# Lectures on Geometric Group Theory

Cornelia Druţu and Michael Kapovich

# Preface

The goal of this book is to present several central topics in geometric group theory, primarily related to the large scale geometry of infinite groups and spaces on which such groups act, and to illustrate them with fundamental theorems such as Gromov's Theorem on groups of polynomial growth, Tits' Alternative, Mostow Rigidity Theorem, Stallings' theorem on ends of groups, theorems of Tukia and Schwartz on quasi-isometric rigidity for lattices in real-hyperbolic spaces, etc. We give essentially self-contained proofs of all the above mentioned results, and we use the opportunity to describe several powerful tools/toolkits of geometric group theory, such as coarse topology, ultralimits and quasiconformal mappings. We also discuss three classes of groups central in geometric group theory: Amenable groups, (relatively) hyperbolic groups, and groups with Property (T).

The key idea in geometric group theory is to study groups by endowing them with a metric and treating them as *geometric objects*. This can be done for groups that are *finitely generated*, i.e. that can be reconstructed from a finite subset, *via* multiplication and inversion. Many groups naturally appearing in topology, geometry and algebra (e.g. fundamental groups of manifolds, groups of matrices with integer coefficients) are finitely generated. Given a finite generating set $S$ of a group $G$, on can define a metric on $G$ by constructing a connected graph, the *Cayley graph* of $G$, with $G$ serving as the set of vertices and the oriented edges labeled by elements in $S$. A Cayley graph $\mathcal{G}$, as any other connected graph, admits a natural metric invariant under automorphisms of $\mathcal{G}$: The distance between two points is the length of the shortest path in the graph joining these points (see Section 1.3.4). The restriction of this metric to the vertex set $G$ is called the *word metric* $\mathrm{dist}_S$ on the group $G$. The first obstacle to "geometrizing" groups in this fashion is the fact that a Cayley graph depends not only on the group but also on a particular choice of finite generating set. Cayley graphs associated with different generating sets are not isometric but merely *quasi-isometric*.

Another typical situation in which a group $G$ is naturally endowed with a (pseudo)metric is when $G$ acts on a metric space $X$: In this case the group $G$ maps to $X$ *via* the *orbit map* $g \mapsto gx$. The pull-back of the metric on $G$ is then a pseudo-metric on $G$. If $G$ acts on $X$ isometrically, then the resulting pseudometric on $G$ is $G$-invariant. If, furthermore, the space $X$ is proper and geodesic and the action of $G$ is *geometric* (i.e., properly discontinuous and cocompact), then the resulting (pseudo)metric is quasi-isometric to word metrics on $G$ (Theorem 5.29). For example, if the group $G$ is the fundamental group of a closed Riemannian manifold $M$, the action of $G$ on the universal cover $\widetilde{M}$ of $M$ satisfies all these properties. The second class of examples of isometric actions (whose origin lies in functional analysis and representation theory) comes from isometric actions of a group $G$ on Hilbert spaces. The square of the corresponding metric on $G$ is known in the literature as a *conditionally negative definite kernel*. In this case, the relation between the word metric and the metric induced from the Hilbert space is more loose than quasi-isometry; nevertheless, the mere existence of such a metric has many interesting implications, detailed in Chapter 17.

In the setting of geometric view of groups, the following questions become fundamental:

QUESTIONS. (A) *If $G$ and $G'$ are quasi-isometric groups, to what extent do $G$ and $G'$ share the same algebraic properties?*

(B) *If a group $G$ is quasi-isometric to a metric space $X$, what geometric properties (or structures) on $X$ translate to interesting algebraic properties of $G$?*

Addressing these questions is the primary focus of this book. Several striking results (like Gromov's Polynomial Growth Theorem) state that certain algebraic properties of a group can be reconstructed from its loose geometric features.

Closely connected to these considerations are two foundational conjectures which appeared in different contexts but both render the same sense of existence of a "demarkation line" dividing the class of infinite groups into "abelian-like" groups and "free-like" groups. The invariants used to draw the line are quite different (existence of a finitely-additive invariant measure in one case and behavior of the growth function in the other); nevertheless, the two conjectures and the classification results that grew out of these conjectures, have much in common.

The first of these conjectures was inspired by work investigating the existence of various types of group-invariant measures, that originally appeared in the context of Euclidean spaces. Namely, the *Banach-Tarski paradox* (see Chapter 15), while denying the existence of such measures on the Euclidean plane, inspired J. von Neumann to formulate two important concepts: That of *amenable groups* and that of *paradoxical decompositions and groups* [**vN28**]. In an attempt to connect amenability to the algebraic propeties of a group, von Neumann made the observation, in the same paper, that the existence of a free subgroup excludes amenability. This was later formulated explicitly as a conjecture by M. Day [**Day57**, §4]:

CONJECTURE (The von Neumann–Day problem). *Is non-amenability of a group equivalent to the existence of a free non-abelian subgroup?*

The second conjecture appeared in the context of Riemannian geometry, in connection to various attempts to relate, for a compact Riemannian manifold $M$, the geometric features of its universal cover $\widetilde{M}$ to the behavior of its fundamental group $G = \pi_1(M)$. Two of the most basic objects in Riemannian geometry are the *volume* and the *volume growth rate*. The notion of volume growth extends naturally to discrete metric spaces, such as finitely generated groups. The *growth function* of a finitely generated group $G$ (with a fixed finite generating set $S$) is the cardinality $\mathfrak{G}(n)$ of the ball of radius $n$ in the metric space $(G, \mathrm{dist}_S)$. While the function $\mathfrak{G}(n)$ depends on the choice of the finite generating set $S$, the *growth rate* of $\mathfrak{G}(n)$ is independent of $S$. In particular, one can speak of groups of linear, polynomial, exponential growth, etc. More importantly, the growth rate is preserved by quasi-isometries, which allows to establish a close connection between the Riemannian growth of a manifold $\widetilde{M}$ as above, and the growth of $G = \pi_1(M)$.

One can easily see that every abelian group has polynomial growth. It is a more difficult theorem (proven independently by Hyman Bass [**Bas72**] and Yves Guivarc'h [**Gui70, Gui73**]) that all nilpotent groups also have polynomial growth. We prove this result in Section 12.5. In this context, John Milnor formulated the following conjecture

CONJECTURE (Milnor's conjecture). *The growth of any finitely generated group is either polynomial (i.e. $\mathfrak{G}(n) \leqslant Cn^d$ for some fixed $C$ and $d$) or exponential (i.e. $\mathfrak{G}(n) \geqslant Ca^n$ for some fixed $a > 1$ and $C > 0$).*

Milnor's conjecture is *true for solvable groups:* This is the *Milnor–Wolf Theorem*, which states that *solvable groups of polynomial growth are virtually nilpotent.* This theorem still holds for the larger class of *elementary amenable groups* (see Theorem 16.33); moreover, such groups with non-polynomial growth must contain a free non-abelian subsemigroup.

The proof of the Milnor–Wolf Theorem essentially consists of a careful examination of increasing/decreasing sequences of subgroups in nilpotent and solvable groups. Along the way, one discovers other features that nilpotent groups share with abelian groups, but not with solvable groups. For instance, *in a nilpotent group all finite subgroups are contained in a maximal finite subgroup, while solvable groups may contain infinite strictly increasing sequences of finite subgroups.* Furthermore, *all subgroups of a nilpotent group are finitely generated, but this is no longer true for solvable groups.* One step further into the study of a finitely generated subgroup $H$ in a group $G$ is to compare a word metric $\text{dist}_H$ on the subgroup $H$ to the restriction to $H$ of a word metric $\text{dist}_G$ on the ambient group $G$. With an appropriate choice of generating sets, the inequality $\text{dist}_G \leqslant \text{dist}_H$ is immediate: All the paths in $H$ joining $h, h' \in H$ are also paths in $G$, but there might be some other, shorter paths in $G$ joining $h, h'$. The problem is to find an upper bound on $\text{dist}_H$ in terms of $\text{dist}_G$. If $G$ is abelian, the upper bound is linear as a function of $\text{dist}_G$. If $\text{dist}_H$ is bounded by a polynomial in $\text{dist}_G$, then the subgroup $H$ is said to be *polynomially distorted* in $G$, while if $\text{dist}_H$ is approximately $\exp(\lambda \text{dist}_G)$ for some $\lambda > 0$, the subgroup $H$ is said to be *exponentially distorted.* It turns out that *all subgroups in a nilpotent group are polynomially distorted, while in solvable groups there exist finitely generated subgroups with exponential distortion.*

Both the von Neumann-Day conjecture and the Milnor conjecture were answered in the affirmative for linear groups by Jacques Tits:

THEOREM (Tits' Alternative). *Let $F$ be a field of zero characteristic and let $\Gamma$ be a subgroup of $GL(n, F)$. Then either $\Gamma$ is virtually solvable or $\Gamma$ contains a free nonabelian subgroup.*

We prove Tits' Alternative in Chapter 13. Note that this alternative also holds for fields of positive characteristic, provided that $\Gamma$ is finitely generated; we decided to limit the discussion to the zero characteristic case in order to avoid algebraic technicalities and because this is the only case of Tits' Alternative used in the proof of Gromov's theorem below.

There are other classes of groups in which both von Neumann-Day and Milnor conjectures are true, they include: Subgroups of Gromov–hyperbolic groups ([**Gro87**, §8.2.$F$], [**GdlH90**, Chapter 8]), fundamental groups of closed Riemannian manifolds of nonpositive curvature [**Bal95**], subgroups of the mapping class group [**Iva92**] and the groups of outer automorphisms of free groups [**BFH00, BFH05**].

The von Neumann-Day conjecture is not true in general: The first counter-examples were given by A. Olshansky in [**Ol'80**]. In [**Ady82**] it was shown that the free Burnside groups $B(n, m)$ with $n \geqslant 2$ and $m \geqslant 665$, $m$ odd, are also counter-examples. Finally, finitely presented counter-examples were constructed by Ol'shansky and Sapir in [**OS02**]. These papers have lead to the development of certain techniques of constructing "infinite finitely generated monsters". While the negation of amenability (i.e. the paradoxical behavior) is, thus, still not completely understood algebraically, several stronger properties implying non-amenability were introduced, among which are various fixed-point properties, most importantly *Kazhdan's Property* (T) (Chapter 17). Remarkably, amenability (hence paradoxical behavior) is a quasi-isometry invariant, while Property (T) is not.

Milnor's conjecture in full generality is, likewise, false: The first groups of *intermediate growth*, i.e. growth which is super-polynomial but subexponential, were constructed by Rostislav Grigorchuk. Moreover, he proved the following:

Theorem (Grigorchuk's Subexponential Growth theorem). *Let $f$ be an arbitrary sub-exponential function larger than $2^{\sqrt{n}}$. Then there exists a finitely generated group $\Gamma$ with subexponential growth function $\mathfrak{G}(n)$ so that:*

$$f(n) \leqslant \mathfrak{G}(n)$$

*for infinitely many $n \in \mathbb{N}$.*

Later on, Anna Erschler [**Ers04**] adapted Grigorchuk's arguments to improve the above result with the inequality $f(n) \leqslant \mathfrak{G}(n)$ for all but finitely many $n$. In the above examples, the exact growth function was unknown. However, Laurent Bartholdi and Anna Erschler [**BE12**] constructed examples of groups of intermediate growth, where they actually compute $\mathfrak{G}(n)$, up to the appropriate equivalence relation. Note, however, that Milnor's conjecture is still open for finitely presented groups.

On the other hand, Mikhael Gromov proved an even more striking result:

Theorem (Gromov's Polynomial Growth Theorem, [**Gro81**]). *Every finitely generated group of polynomial growth is virtually nilpotent.*

This is a typical example of an algebraic property that may be recognized *via* a, seemingly, weak geometric information. A corollary of Gromov's theorem is *quasi-isometric rigidity* for virtually nilpotent groups:

Corollary. *Suppose that $G$ is a group quasi-isometric to a nilpotent group. Then $G$ itself is* virtually nilpotent*, i.e. it contains a nilpotent subgroup of finite index.*

Gromov's theorem and its corollary will be proven in Chapter 14. Since the first version of these notes was written, Bruce Kleiner [**Kle10**] gave a completely different (and much shorter) proof of Gromov's polynomial growth theorem, using harmonic functions on graphs (his proof, however, still requires Tits' Alternative). Kleiner's techniques provided the starting point for Y. Shalom and T. Tao, who proved the following effective version of Gromov's Theorem [**ST10**]:

Theorem (Shalom–Tao Effective Polynomial Growth Theorem). *There exists a constant $C$ such that for any finitely generated group $G$ and $d > 0$, if for some $R \geqslant \exp\left(\exp\left(Cd^C\right)\right)$, the ball of radius $R$ in $G$ has at most $R^d$ elements, then $G$ has a finite index nilpotent subgroup of class less than $C^d$.*

We decided to retain, however, Gromov's original proof since it contains a wealth of ideas that generated in their turn new areas of research. Remarkably, the same piece of logic (a weak version of the axiom of choice) that makes the Banach-Tarski paradox possible also allows to construct *ultralimits*, a powerful tool in the proof of Gromov's theorem and that of many rigidity theorems (e.g, quasi-isometric rigidity theorems of Kapovich, Kleiner and Leeb) as well as in the investigation of fixed point properties.

