A natural choice is the following:

$$\lambda^\partial(X) := \sum_{x \in X, y \notin X} c(x, y).$$

We say that a reversible Markov process satisfies a linear isoperimetric inequality of constant $\varsigma$ with $\varsigma > 0$ if

$$\lambda^\partial(X) \geq \varsigma \lambda(X)$$

for every finite set $X \subseteq S$.

**Remark 3.2.26.** Notice that the definition of linear isoperimetric inequality is independent of the choice of the reversible measure $\lambda$ because this is unique up to a scale factor. The same is true for all the other inequalities we are going to encounter in what follows (i.e. Dirichlet and Sobolev inequalities). In fact, all the objects we will be dealing with in this section scale linearly with $\lambda$. Keeping this in mind, we will generally fix a reversible measure to carry out calculations without explicitly mention it.
Remark 3.2.27. Recall that a simple random walk on a graph can be represented as the electric network with conductance 1 on every edge. Notice that with this normalization $\lambda^\partial(X)$ coincides with $|\partial E X|$, where $\partial E X$ denotes the set of edges with one endpoint in $X$ and the other in the complement (see Subsection 1.2.1). Moreover, if the graph is connected and of bounded degree $C$, then the reversible measures of a set is comparable with its cardinality: $|X| \leq \lambda(X) \leq C|X|$.

In a similar fashion, for any irreducible symmetric random walk on a group the counting measure $\lambda(X) = |X|$ is a reversible measure for the random walk. Moreover, if the support $S = \text{Supp}(\mu)$ of the generating measure $\mu$ is finite, we have that the Cayley graph $C_S(\Gamma)$ satisfies $m|\partial E X| \leq \lambda^\partial(X) \leq M|\partial E X|$, where $m = \min_{g \in S} \mu(g)$ and $M = \max_{g \in S} \mu(g)$.

Thus, in both cases the random walk satisfies a linear isoperimetric inequality in the sense of Markov processes if and only if the underlying graph satisfies a linear isoperimetric inequality in the sense of graphs. (See Section 1.2.1).

The main result of this section is the following criterion for the existence of isoperimetric inequalities:

**Theorem 3.2.28.** A reversible Markov process satisfies a linear isoperimetric inequality if and only if there exists a constant $\kappa > 0$ such that every function $f \in \ell_c(S)$ satisfies the Dirichlet inequality

$$\mathcal{E}(f) \geq \kappa \|f\|^2.$$

It is easy to see that the Dirichlet inequality implies the isoperimetric inequality because

$$\lambda^\partial(X) = \mathcal{E}(1_X) \geq \kappa \|1_X\|^2 = \kappa \lambda(X),$$

where $1_X$ is the indicator function. Still, the converse implication is more involved and requires an intermediate result.

We define the total flow of a function $f \in \ell_c$ as

$$\mathcal{F}(f) := \frac{1}{2} \sum_{e \in \mathcal{E}} |f(y) - f(x)|c(x, y)$$

and recall that the $\ell^1$-norm of $f$ is defined as

$$\|f\|_1 = \sum_{x \in S} |f(x)|C_x.$$

These two quantities are related as follows.

**Proposition 3.2.29.** A reversible Markov process $(S, P)$ satisfies the linear isoperimetric inequality of constant $\varsigma > 0$ if and only if every function with compact support $f \in \ell_c$ satisfies the Sobolev inequality

$$\mathcal{F}(f) \geq \varsigma \|f\|_1.$$
Proof. Again, to see that the Sobolev inequality implies the isoperimetric one it suffices to apply it to the indicator functions.

To prove the converse, we can restrict ourselves to the case $f \geq 0$ because $\mathcal{F}(f) \geq \mathcal{F}(|f|)$ while the norm $\|f\|_1$ remains unchanged.

Notice that since $f$ has finite support we can write it as $f = \sum_{i=1}^{n} \alpha_i 1_{X_i}$ with $\alpha_i > 0$ and $X_1 \supset X_2 \supset \cdots \supset X_n$.

Now we have:

$$
\mathcal{F}(f) = \frac{1}{2} \sum_{x,y \in S} \left| \sum_{i=1}^{n} \alpha_i 1_{X_i}(y) - \alpha_i 1_{X_i}(x) \right| c(x, y)
= \frac{1}{2} \sum_{x,y \in S} \left( \sum_{i=1}^{n} \alpha_i |1_{X_i}(y) - 1_{X_i}(x)| \right) c(x, y)
= \sum_{i=1}^{n} \alpha_i \frac{1}{2} \sum_{x,y \in S} |1_{X_i}(y) - 1_{X_i}(x)| c(x, y)
= \sum_{i=1}^{n} \alpha_i \left( \sum_{x \in X_i, y \notin X_i} c(x, y) \right) = \sum_{i=1}^{n} \alpha_i \lambda^0(X_i).
$$

Hence the isoperimetric inequality yields:

$$
\mathcal{F}(f) = \sum_{i=1}^{n} \alpha_i \lambda^0(X_i) \geq \varsigma \sum_{i=1}^{n} \alpha_i \lambda(X_i) = \varsigma \|f\|_1.
$$

We can now complete the proof of Theorem 3.2.28.

Proof of Theorem 3.2.28. We only need to show that isoperimetric inequalities imply Dirichlet inequalities. We begin noticing that $\|f\|_2^2 = \|f^2\|_1$. Then we apply Proposition 3.2.29 and obtain:

$$
\|f\|_2^2 \leq \frac{1}{\varsigma} \mathcal{F}(f^2)
= \frac{1}{2\varsigma} \sum_{x,y \in S} |f^2(y) - f^2(x)| c(x, y)
\leq \frac{1}{2\varsigma} \sum_{x,y \in S} |f(y) - f(x)| (|f(y)| + |f(x)|) c(x, y)
$$
Now we can apply Cauchy-Schwartz obtaining two factors. The first one is:

\[ \left( \frac{1}{\zeta} \sum_{x,y \in S} |f(y) - f(x)|^2 c(x, y) \right)^{1/2} = \frac{1}{\zeta} \mathcal{E}(f)^{1/2}. \]

The other one is:

\[ \left( \frac{1}{\zeta} \sum_{x,y \in S} (|f(y)| + |f(x)|)^2 c(x, y) \right)^{1/2} \leq \sqrt{2} \|f\|_2, \]

where we used the inequality \((x + y)^2 \leq 2(x^2 + y^2)\).

Thus we have

\[ \|f\|_2^2 \leq \frac{1}{\zeta} \mathcal{E}(f)^{1/2} \sqrt{2} \|f\|_2. \]

Squaring and dividing by \(\|f\|_2^2\), we obtain

\[ \|f\|_2^2 \leq \frac{2}{\zeta^2} \mathcal{E}(f) \]

as it was desired.

Recall that a finitely generated group is non-amenable if its Cayley graph satisfies a linear isoperimetric inequality (see Subsection 1.2.1). Theorem 3.2.28 together with Proposition 3.2.25 can be promptly applied on groups to obtain a criterion for amenability. In fact, let \(\Gamma\) be a finitely generated group and \(\mu\) a symmetric probability measure on \(\Gamma\) whose support \(S\) is a finite generating set. Recall that in subsection 3.2.3 we defined the spectral radius of an irreducible random walk as

\[ \rho(P) := \limsup_{n \to \infty} (p_n(o, o))^{1/n} \]

and we proved that it is equal to the norm of the Markov operator \(\|P\|_2\) (Theorem 3.2.16). Then, we have the following:

**Corollary 3.2.30.** The group \(\Gamma\) is non-amenable if and only if \(\rho(\mu) < 1\), where \(\rho(\mu)\) is the spectral radius of the random walk on \(\Gamma\) induced by \(\mu\).

**Proof.** It is enough to note that linear isoperimetric inequalities on \(\Gamma\) in the sense of electric networks imply linear isoperimetric inequalities on \(C_S(\Gamma)\) in the sense of graphs and vice versa (Remark 3.2.27). Then the result follows combining Theorem 3.2.28 with Proposition 3.2.25.

Paying a little more attention, it is possible to weaken the hypotheses on the support of \(\mu\) obtaining the following:
3.2. REVERSIBLE RANDOM WALKS

**Theorem 3.2.31.** If $\Gamma$ is a finitely generated amenable group, then for every symmetric probability measure $\mu$ the generated random walk has spectral radius $\rho(\mu) = 1$. Conversely, if on a group $\Gamma$ there exists a symmetric probability measure $\mu$ whose support generates $\Gamma$ and the spectral radius $\rho(\mu)$ is equal to 1, then $\Gamma$ is amenable.

**Proof.** Let $\Gamma$ be amenable and $\mu$ a symmetric probability measure on $\Gamma$. We proved in Theorem 3.2.16 that the spectral radius $\rho(\mu)$ of the induced random walk is equal to the norm of the Markov operator $P : \ell^2(\Gamma) \to \ell^2(\Gamma)$, hence to prove that $\rho(\mu) = 1$ it is enough to show that there are functions that are ‘almost invariant’ under $P$.

For every $\varepsilon > 0$ there exists a finite set $S \subset \Gamma$ where it is concentrated the main part of the measure $\mu$, i.e. $\mu(S) > 1 - \varepsilon$. Consider then the Cayley graph of $C_S(\Gamma)$. By hypothesis, there exist a Følner exhaustion $F_n$ for $C_S(\Gamma)$ (we are using the notations of Section 1.2). Applying the Markov operator to the indicator functions of these sets, we find that for every $x$ in the interior of $F_n$ (i.e. for every $x \in F_n \setminus \partial_X^F(F_n)$)

$$P(1_{F_n})(x) = \sum_{y \in S} \mu(y)1_{F_n}(xy) + \sum_{y \in \Gamma \setminus S} \mu(y)1_{F_n}(xy)$$

$$\geq \sum_{y \in S} \mu(y)1_{F_n}(xy)$$

$$= \mu(S) > 1 - \varepsilon.$$

Thus the norm of $P(1_{F_n})$ satisfies

$$\|P(1_{F_n})\|_2 \geq (1 - \varepsilon)^2 |F_n \setminus \partial_X^F(F_n)| = (1 - \varepsilon)^2 (\|1_{F_n}\|_2 - |\partial_X^F(F_n)|).$$

Since $|\partial_X^F(F_n)|/|F_n|$ tends to zero by construction, we obtain that $\|P\|_2 \geq (1 - \varepsilon)$. Whence we conclude letting $\varepsilon$ go to zero.

The idea to prove the converse is to reduce the theorem to the case of measures with finite support. Let $\mu$ be a symmetric probability measure and choose a symmetric finite generating set $S = S^{-1}$ that is contained in the support of $\mu$. We define two auxiliary symmetric probability measures obtained normalizing the restriction of $\mu$ to $S$ and to its complement:

$$\mu_S = \frac{1}{\mu(S)}\mu|_S,$$

$$\mu_C = \frac{1}{1 - \mu(S)}\mu|_{\Gamma \setminus S}.$$

Then we clearly have that $\mu$ is equal to the convex combination $\mu(S)\mu_S + (1 - \mu(S))\mu_C$. In particular, if we denote their Markov operators by $P, P_S$.
and $P_C$, then $P = \mu(S)P_S + (1 - \mu(S))P_C$. By Lemma 3.2.17, we have:

$$\|P\|_2 = \sup_{\|f\|_2 = 1} |\langle Pf, f \rangle|$$
$$= \sup_{\|f\|_2 = 1} |\mu(S)\langle PSf, f \rangle + (1 - \mu(S))\langle PCf, f \rangle|$$
$$\leq \mu(S)\|PS\|_2 + (1 - \mu(S))\|PC\|_2.$$

Since $\|P\|_2 = 1$ by hypothesis and Markov operators have norm smaller than or equal to 1, we deduce that both $P_S$ and $P_C$ must have norm equal to 1. The thesis follows applying Corollary 3.2.30 to $\mu_S$.

\[\square\]

Remark 3.2.32. In the first part of Theorem 3.2.31 we purposely omitted to require the random walk to be irreducible notwithstanding that we defined the spectral radius only for irreducible random walks. That is because the definition of spectral radius makes perfectly sense for generic random walks on groups (also non reversible). We used to require irreducibility only to remove the dependency on the base point $o \in S$, but for random walks on group $p_n(o, o) = \mu_n(e)$ is automatically independent of any choice.

Remark 3.2.33. It is possible to define amenability also for groups that are non-finitely generated (the notion of invariant mean does not depend on a choice of generator). Using a generalization of Følner criterion, it is possible to show that Theorem 3.2.31 holds in this broader context.

### 3.3 Harmonic functions and Poisson boundary

In this section we will show that with a limit process it is possible to pass from a bounded harmonic functions to a particular set of real valued functions on the space of sample paths. Vice versa, it is possible to invert such a limit process taking the expected value of functions along sample paths. To understand the meaning of this relation it is useful to imagine that sample paths converge in some sense to a boundary. Then taking the limit of harmonic functions identifies a real valued function on the boundary. Conversely, once a function on the boundary is fixed, one can associate to each state $x$ the expected value at infinity of the sample paths of the random walk starting in $x$.

In Subsection 3.3.3 we will refine the above relation in the case of random walks on groups and we will state some powerful theorems relating non-amenability with the presence of non-constant bounded harmonic functions. We will conclude Subsection 3.3.3 stating some useful entropic criteria for the triviality of Poisson boundaries.

Also in this section the general outline is that of [Pet13]. The necessary probabilistic tools are more refined than those required from the previous section, hence we will not be able to prove all of them. A good reference for
3.3. HARMONIC FUNCTIONS AND POISSON BOUNDARY

what we do not prove is [Dur13]. The last two subsections list results from [KV83] and unify the results obtained so far.

3.3.1 Martingales

Let \((\Omega, \mathcal{F}, \mathbf{P})\) be a probability space with probability measure \(\mathbf{P}\) defined on the \(\sigma\)-field \(\mathcal{F}\) and let \(X: \Omega \to \mathbb{R}\) be a random variable with \(\mathbf{E}[|X|] < \infty\). Given a \(\sigma\)-field \(\mathcal{E} \subseteq \mathcal{F}\), we say that a random variable \(Y: \Omega \to \mathbb{R}\) is a conditional expectation of \(X\) given \(\mathcal{E}\) if

(i) \(Y\) is \(\mathcal{E}\)-measurable,

(ii) for any \(A \in \mathcal{E}\) we have

\[
\int_A X \, d\mathbf{P} = \int_A Y \, d\mathbf{P}.
\]

The intuitive meaning of conditional expectations is that they are our best guess for \(X\) knowing that \(\mathcal{E}\) has occurred. We will often use the following theorem whose proof we omit.

**Theorem 3.3.1.** Given a probability space \((\Omega, \mathcal{F}, \mathbf{P})\) and a random variable \(X: \Omega \to \mathbb{R}\), then for any \(\sigma\)-field \(\mathcal{E} \subseteq \mathcal{F}\) there exists a conditional expectation of \(X\) given \(\mathcal{E}\). Moreover, if \(Y\) and \(Y'\) are both conditional expectations of \(X\) given \(\mathcal{E}\) then \(Y = Y'\) almost surely.

The uniqueness of conditional expectations is quite simple, while their existence is not trivial and is usually proved using the Radon-Nicodym Theorem.

In literature, the conditional expectation of \(X\) given \(\mathcal{E}\) is usually denoted with \(\mathbf{E}[X|\mathcal{E}]\). Notice that in view of Theorem 3.3.1 \(\mathbf{E}[X|\mathcal{E}]\) is well-defined up to measure 0 subsets.