Regarding Questions (A) and (B), the best one can hope for is that the geometry of a group (up to quasi-isometric equivalence) allows to recover, not just some of its algebraic features, but the group itself, *up to virtual isomorphism*. Two groups $G_1$ and $G_2$ are said to be *virtually isomorphic* if there exist subgroups

$$F_i \triangleleft H_i \leqslant G_i, i = 1, 2,$$

so that $H_i$ has finite index in $G_i$, $F_i$ is a finite normal subgroup in $H_i$, $i = 1, 2$, and $H_1/F_1$ is isomorphic to $H_2/F_2$. Virtual isomorphism implies quasi-isometry but, in general, the converse is false, see Example 5.37. In the situation when the converse implication also holds, one says that the group $G_1$ is *quasi-isometrically rigid*.

An example of quasi-isometric rigidity is given by the following theorem proven by Richard Schwartz [**Sch96b**]:

THEOREM (Schwartz QI rigidity theorem). *Suppose that $\Gamma$ is a nonuniform lattice of isometries of the hyperbolic space $\mathbb{H}^n, n \geqslant 3$. Then each group quasi-isometric to $\Gamma$ must be virtually isomorphic to $\Gamma$.*

We will present a proof of this theorem in Chapter 22. In the same chapter we use similar "zooming" arguments to prove the special case of *Mostow Rigidity Theorem*:

THEOREM (Mostow Rigidity Theorem). *Let $\Gamma_1$ and $\Gamma_2$ be lattices of isometries of $\mathbb{H}^n, n \geqslant 3$, and let $\varphi : \Gamma_1 \to \Gamma_2$ be a group isomorphism. Then $\varphi$ is given by conjugation* via *an isometry of $\mathbb{H}^n$.*

Note that the proof of Schwartz' theorem fails for $n = 2$, where non-uniform lattices are virtually free. (Here and in what follows when we say that a group has a certain property *virtually* we mean that it has a finite index subgroup with that property.) However, in this case, quasi-isometric rigidity still holds as a corollary of Stallings' theorem on ends of groups:

THEOREM. *Let $\Gamma$ be a group quasi-isometric to a free group of finite rank. Then $\Gamma$ is itself virtually free.*

This theorem will be proven in Chapter 18. We also prove:

THEOREM (Stallings "Ends of groups" theorem). *If $G$ is a finitely generated group with infinitely many ends, then $G$ splits as a graph of groups with finite edge–groups.*

In this book we provide two proofs of the above theorem, which, while quite different, are both inspired by the original argument of Stallings. In Chapter 18 we prove Stallings' theorem for *almost finitely presented* groups. This proof follows the ideas of Dunwoody, Jaco and Rubinstein: We will be using *minimal Dunwoody tracks*, where minimality is defined with respect to a certain hyperbolic metric on the presentation complex (unlike combinatorial minimality used by Dunwoody). In

Chapter 19, we will give another proof, which works for all finitely generated groups and follows a proof sketched by Gromov in [**Gro87**], using least energy harmonic functions. We decided to present both proofs, since they use different machinery (the first is more geometric and the second more analytical) and different (although related) geometric ideas.

In Chapter 18 we also prove:

THEOREM (Dunwoody's Accessibility Theorem). *Let $G$ be an almost finitely presented group. Then $G$ is* accessible*, i.e. the decomposition process of $G$ as a graph of groups with finite edge groups eventually terminates.*

In Chapter 21 we prove Tukia's theorem, which establishes quasi-isometric rigidity of the class of fundamental groups of compact hyperbolic $n$-manifolds, and, thus, complements Schwartz' Theorem above:

THEOREM (Tukia's QI Rigidity Theorem). *If a group $\Gamma$ is quasi-isometric to the hyperbolic $n$-space, then $\Gamma$ is virtually isomorphic to the fundamental group of a compact hyperbolic $n$-manifold.*

Note that the proofs of the theorems of Mostow, Schwartz and Tukia all rely upon the same analytical tool: Quasiconformal mappings of Euclidean spaces. In contrast, the analytical proofs of Stallings' theorem presented in the book are mostly motivated by another branch of geometric analysis, namely, the theory of minimal submanifolds and harmonic functions. In the end of the book we also give a survey of quasi-isometric rigidity results.

In regard to Question (B), we investigate two closely related classes of groups: Hyperbolic and relatively hyperbolic groups. These classes generalize fundamental groups of compact negatively curved Riemannian manifolds and, respectively, complete Riemannian manifolds of finite volume. To this end, in Chapters 8, 9 we cover basics of hyperbolic geometry and theory of hyperbolic and relatively hyperbolic groups.

**Other sources.** Our choice of topics in geometric group theory is far from exhaustive. We refer the reader to [**Aea91**],[**Bal95**], [**Bow91**], [**VSCC92**], [**Bow06**], [**BH99**], [**CDP90**], [**Dav08**], [**Geo08**], [**GdlH90**], [**dlH00**], [**NY11**], [**PB03**], [**Roe03**], [**Sap13**], [**Väi05**], for the discussion of other parts of the theory.

**Requirements.** The book is intended as a reference for graduate students and more experienced researchers, it can be used as a basis for a graduate course and as a first reading for a researcher wishing to learn more about geometric group theory. This book is partly based on lectures which we were teaching at Oxford University (C.D.) and University of Utah and University of California, Davis (M.K.). We expect the reader to be familiar with basics of group theory, algebraic topology (fundamental groups, covering spaces, (co)homology, Poincaré duality) and elements of differential topology and Riemannian geometry. Some of the background material is covered in Chapters 1, 2 and 3.

**Acknowledgments.** The work of the first author was supported in part by the ANR projects "Groupe de recherche de Géométrie et Probabilités dans les Groupes", by the EPSRC grant "Geometric and analytic aspects of infinite groups" and by the ANR-10-BLAN 0116, acronym GGAA. The second author was supported by the NSF grants DMS-02-03045, DMS-04-05180, DMS-05-54349, DMS-09-05802 and

DMS-12-05312. During the work on this book the authors visited the Max Planck Institute for Mathematics (Bonn), and wish to thank this institution for its hospitality. We are grateful to Thomas Delzant, Bruce Kleiner and Mohan Ramachandran for supplying some proofs, to Ilya Kapovich, Andrew Sale, Mark Sapir, Alessandro Sisto for the numerous corrections and suggestions. We also thank Yves de Cornulier, Rostislav Grigorchuk, Pierre Pansu and Romain Tessera for useful references.

Addresses:

Cornelia Druţu: Mathematical Institute, 24-29 St Giles, Oxford OX1 3LB, United Kingdom. email: Cornelia.Drutu@maths.ox.ac.uk

Michael Kapovich: Department of Mathematics, University of California, Davis, CA 95616, USA email: kapovich@math.ucdavis.edu

# Contents

CHAPTER 1

# General preliminaries

## 1.1. Notation and terminology

**1.1.1. General notation.** Given a set $X$ we denote by $\mathcal{P}(X)$ the power set of $X$, i.e., the set of all subsets of $X$. If two subsets $A, B$ in $X$ have the property that $A \cap B = \emptyset$ then we denote their union by $A \sqcup B$, and we call it the *disjoint union.* A *pointed set* is a pair $(X, x)$, where $x$ is an element of $X$. The composition of two maps $f : X \to Y$ and $g : Y \to Z$ is denoted either by $g \circ f$ or by $gf$ . We will use the notation $Id_X$ or simply $Id$ (when $X$ is clear) to the denote the identity map $X \to X$. For a map $f : X \to Y$ and a subset $A \subset X$, we let $f|A$ or $f|_A$ denote the restriction of $f$ to $A$. We will use the notation $|E|$ or card $(E)$ to denote cardinality of a set $E$.

The Axiom of Choice (AC) plays an important part in many of the arguments of this book. We discuss AC in more detail in Section 7.1, where we also list equivalent and weaker forms of AC. Throughout the book we make the following convention:

Convention 1.1. We always assume ZFC: The Zermelo–Fraenkel axioms and the Axiom of Choice.

We will use the notation $\overline{A}$ and $cl(A)$ for the closure of a subset $A$ in a topological space $X$. The *wedge of* a family of pointed topological spaces $(X_i, x_i), i \in I$, denoted by $\vee_{i\in I} X_i$, is the quotient of the disjoint union $\sqcup_{i\in I} X_i$, where we identify all the points $x_i$.

If $f : X \to \mathbb{R}$ is a function on a topological space $X$, then we will denote by $\mathrm{Supp}(f)$ the *support* of $f$, i.e., the set

$$cl\{x \in X : f(x) \neq 0\}.$$

Given a non-empty set $X$, we denote by $\mathrm{Bij}(X)$ the group of bijections $X \to X$ , with composition as the binary operation.

Convention 1.2. Throughout the paper we denote by $\mathbf{1}_A$ the characteristic function of a subset $A$ in a set $X$, i.e. the function $\mathbf{1}_A : X \to \{0, 1\}$, $\mathbf{1}_A(x) = 1$ if and only if $x \in A$.

We will use the notation $d$ or dist to denote the metric on a metric space $X$. For $x \in X$ and $A \subset X$ we will use the notation $\mathrm{dist}(x, A)$ for the *minimal distance* from $x$ to $A$, i.e.,

$$\mathrm{dist}(x, A) = \inf\{d(x, a) : a \in A\}.$$

If $A, B \subset X$ are two subsets $A, B$, we let

$$\mathrm{dist}_{Haus}(A, B) = \max\left(\sup_{a\in A} \mathrm{dist}(a, B), \sup_{b\in B} \mathrm{dist}(b, A)\right)$$

denote the *Hausdorff distance* between $A$ and $B$ in $X$. See Section 1.4 for further details on this distance and its generalizations.

Let $(X, \mathrm{dist})$ be a metric space. We will use the notation $\mathcal{N}_R(A)$ to denote the *open $R$-neighborhood* of a subset $A \subset X$, i.e. $\mathcal{N}_R(A) = \{x \in X : \mathrm{dist}(x, A) < R\}$. In particular, if $A = \{a\}$ then $\mathcal{N}_R(A) = B(a, R)$ is the open $R$-ball centered at $a$.

We will use the notation $\overline{\mathcal{N}}_R(A)$, $\overline{B}(a, R)$ to denote the corresponding *closed neighborhoods* and *closed balls* defined by non-strict inequalities.

We denote by $S(x, r)$ the *sphere with center $x$ and radius $r$*, i.e. the set

$$\{y \in X \ : \ \mathrm{dist}(y, x) = r\}.$$

We will use the notation $[A, B]$ to denote a geodesic segment connecting point $A$ to point $B$ in $X$: Note that such segment may be non-unique, so our notation is slightly ambiguous. Similarly, we will use the notation $\triangle(A, B, C)$ or $T(A, B, C)$ for a geodesic triangle with the vertices $A, B, C$. The *perimeter* of a triangle is the sum of its side-lengths (lengths of its edges). Lastly, we will use the notation $\blacktriangle(A, B, C)$ for a solid triangle with the given vertices. Precise definitions of geodesic segments and triangles will be given in Section 1.3.3.

By the *codimension of a subspace $X$ in a space $Y$* we mean the difference between the dimension of $Y$ and the dimension of $X$, whatever the notion of dimension that we use.

With very few exceptions, in a group $G$ we use the multiplication sign $\cdot$ to denote its binary operation. We denote its identity element either by $e$ or by 1. We denote the inverse of an element $g \in G$ by $g^{-1}$. Given a subset $S$ in $G$ we denote by $S^{-1}$ the subset $\{g^{-1} \mid g \in S\}$. Note that for abelian groups the neutral element is usually denoted 0, the inverse of $x$ by $-x$ and the binary operation by $+$.

If two groups $G$ and $G'$ are isomorphic we write $G \simeq G'$.

A surjective homomorphism is called an *epimorphism*, while an injective homomorphism is called a *monomorphism*. An isomorphism of groups $\varphi : G \to G$ is also called an *automorphism*. In what follows, we denote by $\mathrm{Aut}(G)$ the group of automorphisms of $G$.

We use the notation $H < G$ or $H \leqslant G$ to denote that $H$ is a subgroup in $G$. Given a subgroup $H$ in $G$:

- the *order* $|H|$ of $H$ is its cardinality;
- the *index* of $H$ in $G$, denoted $|G : H|$, is the common cardinality of the quotients $G/H$ and $H\backslash G$.

The *order* of an element $g$ in a group $(G, \cdot)$ is the order of the subgroup $\langle g\rangle$ of $G$ generated by $g$. In other words, the order of $g$ is the minimal positive integer $n$ such that $g^n = 1$. If no such integer exists then $g$ is said to be *of infinite order*. In this case, $\langle g\rangle$ is isomorphic to $\mathbb{Z}$.

For every positive integer $m$ we denote by $\mathbb{Z}_m$ *the cyclic group of order $m$*, $\mathbb{Z}/m\mathbb{Z}$. Given $x, y \in G$ we let $x^y$ denote the conjugation of $x$ by $y$, i.e. $yxy^{-1}$.

**1.1.2. Direct and inverse limits of spaces and groups.** Let $I$ be a *directed set*, i.e., a partially ordered set, where every two elements $i, j$ have an *upper bound*, which is some $k \in I$ such that $i \leqslant k, j \leqslant k$. The reader should think of the set of real numbers, or positive real numbers, or natural numbers, as the main examples of directed sets. A *directed system* of sets (or topological spaces, or groups) indexed by $I$ is a collection of sets (or topological spaces, or groups) $A_i, i \in I$, and maps (or continuous maps, or homomorphisms) $f_{ij} : A_i \to A_j, i \leqslant j$, satisfying the following compatibility conditions:

(1) $f_{ik} = f_{jk} \circ f_{ij}, \forall i \leqslant j \leqslant k$,
(2) $f_{ii} = Id$.

An *inverse system* is defined similarly, except $f_{ij} : A_j \to A_i, i \leqslant j$, and, accordingly, in the first condition we use $f_{ij} \circ f_{jk}$.

The *direct limit* of the direct system is the set

$$A = \varinjlim A_i = \left( \coprod_{i \in I} A_i \right) / \sim$$

where $a_i \sim a_j$ whenever $f_{ik}(a_i) = f_{jk}(a_j)$ for some $k \in I$. In particular, we have maps $f_m : A_m \to A$ given by $f_m(a_m) = [a_m]$, where $[a_m]$ is the equivalence class in $A$ represented by $a_m \in A_m$. Note that

$$A = \bigcup_{i \in I} f_m(A_m).$$

If $A_i$'s are groups, then we equip the direct limit with the group operation:

$$[a_i] \cdot [a_j] = [f_{ik}(a_i)] \cdot [f_{jk}(a_j)],$$

where $k \in I$ is an upper bound for $i$ and $j$.

If $A_i$'s are topological spaces, we equip the direct limit with the *final topology*, i.e., the topology where $U \subset \varinjlim A_i$ is open if and only if $f_i^{-1}(U)$ is open for every $i$. In other words, this is the quotient topology descending from the disjoint union of $A_i$'s.

Similarly, the *inverse limit* of an inverse system is

$$\varprojlim A_i = \left\{ (a_i) \in \prod_{i \in I} A_i : a_i = f_{ij}(a_j), \forall i \leqslant j \right\}.$$

If $A_i$'s are groups, we equip the inverse limit with the group operation induced from the direct product of the groups $A_i$. If $A_i$'s are topological spaces, we equip the inverse limit the *initial topology*, i.e., the subset topology of the Tychonoff topology on the direct product. Explicitly, this is the topology generated by the open sets of the form $f_m^{-1}(U_m)$, $U_m \subset X_m$ are open subsets and $f_m : \varprojlim A_i \to A_m$ is the restriction of the coordinate projection.

EXERCISE 1.3. Every group $G$ is the direct limit of the directed family $G_i, i \in I$, consisting of all finitely generated subgroups of $G$. Here the partial order on $I$ is given by inclusion and homomorphisms $f_{ij} : G_i \to G_j$ are tautological embeddings.

EXERCISE 1.4. Suppose that $G$ is the direct limit of a direct system of groups $\{G_i, f_{ij} : i, j \in I\}$. Assume also that for every $i$ we are given a subgroup $H_i \leqslant G_i$ satisfying

$$f_{ij}(H_i) \leqslant H_j, \quad \forall i \leqslant j.$$

Then the family $\{H_i, f_{ij} : i, j \in I\}$ is again a direct system; let $H$ denote the direct limit of this system. Show that there exists a monomorphism $\phi : H \to G$, so that for every $i \in I$,

$$f_i|_{H_i} = \phi \circ f_i|_{H_i} : H_i \to G.$$

EXERCISE 1.5. 1. Let $H \leqslant G$ be a subgroup. Then $|G : H| \leq n$ if and only if the following holds: For every subset $\{g_0, \dots, g_n\} \subset G$, there exist $g_i, g_j$ so that $g_i g_j^{-1} \in H$.

2. Suppose that $G$ is the direct limit of a family of groups $G_i, i \in I$. Assume also that there exist $n \in \mathbb{N}$ so that for every $i \in I$, the group $G_i$ contains a subgroup $H_i$ of index $\leq n$. Let the group $H$ be the direct limit of the family $\{H_i : i \in I\}$ and $\phi : H \to G$ be the monomorphism as in Exercise 1.4. Show that

$$|G : \phi(H)| \leq n.$$

**1.1.3. Growth rates of functions.** We will be using in this book two different *asymptotic* inequalities and equivalences for functions: One is used to compare Dehn functions of groups and the other to compare growth rates of groups.

DEFINITION 1.6. Let $X$ be a subset of $\mathbb{R}$. Given two functions $f, g : X \to \mathbb{R}$, we say that *the order of the function $f$ is at most the order of the function $g$* and we write $f \precsim g$, if there exist $a, b, c, d, e > 0$ such that

$$f(x) \leqslant ag(bx + c) + dx + e$$

for every $x \in X$, $x \geqslant x_0$, for some fixed $x_0$.

If $f \precsim g$ and $g \precsim f$ then we write $f \approx g$ and we say that $f$ and $g$ are *approximately equivalent.*

The equivalence class of a numerical function with respect to equivalence relation $\approx$ is called *the order of the function.* If a function $f$ has (at most) the same order as the function $x$, $x^2$, $x^3$, $x^d$ or $\exp(x)$ it is said that *the order of the function $f$ is (at most) linear, quadratic, cubic, polynomial, or exponential,* respectively. A function $f$ is said to have *subexponential* order if it has order at most $\exp(x)$ and is not approximately equivalent to $\exp(x)$. A function $f$ is said to have *intermediate order* if it has subexponential order and $x^n \precsim f(x)$ for every $n$.

DEFINITION 1.7. We introduce the following *asymptotic inequality* between functions $f, g : X \to \mathbb{R}$ with $X \subset \mathbb{R}$: We write $f \preceq g$ if there exist $a, b > 0$ such that $f(x) \leq ag(bx)$ for every $x \in X$, $x \geq x_0$ for some fixed $x_0$.

If $f \preceq g$ and $g \preceq f$ then we write $f \asymp g$ and we say that $f$ and $g$ are *asymptotically equal.*

Note that this definition is more refined than the *order notion* $\approx$. For instance, $x \approx 0$ while these functions are not asymptotically equal. This situation arises, for instance, in the case of free groups (which are given free presentation): The Dehn function is zero, while the area filling function of the Cayley graph is $A(\ell) \asymp \ell$. The equivalence relation $\approx$ is more appropriate for Dehn functions than the relation $\asymp$, because in the case of a free group one may consider either a presentation with no relation, in which case the Dehn function is zero, or another presentation that yields a linear Dehn function.

EXERCISE 1.8. 1. Show that $\approx$ and $\asymp$ are equivalence relations.

2. Suppose that $x \preceq f$, $x \preceq g$. Then $f \approx g$ if and only if $f \asymp g$.

## 1.2. Graphs

An *unoriented graph* $\Gamma$ consists of the following data:

- a set $V$ called the *set of vertices* of the graph;

- a set $E$ called the *set of edges* of the graph;
- a map $\iota$ called *incidence map* defined on $E$ and taking values in the set of subsets of $V$ of cardinality one or two.

We will use the notation $V = V(\Gamma)$ and $E = E(\Gamma)$ for the vertex and edge sets of the graph $\Gamma$. Two vertices $u, v$ such that $\{u, v\} = \iota(e)$ for some edge $e$, are called *adjacent*. In this case, $u$ and $v$ are called the *endpoints* of the edge $e$.

An unoriented graph can also be seen as a 1-dimensional cell complex, with 0-skeleton $V$ and with 1-dimensional cells/edges labeled by elements of $E$, such that the boundary of each 1-cell $e \in E$ is the set $\iota(e)$. As with general cell complexes and simplicial complexes, we will frequently conflate a graph with its *geometric realization*, i.e., the underlying topological space.

Convention 1.9. In this book, unless we state otherwise, all graphs are assumed to be unoriented.

Note that in the definition of a graph we allow for *monogons*[1] (i.e. edges connecting a vertex to itself) and *bigons*[2] (distinct edges connecting the same pair of vertices). A graph is *simplicial* if the corresponding cell complex is a simplicial complex. In other words, a graph is simplicial if and only if it contains no monogons and bigons.

An edge connecting vertices $u, v$ of $\Gamma$ is denoted $[u, v]$: This is unambiguous if $\Gamma$ is simplicial. A finite ordered set $[v_1, v_2], [v_2, v_3], \ldots, [v_n, v_{n+1}]$ is called an *edge-path* in $\Gamma$. The number $n$ is called the *combinatorial length* of the edge-path. An edge-path in $\Gamma$ is a *cycle* if $v_{n+1} = v_1$. A *simple* cycle (or a *circuit*), is a cycle where all vertices $v_i, i = 1, \ldots, n$, are distinct. In other words, a simple cycle is a cycle homeomorphic to the circle, i.e., a simple loop in $\Gamma$.

A *simplicial tree* is a simply-connected simplicial graph.

An *isomorphism* of graphs is an isomorphism of the corresponding cell complexes, i.e., it is a homeomorphism $f : \Gamma \to \Gamma'$ so that the images of the edges of $\Gamma$ are edges of $\Gamma'$ and images of vertices are vertices. We use the notation $\mathrm{Aut}(\Gamma)$ for the group of automorphisms of a graph $\Gamma$.

The *valency (or valence, or degree) of a vertex* $v$ of a graph $\Gamma$ is the number of edges having $v$ as one of its endpoints, where every monogon with both vertices equal to $v$ is counted twice.

A *directed (or oriented) graph* $\Gamma$ consists of the following data:

- a set $V$ called *set of vertices* of the graph;
- a set $\overline{E}$ called the *set of edges* of the graph;
- two maps $o : \overline{E} \to V$ and $t : \overline{E} \to V$, called respectively the *head (or origin) map* and the *tail map.*

Then, for every $x, y \in V$ we define the set of oriented edges connecting $x$ to $y$:

$$E_{(x,y)} = \{\bar{e} : (o(\bar{e}), t(\bar{e})) = (x, y)\}.$$

A directed graph is called *symmetric* if for every subset $\{u, v\}$ of $V$ the sets $E_{(x,y)}$ and $E_{(y,x)}$ have the same cardinality. For such graphs, interchanging the maps $t$ and $o$ induces an automorphism of the directed graph, which fixes $V$.

[1]Not to be confused with *unigons*, which are hybrids of unicorns and dragons.

[2]Also known as *digons*.

A symmetric directed graph $\overline{\Gamma}$ is equivalent to a unoriented graph $\Gamma$ with the same vertex set, *via* the following replacement procedure: Pick an involutive bijection $\beta : \overline{E} \to \overline{E}$, which induces bijections $\beta : E_{(x,y)} \to E_{(y,x)}$ for all $x, y \in V$. We then get the equivalence relation $e \sim \beta(e)$. The quotient $E = \overline{E}/\sim$ is the edge-set of the graph $\Gamma$, where the incidence map $\iota$ is defined by $\iota([e]) = \{o(e), t(e)\}$. The unoriented graph $\Gamma$ thus obtained, is called the *underlying unoriented graph* of the given directed graph.

EXERCISE 1.10. Describe the converse to this procedure: Given a graph $\Gamma$, construct a symmetric directed graph $\overline{\Gamma}$, so that $\Gamma$ is the underlying graph of $\overline{\Gamma}$.

DEFINITION 1.11. Let $F \subset V = V(\Gamma)$ be a set of vertices in a (unoriented) graph $\Gamma$. The *vertex-boundary* of $F$, denoted by $\partial_V F$, is the set of vertices in $F$ each of which is adjacent to a vertex in $V \setminus F$.

The *edge-boundary* of $F$, denoted by $E(F, F^c)$, is the set of edges $e$ such that the set of endpoints $\iota(e)$ intersects both $F$ and its complement $F^c = V \setminus F$ in exactly one element.

Unlike the vertex-boundary, the edge boundary is the same for $F$ as for its complement $F^c$. For graphs without bigons, the edge-boundary can be identified with the set of vertices $v \in V \setminus F$ adjacent to a vertex in $F$, in other words, with $\partial_V (V \setminus F)$.

For graphs having a uniform upper bound $C$ on the valency of vertices, cardinalities of the two types of boundaries are *comparable*

$$|\partial_V F| \leqslant |E(F, F^c)| \leqslant C|\partial_V F| \,. \tag{1.1}$$

DEFINITION 1.12. A simplicial graph $\Gamma$ is *bipartite* if the vertex set $V$ splits as $V = Y \sqcup Z$, so that each edge $e \in E$ has one endpoint in $Y$ and one endpoint in $Z$. In this case, we write $\Gamma = Bip(Y, Z; E)$.

EXERCISE 1.13. Let $W$ be an $n$-dimensional vector space over a field $K$ ($n \geqslant 3$). Let $Y$ be the set of 1-dimensional subspaces of $W$ and let $Z$ be the set of 2-dimensional subspaces of $W$. Define the bipartite graph $\Gamma = Bip(Y, Z, E)$, where $y \in Y$ is adjacent to $z \in Z$ if, as subspaces in $W$, $y \subset z$.

1. Compute (in terms of $K$ and $n$) the valence of $\Gamma$, the (combinatorial) length of the shortest circuit in $\Gamma$, and show that $\Gamma$ is connected. 2. Estimate from above the length of the shortest path between any pair of vertices of $\Gamma$. Can you get a bound independent of $K$ and $n$?