Conditional expectations have many useful properties. We state some in the following proposition.

**Proposition 3.3.2.** Let \(X\) and \(Y\) be random variables on \((\Omega, \mathcal{F}, \mathbf{P})\) with finite first moment \(\mathbf{E}[|X|], \mathbf{E}[|Y|] < \infty\). Then conditional expectations have the following properties:

(i) \(\mathbf{E}[aX + bY|\mathcal{E}] = a\mathbf{E}[X|\mathcal{E}] + b\mathbf{E}[Y|\mathcal{E}]\) for every \(a, b \in \mathbb{R}\);

(ii) if \(\mathcal{E}' \subseteq \mathcal{E}\) then \(\left[\mathbf{E}[X|\mathcal{E}] \mid \mathcal{E}'\right] = \mathbf{E}[X|\mathcal{E}']\);

(iii) if \(\sigma(X) \subseteq \mathcal{E}\) then \(\mathbf{E}[X|\mathcal{E}] = X\) (where \(\sigma(X)\) is the smallest \(\sigma\)-fields that makes \(X\) measurable);

(iv) if \(X\) is independent of \(\mathcal{E}\) (i.e \(\mathbf{P}[A \cap B] = \mathbf{P}[A]\mathbf{P}[B]\) for every \(A \in \sigma(X)\) and \(B \in \mathcal{E}\)), then \(\mathbf{E}[X|\mathcal{E}] = \mathbf{E}[X]\).
The proof of these properties is quite straightforward and is generally based on the uniqueness statement of Theorem 3.3.1. For a complete treatment on conditional expectations see [Dur13].

**Remark 3.3.3.** Given a measurable partition \( \Omega = \bigsqcup_{i \in I} A_i \) with \( A_i \in \mathcal{F} \) for every \( i \) in \( I \), let \( \mathcal{E} := \sigma(A_i \mid i \in I) \) be the generated \( \sigma \)-field. Then if \( X \) is an integrable random variable on \((\Omega, \mathcal{F}, P)\) we can express the conditional expectation as

\[
    \mathbb{E}[X|\mathcal{E}](x) = \frac{1}{P(A_i(x))} \int_{A_i(x)} X \, dP
\]

where \( i(x) \) is the unique index in \( I \) such that \( x \in A_i(x) \).

Indeed, such a function is clearly \( \mathcal{E} \)-measurable and by definition its integral on any set \( A_i \) coincides with \( \int_{A_i} X \, dP \) for every \( i \in I \) and hence the same holds for every subset \( B \in \mathcal{E} \).

Now we define another concept which arises naturally in the theory of Markov processes.

**Definition 3.3.4.** A *filtration* is a sequence of \( \sigma \)-fields \((\mathcal{F}_n)_{n \in \mathbb{N}}\) with \( \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \) for every \( n \in \mathbb{N} \).

A sequence of random variables \( X_n \) is a **martingale** with respect to a filtration \( \mathcal{F}_n \) if for every \( n \in \mathbb{N} \) it satisfies:

(i) \( \mathbb{E}[|X_n|] < \infty \),

(ii) \( X_n \) is \( \mathcal{F}_n \)-measurable,

(iii) \( \mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n \).

If \( \{X_i \mid i \in I\} \) is a family of random variables, \( \sigma(X_i \mid i \in I) \) will denote the minimal \( \sigma \)-field which makes all the \( X_i \) measurable. Similarly, the conditional expectation \( \mathbb{E}[X|\sigma(X_i \mid i \in I)] \) will be denoted by \( \mathbb{E}[X|X_i, i \in I] \).

In what follows, when we say that a sequence of random variables \( X_n \) is a martingale (submartingale, supermartingale) without specifying a filtration, we mean that it is a martingale (submartingale, supermartingale) with respect to the filtration \( \mathcal{F}_n = \sigma(X_0, \ldots, X_n) \). With this notation, condition (iii) of Definition 3.3.4 becomes \( \mathbb{E}[X_{n+1}|X_0, \ldots, X_n] = X_n \).

Let \((\mathcal{S}, P)\) be a Markov process, recall that in subsection 3.2.4 we defined harmonic functions for reversible Markov processes and it turned out that a function \( u: \mathcal{S} \to \mathbb{R} \) is harmonic if and only if it is fixed by the Markov operator:

\[
    u(x) = Pu(x) = \sum_{y \in \mathcal{S}} p(x, y)u(y).
\]

(see Equation (3.4)). The condition \( u = Pu \) makes perfectly sense also for non-reversible Markov processes thus we define the *harmonic functions* as
those functions that are preserved by the Markov operator. As we will see in what follows, harmonic functions and martingales are closely related.

**Proposition 3.3.5.** If a bounded function \( u: S \to \mathbb{R} \) is harmonic, then for every random walk \( X_n \) on \( S \) with transition probabilities \( P \) the sequence \( u(X_n) \) is a martingale.

**Proof.** Notice that \( u(X_n) \) is \( \sigma(X_1, \ldots, X_n) \)-measurable by definition and that \( u(X_n) \) is clearly integrable since \( u \) is bounded. Therefore we only need to prove that

\[
E[u(X_{n+1}) | X_0, \ldots, X_n] = u(X_n).
\]

Let \( A(x_0, \ldots, x_n) \subset \Omega \) be the subset \( X_0^{-1}(x_0) \cap \cdots \cap X_n^{-1}(x_n) \). Then, as the \( x_i \)'s vary in \( S \), the sets \( A(x_0, \ldots, x_n) \) form a partition of \( \Omega \) generating the \( \sigma \)-field \( \sigma(X_0, \ldots, X_n) \). By the uniqueness statement in Theorem 3.3.1 it is hence enough to check that

\[
\int_{A(x_0, \ldots, x_n)} u(X_{n+1})dP = \int_{A(x_0, \ldots, x_n)} u(X_n)dP = u(x_n)P(A(x_0, \ldots, x_n)) = \theta(x_0)p(x_0, x_1) \cdots p(x_{n-1}, x_n)u(x_n),
\]

where \( \theta \) is the initial probability distribution \( P_{X_0} \) on \( S \).

A direct calculation leads to

\[
\int_{A(x_0, \ldots, x_n)} u(X_{n+1})dP = \theta(x_0)p(x_0, x_1) \cdots p(x_{n-1}, x_n) \sum_{y \in S} p(x_n, y)u(y),
\]

whence the thesis by the harmonicity of \( u \).

**Remark 3.3.6.** Some sort of inverse holds. For example, let \( X_n \) be an irreducible random walk and \( f \) a real valued function on \( S \). The same calculation shows that if \( f(X_n) \) is a martingale then \( f \) is harmonic.

### 3.3.2 Invariant functions

From now on every time we speak about random walks \( X_n: \Omega \to S \) we assume \( \Omega \) to be the space of sample paths \( \Omega = S^N \) with the product \( \sigma \)-field \( \mathcal{F} \) and the random variables to be the projections on \( S \). As seen in Section 3.1, this assumption is not restrictive. Moreover, for every choice of transition probabilities \( P \) and initial distribution \( \theta \) on \( S \) there exists a unique probability measure \( P_\theta \) on \( \Omega \) such that the projections \( X_n \) form the random walk with these initial distribution and transition matrix. Recall that for every \( x \in S \) we denote by \( P_x \) the probability induced on \( \Omega \) by the starting distribution \( \theta = \delta_x \).

The fundamental tool in this section is the following result (for the proof see Theorem 5.2.8 in [Dur13]).
Theorem 3.3.7. If $X_n$ is a bounded martingale, then there exists a bounded random variable $X_\infty$ such that

$$\lim_{n \to \infty} X_n = X_\infty$$

almost surely.

Let $X_n$ be a random walk on a state space $S$ with starting distribution $\theta$. For every bounded harmonic function $u : S \to \mathbb{R}$ the sequence $u(X_n)$ is a bounded martingale by Proposition 3.3.5. Hence we can apply Theorem 3.3.7 and we obtain a random variable defined on the space of sample paths $\hat{u}_\theta : \Omega \to \mathbb{R}$ such that $u(X_n) \to \hat{u}_\theta$ almost surely on $(\Omega, P_\theta)$. This means that for $P_\theta$-almost every $x = (x_n)_{n \in \mathbb{N}}$ there exists the limit

$$\lim_{n \to \infty} u(x_n) = \hat{u}_\theta(x).$$

We define an equivalence relation on the space of sample paths $\Omega$ letting $x \sim y$ if there is a (time) translation of $x$ that coincides with $y$ apart from finitely many indices. (i.e. $x \sim y$ if and only if there exist integers $k, n_0$ such that $y_{n+k} = x_n$ for every $n > n_0$). Notice that if $u(x_n)$ admits a limit then $u(y_n)$ converges to the same limit for every sample path $y \sim x$. In particular, up to modifying $\hat{u}_\theta$ on a set of $P_\theta$-measure zero, we can assume that if $x \sim y$ then $\hat{u}_\theta(x) = \hat{u}_\theta(y)$. For example, we can choose a representative for $\hat{u}_\theta$ setting $\hat{u}_\theta(x) := \lim \sup_{n \to \infty} u(x_n)$. This coincides with $\hat{u}_\theta$ $P_\theta$-almost everywhere and it is clearly invariant with respect to translations.

Notice that $\lim \sup u(x_n)$ does not depend on the choice of the initial distribution $\theta$. This means that we have found a common representative $\hat{u}$ for all the $\hat{u}_\theta$. This is not trivial because the sets of measure zero accordingly to a measure $P_\theta$ need not have measure zero accordingly to a different measure $P_{\theta'}$ (for example, given two states $x \neq y$ the probability measures $P_x$ and $P_y$ have disjoint support).

Actually, the independence on $\theta$ does not come as a surprise either. It is natural to consider the (infinite) measure $\tilde{\nu}$ on $\Omega$ given by

$$\tilde{\nu} := \sum_{x \in S} P_x$$

(this is a natural measure for $\Omega$ in that when asking how probable it is to walk along a certain path $(x_0, x_1, \ldots)$ one unconsciously mean ‘taking for granted that we start from $x_0$’, otherwise the answer is trivially 0). We claim that for $\tilde{\nu}$-almost-every sample path $x$ the limit $\lim_{n \to \infty} u(x_n)$ exists. Indeed, let $B \subseteq \Omega$ be the set of paths where $u$ does not converge. Then

$$B = \bigcup_{y \in S} B_y = \bigcup_{y \in S} \{ x \in B \mid x_0 = y \},$$
but
\[ \tilde{\nu}(B_y) = P_y(B_y) = P_y(B) \]
and we have already proved that the latter is zero.

In particular, we have that \( \hat{u}(x) = \lim_{n \to \infty} u(x_n) \) is well-defined up to \( \tilde{\nu} \)-measure-zero sets and \( \hat{u} \) is \( P_\theta \)-equivalent to \( \hat{u}_\theta \) for every initial distribution \( \theta \).

To sum up, we proved that taking the limit along sample paths gives us a map
\[ L: \mathcal{H}_\infty(\mathcal{S}, P) \to L_\infty^{\text{inv}}(\Omega) \]
where \( \mathcal{H}_\infty(\mathcal{S}, P) \) is the set of bounded harmonic functions and \( L_\infty^{\text{inv}}(\Omega) \) is the space of bounded invariant function on \( \Omega \) (considered up to \( \tilde{\nu} \)-measure zero sets)
\[ L_\infty^{\text{inv}}(\Omega) := \{ f \in L_\infty(\Omega) \mid f(x) = f(y) \text{ if } x \sim y \} \].

**Remark 3.3.8.** Notice that when using the notation \( L^p(\text{measure space}) \) we are considering only measurable functions. While measurability is trivial on \( S \) because it is endowed with the discrete \( \sigma \)-field, the same is not true in \( \Omega \) where we are using the product \( \sigma \)-field. Nevertheless it is easy to see that the functions \( \hat{u} \) are measurable because they are limits of the measurable functions \( u(X_n) \).

We will now prove that the map \( L: \mathcal{H}_\infty(\mathcal{S}, P) \to L_\infty^{\text{inv}}(\Omega) \) is injective.

If \( u: \mathcal{S} \to \mathbb{R} \) is a harmonic function, then for every \( x \in \mathcal{S} \) and for every \( n \in \mathbb{N} \)
\[ u(x) = \sum_{y \in \mathcal{S}} p_n(x, y) u(y) = \int_{\Omega} u(X_n) dP_x. \]
In particular, for every \( u \in \mathcal{H}(\mathcal{S}, P) \) the Dominated Convergence Theorem implies that
\[ u(x) = \int_{\Omega} u(X_n) dP_x \xrightarrow{n \to \infty} \int_{\Omega} \hat{u}(x) dP_x(x). \]
This means that the expectation function \( E: L_\infty(\Omega) \to L_\infty(\mathcal{S}) \) defined as
\[ E(f)(x) = E_x[f] = \int_{\Omega} f(x) dP_x(x) \]
gives us a left inverse for \( L \).

Much more is true. Indeed, we will soon find out that images of invariant functions under \( E \) are actually harmonic functions and hence \( E(L_\infty^{\text{inv}}(\Omega)) = \mathcal{H}_\infty(\mathcal{S}, P) \). Moreover, we will show that \( L \) is also a right inverse for \( E \).

**Remark 3.3.9.** Notice that \( L \) is an isometric embedding (with respect to the sup norm). It is clear that \( \|\hat{u}\|_\infty \leq \|u\|_\infty \) since the former is the limit of the latter. Conversely, let \( x_0 \in \mathcal{S} \) such that \( |u(x_0)| > \|u\| - \varepsilon \), then
\[ u(x_0) = \int_{\Omega} \hat{u}(x) dP_{x_0}(x) \]
implies \( \|\hat{u}\|_\infty > \|u\| - \varepsilon \).
Denote by $T : \Omega \to \Omega$ the left shift, \textit{i.e.} the map which sends $\mathbf{x} = (x_n)_{n \in \mathbb{N}}$ to the element $T(\mathbf{x})$ given by $T(\mathbf{x})_n = x_{n+1}$. Notice that the push forward of the probability measure $P_x$ is given by

$$T_*P_x = \sum_{y \in \mathcal{S}} p(x,y)P_y.$$ 

In fact, $T_*P_x$ is the probability measure induced on $\mathcal{S}^{\mathbb{N}}$ by the random walk with starting distribution $\sum_{y \in \mathcal{S}} p(x,y) \delta_y$.

\textbf{Remark 3.3.10.} Notice that $\mathbf{x} \sim \mathbf{y}$ if and only if there exist $k, k' \in \mathbb{N}$ such that $T^k(\mathbf{x}) = T^{k'}(\mathbf{y})$.

Now it is easy to see the reason why images of invariant functions are harmonic. Let $f \in L_\infty^{\text{inv}}(\Omega)$, then $f(\mathbf{x}) = f(T\mathbf{x})$ and hence

$$E(f)(\mathbf{x}) = \int_{\Omega} f(\mathbf{x})dP_x(\mathbf{x}) = \int_{\Omega} f(T\mathbf{x})dP_x(\mathbf{x}) = \int_{\Omega} f(\mathbf{y})d(T_*P_x)(\mathbf{y}) = \sum_{y \in \mathcal{S}} p(x,y) \int_{\Omega} f(\mathbf{y})dP_y(\mathbf{y}) = P(E(f))(\mathbf{x}).$$

To prove that taking the limit along sample paths gives the inverse function of $E$ we need a refinement of Theorem 3.3.7.

\textbf{Theorem 3.3.11 (Lévy's Zero-One Law).} \textit{Given a filtration of $\sigma$-fields $\mathcal{F}_n \nearrow \mathcal{F}_\infty$ and a bounded random variable $X : \Omega \to \mathbb{R}$, then $E[X|\mathcal{F}_n] \xrightarrow{a.s.} E[X|\mathcal{F}_\infty]$.}

\textit{Proof.} Notice that $X_n = E[X|\mathcal{F}_n]$ is a martingale, hence by Theorem 3.3.7 there exists $X_\infty$ such that $X_n \to X_\infty$ almost surely. All it remains to do is to show that $X_\infty$ is the conditional expectation of $X$ given $\mathcal{F}_\infty$.

It is clear that $X_\infty$ is $\mathcal{F}_\infty$-measurable because it is the limit of $\mathcal{F}_\infty$-measurable random variables.

Since $\mathcal{F}_\infty$ is generated by $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$, to prove that $\int_A X_\infty dP = \int_A X dP$ for every $A \in \mathcal{F}_\infty$ it is enough to check that the equality holds for $A \in \mathcal{F}_n$ and this is trivially true by dominated convergence. $\square$

As we promised before we will now see that the map $L : \mathcal{H}_\infty(S,P) \to L_\infty^{\text{inv}}(\Omega)$ is the inverse of $E$. Let $f \in L_\infty^{\text{inv}}(\Omega)$ be an invariant bounded function. Then for $\tilde{\nu}$-almost every $\mathbf{x} = (x_n)_{n \in \mathbb{N}} \in \Omega$ there exists the limit of $E(f)(x_n)$ as $n$ goes to infinity and

$$L(E(f))(\mathbf{x}) = \lim_{n \to \infty} E(f)(x_n) = \lim_{n \to \infty} \int_{\Omega} f(y)dP_{x_n}.$$
Hence what we would like to prove is that \( f \) evaluated on the sample path \( x \) can be approximated by the average value of \( f \) obtained starting a random walk at a point \( x_n \) for large \( n \).

Now, consider the filtration \( F_n \uparrow F \) given by

\[
F_n = \left\{ A_0, \ldots, A_n \times S^{[n+1, \infty)} \mid A_i \subseteq S, \ i = 0, \ldots, n \right\}.
\]

Since the partition of \( \Omega \) given by the sets \( A_{(a_0, \ldots, a_n)} = (a_0, \ldots, a_n) \times S^{[n+1, \infty)} \) generates \( F_n \), by Remark 3.3.3 we have:

\[
E_{x_0}[f|F_n](x) = \frac{1}{P_{x_0}(A_{(x_0, \ldots, x_n)})} \int_{A_{(x_0, \ldots, x_n)}} f(y) dP_{x_0}(y).
\]

Notice that the following holds:

\[
\int_{A_{(a_0, \ldots, a_n)}} f(y) dP_\theta(y) = \theta(a_0) \prod_{i=1}^n p(a_{i-1}, a_i) \int_{A_{(a_0, \ldots, a_n-1, z)}} f(a_0, \ldots, a_{n-1}, z) dP_{a_n}(z)
\]

and that

\[
\theta(a_0) \prod_{i=1}^n p(a_{i-1}, a_i) = P_\theta(A_{(a_0, \ldots, a_n)}).
\]

Hence we have

\[
E_{x_0}[f|F_n](x) = \int_{\Omega} f(x_0, \ldots, x_{n-1}, z) dP_{x_n}(z) = \int_{\Omega} f(z) dP_{x_n}(z),
\]

where the last equality holds because \( f \) invariant.

Since \( E_{x_0}[f|F] = f \), by Lévy’s Zero-One Law 3.3.11

\[
\int_{\Omega} f(z) dP_{x_n}(z) = E_{x_0}[f|F_n](x) \xrightarrow{n \to \infty} f(x)
\]

for \( P_{x_0} \)-almost every \( x \in \Omega \). Being the last true for every \( x_0 \) in \( S \), we conclude that the convergence is actually \( \tilde{\nu} \)-almost everywhere.

We collect the results obtained so far in the following:

**Theorem 3.3.12.** For any Markov process \( (S, P) \) there is a natural isometry between the space of bounded harmonic functions on \( S \) and that of bounded invariant functions on the state space \( \Omega \).

\[
\mathcal{H}_\infty(S, P) \overset{L}{\xrightarrow{E}} L_\text{inv}^\infty(\Omega).
\]
CHAPTER 3. RANDOM WALKS

3.3.3 The Poisson boundary

As before, let $\Omega = S^\mathbb{N}$ be a space of sample paths and $\mathcal{F}$ the product $\sigma$-field. We define the invariant $\sigma$-field $I$ on $\Omega$ as

$$I := \{ A \in \mathcal{F} \mid x \in A \iff y \in A \text{ for every } x \sim y \}.$$ 

Notice that a measurable function on $\Omega$ is invariant if and only if it is $I$-measurable.

**Definition 3.3.13.** Let $X_n$ be a random walk on a state space $S$. The Poisson boundary $B(X_n)$ of the random walk is the probability space $(\Omega, I, P)$, where $P$ is the probability induced on $\Omega$ by the random walk.

**Remark 3.3.14.** The probability measure $P$ is well-defined on $I$ because this is a subalgebra of $\mathcal{F}$ by definition.

We say that a random walk has trivial Poisson boundary if the probability measure takes only values 0 or 1 on the invariant $\sigma$-field. That is, for every $A \in I$ the probability $P(A)$ is either 1 or 0.

When the state space $S$ is a countable group $\Gamma$, Poisson boundaries acquire a great importance. The fact is that $\Gamma$ acts on the space of paths $\Omega = \Gamma^\mathbb{N}$ by left multiplication $h \cdot (g_n)_{n \in \mathbb{N}} := (hg_n)_{n \in \mathbb{N}}$ and this will induce an interesting action on Poisson boundaries.

Recall that for any probability measure $\mu$ on $\Gamma$ and starting distribution $\theta$ we can consider the random walk $(X^n_\mu)_{n}$ starting with distribution $\theta$ and moving accordingly to the transition probability $\mu$. We write $P^n_\mu$ for the induced probability measure on $\Omega$. If $\theta$ is concentrated on a point $g \in \Gamma$ we write $P^n_\mu g$ or simply $P^n_\mu$ if the starting position is the identity $\theta = \delta_e$.

The action $\Gamma \curvearrowright \Omega$ induces an action on the probability measures on $(\Omega, \mathcal{F})$ via push forward and the key point is that $h \cdot (P^n_\mu g) = P^n_\mu hg$. Since the left multiplication is transitive, this is telling us that to describe the measure space $(\Omega, \tilde{\nu})$ of Subsection 3.3.2 it is enough to know a single probability measure $P^n_g$ and then all the other can be retrieved by means of the action $\Gamma \curvearrowright \Omega$.

**Remark 3.3.15.** Notice that the action $\Gamma \curvearrowright \Omega$ is measurable also with respect to the invariant $\sigma$-field $I$. Thus we have that it induces a map between Poisson boundaries and we have

$$h \cdot \mathcal{B}((X^n_g)_n) = \mathcal{B}((X^n_{hg})_n).$$

We will now focus our interest on the invariant $\sigma$-field. Recall that we denoted by $T: \Omega \to \Omega$ the left shift $T((g_m))_n := g_{n+1}$ and that the push forward under $T$ is given by

$$T_* P_g = \sum_{h \in \Gamma} \mu(h) P_{hg}.$$
Notice that every invariant set $A \in \mathcal{I}$ is by definition invariant under $T$. In particular, the push forward $T_\ast \mathcal{L}_\mu$ coincides with $\mathcal{L}_\mu$ when restricted to the invariant $\sigma$-field. An obvious but remarkable consequence of this fact, is that the probability measures $\mathcal{L}_\mu$ are $\mu$-stationary under the action of $\Gamma$ on the Poisson boundaries $(\Omega, \mathcal{I}, \mathcal{L}_\mu)$. That is, for any starting point $g \in \Gamma$ the probability measure $\mathcal{L}_\mu$ on the $\sigma$-field $\mathcal{I}$ is equal to the mean of its translates:

$$
\mathcal{L}_\mu = T_\ast \mathcal{L}_\mu = \sum_{h \in \Gamma} \mu(h) \mathcal{L}_{hg} = \sum_{h \in \Gamma} \mu(h)(h \cdot \mathcal{L}_\mu).
$$

**Remark 3.3.16.** Observe that the above equality is obviously false in the product $\sigma$-field $\mathcal{F}$.

As an immediate consequence, we obtain that whenever an element $h_0$ belongs to the semi-group generated by the support of the probability measure $\mu$, then for every probability measure $\mathcal{L}_\mu$ its translate $h_0 \cdot (\mathcal{L}_\mu)$ is absolutely continuous with respect to $\mathcal{L}_\mu$. That is, any invariant measurable set $A \in \mathcal{I}$ which is negligible with respect to $\mathcal{L}_\mu$ is also negligible with respect to $h_0 \cdot (\mathcal{L}_\mu)$:

$$
\mathcal{L}_\mu(A) = 0 \implies \mathcal{L}_{h_0 \cdot \mu}(A) = 0.
$$

In fact, for any such $h_0$ there exists an integer $n$ large enough so that $\mu_n(h_0) > 0$ (recall that $\mu_n$ denotes the $n$-th convolution of $\mu$). Thus, iterating the $\mu$-stationarity we have

$$
\mathcal{L}_\mu(A) = \sum_{h \in \Gamma} \mu_n(h)(h \cdot \mathcal{L}_\mu(A)) \geq \mu_n(h_0)(h_0 \cdot \mathcal{L}_\mu(A)).
$$

It follows that if both $h_0$ and $h_0^{-1}$ belong to the semigroup generated by $\text{Supp}(\mu)$, then for every $g \in \Gamma$ the probability measures $\mathcal{L}_\mu$ and $h_0 \cdot \mathcal{L}_\mu$ have the same measure-zero invariant subsets. When this happen, the probability measure $\mathcal{L}_\mu$ is called quasi-invariant under the action of $h_0$.

Now it is easy to relate Poisson boundaries with harmonic functions. Recall that a random walk is irreducible if for every pair of states $x$ and $y$ there exists an integer $n$ such that $p_n(x, y) > 0$.

**Proposition 3.3.17.** If a probability measure $\mu$ on a countable group $\Gamma$ generates an irreducible random walk $(X^n)_n$, then the space of bounded invariant functions on $\Omega$ and the space of bounded functions on the Poisson boundary are naturally isomorphic

$$
L^\infty_{\text{inv}}(\Omega) \cong L^\infty(\mathcal{B}((X^n)_n)).
$$

**Proof.** Since both those spaces are equivalence classes of bounded real valued functions of $\Omega$, to prove the Proposition it is enough to show that a set is negligible under $\tilde{\nu}$ if and only if it is negligible under $\mathcal{L}_\mu$. 

For any set $A \subset \Omega$, it is clear that $\tilde{\nu}(A) = 0$ implies $P^\mu(A) = 0$. Conversely, if $P^\mu(A) = 0$ then by irreducibility we have that $P^\mu$ is quasi-invariant under the action of the whole group $\Gamma$. Thus we deduce that $P^\mu_g(A) = 0$ for every $g \in \Gamma$ and we conclude because we have that
\[
\nu(A) = \sum_{g \in \Gamma} P^\mu_g(A) = 0.
\]

Denoting by $\mathcal{H}^\infty(\Gamma, \mu)$ the space of bounded harmonic functions with respect to the Markov process induced by $\mu$, Theorem 3.3.12 yields:

**Corollary 3.3.18.** If a probability measure $\mu$ on a countable group $\Gamma$ generates an irreducible random walk $(X^\mu)_n$, then there is a one to one correspondence
\[
\mathcal{H}^\infty(\Gamma, \mu) \leftrightarrow L^\infty(\mathcal{B}((X^\mu)_n)).
\]
In particular, the Poisson boundary $\mathcal{B}((X^\mu)_n)$ is trivial if and only if every bounded harmonic function is constant.

**Remark 3.3.19.** It is also possible to prove Corollary 3.3.18 directly defining the inverse map $E : L^\infty(\mathcal{B}((X^\mu)_n)) \to \mathcal{H}^\infty(\Gamma, \mu)$ as
\[
E(f)(h) := \int_\Omega f(g)d(h \cdot P^\mu).
\]

**Remark 3.3.20.** We found convenient to assume $\mu$ to generate an irreducible random walk. Still, paying some extra care to measure-zero subsets, it is possible to extend this theory also for those probability measures $\mu$ whose support generates $\Gamma$ as a group (and non-necessarily as a semi-group). In this context, the correspondence of Corollary 3.3.18 has to be modified, but it is still true that the generated Poisson boundary is trivial if and only if every bounded harmonic function is constant.

We will now state some results from [KV83] that link Poisson boundaries and amenability. Recall that in Subsection 1.2.2 we characterized amenability for groups in terms of invariant means. Notice that we can view probability measures on $\Gamma$ as functionals on $L^\infty(\Gamma)$ sending a function to its integral. In particular, it is well-defined the convergence of probability measures in the space of functionals $L^\infty(\Gamma)^*$. We say that a random walk on a group with transition probability $\mu$ is **aperiodic** if the greatest common divisor of the set of integers $\{n \mid \mu_n(e) > 0\}$ is 1. Notice that if there is a non-trivial probability to stand still (i.e. $\mu(e) > 0$) then the random walk is trivially aperiodic. The following holds:

**Theorem 3.3.21.** If a probability measure $\mu$ on a countable group $\Gamma$ generates an irreducible aperiodic random walk $(X^\mu)_n$, then the Poisson boundary $\mathcal{B}((X^\mu)_n)$ is trivial if and only if the convolutions $\mu_n$ converge to a left-invariant mean on $L^\infty(\Gamma)$. 
3.3. HARMONIC FUNCTIONS AND POISSON BOUNDARY

See [KV83, Theorem 4.2] for a proof. There it is also shown how to weaken the hypothesis on aperiodicity in order to obtain the following.

**Corollary 3.3.22.** If a countable group $\Gamma$ is non-amenable then every irreducible random walk on $\Gamma$ has non-trivial Poisson boundary.

Another interesting result is the following. (See [KV83, Theorem 4.4].)

**Theorem 3.3.23.** A countable group $\Gamma$ admits a probability measure $\mu$ whose support is the whole $\Gamma$ and with trivial Poisson boundary if and only if $\Gamma$ is amenable.

**Remark 3.3.24.** Both Corollary 3.3.22 and Theorem 3.3.23 work also for countable groups that are not finitely generated. (Recall that Von Neumann’s condition for amenability makes sense also if a group is not finitely generated.)

### 3.3.4 Entropy

An interesting concept that has important consequences in various fields (not only of mathematics) is that of information. Sadly enough, we cannot indulge in this subject and we will limit ourselves to some consequences of entropic criteria.

**Definition 3.3.25.** The entropy of a probability measure $\mu$ on a countable set $S$ is defined as

$$H(\mu) := -\sum_{x \in S} \mu(x) \log(\mu(x))$$

(recall that we use the convention that $t \log(t)$ is zero when $t = 0$).

Notice that the entropy is always positive (possibly infinite) and is zero if and only if the probability measure is a delta function. If the state space $S$ is finite, then the probability distribution which maximizes the entropy is the uniform distribution. Indeed, applying Jensen inequality we obtain:

$$H(\mu) = \sum_{x \in S} \mu(x) \log \left( \frac{1}{\mu(x)} \right) \leq \log \left( \sum_{x \in S} \mu(x) \frac{1}{\mu(x)} \right) = \log(\lvert S \rvert)$$

and the maximum is attained by the uniform distribution $\mu(x) := 1/\lvert S \rvert$.