## 1.3. Topological and metric spaces

**1.3.1. Topological spaces. Lebesgue covering dimension.** Given two topological spaces, we let $C(X; Y)$ denote the space of all continuous maps $X \to Y$; set $C(X) := C(X; \mathbb{R})$. We always endow the space $C(X; Y)$ with the compact-open topology.

DEFINITION 1.14. Two subsets $A, V$ of a topological space $X$ are said to be *separated by a function* if there exists a continuous function $\rho = \rho_{A,V} : X \to [0, 1]$ so that

1. $\rho | A \equiv 0$
2. $\rho | V \equiv 1$.

A topological space $X$ is called *perfectly normal* if every two disjoint closed subsets of $X$ can be separated by a function.

An open covering $\mathcal{U} = \{U_i : i \in I\}$ of a topological space $X$ is called *locally finite* if every subset $J \subset I$ such that

$$\bigcap_{i \in J} U_i \neq \emptyset$$

is finite. Equivalently, every point $x \in X$ has a neighborhood which intersects only finitely many $U_i$'s.

The *multiplicity* of an open covering $\mathcal{U} = \{U_i : i \in I\}$ of a space $X$ is the supremum of cardinalities of subsets $J \subset I$ so that

$$\bigcap_{i \in J} U_i \neq \emptyset.$$

A covering $\mathcal{V}$ is called a *refinement* of a covering $\mathcal{U}$ if every $V \in \mathcal{V}$ is contained in some $U \in \mathcal{U}$.

DEFINITION 1.15. The *(Lebesgue) covering dimension* of a topological space $Y$ is the least number $n$ such that the following holds: Every open cover $\mathcal{U}$ of $Y$ admits a refinement $\mathcal{V}$ which has multiplicity at most $n + 1$.

The following example shows that covering dimension is consistent with our "intuitive" notion of dimension:

EXAMPLE 1.16. If $M$ is a $n$-dimensional topological manifold, then $n$ equals the covering dimension of $M$. See e.g. [**Nag83**].

**1.3.2. General metric spaces.** A *metric space* is a set $X$ endowed with a function $\mathrm{dist} : X \times X \to \mathbb{R}$ with the following properties:

(M1) $\mathrm{dist}(x, y) \geqslant 0$ for all $x, y \in X$; $\mathrm{dist}(x, y) = 0$ if and only if $x = y$;

(M2) (Symmetry) for all $x, y \in X$, $\mathrm{dist}(y, x) = \mathrm{dist}(x, y)$;

(M3) (Triangle inequality) for all $x, y, z \in X$, $\mathrm{dist}(x, z) \leqslant \mathrm{dist}(x, y) + \mathrm{dist}(y, z)$.

The function dist is called *metric* or *distance function.* Occasionally, it will be convenient to allow dist to take infinite values, in this case, we interpret triangle inequalities following the usual calculus conventions ($a + \infty = \infty$ for every $a \in \mathbb{R} \cup \infty$, etc.).

A metric space is said to satisfy the *ultrametric inequality* if

$$\mathrm{dist}(x, z) \leqslant \max(\mathrm{dist}(x, y), \mathrm{dist}(y, z)), \forall x, y, z \in X.$$

We will see some examples of ultrametric spaces in Section 1.8.

Every norm $|\cdot|$ on a vector space $V$ defines a metric on $V$:

$$\mathrm{dist}(u, v) = |u - v|.$$

The standard examples of norms on the $n$-dimensional real vector space $V$ are:

$$|v|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}, 1 \leqslant p < \infty,$$

and

$$|v|_{max} = |v|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

Exercise 1.17. Show that the Euclidean plane $E^2$ satisfies the *parallelogram identity*: If $A, B, C, D$ are vertices of a parallelogram $P$ in $E^2$ with the diagonals $[AC]$ and $[BD]$, then

$$d^2(A,B) + d^2(B,C) + d^2(C,D) + d^2(D,A) = d^2(A,C) + d^2(B,D), \tag{1.2}$$

i.e., sum of squares of the sides of $P$ equals the sum of squares of the diagonals of $P$.

If $X, Y$ are metric spaces, the *product metric* on the direct product $X \times Y$ is defined by the formula

$$d((x_1,y_1),(x_2,y_2))^2 = d(x_1,x_2)^2 + d(y_1,y_2)^2. \tag{1.3}$$

We will need a *separation* lemma which is standard (see for instance [**Mun75**, §32]), but we include a proof for the convenience of the reader.

Lemma 1.18. *Every metric space $X$ is perfectly normal.*

Proof. Let $A, V \subset X$ be disjoint closed subsets. Both functions $\mathrm{dist}_A$, $\mathrm{dist}_V$, which assign to $x \in X$ its minimal distance to $A$ and to $V$ respectively, are clearly continuous. Therefore the ratio

$$\sigma(x) := \frac{\mathrm{dist}_A(x)}{\mathrm{dist}_V(x)}, \quad \sigma : X \to [0,\infty]$$

is continuous as well. Let $\tau : [0,\infty] \to [0,1]$ be a continuous monotone function such that $\tau(0) = 0, \tau(\infty) = 1$, e.g.

$$\tau(y) = \frac{2}{\pi}\arctan(y), \quad y \neq \infty, \quad \tau(\infty) := 1.$$

Then the composition $\rho := \tau \circ \sigma$ satisfies the required properties. □

A metric space $(X, \mathrm{dist})$ is called *proper* if for every $p \in X$ and $R > 0$ the closed ball $\overline{B}(p,R)$ is compact. In other words, the distance function $d_p(x) = d(p,x)$ is proper.

A topological space is called *locally compact* if for every $x \in X$ there exists a basis of neighborhoods of $x$ consisting of *relatively compact* subsets of $X$, i.e., subsets with compact closure. A metric space is locally compact if and only if for every $x \in X$ there exists $\varepsilon = \varepsilon(x) > 0$ such that the closed ball $\overline{B}(x,\varepsilon)$ is compact.

Definition 1.19. Given a function $\phi : \mathbb{R}_+ \to \mathbb{N}$, a metric space $X$ is called *$\phi$–uniformly discrete* if each ball $\overline{B}(x,r) \subset X$ contains at most $\phi(r)$ points. A metric space is called *uniformly discrete* if it is $\phi$–uniformly discrete for some function $\phi$.

Note that every uniformly discrete metric space necessarily has discrete topology.

Given two metric spaces $(X, \mathrm{dist}_X)$, $(Y, \mathrm{dist}_Y)$, a map $f : X \to Y$ is an *isometric embedding* if for every $x, x' \in X$

$$\mathrm{dist}_Y(f(x), f(x')) = \mathrm{dist}_X(x,x').$$

The image $f(X)$ of an isometric embedding is called an *isometric copy of $X$ in $Y$*.

A surjective isometric embedding is called an *isometry*, and the metric spaces $X$ and $Y$ are called *isometric*. A surjective map $f : X \to Y$ is called *a similarity with the factor $\lambda$* if for all $x, x' \in X$,

$$\mathrm{dist}_Y(f(x), f(x')) = \lambda \mathrm{dist}_X(x,x').$$

The group of isometries of a metric space $X$ is denoted $\mathrm{Isom}(X)$. A metric space is called *homogeneous* if the group $\mathrm{Isom}(X)$ acts transitively on $X$, i.e., for every $x, y \in X$ there exists an isometry $f : X \to X$ such that $f(x) = y$.

**1.3.3. Length metric spaces.** Throughout these notes by a *path* in a topological space $X$ we mean a continuous map $\mathfrak{p} : [a, b] \to X$. A path is said to *join* (or *connect*) two points $x, y$ if $\mathfrak{p}(a) = x$, $\mathfrak{p}(b) = y$. We will frequently conflate a path and its image.

Given a path $\mathfrak{p}$ in a metric space $X$, one defines the *length* of $\mathfrak{p}$ as follows. A partition

$$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b$$

of the interval $[a, b]$ defines a finite collection of points $\mathfrak{p}(t_0), \mathfrak{p}(t_1), \ldots, \mathfrak{p}(t_{n-1}), \mathfrak{p}(t_n)$ in the space $X$. The *length of* $\mathfrak{p}$ is then defined to be

$$\mathrm{length}(\mathfrak{p}) = \sup_{a=t_0<t_1<\cdots<t_n=b} \sum_{i=0}^{n-1} \mathrm{dist}(\mathfrak{p}(t_i), \mathfrak{p}(t_{i+1})) \tag{1.4}$$

where the supremum is taken over all possible partitions of $[a, b]$ and all integers $n$. By the definition and triangle inequalities in $X$, $\mathrm{length}(\mathfrak{p}) \geqslant \mathrm{dist}(\mathfrak{p}(a), \mathfrak{p}(b))$.

If the length of $\mathfrak{p}$ is finite then $\mathfrak{p}$ is called *rectifiable*, and we say that $\mathfrak{p}$ is *non-rectifiable* otherwise.

Exercise 1.20. Consider a $C^1$-smooth path in the Euclidean space $\mathfrak{p} : [a, b] \to \mathbb{R}^n$, $\mathfrak{p}(t) = (x_1(t), \ldots, x_n(t))$. Prove that its length (defined above) is given by the familiar formula

$$\mathrm{length}(\mathfrak{p}) = \int_a^b \sqrt{[x_1'(t)]^2 + \ldots + [x_n'(t)]^2}\, dt.$$

Similarly, if $(M, g)$ is a connected Riemannian manifold and dist is the Riemannian distance function, then the two notions of length, given by equations (2.1) and (1.4), coincide for smooth paths.

Exercise 1.21. Prove that the graph of the function $f : [0, 1] \to \mathbb{R}$,

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if} \quad 0 < x \leqslant 1\,, \\ 0 & \text{if} \quad x = 0\,, \end{cases}$$

is a non-rectifiable path joining $(0, 0)$ and $(1, \sin(1))$.

Let $(X, \mathrm{dist})$ be a metric space. We define a new metric $\mathrm{dist}_\ell$ on $X$, known as the *induced intrinsic metric:* $\mathrm{dist}_\ell(x, y)$ is the infimum of the lengths of all rectifiable paths joining $x$ to $y$.

Exercise 1.22. Show that $\mathrm{dist}_\ell$ is a metric on $X$ with values in $[0, \infty]$.

Suppose that $\mathfrak{p}$ is a path realizing the infimum in the definition of distance $\mathrm{dist}_\ell(x, y)$. We will (re)parameterize such $\mathfrak{p}$ by its arc-length; the resulting path $\mathfrak{p} : [0, D] \to (X, \mathrm{dist}_\ell)$ is called a *geodesic segment* in $(X, \mathrm{dist}_\ell)$.

Exercise 1.23. $\mathrm{dist} \leqslant \mathrm{dist}_\ell$.

Definition 1.24. A metric space $(X, \mathrm{dist})$ such that $\mathrm{dist} = \mathrm{dist}_\ell$ is called a *length (or path) metric space*.

Note that in a path metric space, *a priori*, not every two points are connected by a geodesic. We extend the notion of geodesic to general metric spaces: A *geodesic* in a metric space $X$ is an isometric embedding $\mathfrak{g}$ of an interval in $\mathbb{R}$ into $X$. Note that this notion is different from the one in Riemannian geometry, where geodesics are isometric embeddings only *locally*, and need not be arc-length parameterized. A geodesic is called a *geodesic ray* if it is defined on an interval $(-\infty, a]$ or $[a, +\infty)$, and it is called *bi-infinite* or *complete* if it is defined on $\mathbb{R}$.

DEFINITION 1.25. A metric space $X$ is called *geodesic* if every two points in $X$ are connected by a geodesic path. A subset $A$ in a metric space $X$ is called *convex* if for every two points $x, y \in A$ there exists a geodesic $\gamma \subset X$ connecting $x$ and $y$.

EXERCISE 1.26. Prove that for $(X, \mathrm{dist}_\ell)$ the two notions of geodesics agree.

A *geodesic triangle* $T = T(A, B, C)$ or $\Delta(A, B, C)$ with vertices $A, B, C$ in a metric space $X$ is a collection of geodesic segments $[A, B], [B, C], [C, A]$ in $X$. These segments are called *edges* of $T$. Later on, in Chapters 8 and 9 we will use *generalized* triangles, where some edges are geodesic rays or, even, complete geodesics. The corresponding vertices generalized triangles will be *points of the ideal boundary* of $X$.

EXAMPLES 1.27.
(1) $\mathbb{R}^n$ with the Euclidean metric is a geodesic metric space.
(2) $\mathbb{R}^n \setminus \{0\}$ with the Euclidean metric is a length metric space, but not a geodesic metric space.
(3) The unit circle $\mathbb{S}^1$ with the metric inherited from the Euclidean metric of $\mathbb{R}^2$ (the chordal metric) is not a length metric space. The induced intrinsic metric on $\mathbb{S}^1$ is the one that measures distances as angles in radians, it is the distance function of the Riemannian metric induced by the embedding $\mathbb{S}^1 \to \mathbb{R}^2$.
(4) The Riemannian distance function dist defined for a connected Riemannian manifold $(M, g)$ (see Section 2.1.3) is a path-metric. If this metric is complete, then the path-metric is geodesic.
(5) Every connected graph equipped with the standard distance function (see Section 1.3.4) is a geodesic metric space.