**Remark 3.3.26.** Actually we have just proven that whenever the support of $\mu$ is finite, then also the entropy is finite and satisfies $H(\mu) \leq \log(\lvert \text{Supp}(\mu) \rvert)$

The entropy is a very useful concept and has profound implication in the theory of Markov processes. Still, we will restrict our attention to some of its applications on the theory of random walks on groups. From now on the state space will be a countable group $\Gamma$. An easy computation leads to the following:
Lemma 3.3.27. If $\mu$ and $\nu$ are two probability measures on $\Gamma$, then the entropy of their convolution is smaller than or equal to the sum of their entropies

$$H(\mu * \nu) \leq H(\mu) + H(\nu).$$

Proof. We have:

$$H(\mu * \nu) = - \sum_{x,y \in \Gamma} \mu(y)\nu(y^{-1}x) \log (\mu * \nu(x))$$

$$\leq - \sum_{x,y \in \Gamma} \mu(y)\nu(y^{-1}x) \left[ \log (\mu(y)) + \log (\nu(y^{-1}x)) \right].$$

Splitting the sum and letting $z = y^{-1}x$ we obtain that the latter is equal to

$$- \sum_{y \in \Gamma} \mu(y) \log (\mu(y)) \sum_{z \in \Gamma} \nu(z) - \sum_{z \in \Gamma} \nu(z) \log (\nu(z)) \sum_{y \in \Gamma} \mu(y)$$

that is exactly $H(\mu) + H(\nu)$. \qed

As a corollary, we obtain that the entropy of the $n$-fold convolution $\mu_n = \mu * \cdots * \mu$ is a subadditive function of $n$. In particular, the sequence $H(\mu_n)/n$ admits a limit by the Fekete Lemma (see Lemma 1.1.18). Thus we give the following definition:

Definition 3.3.28. The asymptotic entropy $h(\mu)$ of a random walk with transition probability $\mu$ on a group $\Gamma$ is the limit

$$h(\mu) := \lim_{n \to \infty} \frac{H(\mu_n)}{n}.$$  

Remark 3.3.29. It is clear that $H(\mu_n)/n$ is smaller than or equal to $H(\mu)$ and one can also see that if $H(\mu)$ is infinite then also $H(\mu_n)$ is infinite for every $n$. Thus the asymptotic entropy is finite if and only if so is the entropy of $\mu$. If this is the case we have $h(\mu) \leq H(\mu)$.

Remark 3.3.30. Since $H(\mu_n) \leq \log (|\text{Supp}(\mu_n)|)$, we have that whenever $\Gamma$ is a finitely generated group with subexponential growth and $\mu$ has finite support the asymptotic entropy $h(\mu)$ is zero.

The asymptotic entropy is a useful tool in the study of Poisson boundaries. For example, the following holds. (See [KV83, Theorem 1.1].)

Theorem 3.3.31. Let $\mu$ be a probability measure with finite entropy on a countable group $\Gamma$. Then the Poisson boundary of the random walk with transition probability $\mu$ is trivial if and only if $h(\mu) = 0$.

Corollary 3.3.32. If $\Gamma$ is a group with subexponential growth and $\mu$ is a probability measure with finite entropy, then the induced random walk has trivial boundary.
3.3. HARMONIC FUNCTIONS AND POISSON BOUNDARY

In particular, we have just provided another proof of the fact that groups with subexponential growth are amenable. Actually, the proof we gave in Subsection 1.2.2 is much more direct. Still the approach with random walks could be useful when using Von Neumann’s condition for amenability.

Remark 3.3.33. If \( \Gamma \) has polynomial growth then Corollary 3.3.32 holds also without the hypothesis of finite entropy. Indeed, there is a profound theorem that implies that \( \Gamma \) is virtually nilpotent and it is known that every random walk on a virtually nilpotent group has trivial Poisson boundary. In general though, finiteness of entropy is essential.

Notice that combining Theorem 3.3.31 with Corollary 3.3.22 we obtain that if a group \( \Gamma \) has a probability measure \( \mu \) whose support generates \( \Gamma \) (as a semi-group) and such that \( h(\mu) = 0 \), then \( \Gamma \) is amenable.

The following theorem can be used to compute the asymptotic entropy:

**Theorem 3.3.34.** Let \( \mu \) be a probability measure with finite entropy \( H(\mu) \) on a countable group \( \Gamma \). Then for \( \mathbb{P}^\mu \)-almost every sample path \( g = (g_n)_{n \in \mathbb{N}} \in \Omega \) the following holds:

\[
\lim_{n \to \infty} \frac{\log \left( \mu_n(g_n) \right)}{n} = -h(\mu).
\]

**Remark 3.3.35.** Combining Theorems 3.3.34 and 3.3.31 we obtain that the Poisson boundary of a random walk with finite entropy is trivial if and only if

\[
\lim_{n \to \infty} \frac{\log \left( \mu_n(g_n) \right)}{n} = 0
\]

for almost every sample path \( g \). It is interesting to compare this result with the amenability criterion given by the spectral radius (see Definition 3.2.15 and Theorem 3.2.31), because an easy computation shows that the logarithm of the spectral radius is exactly equal to

\[
\log \left( \rho(\mu) \right) = \limsup_{n \to \infty} \frac{\log \left( \mu_n(e) \right)}{n}.
\]
Chapter 4

Geometric boundaries of random walks

In this chapter we will finally use the theory we developed so far. The general scheme will be to compose random walks on groups with actions of groups on topological spaces in order to obtain random walks on topological spaces. Then we will try to study the latter to obtain information both on the topological space and on the group.

In the first section, we will talk about random walks on topological spaces in general and we will try to give some example to justify the theorems that we will not be able to prove later on. Here we will use various results from Chapters 1 and 3. In the last section, we specifically study random walks on the mapping class group with its relative metric (this is somewhat equivalent to studying random walks on the curve complex). The final result will be the proof of the fact that a random walk generally ends up walking among pseudo-Anosov elements. Many results from Chapter 2 are needed here.

4.1 Motivating examples

Here we introduce random walks on topological spaces as compositions of random walks on groups and topological actions. Our main objective is to make apparent some relations between boundaries of random walks as they are defined in Section 3.3 and convergence at infinity in topological spaces.

After some general discussion, we will focus on the case of random walks on the hyperbolic plane $\mathbb{H}^2$, where we can use various results from Chapters 1 and 3 to easily justify our claims.

As this section is more of exemplificative character, many facts are not completely proved and the formalism is somewhat relaxed. Our discussion is mainly inspired from remarks from [Pet13] and [KM96].
4.1.1 General observations

Let $\Gamma$ be a countable group acting (on the left) on a topological space $X$ and assume that there is some sort of boundary $\partial X$ for $X$ and a topology on the complete space $\overline{X} = X \cup \partial X$ such that the action $\Gamma \curvearrowright X$ naturally extends to an action on the whole $\overline{X}$. (The example that one should keep in mind is that of a $\delta$-hyperbolic space $X$ with the Gromov boundary $\partial_{\infty} X$. See Subsection 1.3.2.)

With the notations of Subsection 3.3.3, let $\mu$ be a probability measure on $\Gamma$, and $\Omega = \Gamma^\mathbb{N}$ the state space. Denote by $(X^\mu_\theta)_n$ the random walk with starting distribution $\theta$ and transition probability $\mu$ and denote by $P^\mu_\theta$ the induced probability measure on $\Omega$ (if $\theta$ is not specified we assume the random walk to start from the identity $e \in \Gamma$). Choosing a base point $\bar{x} \in X$, we obtain a map $\Gamma \to X$ sending an element $g$ to $g(\bar{x})$. In particular, we obtain a random walk on the topological space $\overline{X}$ by composition. Assume now to know that for almost every sample path $g = (g_n)_{n \in \mathbb{N}} \in \Omega$ the sequence $g_n(\bar{x})$ converges to a point $g_\infty \in \partial X$. Then we obtain a probability measure $\nu$ on $\partial X$ letting

$$\nu(A) := P^\mu_\theta \left\{ g \in \Omega \mid g_n(\bar{x}) \to g_\infty, \ g_\infty \in A \right\}$$

for every Borel subset $A \subseteq \partial X$ (actually, here we are assuming the topology on $\overline{X}$ to be somewhat decent, e.g. Hausdorff and first countable). We will call such $\nu$ hitting measure or harmonic measure. It is very interesting to compare the measure space $(\partial X, \nu)$ with the Poisson boundary $B((X^\mu)_n)$.

Notice that the hitting measure can be described as a push-forward. In fact, our assumptions imply that the map $\pi: \Omega \to \partial X$ sending $(g_n)_{n \in \mathbb{N}}$ to $g_\infty$ is well-defined up to a $P^\mu_\theta$-measure zero set and the hitting measure $\nu$ is equal to $\pi^*_\mu(P^\mu_\theta)$.

The group $\Gamma$ acts on the probability space $(\partial X, \mu)$ in two ways. That is, for every $g \in \Gamma$ one can both consider the push forward of the boundary map $\partial g: \partial X \to \partial X$ (we assumed that the action of $\Gamma$ extends to the whole $\overline{X}$)

$$\nu \mapsto \partial g_*(\nu);$$

or the pre-composition of the push-forward $\pi_\mu$ with the left multiplication by $g$ in $\Omega$

$$\nu \mapsto \pi_\mu[(L_g)_*(P^\mu)].$$

These actions coincide because for every Borel subset $A \subseteq \partial X$ we have

$$\partial g_*\nu(A) = \nu(g^{-1} \cdot A) = P^\mu[\pi^{-1}(g^{-1} \cdot A)]$$

and

$$\pi_*[(L_g)_*(P^\mu)](A) = P^\mu[L_{g^{-1}}(\pi^{-1}(A))]$$

and these two expressions are equivalent because the action of $\Gamma$ is compatible with $\pi$. 

4.1. MOTIVATING EXAMPLES

Notice that \( \pi_*[(L_g)_*(P^\mu)] = \pi_* (P_g^\mu) \) (where \( P_g^\mu \) is \( P^\mu \) with \( \theta = \delta_g \)). Thus, if \( \nu \) is the hitting measure induced by the random walk starting at \( \bar{x} \), then \( g \cdot \nu \) is the hitting measure induced by the random walk starting at \( g(\bar{x}) \).

On our way to compare \( (\partial X, \nu) \) with the Poisson boundary \( B((X^\mu)_n) \), we notice that the hitting measure is \( \mu \)-stationary. Indeed, it is clear that the hitting measure induced by the random walk \( (X^\mu)_n=0,1,2,... \) is equal to the hitting measure induced by the random walk \( (X^\mu)_n=1,2,... \) and the latter is equal to \( (X_\theta^\mu)_n \) with \( \theta = \sum_{h \in \Gamma} \mu(h) \delta_h \). Thus we have

\[
\sum_{h \in \Gamma} \mu(h)(h \cdot \nu) = \sum_{h \in \Gamma} \mu(h) \pi_*[(L_h)_*(P^\mu)] = \pi_* \left[ \sum_{h \in \Gamma} \mu(h)(L_h)_*(P^\mu) \right],
\]

where the latter is equal to \( \pi_* (P_\theta^\mu) \) and we already noticed that this is equal to the hitting measure \( \pi_* (P^\mu) \).

Recall that in Subsection 3.3.3 we found a map from \( L^\infty(B((X^\mu)_n)) \) to the space of bounded harmonic functions \( H^\infty(\Gamma, \mu) \). The same idea works also in this context. For any bounded measurable function \( f: \partial X \to \mathbb{R} \) and for any element \( g \in \Gamma \) we can start a random walk at the point \( g(\bar{x}) \) and then look at the expected exit value. That is, we define a function \( u: \Gamma \to \mathbb{R} \) letting \( u(g) \) be expectation of \( f \) under the hitting measure \( g \cdot \nu \):

\[
u(g) := \int_{\partial X} f(x) d(g \cdot \nu) = \int_{\partial X} f(g^{-1}(x)) d\nu.
\]

Also in this case the function \( u \) is harmonic. Indeed, we have:

\[
\sum_{h \in \Gamma} \mu(h) u(gh) = \sum_{h \in \Gamma} \mu(h) \int_{\partial X} f(x) d(gh \cdot \nu) \\
= \int_{\partial X} f(g^{-1}(x)) d\left( \sum_{h \in \Gamma} \mu(h)(h \cdot \nu) \right) \\
= \int_{\partial X} f(g^{-1}(x)) d\nu = u(g).
\]

Paying some extra care to measure-zero subsets of \( \partial X \), we obtain a map \( E: L^\infty(\partial X) \to H^\infty(\Gamma, \mu) \).

Remark 4.1.1. These similarities between \( (\partial X, \nu) \) and \( B((X^\mu)_n) \) are not casual. Indeed, the Poisson boundary is a very profound object and it is possible to define it in many different ways. One of the possible definitions is exactly as a sort of ‘maximal’ measure space \( (\tilde{B}, \tilde{\nu}) \) with an action \( \Gamma \curvearrowleft \tilde{B} \) such that \( \tilde{\nu} \) is \( \mu \)-stationary. Then, any other space like \( (\partial X, \nu) \) can be identified with a quotient of the Poisson boundary \( (\tilde{B}, \tilde{\nu}) \). See [KM96] and [KV83] for a broader introduction to the Poisson boundary.
CHAPTER 4. GEOMETRIC BOUNDARIES OF RANDOM WALKS

4.1.2 Random walks on the hyperbolic plane

We now want to give more concrete examples by studying actions of groups on the hyperbolic plane. So, $\Gamma$ will continue to be a countable group acting on a topological space and this time the topological space will be the hyperbolic plane $X = H^2$. Here we have a natural boundary $\partial X = \partial_\infty H^2 \cong S^1$ and the total space is homeomorphic to the closure of the Poincaré disk $\overline{X} = \mathbb{H}^2 \cong \mathbb{D}^2$. Moreover, we will assume the action $\Gamma \curvearrowright \mathbb{H}^2$ to be an action by (orientation-preserving) isometries, so that it naturally extends to the Gromov compactification $\overline{\mathbb{H}^2}$. (See Subsection 1.3.2.)

As a warming up we start by considering cyclic groups $\Gamma = \langle \gamma \rangle$ with $\gamma \in \text{Isom}^+(\mathbb{H}^2)$. Thus we have a map $\Gamma \to \mathbb{H}^2$ sending $\gamma^n$ to $\gamma^n(0)$ (if $\gamma$ is not a rotation of centre 0 this map is also an embedding). We ask now if it is possible that a random walk on $\Gamma$ gives rise to a non-trivial hitting measure on $\partial_\infty \mathbb{H}^2$.

It is clear that if $\gamma$ is an elliptic isometry then the set $\gamma^n(0)$ is at bounded distance from the origin and hence no random walks can converge to the boundary at infinity. If $\gamma$ is parabolic or hyperbolic the sequence $\gamma^n(0)$ has respectively one or two points in $\partial_\infty \mathbb{H}^2$. Now, it is easy to see that for any probability measure $\mu$, the random walk on $\Gamma$ will converge to $\partial_\infty \mathbb{H}^2$ if and only if it is not recurrent. That is, the set of sample paths $(g_n)_{n \in \mathbb{N}}$ such that $g_n = e$ infinitely many times is negligible.

Thus we need to study random walks on the cyclic group $\mathbb{Z}$ to see if they are recurrent or not. Such a random walk is given by a probability measure $\mu$ on $\mathbb{Z}$ not necessarily symmetric. Seeing $\mathbb{Z}$ as a subset of $\mathbb{R}$, we can compute the expected position after the first step of the random walk as

$$\epsilon = \sum_{k \in \mathbb{Z}} k \mu(k)$$

(we are assuming $\mu$ to be integrable). Iterating, we have that the expected position at the $n$-th step of the random walk is $n \epsilon$. It is quite easy to see that as the random walk goes on the distance between the actual position and the expected position is bounded almost surely by a logarithmic function. In particular, we have that if $\mu$ is not balanced (i.e. $\epsilon$ is not 0) then the corresponding random walk is not recurrent and it can be seen as flowing in direction of $\epsilon$. If that is the case, the random walks on $\Gamma$ gives rise to a hitting measure $\nu$ on $\partial_\infty \mathbb{H}^2$ and such a measure will be concentrated on a single point at the infinity. On the contrary, if $\mu$ is balanced then one can show that the random walk is recurrent and hence it does not generate a harmonic measure.

Let $\Gamma < \text{Isom}^+(\mathbb{H}^2)$ be a non-cyclic subgroup of isometries and $\mu$ a probability measure on $\Gamma$. Denote by $\text{sgr}(\mu) \subseteq \Gamma$ the semi-group generated by the support of $\mu$ and by $\text{gr}(\mu) < \Gamma$ the generated group. Now, a priori we do not know if a random walk on $\Gamma$ will converge almost surely to the
boundary $\partial_\infty \mathbb{H}^2$ and hence we do not know whether it generates a hitting measure $\nu$ or not. Still, if we assume that $\text{gr}(\mu)$ contains two hyperbolic isometries $\gamma_1$ and $\gamma_2$ whose axes have disjoint endpoints, then we do know that if such a $\nu$ exists it must be non-atomic. That is, if $\text{gr}(\mu)$ is large enough there cannot be a single point $x_\infty \in \partial_\infty \mathbb{H}^2$ that has strictly positive measure $\nu(\{x_\infty\}) > 0$.

The reason why this should be true is the following lemma:

**Lemma 4.1.2.** Let $\mu$ be a probability measure on a group $\Gamma$ and let $\Gamma$ act on a probability space $(X, \nu)$ in such a way that $\nu$ is $\mu$-stationary. If a measurable set $A \subset X$ has infinitely many images under the action of $\text{gr}(\mu)$ and for each $g \in \text{gr}(\mu)$ either $A = gA$ or $\nu(A \cap gA) = 0$, then $\nu(A) = 0$.

Taking Lemma 4.1.2 for granted, our claim is clearly true. Indeed, we have seen in Subsection 4.1.1 that any boundary $\partial_\infty X$ equipped with a hitting measure $\nu$ is $\mu$-stationary. Moreover it is easy to see that any point $x_\infty$ in the boundary $\partial_\infty \mathbb{H}^2$ has infinitely many images under the group $\langle \gamma_1, \gamma_2 \rangle$ generated by two hyperbolic isometries whose axes have disjoint endpoints. Hence the Lemma applies.

**Remark 4.1.3.** It is not enough to ask that $\gamma_1$ and $\gamma_2$ have different axes. Consider for example the group $\Gamma = \langle \gamma, \eta \rangle$ generated by an hyperbolic isometry $\gamma$ and a parabolic isometry $\eta$ whose fixed point $x_\infty$ is contained in the endpoints of the axis of $\gamma$. Then conjugating $\gamma$ yields infinitely many hyperbolic isometries but all of them contain $x_\infty$ in their axis. It is easy to exhibit a probability measure $\mu$ such that $\text{gr}(\mu) = \Gamma$ and the hitting measure is concentrated on $\{x_\infty\}$.

Before giving the proof of Lemma 4.1.2 we need a simple technical result that allows us to deal with asymmetric random walks as well.

**Lemma 4.1.4.** Let $\Gamma$ a group generated by a set $S$ and let $\text{sgr}(S) \subseteq \Gamma$ be the semigroup generated by $S$. If $\Gamma$ acts on a set $X$ and $x$ is any element of $X$, then the orbit of $x$ under the action of $\text{sgr}(S)$ is finite if and only if it is finite under the action of the whole group $\Gamma$.

**Proof.** Denote the orbit under the action of $\text{sgr}(S)$ by

$$O(x) = \{h(x) \mid h \in \text{sgr}(S)\}.$$

If $O(x)$ is finite, then every $h$ is $\text{sgr}(S)$ acts as a permutation on $O(x)$ and hence also $h^{-1}$ acts as a permutation on $O(x)$. Thus it is easy to conclude because by definition any element $g \in \Gamma$ can be written as $g = s_1^{\varepsilon_1} s_2^{\varepsilon_2} \cdots s_n^{\varepsilon_n}$ with $s_i \in S$ and $\varepsilon_i = \pm$. Since $s_i^{\varepsilon_i}$ acts as a permutation on $O(x)$, so does $g$. In particular, the orbit of $x$ under $\Gamma$ is exactly $O(x)$. \qed
CHAPTER 4. GEOMETRIC BOUNDARIES OF RANDOM WALKS

Proof of Lemma 4.1.2. Up to restricting ourselves to \( \text{gr}(\mu) \) we can assume \( \text{gr}(\mu) = \Gamma \). Let

\[
A' = A \setminus \left( \bigcup_{gA \neq A} gA \right).
\]

By hypothesis \( \nu(A') = \nu(A) \), hence up to replacing \( A \) with \( A' \) we can assume that for every \( h, g \) in \( \Gamma \) the sets \( hA \) and \( gA \) either coincide or are disjoint.

Consider now the supremum of the probabilities of translates of \( A \)

\[
K = \sup_{g \in \Gamma} \nu(gA).
\]

If the supremum is not attained, there exists a sequence of disjoint translates \( \{g_1A, g_2A, \ldots \} \) such that \( \nu(g_nA) \) tends to \( K \). In particular if \( K > 0 \) there would be infinitely many disjoint sets \( g_iA \) with probability greater than \( K/2 \) which gives a contradiction because \( \nu \) is finite.

Suppose now that the supremum is attained for a certain \( g \in \Gamma \). Since \( \nu \) is \( \mu \)-stationary we have

\[
K = \nu(gA) = \sum_{h \in \Gamma} \mu(h) h \cdot \nu(gA) = \sum_{h \in \Gamma} \mu(h) \nu(h^{-1}gA).
\]

Applying the equality \( n \) times yields

\[
K = \sum_{h \in \Gamma} \mu_\nu(h) \nu(h^{-1}gA),
\]

whence we deduce by maximality of \( K \) that \( \nu(h^{-1}gA) = K \) whenever \( \mu_\nu(h) > 0 \). If we knew that there are infinitely many disjoint \( h^{-1}gA \) with \( h \in \text{sgr}(\mu) \) we would be done because

\[
1 \geq \nu \left( \bigcup_{h \in \text{sgr}(\mu)} h^{-1}gA \right) = K \cdot \# \{ \text{disjoint } h^{-1}gA \}.
\]

Thus we only need to show that if \( gA \) has infinitely many images under the action of \( \text{gr}(\mu) \), then it must also have infinitely many images under the action of \( \text{sgr}(\mu) \). This follows easily from Lemma 4.1.4 considering the induced action on the set of parts \( \mathcal{P}(X) \).

4.1.3 Convergence at infinity in the hyperbolic plane

We wish to conclude this section giving some examples of random walks on groups of isometries \( \Gamma < \text{Isom}^+(\mathbb{H}^2) \) whose images do converge to the boundary at infinity \( \partial_\infty \mathbb{H}^2 \). To do so, it is convenient to speak about exit speed for random walks in general. In particular, we will prove the following:
Proposition 4.1.5. Let $\mu$ be a symmetric probability measure on a finitely generated group $\Gamma$. If the spectral radius $\rho(\mu)$ of the induced random walk is strictly smaller then 1, then for any finite set of generators $S$ there exists a constant $\alpha > 0$ such that for almost every sample path $g = (g_n)_{n \in \mathbb{N}}$

$$\liminf_{n \to \infty} \frac{d_w(e, g_n)}{n} \geq \alpha,$$

where $d_w$ represent the word distance with respect to $S$. 

Proof. Notice that since the random walk $(X_n^\mu)$ is symmetric, the counting measure $\lambda(A) = |A|$ is a reversible measure. In particular, we can apply all the results of Section 3.2. Recall that the Markov operator $P: \ell^2(\Gamma) \to \ell^2(\Gamma)$ is defined as

$$P(f)(g) = \sum_{h \in \Gamma} \mu(g^{-1}h) f(h)$$

and its norm is equal to the spectral radius $\|P\|_2 = \rho(\mu)$. (We adapted the notations to random walks on groups.)

For any $g, h \in \Gamma$ we have that

$$\mu_n(g) = \langle 1_e, P^n 1_{g^{-1}} \rangle$$

$$\leq \|1_e\|_2 \|P^n\|_2 \|1_g\|_2$$

$$\leq \|P\|_2^n = \rho(\mu)^n,$$

where $\mu_n$ is the $n$-th convolution of $\mu$. Choose $\alpha > 0$ small enough such that the constant

$$C := (2|S|)^\alpha \rho(\mu)$$

is strictly smaller then 1 and write $A_n$ for the set of sample paths whose $n$-th component stays at most at distance $\alpha n$ from the identity

$$A_n = \{ g \in \Omega \mid d_w(e, g_n) \leq \alpha n \}.$$

Then we have that

$$P^\mu(A_n) = \sum_{g \in B_{\alpha n}(e)} \mu_n(g)$$

$$\leq |B_{\alpha n}(e)| \rho(\mu)^n$$

$$\leq [(2|S|)^\alpha \rho(\mu)]^n \leq C^n.$$

This implies that the summation over $n \in \mathbb{N}$ of $P^\mu(A_n)$ is finite. In particular, the set $A$ of sample paths that lie in infinitely many $A_n$’s must have zero probability. In fact, $A$ can be written as

$$A = \bigcap_{n \in \mathbb{N}} \left( \bigcup_{m > n} A_m \right).$$
Thus for every $n$ we have

$$P^\mu(A) \leq P^\mu\left(\bigcup_{m>n} A_m\right) \leq \sum_{m>n} P^\mu(A_m)$$

and this goes to 0 as $n$ goes to infinity.

Recalling that Theorem 3.2.31 implies that every symmetric irreducible random walk $X_n$ on a non-amenable group $\Gamma$ has spectral radius $\rho(X_n) < 1$, we obtain the following corollary. (Recall that a random walk $(X^\mu)_n$ on a group $\Gamma$ is irreducible if sgr($\mu$) = $\Gamma$.)

**Corollary 4.1.6.** Let $\Gamma$ be a finitely generated group with a word distance $d_w$. If $\Gamma$ is non-amenable, then for every symmetric irreducible random walk $X_n$ on $\Gamma$ there exist a constant $\alpha > 0$ such that

$$\liminf_{n \to \infty} \frac{d_w(e, X_n)}{n} \geq \alpha$$

almost surely.

**Remark 4.1.7.** The proof of Proposition 4.1.5 works also if the state space is not a group. The only things that are truly needed are the reversible measure and a distance such that the volume of metric balls grows at most exponentially with the radius.

**Remark 4.1.8.** For a random walk $X_n$ on a finitely generated group it is easy to see that the expected value of the distance from the origin at the $n$-th step $E[d_w(e, X_n)]$ is a subadditive function. Thus by the Fekete Lemma (Lemma 1.1.18) there exists the limit

$$\lim_{n \to \infty} \frac{E[d_w(e, X_n)]}{n}.$$

This limit is the **linear exit speed** of the random walk. By Corollary 4.1.6 we deduce that an irreducible random walk on a non-amenable group must have positive linear exit speed.

Now we can use Corollary 4.1.6 to study some random walks on the hyperbolic plane. Let $\Gamma$ be the fundamental group of a closed surface $S_g$ with genus $g \geq 2$. Then $S_g$ admits hyperbolic metrics. For any choice of a hyperbolic structure on $S_g$, we obtain a Riemannian cover $\mathbb{H}^2 \to S_g$ and the group of deck transformation is isomorphic to $\Gamma$. (The isomorphism being unique up to conjugation.)

In other words, for any choice of hyperbolic metrics on $S_g$ we obtain an embedding $\Gamma \to \mathbb{H}^2$ such that $\mathbb{H}^2/\Gamma$ is isometric to $S_g$. Then, by the Milnor-Švarc Theorem (Theorem 1.1.10) the map sending an element $g \in \Gamma$ to $g(0) \in \mathbb{H}^2$ is a quasi-isometry between $\Gamma$ and $\mathbb{H}^2$. 
It is easy to see that the fundamental group of a closed surface with genus \( g \geq 2 \) is non-amenable. For example, it contains a copy of the free group of rank 2 and subgroups of amenable groups are amenable (Proposition 1.2.20). Now, Corollary 4.1.6 implies that for any symmetric irreducible random walk \((X^n)\) on \( \Gamma \) the distance from the origin \( \mathbf{e} \) increases linearly almost surely. Since we know that the map \( g \mapsto g(0) \) is a quasi-isometry, we deduce that also for the induced random walk on \( \mathbb{H}^2 \) the distance from the origin \( 0 \) increases linearly almost surely.

To conclude that \( g_n(0) \) converges almost surely to the boundary at infinity \( \partial_\infty \mathbb{H}^2 \), it is hence enough to prove that it does converge somewhere in the Gromov compactification \( \mathbb{H}^2 = \mathbb{H}^2 \cup \partial_\infty \mathbb{H}^2 \). To do so, one can notice that the random walk travels at finite speed (almost surely). Thus, when a sample path is bound to stay outside from a very large ball centred at 0 then it cannot move too badly in the Poincaré disc \( \mathcal{D}^2 \). Notice that this fact is not trivial but one has to make some estimates.

**Remark 4.1.9.** Essentially the same idea works also if \( \Gamma \) is the fundamental group of a compact surface \( S_{g,b} \) (i.e. with boundary but without punctures) with \( \chi(S_{g,b}) < 0 \). Indeed, also in this case we can choose a hyperbolic structure. The Riemannian cover then is a convex subset of \( \mathbb{H}^2 \) and it is again quasi-isometric to \( \Gamma \).

### 4.2 Random walks on the mapping class group

In this final section we will prove that ‘most’ random walks on mapping class groups tend to walk among pseudo-Anosov elements with asymptotic probability 1. The general scheme will be to prove that random walks tend to exit from certain sets by looking at their asymptotic behaviour. That is, we will state a theorem asserting that sample paths converge to the relative boundary of the mapping class group and the study of the induced hitting measure will be an important tool throughout the section.

The theorem that we are going to prove is quite general and works for every non-sporadic surface. Actually, the idea is to prove a less general result first and then to generalize it taking some quotient under finite groups of isometries. Doing so however, it may happen that we end up working with sporadic surfaces. In that case, we conclude appealing to results from [Fur71].

Our proof follows that of J. Maher and relies heavily on many non-trivial results from Chapter 2. See [Mah11].

#### 4.2.1 Statement of the main theorem

If it is not differently specified, \( G \) will always denote the mapping class group of a non-sporadic surface \( S \). We will sometimes use the word metric \( d_w \) on
G; if that is the case, suppose a finite set of generator is fixed. The actual choice of generating set is of no influence.