EXERCISE 1.28. If $X, Y$ are geodesic metric spaces, so is $X \times Y$. If $X, Y$ are path-metric spaces, so is $X \times Y$. Here $X \times Y$ is equipped with the product metric defined by (1.3).

THEOREM 1.29 (Hopf–Rinow Theorem [**Gro07**]). *If a length metric space is complete and locally compact, then it is geodesic and proper.*

EXERCISE 1.30. Construct an example of a metric space $X$ which is not a length metric space, so that $X$ is complete, locally compact, but is not proper.

**1.3.4. Graphs as length spaces.** Let $\Gamma$ be a connected graph. Recall that we are conflating $\Gamma$ and its geometric realization, so the notation $x \in \Gamma$ below will simply mean that $x$ is a point of the geometric realization.

We introduce a path-metric dist on the geometric realization of $\Gamma$ as follows. We declare every edge of $\Gamma$ to be isometric to the unit interval in $\mathbb{R}$. Then, the distance between any vertices of $\Gamma$ is the combinatorial length of the shortest edge-path connecting these vertices. Of course, points of the interiors of edges of $\Gamma$ are not connected by any edge-paths. Thus, we consider *fractional* edge-paths, where in addition to the edges of $\Gamma$ we allow intervals contained in the edges. The length of such a fractional path is the sum of lengths of the intervals in the path. Then, for $x, y \in \Gamma$, $\text{dist}(x, y)$ is

$$\inf_{\mathfrak{p}} (\text{length}(\mathfrak{p})),$$

where the infimum is taken over all fractional edge-paths $\mathfrak{p}$ in $\Gamma$ connecting $x$ to $y$.

EXERCISE 1.31. a. Show that infimum is the same as minimum in this definition.

b. Show that every edge of $\Gamma$ (treated as a unit interval) is isometrically embedded in $(\Gamma, \text{dist})$.

c. Show that dist is a path-metric.

d. Show that dist is a complete metric.

The metric dist is called the *standard* metric on $\Gamma$.

The notion of a standard metric on a graph generalizes to the concept of a *metric graph*, which is a connected graph $\Gamma$ equipped with a path-metric $dist_\ell$. Such path-metric is, of course, uniquely determined by the lengths of edges of $\Gamma$ with respect to the metric $d$.

EXAMPLE 1.32. Consider $\Gamma$ which is the complete graph on 3 vertices (a triangle) and declare that two edges $e_1, e_2$ of $\Gamma$ are unit intervals and the remaining edge $e_3$ of $\Gamma$ has length 3. Let $\text{dist}_\ell$ be the corresponding path-metric on $\Gamma$. Then $e_3$ is not isometrically embedded in $(\Gamma, \text{dist}_\ell)$.

### 1.4. Hausdorff and Gromov-Hausdorff distances. Nets

Given subsets $A_1, A_2$ in a metric space $(X, d)$, define the *minimal distance* between these sets as

$$\text{dist}(A_1, A_2) = \inf\{d(a_1, a_2) : a_i \in A_i, i = 1, 2\}.$$

The *Hausdorff (pseudo)distance* between subsets $A_1, A_2 \subset X$ is defined as

$$\text{dist}_{Haus}(A_1, A_2) := \inf\{R : A_1 \subset \mathcal{N}_R(A_2), A_2 \subset \mathcal{N}_R(A_1)\}.$$

Two subsets of $X$ are called *Hausdorff-close* if they are within finite Hausdorff distance from each other.

The Hausdorff distance between two distinct spaces (for instance, between a space and a dense subspace in it) can be zero. The Hausdorff distance becomes a genuine distance only when restricted to certain classes of subsets, for instance, to the class of compact subsets of a metric space. Still, for simplicity, we call it a *distance* or a *metric* in all cases.

Hausdorff distance defines the topology of *Hausdorff–convergence* on the set $K(X)$ of compact subsets of a metric space $X$. This topology extends to the set $C(X)$ of closed subsets of $X$ as follows. Given $\epsilon > 0$ and a compact $K \subset X$ we define the neighborhood $U_{\epsilon,K}$ of a closed subset $C \in C(X)$ to be

$$\{Z \in C(X) : \text{dist}_{Haus}(Z \cap K, C \cap K) < \epsilon\}.$$

This system of neighborhoods generates a topology on $C(X)$, called *Chabauty topology*. Thus, a sequence $C_i \in C(X)$ converges to a closed subset $C \in C(X)$ if and only if for every compact subset $K \subset X$,

$$\lim_{i\to\infty} C_i \cap K = C \cap K,$$

where the limit is in topology of Hausdorff–convergence.

M. Gromov defined in [**Gro81**, section 6] the *modified Hausdorff pseudo-distance* (also called the *Gromov–Hausdorff pseudo-distance*) on the class of proper metric spaces:

$$\text{(1.5)} \qquad \mathrm{dist}_{GHaus}((X,d_X),(Y,d_Y)) = \inf_{(x,y)\in X\times Y} \inf\{\varepsilon > 0 \mid \exists \text{ a pseudo-metric}$$

$$\mathrm{dist} \text{ on } M = X \sqcup Y, \text{ such that } \mathrm{dist}(x,y) < \varepsilon, \mathrm{dist}|_X = d_X, \mathrm{dist}|_Y = d_Y \text{ and}$$

$$B(x, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(Y), B(y,1/\varepsilon) \subset \mathcal{N}_\varepsilon(X)\}\,.$$

For homogeneous metric spaces the modified Hausdorff pseudo-distance coincides with the pseudo-distance for the pointed metric spaces:

$$\text{(1.6)} \qquad \mathrm{dist}_{\tilde{H}}((X,d_X,x_0),(Y,d_Y,y_0)) = \inf\{\varepsilon > 0 \mid \exists \text{ a pseudo-metric}$$

$$\mathrm{dist} \text{ on } M = X \sqcup Y \text{ such that } \mathrm{dist}(x_0,y_0) < \varepsilon,\ \mathrm{dist}|_X = d_X, \mathrm{dist}|_Y = d_Y,$$

$$B(x_0, 1/\varepsilon) \subset \mathcal{N}_\varepsilon(Y), B(y_0,1/\varepsilon) \subset \mathcal{N}_\varepsilon(X)\}\,.$$

This pseudo-distance becomes a metric when restricted to the class of proper pointed metric spaces.

Still, as before, to simplify the terminology we shall refer to all three pseudo-distances as 'distances' or 'metrics.'

EXAMPLE 1.33. The real line $\mathbb{R}$ with the standard metric and the planar circle of radius $r$, $\mathcal{C}(O,r)$, with the length metric, are at modified Hausdorff distance

$$\varepsilon_0 := \frac{4}{\sqrt{\pi^2 r^2 + 16} + \pi r}.$$

Since both are homogeneous spaces, it suffices to prove that the pointed metric spaces $(\mathbb{R},0)$ and $(\mathcal{C}(O,r),N)$, where $N$ is the North pole, are at the distance $\varepsilon_0$ with respect to the modified Hausdorff distance with respect to these base-points.

To prove the upper bound we glue $\mathbb{R}$ and $\mathcal{C}(O,r)$ by identifying isometrically the interval $\left[-\frac{\pi}{2}r\,,\,\frac{\pi}{2}r\right]$ in $\mathbb{R}$ to the upper semi-circle (see Figure 1.1), and we endow the graph $M$ thus obtained with its length metric dist. Note that the use of pseudo-metrics on $M$ in the definition of the modified Hausdorff pseudo-distance allows for points $x \in X$ and $y \in Y$ to be identified. The minimal $\varepsilon > 0$ such that in $(M, \mathrm{dist})$

$$\left[-\frac{1}{\varepsilon}\,,\,\frac{1}{\varepsilon}\right] \subset \mathcal{N}_\varepsilon(\mathcal{C}(O,r)) \text{ and } B(N,1/\varepsilon) \subset \mathcal{N}_\varepsilon(\mathbb{R})$$

is $\varepsilon_0$ defined above. This value is the positive solution of the equation

$$\text{(1.7)} \qquad \frac{\pi}{2}r + \varepsilon = \frac{1}{\varepsilon}.$$

For the lower bound consider another metric $\mathrm{dist}'$ on $\mathbb{R}\vee\mathcal{C}(O,r)$ which coincides with the length metrics on both $\mathbb{R}$ and $\mathcal{C}(O,r)$. Let $\varepsilon'$ be the smallest $\varepsilon > 0$ such that $\mathrm{dist}'(0,N) < \varepsilon$ and $\left[-\frac{1}{\varepsilon}\,,\,\frac{1}{\varepsilon}\right] \subset \mathcal{N}_\varepsilon(\mathcal{C}(O,r))$, $B(N,1/\varepsilon) \subset \mathcal{N}_\varepsilon(\mathbb{R})$ in the metric dist′. Let $x', y'$ be the nearest points in $\mathcal{C}(O,r)$ to $-\frac{1}{\varepsilon'}$ and $\frac{1}{\varepsilon'}$, respectively. Since $\mathrm{dist}'(x',y') \leqslant \pi r$, it follows that $\frac{2}{\varepsilon'} \leqslant \pi r + 2\varepsilon'$. The previous inequality implies that $\varepsilon' \geqslant \varepsilon_0$.

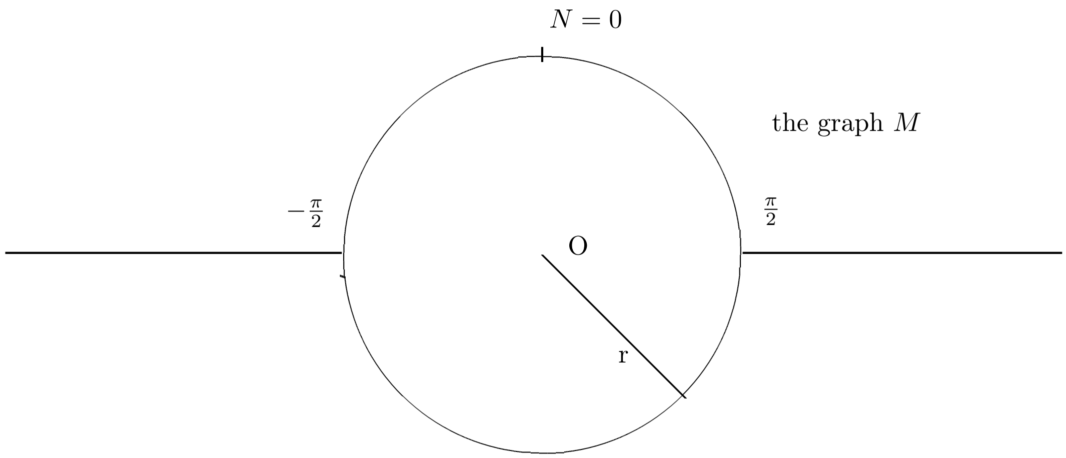



FIGURE 1.1. Circle and real line glued along an arc of length $\pi r$.

One can associate to every metric space $(X, \mathrm{dist})$ a discrete metric space that is at finite Hausdorff distance from $X$, as follows.

DEFINITION 1.34. An *$\varepsilon$–separated subset* $A$ in $X$ is a subset such that

$$\mathrm{dist}(a_1, a_2) \geqslant \varepsilon\,,\ \forall a_1, a_2 \in A,\ a_1 \neq a_2\,.$$

A subset $S$ of a metric space $X$ is said to be $r$-*d*ense in $X$ if the Hausdorff distance between $S$ and $X$ is at most $r$.

DEFINITION 1.35. An *$\varepsilon$-separated $\delta$–net* in a metric space $X$ is a subset of $X$ that is $\varepsilon$–separated and $\delta$–dense.

An *$\varepsilon$-separated net* in $X$ is a subset that is $\varepsilon$–separated and $2\varepsilon$–dense.

When the constants $\varepsilon$ and $\delta$ are not relevant we shall not mention them and simply speak of separated nets.

LEMMA 1.36. *A maximal $\delta$–separated set in $X$ is a $\delta$–separated net in $X$.*

PROOF. Let $N$ be a maximal $\delta$–separated set in $X$. For every $x \in X \setminus N$, the set $N \cup \{x\}$ is no longer $\delta$–separated, by maximality of $N$. Hence there exists $y \in N$ such that $\mathrm{dist}(x, y) < \delta$. □

By Zorn's lemma a maximal $\delta$–separated set always exists. Thus, every metric space contains a $\delta$–separated net, for any $\delta > 0$.

EXERCISE 1.37. Prove that if $(X, \mathrm{dist})$ is compact then every separated net in $X$ is finite; hence, every separated set in $X$ is finite.

DEFINITION 1.38 (Rips complex). Let $(X, d)$ be a metric space. For $R \geqslant 0$ we define a simplicial complex $\text{Rips}_R(X)$; its vertices are points of $X$; vertices $x_0, x_1, ..., x_n$ span a simplex if and only if for all $i, j$,

$$\text{dist}(x_i, x_j) \leqslant R.$$

The simplicial complex $\text{Rips}_R(X)$ is called the *$R$-Rips complex of $X$*.

We will discuss Rips complexes in more detail in §6.2.1.