We will usually denote a random walk with $X_n$ and its transition probability with $\mu$. From now on we will always assume random walks to start from the identity element $e$. We will use the notation 

$$\mu_n = P_{X_n} = \mu \ast \cdots \ast \mu$$

for the distribution probability of a random walk at the $n$-th step. We will write $\text{sgr}(\mu)$ to indicate the semi-group generated by the support of $\mu$ and $\text{gr}(\mu)$ will denote the generated subgroup. Notice that

$$\text{sgr}(\mu) = \bigcup_{n \in \mathbb{N}} \text{Supp}(\mu_n).$$

We can now state the main theorem we are going to prove in this chapter.

**Theorem 4.2.1 (Main Theorem).** Let $G$ be the mapping class group of a non-sporadic surface of finite type $S$. If $X_n$ is a random walk on $G$ with transition probability $\mu$ such that $\text{gr}(\mu)$ is a non-elementary subgroup of $G$, then

$$\lim_{n \to \infty} P_{X_n}\{g \in G \mid g \text{ is periodic or reducible}\} = 0.$$ 

Recall that on $G$ there is a relative metric which makes $\hat{G}$ quasi-isometric to the complex of curves of $S$. From now on we will consider such a relative metric to be fixed. In particular, $\hat{G}$ is hyperbolic and hence the Gromov boundary $\partial_{\infty} \hat{G}$ is defined. When we want to stress that an element $g$ or a subset $A$ of $G$ are considered with respect to the relative metric we use the notation $\hat{g}$ and $\hat{A}$. The quasi-isometry $\Psi_c: G \to \mathcal{C}(S)$ is obtained considering the map $g \mapsto g(c)$ where $c$ is the class of an essential closed curve (see Subsection 2.2.3). For simplicity we will fix an essential curve $\tilde{c}$ and always consider the quasi-isometry $\Psi_{\tilde{c}}$.

**Remark 4.2.2.** Since we are mainly interested in asymptotic properties of relative metrics we do not truly need to fix one of them nor a quasi-isometry (see Remark 2.2.23).

The following lemma will be crucial for the proof of the Main Theorem.

**Lemma 4.2.3.** If $g$ is a periodic or reducible element of the mapping class group of a non-sporadic surface $S$, then it is conjugated to an element of relative length at most $K$ where the constant $K$ is independent of $g$.

**Proof.** First of all, if $g$ is periodic then the statement is an obvious consequence of the finiteness of the conjugacy classes of finite subgroups of the mapping class group (Proposition 2.4.22).

If $g$ is reducible, then by definition it fixes a set of classes of disjoint curves $S = \{a_1, \ldots, a_n\}$. Since there are only finitely many curves up to
4.2. RANDOM WALKS ON THE MAPPING CLASS GROUP

homeomorphism, there exists an element \( h \in G \) such that \( d(h(\tilde{c}), a_i) \leq 1 \) for every \( i = 1, \ldots, n \) (where \( \tilde{c} \) is the curve used to define the quasi-isometry \( \Psi_{\tilde{c}} \)). Then, the conjugate element \( h^{-1} \circ g \circ h \) fixes the set \( h^{-1}S \) and the distance of \( \tilde{c} \) from any of the \( h^{-1}(a_i) \)'s is smaller than or equal to 1.

Thus, the image \( h^{-1}gh(\tilde{c}) \) is at most at distance 3 from \( \tilde{c} \). We deduce that the relative length of \( h^{-1}gh \) in \( \hat{\text{Mod}}(S) \) is at most \( L(3 + A) \) where \( L \) and \( A \) are the constants of a quasi-inverse of \( \Psi_{\tilde{c}} \) (since we have fixed \( \tilde{c} \) it is clear that we obtain a bound independent of \( g \)).

Now, let \( X_n \) be a random walk on \( G \) such that \( \text{gr}(\mu) \) is a non-elementary subgroup of \( G \). In view of Lemma 4.2.3, to prove Theorem 4.2.1 it would be enough to prove that if \( R \) is a subset of \( G \) such that every element of \( R \) is conjugated to an element of relative length at most \( K \) with \( K \) fixed then

\[
\lim_{n \to \infty} P_{X_n}(R) = 0.
\]

In the remainder of the chapter we will actually prove that adding some extra hypothesis on the support of the random walk the probability \( P_{X_n}(R) \) does tend to zero as \( n \) goes to infinity. Then we will show how to deduce Theorem 4.2.1 from this fact.

Remark 4.2.4. In [Mah11, Theorem 5.4] it is stated that the asymptotic probability for a random walk to stay in such a region \( R \) is actually zero without any additional hypothesis. Still, in order to avoid some technicalities we will not prove such a general theorem.

4.2.2 Relative convergence at infinity

Recall that if \( \mu \) is a probability measure on a group \( \Gamma \) and \( \Gamma \) act on a probability space \((X, \nu)\) then \( \nu \) is \( \mu \)-stationary if

\[
\nu(A) = \sum_{g \in \Gamma} \mu(g)\nu(A) = \sum_{g \in \Gamma} \mu(g)\nu(g^{-1}A)
\]

for every measurable subset \( A \subset X \).

In analogy with the examples of Subsection 4.1.3, it turns out that the following holds.

**Theorem 4.2.5.** Let \( G \) be the mapping class group of a non-sporadic surface and \( \mu \) a probability measure on \( G \) such that \( \text{gr}(\mu) \) is non-elementary and consider the random walk on \( G \) with transition probability \( \mu \). Then for almost every sample path \( g = (g_n)_{n \in \mathbb{N}} \) the sequence \( g_n \) converges in the relative metric to the Gromov boundary \( \partial_{\infty}\hat{G} \).

Moreover, the hitting distribution on \( \partial_{\infty}\hat{G} \)

\[
\nu(A) = \mathbb{P}\left\{ g \mid \lim_{n \to \infty} g_n \in A \right\}
\]
is given by a unique \( \mu \)-stationary non-atomic probability measure \( \nu \), called the harmonic measure (or hitting measure).

In [Mah11, Theorem 5.1] it is shown how to deduce Theorem 4.2.5 from an analogue result for random walks on Teichmüller spaces (see [KM96]) and a characterisation of the Gromov boundary \( \partial_\infty \hat{G} \) in terms of minimal foliations (see [Kla99]).

We say that a set \( A \) is transient for a random walk if almost every sample path intersect \( A \) only finitely many times. Then Theorem 4.2.5 implies the following.

**Corollary 4.2.6.** Given a subset \( A \) of the mapping class group \( G \), let \( \partial_\infty \hat{A} \subset \partial_\infty \hat{G} \) be its boundary at infinity in the relative metric. If \( \nu(\partial_\infty \hat{A}) = 0 \) then \( A \) is transient.

**Proof.** If a sample path is recurrent in \( A \) and relatively converges to a point at infinity, then this point must clearly be in \( \partial_\infty \hat{A} \). Since by Theorem 4.2.5 almost every sample path converges at infinity, then

\[
P\{g \in \Omega \mid g \text{ is recurrent in } A\} \leq \nu(\partial_\infty \hat{A})
\]

because \( \nu \) is equal to the hitting measure.

Notice that the converse of Corollary 4.2.6 is generally not true because there could be many sample paths converging to \( \partial_\infty \hat{A} \) without intersecting \( A \) itself. The above result will be useful when combined with the following general lemma:

**Lemma 4.2.7.** If a set \( A \) is transient then \( P_{X_n}(A) \) tends to zero as \( n \) tends to the infinity.

**Proof.** By contradiction, it is enough to note that if there exist infinitely many \( m \) such that \( P(X_m^{-1}(A)) \geq \varepsilon \), then the probability of the set of recurrent sample paths is

\[
P \left( \bigcap_{n \in \mathbb{N}} \bigcup_{m > n} X_m^{-1}(A) \right) = \lim_{n \to \infty} P \left( \bigcup_{m > n} X_m^{-1}(A) \right) \geq \varepsilon.
\]

Now, if the set of non-pseudo-Anosov elements of \( G \) was bounded in the relative metric, Theorem 4.2.1 would be proved. Still, we only know that reducible and periodic elements are bounded up to conjugacy (actually, one can prove that the whole \( \partial_\infty \hat{G} \) is contained in the closure of the set of non-pseudo-Anosov homeomorphisms). The conclusive idea will be to prove that the random walk tends to stay in regions of \( G \) where reducible and periodic homeomorphisms are more and more sparse.
Given a set $R \subset G$, we define \textit{k-crowded part} of $R$ as the subset $R^{(k)} \subseteq R$ given by

$$R^{(k)} := \{ x \in R \mid \exists y \in R \setminus \{x\}, \ d_w(x, y) \leq k \}.$$ We claim that when $R$ is the set of non-pseudo-Anosov homeomorphisms $P_{X_n}(R^{(k)})$ usually tends to zero as $n$ goes to infinity. This claim seems a very reasonable step toward the proof of the Main Theorem. Indeed, in the next subsection we will show that the theorem follows from the claim.

Notice that

$$R^{(k)} = \bigcup_{g \in B(k)} R \cap Rg$$

where $B_k$ is the ball of radius $k$ centred at the origin $e$ of $G$ with respect to the word metric. Since $B(k)$ is finite, to prove that we tend to wander outside $R^{(k)}$ it is enough to prove that sample paths tend to walk away from the sets $R \cap Rg$ for every $g \in G$.

Again, if we knew that the sets $R \cap Rg$ are bounded subsets of $\hat{G}$ then the claim would promptly follow from Theorem 4.2.5. Still, this is not the case because in general the sets $R \cap gR$ are unbounded and hence have non-trivial boundary in $\partial_\infty \hat{G}$.

In what follows, much effort will be put into proving that (with some extra hypothesis) the boundaries at infinity $\partial_\infty (R \cap Rg)$ have zero hitting measure. Once we will know that, the above claim will easily follow from what we have done so far.

### 4.2.3 Random walks and sparse subsets

In this subsection we will show that if a random walk tends to stay outside the \textit{k-crowded part} of a set $A \subset G$ for every $k$, then it tends to stay outside $A$ itself.

**Lemma 4.2.8.** If $gr(\mu)$ is a non-elementary subgroup of $G$, then the supremum

$$s_n := \sup_{g \in G} (\mu_n(g))$$

tends to zero as $n$ goes to infinity.

**Proof.** By contradiction, suppose that for infinitely many $m \in \mathbb{N}$ there exists $h_m \in G$ such that $\mu_m(h_m) \geq \varepsilon$ for some $\varepsilon > 0$. We can find a set with positive measure consisting of sample paths recurrent on $\{h_m\}$ as we did in Lemma 4.2.7. Indeed, let $M = \{ m \mid \mu_m(h_m) \geq \varepsilon \}$ and consider the sets of sample paths which take the value $h_m$ at the $m$-th step

$$A_m = \{ g \in \Omega \mid g_m = h_m \ \forall m \in M \},$$

\footnote{Another technical hypothesis is needed.}
then the limsup

\[ A := \bigcap_{n \in \mathbb{N}} \bigcup_{m > n} A_m \]

consists of paths which take the values \( h_m \) infinitely often and we have

\[ P(A) = \lim_{n \to \infty} P\left( \bigcup_{m > n} A_m \right) \geq \varepsilon. \]

By Theorem 4.2.5 almost every sample path \( g \) converges to a limit point \( g_\infty \in \partial_\infty \hat{G} \) in the relative metric. Choose an element \( \tilde{g} \in A \) which relatively converges at infinity and let \( M' := \{ m' \mid g_{m'} = h_{m'} \} \). By definition \( M' \) is an infinite set and, since \( \tilde{g} \) relatively converges to a point \( \tilde{g}_\infty \), so does the sequence \( (h_{m'})_{m' \in M'} \).

By the same argument as above the set

\[ B = \{ g \mid g_{m'} = h_{m'} \text{ for infinitely many } m' \in M' \} \]

has probability \( P(B) \geq \varepsilon \) and again we know that almost every sample path \( g \in B \) relatively converges at infinity. Still, this time we also know that all these paths must converge to the same limit point

\[ \tilde{g}_\infty = \lim_{m' \to \infty} h_{m'}. \]

By Theorem 4.2.5, the harmonic measure \( \nu \) on \( \partial_\infty \hat{G} \) is non-atomic and coincides with the hitting probability. But we have \( \nu(\tilde{g}_\infty) \geq P(B) \geq \varepsilon \) whence a contradiction.

We can now prove the main probabilistic tool of this subsection.

**Proposition 4.2.9.** Let \( \mu \) be a probability measure on \( G \) generating a non-elementary subgroup \( \text{gr}(\mu) < G \). If a set \( A \subset G \) is such that for every \( k \) the measure of the \( k \)-crowded part \( \mu_n(A^{(k)}) \) tends to zero as \( n \) goes to infinity, then

\[ \lim_{n \to \infty} \mu_n(A) = 0 \]

**Proof.** For every \( k \in \mathbb{N} \) we have

\[ \mu_n(A) = \mu_n(A^{(k)}) + \mu_n(A \setminus A^{(k)}). \quad (4.1) \]

While by hypothesis we know that

\[ \lim_{n \to \infty} \mu_n(A^{(k)}) = 0, \]

we still have to work to bound \( \mu_n(A \setminus A^{(k)}) \). By definition, \( X := A \setminus A^{(k)} \) is a \( k \)-sparse set and so are all its translates \( gX \). In particular, this implies...
that for every \( g \in G \) the set \( gX \) intersects the ball of radius \( k/2 \) centred at the origin at most one time. Thus for every \( m \in \mathbb{N} \) we have
\[
\mu_m(gX) = \mu_m(gX \cap B_{k/2}) + \mu_m(gX \setminus B_{k/2}) \\
\leq \max_{h \in B_{k/2}} (\mu_m(h)) + \mu_m(G \setminus B_{k/2})
\]

Let \( 0 < m < n \), then \( \mu_n \) is the convolution \( \mu_m \ast \mu_{n-m} \). Hence we have
\[
\mu_n(X) = \sum_{g \in G} \mu_{n-m}(g) \mu_m(g^{-1}X)
\]
and, being our estimate for \( \mu_m(gX) \) independent of \( g \), we obtain
\[
\mu_n(X) \leq \max_{h \in B_{k/2}} (\mu_m(h)) + \mu_m(G \setminus B_{k/2}).
\]

Recall that by Lemma 4.2.8
\[
s_m := \sup_{g \in G} (\mu_m(g)) \xrightarrow{n \to \infty} 0.
\]

By (4.1) we have that for every \( k \) and for every \( 0 < m < n \)
\[
\mu_n(A) \leq \mu_n(A^{(k)}) + \max_{h \in B_{k/2}} (\mu_m(h)) + \mu_m(G \setminus B_{k/2}) \\
\leq \mu_n(A^{(k)}) + s_m + \mu_m(G \setminus B_{k/2}).
\]
Hence for every \( k, m \in \mathbb{N} \)
\[
\lim_{n \to \infty} \mu_n(A) \leq s_m + \mu_m(G \setminus B_{k/2}).
\]

Notice that for every \( m \) and for every \( \varepsilon > 0 \) there exists \( k_m \) large enough so that \( \mu_m(G \setminus B_{k_m/2}) \leq 1/m \). The thesis follows since
\[
\lim_{n \to \infty} \mu_n(A) \leq \lim_{m \to \infty} \left[ s_m + \mu_m(G \setminus B_{k_m/2}) \right] = 0.
\]

### 4.2.4 Conjugacy bounds and centralizers

In this subsection we will find a more convenient way to handle the sets \( \partial_\infty(R \cap Rg) \) defined in Section 4.2.2. Specifically, we will relate them to boundaries of centralizers in \( G \). On our way to relate sparse sets and centralizers of the mapping class group, we give the following:

**Definition 4.2.10.** A weakly hyperbolic group \( \Gamma \) has *conjugacy bounds* if for every pair of conjugate elements \( a, b \in \Gamma \) there exists a conjugating word \( w \) (i.e. \( a = wbw^{-1} \)), such that
\[
\|\hat{w}\| \leq \beta (\|\hat{a}\| + \|\hat{b}\|)
\]
where \( \|\cdot\| \) denotes the relative length and \( \beta \) is a constant of \( \Gamma \).
The following is proved in [Mah11, Theorem 3.1].

**Theorem 4.2.11.** The mapping class group of a non-sporadic surface $S$ has relative conjugacy bounds with constant $\beta$ depending only $S$.

The reason why we need the above result is easily explained. Let $\Gamma$ be a weakly relatively hyperbolic group and recall that in Subsection 1.3.3 we defined the $L$-horoball neighbourhood of a subset $\hat{A} \subset \Gamma$ as

$$\hat{\Theta}_L(\hat{A}) := \bigcup_{x \in A} B_{\|\hat{x}\|+L}(x),$$

where $\|\hat{x}\|$ denotes the relative distance of $x$ from the identity $e$ and $B_{\|\hat{x}\|+L}(x)$ is the ball centred at $x$ of radius $\|\hat{x}\|+L$ in the relative metric. Then the following holds:

**Theorem 4.2.12.** Let $\Gamma$ be a weakly relatively hyperbolic group that has conjugacy bounds with constant $\beta$. If $R$ is a subset of $\Gamma$ whose elements are all conjugated to elements of relative length at most $K$, then for every element $g \in \Gamma$ we have

$$R \cap Rg \subseteq \hat{\Theta}_L(C(g)),$$

where the constant $L$ depends only on the relative length $\|\hat{g}\|$, the constant $K$ and the group constants $\delta$ and $\beta$.

Recall now that by Lemma 1.3.22 the boundary at infinity of a set is equal to the boundary at infinity of any of its horoball neighbourhoods. In particular, once Theorem 4.2.12 is proved, combining it with Lemma 1.3.22 yields the following key corollary:

**Corollary 4.2.13.** In the hypotheses of Theorem 4.2.12, the boundary at infinity of $R \cap Rg$ in the relative metric is contained in $\partial_{\infty} C(g)$.

Before proving Theorem 4.2.12, we need to establish a preliminary lemma. If $\Gamma$ is a weakly relatively hyperbolic group, for every pair of elements $x, y \in \Gamma$ the bracket $[x, y]$ will denote a geodesic in the relative metric joining $x$ to $y$. Since we will be dealing only with geodesics with respect to the relative metric, we will no more say it explicitly. Moreover, when we speak about the length of a geodesic we clearly mean it with respect to the relative length. Still, for the sake of coherence we will stress that lengths of elements of $\Gamma$ are considered with respect to the relative distance saying it explicitly and continuing to use the notation $\|\hat{x}\|$. As usual, a geodesic $[x, y]$ needs not be unique. When dealing with metric properties of $\delta$-hyperbolic spaces, the actual choice for such a geodesic is usually uninfluential because different geodesics with the same endpoints have Hausdorff distance at most $\delta$. Still, since we are dealing with a much more structured object, it will be convenient to put some extra care on the choice
of geodesics. In what follows it is hence understood that once a geodesic \([x, y]\) is fixed, the geodesic \([y, x]\) is the same as \([x, y]\) but going backward and the geodesic \([zx, zy]\) is equal to the geodesic \(z[x, y]\) i.e. the geodesic \([x, y]\) translated by \(z\).

Having said that, let \(\Gamma\) be a weakly relatively hyperbolic group with conjugacy bounds and let \(r\) and \(s\) be two conjugated elements of \(\Gamma\). If \(w\) is a conjugating word \(s = wrw^{-1}\), then we have two natural paths in \(\Gamma\) from the identity \(e\) to \(r\): the geodesic \([e, r]\) and a piecewise geodesic \([e, w] \cup [w, ws] \cup [ws, r]\) (equivalently \([e, w] \cup [w, e] \cup ws[e, w^{-1}]\) or \([e, w] \cup [w, e] \cup r[w, e]\)).

We begin the proof of Theorem 4.2.12 with the following.

**Lemma 4.2.14.** Given a group \(\Gamma\) as above and two conjugated elements \(s = wrw^{-1}\), if \(w\) is a conjugating element of minimal relative length, then the path \([e, w] \cup [w, ws] \cup [ws, r]\) is contained in an \(L_1\)-neighbourhood of \([e, r]\) where the constant \(L_1\) depends only on the relative length \(\|s\|\) and the group constants \(\delta\) and \(\beta\).

**Proof.** The idea goes as follows: we have to prove that the word \(w\) does not go exaggeratedly far from the geodesic \([e, r]\). By hypothesis, \(\Gamma\) has conjugacy bounds, thus we already know that the relative length of \(w\) is bounded by those of \(r\) and \(s\). We have to get rid of the dependence on \(r\). Since the geodesic \([ws, r]\) is the inverse of the geodesic \([r, ws] = r[e, w]\), for every point \(x\) in \([e, w]\) its translate \(rxx\) lies in \([w, ws]\). Moreover, \(x^{-1}rx\) is conjugated to \(s\) and \(x^{-1}w\) is a conjugating word of shortest relative length. In particular, the length of \(x^{-1}w\) is bounded by the length of \(s\) and that of \(x^{-1}rx\), thus our objective is to find an \(x\) close both to \(rx\) and \([e, r]\). Now we proceed with a detailed discussion.

Fix a constant \(C\). If the relative distance between the geodesics \([e, r]\) and \([w, ws]\) is smaller than \(C\) then both \(d(w, [e, r])\) and \(d(ws, [e, r])\) are smaller than \(\|s\| + C\) and hence the whole path \([e, w] \cup [w, ws] \cup [ws, r]\) is contained in a neighbourhood of \([e, r]\) of radius \(\|s\| + C + 2\delta\). For, it is easy to see that whenever the extremities of a geodesic segment are at bounded distance from a geodesic \(\gamma\), the whole segment is at bounded distance from \(\gamma\) because geodesic squares are \(2\delta\)-thin.

Thus, for every fixed constant \(C\) we only need to work out the case where the distance of \([w, ws]\) from \([e, r]\) is greater then \(C\). Notice that if \(C\) is large enough, there exists an element \(x\) in the geodesic \([e, w]\) such that the distance \(d(x, [e, r])\) is exactly \(2\delta + 1\) (actually, since \(\Gamma\) is a discrete space, here we are clearly assuming \(\delta\) to be an integer number). Since geodesic squares are \(2\delta\)-slim, we deduce that \(x\) must be close to \([w, ws]\) or \([ws, r]\). Moreover, if we take \(C\) to be larger than \(4\delta + 2\), then \(x\) cannot be close to \([w, ws]\) either. Thus we can assume it is \(2\delta\)-close to \([ws, r]\).

Let \(y\) be a point of \([ws, r]\) nearest to \(x\). We want to show that \(y\) is close to \(rx\) (Figure 4.1). Since they both lie on \([w, ws]\), it is sufficient to show that
the distance $d(y, ws)$ is close to $d(rx, ws)$. Let $a := d(x, w) = d(rx, w)$.
By the triangle inequality we have that

$$a \leq d(x, y) + d(y, ws) + d(ws, w) \leq 2\delta + ||\hat{s}|| + d(y, ws)$$

and

$$d(y, ws) \leq d(y, x) + d(x, w) + d(w, ws) \leq 2\delta + a + ||\hat{s}||$$,

whence

$$|a - d(y, ws)| \leq ||\hat{s}|| + 2\delta.$$ 

We conclude that $d(x, rx) \leq ||\hat{s}|| + 4\delta$.

To complete the proof it is now sufficient to note that the relative length of $x^{-1}rx \in \Gamma$ has relative length at most $||\hat{s}|| + 4\delta$. Since $x^{-1}rx = (x^{-1}w)s(x^{-1}w)^{-1}$, we have that $x^{-1}w$ is an element conjugating $s$ to $x^{-1}rx$. Moreover, $x^{-1}w$ must have minimal relative length among the elements that conjugate $s$ to $x^{-1}rx$ otherwise we could shorten $w$ substituting its final part with one of these. We conclude that $x^{-1}w$ must have relative length at most $\beta(2||\hat{s}|| + 4\delta)$ by the conjugacy bounds condition.

To complete the proof it is now sufficient to note that the relative length of $x^{-1}w$ is equal to the length of the geodesic $[e, x^{-1}w] = [x, w]$: since by definition the distance between $x$ and $[e, r]$ is $2\delta + 1$, we have that the distance between $[w, ws]$ and $[1, r]$ is at most

$$d([w, ws], [1, r]) \leq \beta(2||\hat{s}|| + 4\delta) + 2\delta + 1.$$ 

Hence the path $[e, w] \cup [w, ws] \cup [ws, r]$ is contained in a neighbourhood of $[e, r]$ of radius $\beta(2||\hat{s}|| + 4\delta) + ||\hat{s}|| + 4\delta + 1$. Recalling that we assumed the geodesic $[w, ws]$ to be sufficiently far, we conclude that the Lemma holds with the constant

$$L_1 = \max\left\{ \beta(2||\hat{s}|| + 4\delta) + ||\hat{s}|| + 4\delta + 1 , \ ||\hat{s}|| + C + 2\delta \right\}$$

where $C = 4\delta + 2$.

We can now proceed with the proof of the theorem.
4.2. RANDOM WALKS ON THE MAPPING CLASS GROUP

4.2. RANDOM WALKS ON THE MAPPING CLASS GROUP

Figure 4.2: The points $w$ and $rw$ cannot be too far from the midpoint of $[e, r]$.

Proof of Theorem 4.2.12. For notational convenience, we will actually prove that $R \cap Rg^{-1}$ is contained in a horoball-neighbourhood of $C(g)$ (that is the same of $C(g^{-1})$). Let $r$ be an element of $R \cap Rg^{-1}$, then also $r' = rg$ belongs to $R$. By hypothesis, $r$ and $r'$ are conjugated to relatively short elements $s$ and $s'$. Let $w$ and $w'$ be conjugating words of minimal relative length $r = wsw^{-1}$, $r' = w's'(s')^{-1}$. The first step of the proof is to show that $w$ and $w'$ are at a bounded distance apart.

We notice that the relative length of $w$ is roughly a half of that of $r$. Indeed, consider two piecewise geodesic paths $[e, r]$ and $[e, w] \cup [w, ws] \cup [ws, wsw^{-1}]$ as in Lemma 4.2.14 and let $p$ and $q$ be nearest point projections of $w$ and $ws$ on $[1, r]$ (Figure 4.2).

Recall that by Lemma 4.2.14 the distances $d(w, [1, r])$ and $d(ws, [1, r])$ are both bounded by a constant $L_1$ depending only on the group and $\|\hat{s}\|$. Since the latter is by hypothesis smaller than or equal to a fixed constant $K$, we can assume that also $L_1$ is fixed. By the triangle inequality we get

$$\|\hat{w}\| \leq d(e, p) + L_1$$

$$\|\hat{w}\| = \|\hat{w}^{-1}\| \leq d(q, r) + L_1$$

$$d(p, q) \leq 2L_1 + \|\hat{s}\| \leq 2L_1 + K.$$

Since $\|\hat{r}\| = d(e, p) + d(q, r) \pm d(p, q)$, we obtain $2\|\hat{w}\| \leq d(e, p) + d(q, r) + 2L_1 \leq \|r\| + K + 4L_1$. Moreover, we have

$$\|\hat{r}\| \leq d(e, w) + d(w, ws) + d(ws, r) = 2\|\hat{w}\| + \|\hat{s}\|.$$

Thus

$$\frac{1}{2} (\|\hat{r}\| - K) \leq \|\hat{w}\| \leq \frac{1}{2} (\|\hat{r}\| + K + 4L_1). \quad (4.2)$$

Notice that an analogue of Equation (4.2) holds also for $w'$ because we only used that $\|\hat{s}\|$ is bounded by $K$.

We will now show that we can assume $w$ and $w'$ to be relatively close. Let $p$ and $p'$ be nearest point projections of $w$ and $w'$ to $[e, r]$ and $[e, r']$ respectively. (Recall that a nearest point projection is a point that realizes the distance. Such a point could be non-unique.)
By Equation (4.2) we deduce that $||\hat{p}|| \leq (||\hat{r}|| + K + 4L_1)/2 + L_1$. Thus, if $||\hat{r}||$ is big enough we obtain by $\delta$-hyperbolicity that $p$ is $\delta$-close to the geodesic $[e, r']$. Specifically, it is enough to assume

$$\| \hat{r} \| \geq K + 6L_1 + 2\| \hat{g} \| + 2\delta.$$  

Such an assumption is not restrictive since $g$ is fixed and to prove the Theorem it is clearly sufficient proving it only for elements $r$ with relative length greater than a fixed constant.

Let $p''$ be a nearest point projection of $p$ to $[e, r']$ (Figure 4.3). Then $p'$ and $p''$ lie in the same geodesic and their distance is $||\hat{p}'|| - ||\hat{p}''||$. Such a distance is bounded because both points are roughly at half length of the geodesic. Indeed, applying Equation (4.2) twice we obtain

$$\frac{1}{2} (||\hat{r}'\| - K) - L_1 \leq ||\hat{p}'|| \leq \frac{1}{2} (||\hat{r}'\| + K + 4L_1) + L_1$$

and

$$\frac{1}{2} (||\hat{r}|| - K) - L_1 - \delta \leq ||\hat{p}''|| \leq \frac{1}{2} (||\hat{r}|| + K + 4L_1) + L_1 + \delta.$$  

Using that $||\hat{r}|| - ||\hat{p}'|| \leq ||\hat{g}||$ we get

$$||\hat{p}'|| - ||\hat{p}''|| \leq \frac{1}{2} ||\hat{g}|| + K + 6L_1 + \delta.$$  

By the triangle inequality we conclude that

$$d(w, w') \leq \frac{1}{2} ||\hat{g}|| + K + 8L_1 + 2\delta.$$  

Let $L_2 = ||\hat{g}||/4 + K/2 + 4L_1 + \delta$, so that $d(w, w') \leq 2L_2$. Then there exists an element $x$ in $[w, w']$ such that both $d(x, w)$ and $d(x, w')$ are at most
4.2. RANDOM WALKS ON THE MAPPING CLASS GROUP

Figure 4.4: The element $x$ conjugates $g = r^{-1}r'$ to an element with relatively bounded distance.

$L_2$. Our goal is to show that $x$ is close to the centralizer $C(g)$. Let $y = x^{-1}w$ and $y' = x^{-1}w'$. These two elements have relative length $d(x, w)$ and $d(x, w')$ and are hence bounded by $L_2$. Let $z = wsy^{-1}$ and $z' = w's'(y')^{-1}$ (see Figure 4.4).

Then also the distance between $z$ and $z'$ is bounded. Indeed, by the triangle inequality we have

$$d(z, z') \leq 4L_2 + 2K.$$ 

Notice that $x$ sends the geodesic $r[e, g]$ to a geodesic $[z, z']$. That is, $g$ is equal to $x(z^{-1}z')x^{-1}$. Since $\Gamma$ has conjugacy bounds, there exists an element $v \in \Gamma$ such that $g = v(z^{-1}z')v^{-1}$ and $v$ has relative length at most $L_3 = \beta(\|\hat{g}\| + 4L_2 + 2K)$.

This implies that $x^{-1}gx = v^{-1}gv$ and hence $gxv^{-1} = xv^{-1}g$. That is, $xv^{-1}$ is in the centralizer of $g$ and hence $x$ is at most at distance $L_3$ from $C(g)$.