## 1.5. Lipschitz maps and Banach-Mazur distance

**1.5.1. Lipschitz and locally Lipschitz maps.** A map $f : X \to Y$ between two metric spaces $(X, \text{dist}_X)$, $(Y, \text{dist}_Y)$ is *$L$-Lipschitz* if for all $x, x' \in X$

$$\text{dist}_Y(f(x), f(x')) \leqslant L\text{dist}_X(x, x').$$

A map which is $L$-Lipschitz for some $L$ is called simply *Lipschitz*.

EXERCISE 1.39. Show that every $L$-Lipschitz path $\mathfrak{p} : [0, 1] \to X$ is rectifiable and $\text{length}(\mathfrak{p}) \leqslant L$.

The following is a fundamental theorem about Lipschitz maps between Euclidean spaces:

THEOREM 1.40 (Rademacher Theorem, see Theorem 3.1 in [**Hei01**]). *Let $U$ be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}^m$ be Lipschitz. Then $f$ is differentiable at almost every point in $U$.*

A map $f : X \to Y$ is called *locally Lipschitz* if for every $x \in X$ there exists $\epsilon > 0$ so that the restriction $f|B(x, \epsilon)$ is Lipschitz. We let $\text{Lip}_{\text{loc}}(X; Y)$ denote the space of locally Lipschitz maps $X \to Y$. We set $\text{Lip}_{\text{loc}}(X) := \text{Lip}_{\text{loc}}(X; \mathbb{R})$.

EXERCISE 1.41. Fix a point $p$ in a metric space $(X, \text{dist})$ and define the function $\text{dist}_p$ by $\text{dist}_p(x) := \text{dist}(x, p)$. Show that this function is 1-Lipschitz.

LEMMA 1.42 (Lipschitz bump-function). *Let $0 < R < \infty$. Then there exists a $\frac{1}{R}$–Lipschitz function $\varphi = \varphi_{p,R}$ on $X$ such that*

1. *$\varphi$ is positive on $B(p, R)$ and zero on $X \setminus B(p, R)$.*
2. *$\varphi(p) = 1$.*
3. *$0 \leqslant \varphi \leqslant 1$ on $X$.*

PROOF. We first define the function $\zeta : \mathbb{R}_+ \to [0, 1]$ which vanishes on the interval $[R, \infty)$, is linear on $[0, R]$ and equals 1 at 0. Then $\zeta$ is $\frac{1}{R}$–Lipschitz. Now take $\varphi := \zeta \circ \text{dist}_p$. □

LEMMA 1.43 (Lipschitz partition of unity). *Suppose that we are given a locally finite covering of a metric space $X$ by a countable set of open $R_i$-balls $B_i := B(x_i, R_i)$, $i \in I \subset \mathbb{N}$. Then there exists a collection of Lipschitz functions $\eta_i$, $i \in I$ so that:*

1. *$\sum_i \eta_i \equiv 1$.*
2. *$0 \leqslant \eta_i \leqslant 1, \quad \forall i \in I$.*
3. *$\text{Supp}(\eta_i) \subset \overline{B(x_i, R_i)}, \quad \forall i \in I$.*

Proof. For each $i$ define the bump-function using Lemma 1.42:

$$\varphi_i := \varphi_{x_i,R_i}.$$

Then the function

$$\varphi := \sum_{i\in I} \varphi_i$$

is positive on $X$. Finally, define

$$\eta_i := \frac{\varphi_i}{\varphi}.$$

It is clear that the functions $\eta_i$ satisfy all the required properties. □

Remark 1.44. Since the collection of balls $\{B_i\}$ is locally finite, it is clear that the function

$$L(x) := \sup_{i\in I, \eta_i(x)\neq 0} \mathrm{Lip}(\eta_i)$$

is bounded on compact sets in $X$, however, in general, it is unbounded on $X$. We refer the reader to the equation (1.8) for the definition of $\mathrm{Lip}(\eta_i)$.

From now on, we assume that $X$ is a proper metric space.

Proposition 1.45. $\mathrm{Lip}_{\mathrm{loc}}(X)$ *is a dense subset in* $C(X)$*, the space of continuous functions* $X \to \mathbb{R}$*, equipped with the compact-open topology (topology of uniform convergence on compacts).*

Proof. Fix a base-point $o \in X$ and let $A_n$ denote the annulus

$$\{x \in X : n-1 \leqslant \mathrm{dist}(x, o) \leqslant n\}, n \in \mathbb{N}.$$

Let $f$ be a continuous function on $X$. Pick $\epsilon > 0$. Our goal is to find a locally Lipschitz function $g$ on $X$ so that $|f(x) - g(x)| < \epsilon$ for all $x \in X$. Since $f$ is uniformly continuous on compact sets, for each $n \in \mathbb{N}$ there exists $\delta = \delta(n, \epsilon)$ such that

$$\forall x, x' \in A_n, \quad \mathrm{dist}(x, x') < \delta \Rightarrow |f(x) - f(x')| < \epsilon\,.$$

Therefore for each $n$ we find a finite subset

$$X_n := \{x_{n,1}, \ldots, x_{n,m_n}\} \subset A_n$$

so that for $r := \delta(n, \epsilon)/4, R := 2r$, the open balls $B_{n,j} := B(x_{n,j}, r)$ cover $A_n$. We reindex the set of points $\{x_{n,j}\}$ and the balls $B_{n,j}$ with a countable set $I$. Thus, we obtain an open locally finite covering of $X$ by the balls $B_j, j \in I$. Let $\{\eta_j, j \in I\}$ denote the corresponding Lipschitz partition of unity. It is then clear that

$$g(x) := \sum_{i\in I} \eta_i(x) f(x_i)$$

is a locally Lipschitz function. For $x \in B_i$ let $J \subset I$ be such that

$$x \notin B(x_j, R_j), \quad \forall j \notin J.$$

Then $|f(x) - f(x_j)| < \epsilon$ for all $j \in J$. Therefore

$$|g(x) - f(x)| \leqslant \sum_{j\in J} \eta_j(x)|f(x_j) - f(x)| < \epsilon \sum_{j\in J} \eta_j(x) = \epsilon \sum_{i\in I} \eta_j(x) = \epsilon.$$

It follows that $|f(x) - g(x)| < \epsilon$ for all $x \in X$. □

A relative version of Proposition 1.45 also holds:

PROPOSITION 1.46. *Let $A \subset X$ be a closed subset contained in a subset $U$ which is open in $X$. Then, for every $\epsilon > 0$ and every continuous function $f \in C(X)$ there exists a function $g \in C(X)$ so that:*

1. *$g$ is locally Lipschitz on $X \setminus U$.*
2. *$\|f - g\| < \epsilon$.*
3. *$g|A = f|A$.*

PROOF. For the closed set $V := X \setminus U$ pick a continuous function $\rho = \rho_{A,V}$ separating the sets $A$ and $V$. Such a function exists, by Lemma 1.18. According to Proposition 1.45, there exists $h \in \mathrm{Lip}_{\mathrm{loc}}(X)$ such that $\|f - h\| < \epsilon$. Then take

$$g(x) := \rho(x)h(x) + (1 - \rho(x))f(x).$$

We leave it to the reader to verify that $g$ satisfies all the requirements of the proposition. □

**1.5.2. Bi–Lipschitz maps. The Banach-Mazur distance.** A map $f : X \to Y$ is $L$*−bi-Lipschitz* if it is a bijection and both $f$ and $f^{-1}$ are $L$-Lipschitz for some $L$; equivalently, $f$ is surjective and there exists a constant $L \geqslant 1$ such that for every $x, x' \in X$

$$\frac{1}{L}\mathrm{dist}_X(x, x') \leqslant \mathrm{dist}_Y(f(x), f(x')) \leqslant L\mathrm{dist}_X(x, x')\,.$$

A *bi-Lipschitz embedding* is defined by dropping surjectivity assumption.

EXAMPLE 1.47. Suppose that $X, Y$ are connected Riemannian manifolds $(M, g)$, $(N, h)$ (see Section 2.1.3). Then a diffeomorphism $f : M \to N$ is $L$-bi-Lipschitz if and only if

$$L^{-1} \leqslant \sqrt{\frac{f^*h}{g}} \leqslant L.$$

In other words, for every tangent vector $v \in TM$,

$$L^{-1} \leqslant \frac{|df(v)|}{|v|} \leqslant L.$$

If there exists a bi-Lipschitz map $f : X \to Y$, the metric spaces $(X, \mathrm{dist}_X)$ and $(Y, \mathrm{dist}_Y)$ are called *bi-Lipschitz equivalent* or *bi-Lipschitz homeomorphic*. If $\mathrm{dist}_1$ and $\mathrm{dist}_2$ are two distances on the same metric space $X$ such that the identity map $\mathrm{id} : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$ is bi-Lipschitz, then we say that $\mathrm{dist}_1$ and $\mathrm{dist}_2$ are *bi-Lipschitz equivalent*.

EXAMPLES 1.48.

(1) If $d_1, d_2$ are metrics on $\mathbb{R}^n$ defined by two norms on $\mathbb{R}^n$, then $d_1, d_2$ are bi-Lipschitz equivalent.

(2) Two left-invariant Riemannian metrics on a connected real Lie group define bi-Lipschitz equivalent distance functions.

For a Lipschitz function $f : X \to \mathbb{R}$ let $\mathrm{Lip}(f)$ denote

$$\mathrm{Lip}(f) := \inf\{L : f \text{ is } L\text{–Lipschitz}\} \tag{1.8}$$

EXAMPLE 1.49. If $T : V \to W$ is a continuous linear map between Banach spaces, then

$$Lip(T) = \|T\|,$$

the operator norm of $T$.

The *Banach-Mazur distance* $\mathrm{dist}_{BM}(V,W)$ between two Banach spaces $V$ and $W$ is

$$\log\left(\inf_{T:V\to W}\left(\|T\|\cdot\|T^{-1}\|\right)\right),$$

where the infimum is taken over all invertible linear maps $T:V\to W$.

THEOREM 1.50 (John's Theorem, see e.g. **[Ver11]**, Theorem 2.1). *For every pair of $n$-dimensional normed vector spaces $V,W$,* $\mathrm{dist}_{BM}(V,W)\leqslant\log(n)$.

EXERCISE 1.51. Suppose that $f,g$ are Lipschitz functions on $X$. Let $\|f\|,\|g\|$ denote the sup-norms of $f$ and $g$ on $X$. Show that
1.$\mathrm{Lip}(f+g)\leqslant\mathrm{Lip}(f)+\mathrm{Lip}(g)$.
2. $\mathrm{Lip}(fg)\leqslant\mathrm{Lip}(f)\|g\|+\mathrm{Lip}(g)\|f\|$.
3.

$$\mathrm{Lip}\left(\frac{f}{g}\right)\leqslant\frac{\mathrm{Lip}(f)\|g\|+\mathrm{Lip}(g)\|f\|}{\inf_{x\in X}g^2(x)}.$$

Note that in case when $f$ is a smooth function on a Riemannian manifold, these formulae follow from the formulae for the derivatives of the sum, product and ratio of two functions.

## 1.6. Hausdorff dimension

We recall the concept of *Hausdorff dimension* for metric spaces. Let $K$ be a metric space and $\alpha>0$. The *$\alpha$–Hausdorff measure* $\mu_\alpha(K)$ is defined as

$$\lim_{r\to 0}\inf\sum_{i=1}^{N}r_i^{\alpha},\tag{1.9}$$

where the infimum is taken over all countable coverings of $K$ by balls $B(x_i,r_i)$, $r_i\leqslant r$ $(i=1,\dots,N)$. The motivation for this definition is that the volume of the Euclidean $r$-ball of dimension $a\in\mathbb{N}$ is $r^a$ (up to a uniform constant); hence, Lebesgue measure of a subset of $\mathbb{R}^a$ is (up to a uniform constant) estimated from above by the $a$-Hausdorff measure. Euclidean spaces, of course, have integer dimension, the point of Hausdorff measure and dimension is to extend the definition to the non-integer case.

The *Hausdorff dimension* of the metric space $K$ is defined as:

$$\dim_H(K):=\inf\{\alpha:\mu_\alpha(K)=0\}.$$

EXERCISE 1.52. Verify that the Hausdorff dimension of the Euclidean space $\mathbb{R}^n$ is $n$.

We will need the following theorem:

THEOREM 1.53 (L. Sznirelman; see also **[HW41]**). *Suppose that $X$ is a proper metric space; then the covering dimension* $\dim(X)$ *is at most the Hausdorff dimension* $\dim_H(X)$.

Let $A\subset X$ be a closed subset. Let $B^n:=\bar{B}(0,1)\subset\mathbb{R}^n$ denote the closed unit ball in $\mathbb{R}^n$. Define

$$C(X,A;B^n):=\{f:X\to B^n\ ;\ f(A)\subset S^{n-1}=\partial B^n\}.$$

An immediate consequence of Proposition 1.46 is the following.

Corollary 1.54. *For every function $f \in C(X, A; B^n)$ and an open set $U \subset X$ containing $A$, there exists a sequence of functions $g_i \in C(X, A; B^n)$ so that for all $i \in \mathbb{N}$:*

1. *$g_i|A = f|A$.*
2. *$g_i \in \mathrm{Lip}(X \setminus U; \mathbb{R}^n)$.*

For a continuous map $f : X \to B^n$ define $A = A_f$ as

$$A := f^{-1}(S^{n-1}).$$

Definition 1.55. The map $f$ is *essential* if it is homotopic rel. $A$ to a map $f' : X \to S^{n-1}$. An *inessential map* is the one which is not essential.

We will be using the following characterization of the covering dimension due to Alexandrov:

Theorem 1.56 (P. S. Alexandrov, see Theorem III.5 in [**Nag83**]). $\dim(X) < n$ *if and only if every continuous map $f : X \to B^n$ is inessential.*

We are now ready to prove Theorem 1.53. Suppose that $\dim_H(X) < n$. We will prove that $\dim(X) < n$ as well. We need to show that every continuous map $f : X \to B^n$ is inessential. Let $D$ denote the annulus $\{x \in \mathbb{R}^n : 1/2 \leqslant |x| < 1\}$. Set $A := f^{-1}(S^{n-1})$ and $U := f^{-1}(D)$.

Take the sequence $g_i$ given by Corollary 1.54. Since each $g_i$ is homotopic to $f$ rel. $A$, it suffices to show that some $g_i$ is inessential. Since $f = \lim_i g_i$, it follows that for all sufficiently large $i$,

$$g_i(U) \cap B\left(0, \frac{1}{3}\right) = \emptyset.$$

We claim that the image of every such $g_i$ misses a point in $B\left(0, \frac{1}{3}\right)$. Indeed, since $\dim_H(X) < n$, the $n$-dimensional Hausdorff measure of $X$ is zero. However, $g_i|X \setminus U$ is locally Lipschitz. Therefore $g_i(X \setminus U)$ has zero $n$-dimensional Hausdorff (and hence Lebesgue) measure. It follows that $g_i(X)$ misses a point $y$ in $B\left(0, \frac{1}{3}\right)$. Composing $g_i$ with the retraction $B^n \setminus \{y\} \to S^{n-1}$ we get a map $f' : X \to S^{n-1}$ which is homotopic to $f$ rel. $A$. Thus $f$ is inessential and, therefore, $\dim(X) < n$. □

## 1.7. Norms and valuations

In this and the following section we describe certain metric spaces of algebraic origin that will be used in the proof of the Tits alternative.

A *norm* on a ring $R$ is a function $|\cdot|$ from $R$ to $\mathbb{R}_+$, which satisfies the following axioms:

1. $|x| = 0 \iff x = 0$.
2. $|xy| = |x| \cdot |y|$.
3. $|x + y| \leqslant |x| + |y|$.

An element $x \in R$ such that $|x| = 1$ is called a *unit*.

We will say that a norm $|\cdot|$ is *nonarchimedean* if it satisfies the *ultrametric* inequality

$$|x + y| \leqslant \max(|x|, |y|).$$

We say that $|\cdot|$ is *archimedean* if there exists an isometric monomorphism $R \hookrightarrow \mathbb{C}$. We will be primarily interested in normed archimedean fields which are $\mathbb{R}$ and $\mathbb{C}$

with the usual norms given by the absolute value. (By a theorem of Gelfand–Tornheim, if a normed field $F$ contains $\mathbb{R}$ as subfield then $F$ is isomorphic, as a field, either to $\mathbb{R}$ or to $\mathbb{C}$.)

Below is an alternative approach to nonarchimedean normed rings $R$. A function $\nu : R \to \mathbb{R} \cup \{\infty\}$ is called a *valuation* if it satisfies the following axioms:
1. $\nu(x) = \infty \iff x = 0$.
2. $\nu(xy) = \nu(x) + \nu(y)$.
3. $\nu(x + y) \geqslant \min(\nu(x), \nu(y))$.

Therefore, one converts a valuation to a nonarchimedean norm by setting

$$|x| = c^{-\nu(x)}, x \neq 0, \quad |0| = 0,$$

where $c > 0$ is a fixed real number.

REMARK 1.57. More generally, one also considers valuations with values in arbitrary ordered abelian groups, but we will not need this.

A normed ring $R$ is said to be *local* if it is locally compact as a metric space; a normed ring $R$ is said to be *complete* if it is complete as a metric space. A norm on a field $F$ is said to be *discrete* if the image $\Gamma$ of $|\cdot| : F \setminus \{0\} \to (0, \infty)$ is an infinite cyclic group. If the norm is discrete, then an element $\pi \in F$ such that $|\pi|$ is a generator of $\Gamma$ satisfying $|\pi| < 1$, is called a *uniformizer* of $F$. If $F$ is a field with valuation $\nu$, then the subset

$$O_\nu = \{x \in F : \nu(x) \geqslant 0\}$$

is a subring in $F$, the *valuation ring* or the *ring of integers* in $F$.

EXERCISE 1.58. 1. Verify that every nonzero element of a field $F$ with discrete norm has the form $\pi^k u$, where $u$ is a unit.

2. Verify that every discrete norm is nonarchimedean.

Below are the two main examples of fields with discrete norms:

1. Field $\mathbb{Q}_p$ of $p$-adic numbers. Fix a prime number $p$. For each number $x = q/p^n \in \mathbb{Q}$ (where both numerator and denominator of $q$ are not divisible by $p$) set $|x|_p := p^n$. Then $|\cdot|_p$ is a nonarchimedean norm on $\mathbb{Q}$, called the *$p$-adic norm.* The completion of $\mathbb{Q}$ with respect to the $p$-adic norm is the field of $p$-adic numbers $\mathbb{Q}_p$. The ring of $p$-adic integers $O_p$ intersects $\mathbb{Q}$ along the subset consisting of (reduced) fractions $\frac{n}{m}$ where $m, n \in Z$ and $m$ is not divisible by $p$. Note that $p$ is a uniformizer of $\mathbb{Q}_p$.

REMARK 1.59. We will not use the common notation $\mathbb{Z}_p$ for $O_p$, in order to avoid the confusion with finite cyclic groups.

EXERCISE 1.60. Verify that $O_p$ is open in $\mathbb{Q}_p$. Hint: Use the fact that $|x+y|_p \leqslant 1$ provided that $|x|_p \leqslant 1, |y_p| \leqslant 1$.

Recall that one can describe real numbers using infinite decimal sequences. There is a similar description of $p$-adic numbers using "base $p$ arithmetic." Namely, we can identify $p$-adic numbers with semi-infinite Laurent series

$$\sum_{k=-n}^{\infty} a_k p^k,$$

where $n \in \mathbb{Z}$ and $a_k \in \{0, \dots, p-1\}$. Operations of addition and multiplication here are the usual operations with power series where we treat $p$ as a formal variable, the only difference is that we still have to "carry to the right" as in the usual decimal arithmetic.

With this identification, $|x|_p = p^n$, where $a_{-n}$ is the first nonzero coefficient in the power series. In other words, $\nu(x) = -n$ is the valuation. In particular, the ring $O_p$ is identified with the set of series

$$\sum_{k=0}^{\infty} a_k p^k.$$

REMARK 1.61. In other words, one can describe $p$-adic numbers as left-infinite sequences of (base $p$) digits

$$\cdots a_m a_{m-1} \dots a_0 . a_{-1} \cdots a_{-n}$$

where $\forall i, a_i \in \{0, \dots, p-1\}$, and the algebraic operations require "carrying to the left" instead of carrying to the right.

EXERCISE 1.62. Show that in $\mathbb{Q}_p$,

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}.$$

2. Let $A$ be a field. Consider the ring $R = A[t, t^{-1}]$ of Laurent polynomials

$$f(t) = \sum_{k=n}^{m} a_k t^k.$$

Set $\nu(0) = \infty$ and for nonzero $f$ let $\nu(f)$ be the least $n$ so that $a_n \neq 0$. In other words, $\nu(f)$ is the order of vanishing of $f$ at $0 \in R$.

EXERCISE 1.63. 1. Verify that $\nu$ is a valuation on $R$. Define $|f| := e^{-\nu(f)}$.

2. Verify that the completion $\widehat{R}$ of $R$ with respect to the above norm is naturally isomorphic to the ring of semi-infinite formal Laurent series

$$f = \sum_{k=n}^{\infty} a_k t^k,$$

where $\nu(f)$ is the minimal $n$ such that $a_n \neq 0$.

Let $A(t)$ be the field of rational functions in the variable $t$. We embed $A$ in $\widehat{R}$ by the rule

$$\frac{1}{1-at} = 1 + \sum_{n=1}^{\infty} a^n t^n.$$

If $A$ is algebraically closed, every rational function is a product of a polynomial function and several functions of the form

$$\frac{1}{a_i - t},$$

so we obtain an embedding $A(t) \hookrightarrow \widehat{R}$ in this case. If $A$ is not algebraically closed, proceed as follows. First, construct, as above, an embedding $\iota$ of $\bar{A}(t)$ to the completion of $\bar{A}[t, t^{-1}]$, where $\bar{A}$ is the algebraic closure of $A$. Next, observe that this embedding is equivariant with respect to the Galois group $Gal(\bar{A}/A)$, where $\sigma \in Gal(\bar{A}/A)$ acts on Laurent series

$$f = \sum_{k=n}^{\infty} a_k t^k, a \in \bar{A},$$

by

$$f^\sigma = \sum_{k=n}^{\infty} a_k^\sigma t^k.$$

Therefore, $\iota(A(t)) \subset \widehat{R}, R = A[t, t^{-1}]$.

In any case, we obtain a norm on $A(t)$ by restricting the norm in $\widehat{R}$. Since $R \subset \iota A(t)$, it follows that $\widehat{R}$ is the completion of $\iota A(t)$. In particular, $\widehat{R}$ is a complete normed field.

EXERCISE 1.64. 1. Verify that $\widehat{R}$ is local if and only if $A$ is finite.

2. Show that $t$ is a uniformizer of $\widehat{R}$.

3. At the first glance, it looks like $\mathbb{Q}_p$ is the same as $\widehat{R}$ for $A = \mathbb{Z}_p$, since elements of both are described using formal power series with coefficients in $\{0, \ldots, p-1\}$. What is the difference between these fields?

LEMMA 1.65. *$\mathbb{Q}_p$ is a local field.*

PROOF. It suffices to show that the ring $O_p$ of $p$-adic integers is compact. Since $\mathbb{Q}_p$ is complete, it suffices to show that $O_p$ is closed and totally bounded, i.e., for every $\epsilon > 0$, $O_p$ has a finite cover by closed $\epsilon$-balls. The fact that $O_p$ is closed follows from the fact that $|\cdot|_p : \mathbb{Q}_p \to \mathbb{R}$ is continuous and $O_p$ is given by the inequality $O_p = \{x : |x|_p \leqslant 1\}$.

Let us check that $O_p$ is totally bounded. For $\epsilon > 0$ pick $k \in \mathbb{N}$ such that $p^{-k} < \epsilon$. The ring $\mathbb{Z}/p^k\mathbb{Z}$ is finite, let $z_1, \ldots, z_N \in \mathbb{Z} \setminus \{0\}$ (where $N = p^k$) denote representatives of the cosets in $\mathbb{Z}/p^k\mathbb{Z}$. We claim that the set of fractions

$$w_{ij} = \frac{z_i}{z_j}, 1 \leqslant i, j \leqslant N,$$

forms a $p^{-k}$-net in $O_p \cap \mathbb{Q}$. Indeed, for a rational number $\frac{m}{n} \in O_p \cap \mathbb{Q}$, find $s, t \in \{z_1, \ldots, z_N\}$ such that

$$s \equiv m, t \equiv n, \mod p^k.$$

Then

$$\frac{m}{n} - \frac{s}{t} \in p^k O_p$$

and, hence,

$$\left| \frac{m}{n} - \frac{s}{t} \right|_p \leqslant p^{-k}.$$

Since $O_p \cap \mathbb{Q}$ is dense in $O_p$, it follows that

$$O_p \subset \bigcup_{i,j=1}^{N} \bar{B}\left(w_{ij}, \epsilon\right). \quad \square$$

EXERCISE 1.66. Show that $O_p$ is homeomorphic to the Cantor set. Hint: Verify that $O_p$ is totally disconnected and perfect.

### 1.8. Metrics on affine and projective spaces

In this section we will use normed fields to define metrics on affine and projective spaces. Consider the vector space $V = F^n$ over a normed field $F$, with the standard basis $e_1, \ldots, e_n$. We equip $V$ with the usual Euclidean/hermitian norm in the case $F$ is archimedean and with the max-norm

$$|(x_1, \ldots, x_n)| = \max_i |x_i|$$

if $F$ is nonarchimedean. We let $\langle \cdot, \cdot \rangle$ denote the standard inner/hermitian product on $V$ in the archimedean case.

Exercise 1.67. Suppose that $F$ is nonarchimedean. Show that the metric $|v - w|$ on $V$ satisfies the ultrametric triangle inequality.

If $F$ is nonarchimedean, define the group $K = GL(n, O)$, consisting of matrices $A$ such that $A, A^{-1} \in Mat_n(O)$.

Exercise 1.68. If $F$ is a nonarchimedean local field, show that the group $K$ is compact with respect to the subset topology induced from $Mat_n(F) = F^{n^2}$.