Now we are almost done. We have $d(w, C(g)) \leq L_3 + \|\hat{w}\| = L_3 + L_2$. Thus if $h \in C(g)$ is the nearest element to $w$, it has relative length at least $\|\hat{w}\| - (L_3 + L_2)$. We conclude that

$$d(r, h) \leq d(w, h) + d(w, ws) + d(ws, r) \leq L_3 + L_2 + K + \|\hat{w}\| \leq \|\hat{h}\| + 2L_3 + 2L_2 + K$$

and hence $r$ is contained in the $L$-horoball neighbourhood of $C(g)$ with a constant $L = 2L_3 + 2L_2 + K$ depending only on $K$, $\|\hat{g}\|$ and the group constants $\beta$ and $\delta$.

4.2.5 A partial result

We can now collect the results obtained so far to prove a refinement of the main theorem using some extra hypotheses. Recall that we noticed that
Theorem 4.2.1 would essentially follow if we could prove that
\[ \nu(R \cap Rg) = 0 \]
for every \( g \in G \) whenever \( R \) is a set of homeomorphisms conjugate to elements of bounded relative length. Recall also that by Lemma 4.1.2 a subset \( A \subset \partial_\infty \hat{G} \) has hitting measure zero provided that it has infinitely many images under the action of \( \text{gr} \mu \) and it is such that for each \( g \in \text{gr}(\mu) \) either \( A = gA \) or \( \nu(A \cap gA) = 0 \). Thus, what we would like to find now is a condition on \( \text{gr}(\mu) \) allowing us to apply Lemma 4.1.2 to the sets \( \nu(R \cap Rg) = 0 \).

As before, let \( G \) be the mapping class group of a non-sporadic surface of finite type \( S \). We define now a technical property for subgroups of \( G \). We say that a subgroup \( H \) of \( G \) satisfies property (*) if

\[ (*) \text{ for every finite subgroup } F < H, \ F \neq \{e\}, \text{ the relative boundary of the centralizer of } F \text{ in } G \ (\text{i.e. } \partial_\infty C(F)) \text{ has infinitely many images under the action of } H. \]

Then we have the following:

**Proposition 4.2.15.** Consider a random walk \( X_n \) on \( G \) such that \( \text{gr}(\mu) \) is a non-elementary subgroup of \( G \) and has property (*). If \( R \subset G \) is a subset such that every element of \( R \) is conjugated to a homeomorphism of relative length at most \( K \) with \( K \) fixed, then

\[ \lim_{n \to \infty} P_{X_n}(R) = 0. \]

**Proof.** By Proposition 4.2.9, it is enough to prove that for every \( k \in \mathbb{N} \) the random walk tends to stay outside the \( k \)-crowded part of \( R \). Namely, we have to show that

\[ P \left( X_n^{-1}(R^{(k)}) \right) = \mu_n(R^{(k)}) \xrightarrow{n \to \infty} 0. \]

Recall that

\[ R^{(k)} = \bigcup_{g \in B_k^G} R \cap Rg \]

where \( B_k^G \) is the ball of radius \( k \) in \( G \) with respect to the word metric. Since the random walk is always supported on \( \text{gr}(\mu) \), up to removing a negligible subset we can assume that \( R \) is also contained in \( \text{gr}(\mu) \) and hence we have

\[ R^{(k)} = \bigcup_{g \in B_k^G \cap \text{gr}(\mu)} R \cap Rg. \]

Being this union finite, by Lemma 4.2.7 we only need to prove that for every \( g \in \text{gr}(\mu) \) the set \( R \cap Rg \) is transient.

By Corollary 4.2.6, it is enough to prove that \( \partial_\infty (R \cap Rg) \) is negligible with respect to the harmonic measure \( \nu \) on \( \partial_\infty G \). Moreover, since the elements of \( R \) are conjugated to relatively short elements and by Theorem 4.2.11 the
mapping class group has conjugacy bounds, we are in the hypothesis of Theorem 4.2.12 and hence by Corollary 4.2.13 we have that \( \partial_\infty (R \cap Rg) \subseteq \partial_\infty C(g) \) where \( C(g) \) is the centralizer of \( g \) in \( G \). Therefore, all it remains to do is to show that \( \nu(\partial_\infty C(g)) = 0 \) for every element \( g \) in \( \text{gr}(\mu) \).

By Proposition 2.4.32, if \( g \in \text{gr}(\mu) \) has infinite order then the boundary at infinity of its centralizer consists of at most two points. Hence it is negligible since the hitting measure \( \nu \) is non-atomic.

It only remains to deal with the case of \( g \) periodic. By contradiction, if \( \nu(\partial_\infty C(F)) > 0 \), let \( F < \text{gr}(\mu) \) be a maximal finite subgroup with \( g \in F \) and such that \( \nu(\partial_\infty C(F)) > 0 \) (such a group exists because by Theorem 2.4.16 there is a bound on the cardinality of finite subgroups). We claim that for every \( h \in \text{gr}(\mu) \) either \( h(\partial_\infty C(F)) = \partial_\infty C(F) \) or \( h(\partial_\infty C(F)) \cap \partial_\infty C(F) \) has measure zero.

Notice that \( h(\partial_\infty C(F)) = \partial_\infty C(hFh^{-1}) \) and by Proposition 2.4.30 we have

\[
\partial_\infty C(hFh^{-1}) \cap \partial_\infty C(F) = \partial_\infty C(F')
\]

where \( F' \) is the subgroup generated by \( F \cup hFh^{-1} \).

If \( hFh^{-1} = F \) then we trivially have \( h(\partial_\infty C(F)) = \partial_\infty C(F) \); otherwise \( F < F' \) is a proper subset. If \( F' \) is finite then \( \nu(\partial_\infty C(F')) = 0 \) by maximality of \( F \). If \( F' \) is infinite then by Theorem 2.3.14 there exists an element of \( F' \) of infinite order and hence \( \partial_\infty C(F') \) consists of at most two points. We conclude that \( \partial_\infty C(F') \) has measure zero since \( \nu \) is non-atomic. In either case we have

\[
\nu[\partial_\infty C(hFh^{-1}) \cap \partial_\infty C(F)] = 0,
\]

proving our claim.

Now, since (*) holds, \( \partial_\infty C(F) \) satisfies the hypothesis of Lemma 4.1.2. Hence we deduce that \( \nu(\partial_\infty C(F)) = 0 \), a contradiction.

**Corollary 4.2.16.** Let \( G \) be the mapping class group of a non-sporadic surface and \( R \subset G \) the set of non-pseudo-Anosov homeomorphisms. If \( X_n \) is a random walk on \( G \) and \( \text{gr}(\mu) \) is a non-elementary subgroup which has property (*), then

\[
\lim_{n \to \infty} P_{X_n}(R) = 0.
\]

**4.2.6 Proof of the Main Theorem**

We can finally complete the proof of Theorem 4.2.1. In order to avoid any confusion, in this subsection the mapping class group of a surface \( S \) will be denoted by \( G_S \). Let \( H := \text{gr}(\mu) \), if \( H \) satisfies property (*) the theorem follows by Corollary 4.2.15. Otherwise, let \( F_1 \) be a finite subgroup of \( H \) such that \( \partial_\infty C(F_1) \) has only finitely many images under the action of \( H \). By
Corollary 2.4.38 we deduce that there exists a finite group $F_1'$ containing $F_1$ and such that $H$ is contained in the normalizer $N(F_1')$.

By the Nielsen Realization Theorem there exists a hyperbolic metric on $S$ where $F_1'$ can be realized as a subgroup of $\text{Isom}^+(S)$. Let $\mathcal{O}_1 = S/F_1'$ be the hyperbolic orbifold obtained quotienting $S$ by this realization of $F_1'$. Then by Theorem 2.4.18 we have a map with finite kernel

$$\Phi_1: N(F_1') \to G_{\mathcal{O}_1}$$

where $G_{\mathcal{O}_1}$ is the orbifold-mapping class group of $\mathcal{O}_1$.

Recall that $G_{\mathcal{O}_1}$ is naturally embedded in $G_{\mathcal{O}_1^*}$ where $\mathcal{O}_1^*$ is the surface obtained by $\mathcal{O}_1$ replacing cone points with punctures. We can consider $\Phi_1(H)$ as a subgroup of $G_{\mathcal{O}_1^*}$ and we note that it is still non-elementary. For, it is clear that $\Phi_1(H)$ contains pseudo-Anosov elements (see Proposition 2.4.23) and $\Phi_1(H)$ cannot have a finite index cyclic subgroup because $\Phi_1$ has finite kernel (see Theorem 2.3.13).

If $\Phi_1(H)$ does not satisfy (*), let $E < \Phi_1(H)$ be a finite subgroup whose centralizer in $G_{\mathcal{O}_1^*}$ has only finitely many images at infinity under the action of $\Phi_1(H)$. Again, by Corollary 2.4.38 there exists $E' \supseteq E$ with $\Phi_1(H) < N(E')$. Moreover, we can assume that also $E'$ is contained in $H$ (see Remark 2.4.39).

Let $F_2 = \Phi_1^{-1}(E')$. Since $\ker \Phi_1 = F_1'$, then $F_2$ is a finite group of $G_S$ containing $F_1'$ as a proper subgroup. Moreover, we have that the normalizer of $F_2$ contains the pre-image of the normalizer of $E'$

$$\Phi_1^{-1}(N(E')) \subseteq N(F_2).$$

(Actually, since $E'$ is contained in the image of $\Phi_1$, we have that $\Phi_1^{-1}(N(E'))$ is the normalizer of $F_2$ in $N(F_1')$). In particular, we have that $H$ is contained in $N(F_2)$ and hence we can quotient by a realization of $F_2$ obtaining another map

$$\Phi_2: N(F_2) \to G_{\mathcal{O}_2} \subset G_{\mathcal{O}_2^*}$$

and we can restrict $\Phi_2$ to $H$. Again, if $\Phi_2(H)$ does not satisfy (*) we can find a finite $F_3$ containing $F_2$ as a proper subgroup and such that $H < N(F_3)$, and so on. This process eventually terminates because we have a strictly increasing sequence of finite subgroups $F_1 \leq F_2 \leq \cdots$ and the cardinality of finite subgroups of the mapping class group is bounded.

We have shown that there exists a finite group $F < G_S$ with $\text{gr}(\mu) < N(F)$ inducing an orbifold cover $P: S \to \mathcal{O}$ and a map $\Phi: \text{gr}(\mu) \to G_{\mathcal{O}} \subseteq G_{\mathcal{O}^*}$ such that $\Phi(\text{gr}(\mu))$ is non-elementary and satisfies (*) in $G_{\mathcal{O}^*}$. Notice that if $X_n$ is the random walk on $G_S$ with transition probability $\mu$, then $\Phi \circ X_n$ is the random walk on $G_{\mathcal{O}^*}$ with transition probability $\Phi \circ \mu$ (the random variable $\Phi \circ X_n$ is well-defined because the image of $X_n$ is contained in $\text{gr}(\mu)$). Moreover, for every subset $R \subseteq G_{\mathcal{O}^*}$ we have

$$\mathbf{P}_{X_n}\{\Phi^{-1}(R)\} = \mathbf{P}_{\Phi \circ X_n}\{R\}.$$
4.2. RANDOM WALKS ON THE MAPPING CLASS GROUP

Let $R$ be the set of non-pseudo-Anosov homeomorphisms of $O^*$

$$R := \{ g \in G_{O^*} \mid g \text{ reducible or periodic} \}.$$ 

By Proposition 2.4.23, for every element $g \in \Phi(G_S) \subseteq G_{O^*}$ and for every $g' \in \Phi^{-1}(g)$, $g$ is a pseudo-Anosov element of $G_{O^*}$ if and only if $g'$ is a pseudo-Anosov element of $G_S$.

Now, let $R' \subseteq G_S$ be the set of non-pseudo-Anosov homeomorphisms of $S$. Since $P_{X_n}(R') = P_{X_n}(R' \cap \text{gr}(\mu))$ and the latter set is equal to $\Phi^{-1}(R)$, we have

$$P_{X_n}(R') = P_{X_n}(\Phi^{-1}(R)) = P_{\phi^n, X_n}(R).$$

Hence to complete the proof of Theorem 4.2.1 we only need to show that $P_{\phi^n, X_n}(R)$ tends to zero as $n$ goes to infinity.

Notice that $O^*$ must have negative Euler characteristic and it cannot be a sphere with three punctures because we know that the mapping class group of $O^*$ contains $\Phi(H)$ as a non-elementary subgroup. Now, if $O^*$ is a non-sporadic surface, the theorem follows by Corollary 4.2.16. If $O^*$ is sporadic then it must be the once punctured torus or the sphere with four punctures. In this case $G_{O^*}$ is commensurable with $SL(2, \mathbb{Z})$ and the thesis follows from known facts about random walks on $SL(2, \mathbb{Z})$ (see [Fur71]).
Titoli di coda

Giunti alla fine di questo lungo percorso di studi è d’obbligo ringraziare tutti coloro che mi hanno aiutato e sostenuto in questi anni. Primi fra tutti i miei genitori, che sono sempre stati un punto fermo e una certezza a cui rivolgermi nei momenti di difficoltà. Se sono diventato quello che sono è di sicuro grazie a loro.

Da un punto di vista accademico invece, non posso che ringraziare sentitamente il mio relatore Roberto Frigerio che mi ha pazientemente accompagnato attraverso due tesi ed un colloquio. Non lo ringrazierò mai abbastanza per la sua disponibilità e l’attenzione con cui ha letto tutti i miei elaborati correggendo carrellate di strafalcioni matematici e non. Ci tengo inoltre a ringraziare Alessandro Sisto e Maria Beatrice Pozzetti per la gran quantità di idee e consigli che mi hanno dato negli ultimi tempi. Un ringraziamento va anche al Prof. Ambrosio e al Dott. Martelli per aver acconsentito a scrivermi delle lettere di raccomandazione che hanno evidentemente avuto un buon successo.

Per tutto il tempo passato assieme, non posso che ringraziare la cumpa dei matematici del mio anno: il Duca, Feddy, Tap, Jay, Biello, Ippo e Rigo (è inutile che questi ultimi continuino a far finta di studiare fisica, lo sappiamo che sono matematici dentro). Anche se non è del nostro anno, includerò nella lista anche Giulio, che dopotutto se l’è meritato. Ci tengo a ringraziare e salutare anche quelli che mi hanno dato idee e consigli negli ultimi tempi. Un ringraziamento va alla Valdagno Super Cool Crew, con un grazie particolare agli elementi storici: Irene, Luigi, Federico, Francesco, Mario, Franco, Andrea, Marta, Anna, Diego, Nicola e Marco. Ai tempi del liceo ne abbiamo fatte di belle…

Per finire, un ringraziamento speciale a tutti gli amici e parenti che hanno trovato il tempo e la voglia di venire fino a Pisa per assistere alla mia
discussione di laurea, ho molto apprezzato.

Poiché lo spazio è tiranno, mi devo fermare qui. Se sei una delle persone che dopo aver letto tutti i ringraziamenti ritiene di non essere stata tenuta sufficientemente in considerazione (e di sicuro ce ne sono, basti pensare ad altri amici, ex-professori, etc...), sappi che questo paragrafo è per te. Ti ringrazio per tutto l’aiuto morale e materiale che mi hai dato.

Con affetto,
Federico
Bibliography


[Kla99] E. Klarreich, *The boundary at infinity of the curve complex and the relative Teichmüller space*.