Lemma 1.69. *The group $K$ acts isometrically on $V$.*

Proof. It suffices to show that elements $g \in K$ do not increase the norm on $V$. Let $a_{ij}$ denote the matrix coefficients of $g$. Then, for a vector $v = \sum_i v_i e_i \in V$, the vector $w = g(v)$ has coordinates

$$w_j = \sum_i a_{ji} v_i.$$

Since $|a_{ij}| \leqslant 1$, the ultrametric inequality implies

$$|w| = \max_j |w_j|, \quad |w_j| \leqslant \max_i |a_{ji} v_i| \leqslant |v|.$$

Thus, $|g(v)| \leqslant |v|$. □

If $F$ is archimedean, we let $K < GL(V)$ denote the orthogonal/hermitian subgroup preserving the inner/hermitian product on $V$. The following is a standard fact from the elementary linear algebra:

Theorem 1.70 (Singular Value Decomposition Theorem). *If $F$ is archimedean, then every matrix $M \in End(V)$ admits a* singular valued decomposition

$$M = UDV,$$

*where $U, V \in K$ and $D$ is a diagonal matrix with nonnegative entries arranged in the descending order. The diagonal entries of $D$ are called the* singular values *of $M$.*

We will now prove an analogue of the singular value decomposition in the case of nonarchimedean normed fields:

Theorem 1.71 (Smith Normal Form Theorem). *Let $F$ be a field with discrete norm and uniformizer $\pi$ and ring of integers $O$. Then every matrix $M \in Mat_n(F)$ admits a* Smith Normal Form *decomposition*

$$M = LDU,$$

*where $D$ is diagonal with diagonal entries $(d_1, \ldots, d_n)$, $d_i = \pi^{k_i}$, $i = 1, \ldots, n$,*

$$k_1 \geqslant k_2 \geqslant \ldots \geqslant k_n,$$

*and* $L, U \in K = GL(n, O)$. *The diagonal entries* $d_i \in F$ *are called the* invariant factors *of* $M$.

Proof. First, note that permutation matrices belong to $K$; the group $K$ also contains upper and lower triangular matrices with coefficients in $O$, whose diagonal entries are units in $F$. We now apply Gauss Elimination Algorithm to the matrix $M$. Note that the row operation of adding the $z$-multiple of the $i$-th row to the $j$-th row amounts to multiplication on the left with the lower-triangular elementary matrix $E_{ij}(z)$ with the $ij$-entry equal $z$. If $z \in O$, then $E_{ij} \in K$. Similarly, column operations amount to multiplication on the right by an upper-triangular elementary matrix. Observe also that dividing a row (column) by a unit in $F$ amounts to multiplying a matrix on left (right) by an appropriate diagonal matrix with unit entries on the diagonal.

We now describe row operations for the Gauss Elimination in detail (column operations will be similar). Consider (nonzero) $i$-th column of a matrix $A \in End(F^n)$. We first multiply $M$ on left and right by permutation matrices so that $a_{ii}$ has the largest norm in the $i$-th column. By dividing rows on $A$ by units in $F$, we achieve that every entry in the $i$-th column is a power of $\pi$. Now, eliminating nonzero entries in the $i$-th column will require only row operations involving $\pi^{s_{ij}}$-multiples of the $i$-th row, where $s_{ij} \geqslant 0$, i.e., $\pi^{s_{ij}} \in O$. Applying this form of Gauss Algorithm to $M$, we convert $M$ to a diagonal matrix $A$, whose diagonal entries are powers of $\pi$ and

$$A = L'MU', \quad L', M' \in GL(n, O).$$

Multiplying $A$ on left and right by permutation matrices, we rearrange the diagonal entries to have weakly decreasing exponents. □

Note that both singular value decomposition and Smith normal form decomposition both have the form:

$$M = UDV, \quad U, V \in K,$$

and $D$ is diagonal. Such decomposition of the $Mat_n(F)$ is called the *Cartan decomposition*. To simplify the terminology, we will refer to the diagonal entries of $D$ as *singular values* of $M$ in both archimedean and nonarchimedean cases.

Exercise 1.72. Deduce the Cartan decomposition in $F = \mathbb{R}$ or $F = \mathbb{C}$, from the statement that given any Euclidean/hermitian bilinear form $q$ on $V = F^n$, there exists a basis orthogonal with respect to $q$ and orthonormal with respect to the standard inner product

$$x_1\overline{y}_1 + \ldots + x_n\overline{y}_n.$$

We now turn our discussion to projective spaces. The $F$-projective space $P = F\mathbb{P}^n$ is the quotient of $F^{n+1} \setminus \{0\}$ by the action of $F^\times$ *via* scalar multiplication. We let $[v]$ denote the projection of a nonzero vector $v \in V = F^{n+1}$ to $F\mathbb{P}^n$. The $j$-th affine coordinate patch on $P$ is the affine subspace $A_j \subset V$,

$$A_j = (x_1, \ldots, 1, \ldots, x_{n+1}),$$

where 1 appears in the $j$-th coordinate.

Notation 1.73. Given a nonzero vector $v \in V$ let $[v]$ denote the projection of $v$ to the projective space $\mathbb{P}(V)$; similarly, for a subset $W \subset V$ we let $[W]$ denote the image of $W \setminus \{0\}$ under the canonical projection $V \to \mathbb{P}(V)$. Given an invertible linear map $g : V \to V$, we will retain the notation $g$ for the induced projective map $\mathbb{P}(V) \to \mathbb{P}(V)$.

Suppose now that $F$ is a normed field. Our next goal is to define the *chordal metric* on $F\mathbb{P}^n$. In the case of an archimedean field $F$, we define the Euclidean or hermitian norm on $V \wedge V$ by declaring basis vectors

$$e_i \wedge e_j, 1 \leqslant i < j \leqslant n+1$$

to be orthonormal. Then

$$|v \wedge w|^2 = |v|^2 |w|^2 - \langle v, w\rangle \langle w, v\rangle.$$

Note that if $u$, $v$ are unit vectors with $\angle(v, w) = \varphi$, then $|v \wedge w| = |\sin(\varphi)|$.

In the case when $F$ is nonarchimedean, we equip $V \wedge V$ with the max-norm so that

$$|v \wedge w| = \max_{i,j} |x_i y_j - x_j y_i|$$

where $v = (x_1, \ldots, x_{n+1})$, $w = (y_1, \ldots, y_{n+1})$.

LEMMA 1.74. *Suppose that $u$ is a unit vector and $v \in V$ is such that $|u_i - v_i| \leqslant \epsilon$ for all $i$. Then*

$$|v \wedge w| \leqslant 2(n+1)\epsilon.$$

PROOF. We will consider the archimedean case since the nonarchimedean case is similar. For every $i$ let $\delta_i = v_i - u_i$. Then

$$|u_i v_j - u_j v_i|^2 \leqslant |u_i \delta_j - u_j \delta_i|^2 \leqslant 4\epsilon^2$$

Thus,

$$|u \wedge v|^2 \leqslant 4(n+1)^2 \epsilon^2. \quad \square$$

DEFINITION 1.75. The *chordal metric* on $P = F\mathbb{P}^n$ is defined by

$$d([v], [w]) = \frac{|v \wedge w|}{|v| \cdot |w|}.$$

In the nonarchimedean case this definition is due to A. Néron **[N\'64]**.

EXERCISE 1.76. 1. If $F$ is nonarchimedean, show that the group $GL(n+1, O)$ preserves the chordal metric.

2. If $F = \mathbb{R}$, show that the orthogonal group preserves the chordal metric.

3. If $F = \mathbb{C}$, show that the unitary group preserves the chordal metric.

It is clear that $d(\lambda v, \mu w) = d(v, w)$ for all nonzero scalars $\lambda, \mu$ and nonzero vectors $v, w$. It is also clear that $d(v, w) = d(w, v)$ and $d(v, w) = 0$ if and only if $[v] = [w]$. What is not so obvious is why $d$ satisfies the triangle inequality. Note, however, that in the case of a nonarchimedean field $F$,

$$d([v], [w]) \leqslant 1$$

for all $[v], [w] \in P$. Indeed, pick unit vectors $v, w$ representing $[v], [w]$; in particular, $v_i, w_j$ belong to $O$ for all $i, j$. Then, the denominator in the definition of $d([v], [w])$ equals 1, while the numerator is $\leqslant 1$, since $O$ is a ring.

PROPOSITION 1.77. *If $F$ is nonarchimedean, then $d$ satisfies the triangle inequality.*

Proof. We will verify the triangle inequality by giving an alternative description of the function $d$. We define *affine patches* on $P$ to be the affine hyperplanes

$$A_j = \{x \in V : x_j = 1\} \subset V$$

together with the (injective) projections $A_j \to P$. Every affine patch is, of course, just a translate of $F^n$, so that $e_j$ is the translate of the origin. We, then, equip $A_j$ with the restriction of the metric $|v - w|$ from $V$. Let $B_j \subset A_j$ denote the closed unit ball centered at $e_j$. In other words,

$$B_j = A_j \cap O^{n+1}.$$

We now set $d_j(x, y) = |x - y|$ if $x, y \in B_j$ and $d_j(x, y) = 1$ otherwise. It follows immediately from the ultrametric triangle inequality that $d_j$ is a metric. We, then, define for $[x], [y] \in P$ the function $\operatorname{dist}([x], [y])$ by:

1. If there exists $j$ so that $x, y \in B_j$ project to $[x], [y]$, then $\operatorname{dist}([x], [y]) := d_j(x, y)$.

2. Otherwise, set $\operatorname{dist}([x], [y]) = 1$.

If we knew that dist is well-defined (*a priori*, different indices $j$ give different values of dist), it would be clear that dist satisfies the ultrametric triangle inequality. Proposition will, now, follow from

Lemma 1.78. *$d([x], [y]) = \operatorname{dist}([x], [y])$ for all points in $P$.*

Proof. The proof will break in two cases:

1. There exists $k$ such that $[x], [y]$ lift to $x, y \in B_k$. To simplify the notation, we will assume that $k = n + 1$. Since $x, y \in B_{n+1}$, $|x_i| \leqslant 1, |y_i| \leqslant 1$ for all $i$, and $x_{n+1} = y_{n+1} = 1$. In particular, $|x| = |y| = 1$. Hence, for every $i$,

$$|x_i - y_i| = |x_i y_{n+1} - x_j y_{n+1}| \leqslant \max_j |x_i y_j - x_j y_i| \leqslant d([x], [y]),$$

which implies that

$$\operatorname{dist}([x], [y]) \leqslant d([x], [y]).$$

We will now prove the opposite inequality:

$$\forall i, j \quad |x_i y_j - x_j y_i| \leqslant a := |x - y|.$$

There exist $z_i, z_j \in F$ so that

$$y_i = x_i(1 + z_i), \quad y_j = x_j(1 + z_j),$$

where, if $x_i \neq 0, x_j \neq 0$,

$$z_i = \frac{y_i - x_i}{x_i}, \quad z_j = \frac{y_j - x_j}{x_j}.$$

We will consider the case $x_i x_j \neq 0$, leaving the exceptional cases to the reader. Then,

$$|z_i| \leqslant \frac{a}{|x_i|}, \quad |z_i| \leqslant \frac{a}{|x_j|}.$$

Computing $x_i y_j - x_j y_i$ using the new variables $z_i, z_j$, we obtain:

$$|x_i y_j - x_j y_i| = |x_i x_j (1 + z_j) - x_i x_j (1 + z_i)| = |x_j x_j (z_j - z_i)| \leqslant$$

$$|x_i x_j| \max(|z_i|, |z_i|) \leqslant |x_i x_j| \max\left(\frac{a}{|x_i|}, \frac{a}{|x_j|}\right) \leqslant a \max(|x_i|, |x_j|) \leqslant a,$$

since $x_i, x_j \in O$.

2. Suppose that (1) does not happen. Since $d([x],[y]) \leqslant 1$ and $\operatorname{dist}([x],[y]) = 1$ (in the second case), we just have to prove that

$$d([x],[y]) \geqslant 1.$$

Consider representatives $x, y$ of points $[x], [y]$ and let $i, j$ be the indices such that

$$|x_i| = |x|, \quad |y_j| = |y|.$$

Clearly, $i, j$ are independent of the choices of the vectors $x, y$ representing $[x], [y]$. Therefore, we choose $x$ so that $x_i = 1$, which implies that $x_k \in O$ for all $k$. If $y_i = 0$ then

$$|x_i y_j - x_j y_i| = |y_j|$$

and

$$d([x],[y]) \geqslant \frac{\max_j |1 \cdot y_j|}{|y_j|} = 1.$$

Thus, we assume that $y_i \neq 0$. This allows us to choose $y \in A_i$ as well. Since (1) does not occur, $y \notin O^{n+1}$, which implies that $|y_j| > 1$. Now,

$$d([x],[y]) \geqslant \frac{|x_i y_j - x_j y_i|}{|x_i| \cdot |y_j|} = \frac{|y_j - x_j|}{|y_j|}.$$

Since $x_j \in O$ and $y_j \notin O$, the ultrametric inequality implies that $|y_j - x_j| = |y_j|$. Therefore,

$$\frac{|y_j - x_j|}{|y_j|} = \frac{|y_j|}{|y_j|} = 1$$

and $d([x],[y]) \geqslant 1$. This concludes the proof of lemma and proposition. □

We now consider real and complex projective spaces. Choosing unit vectors $u, v$ as representatives of points $[u], [v] \in P$, we get:

$$d([u],[v]) = \sin(\angle(u,v)),$$

where we normalize the angle to be in the interval $[0, \pi]$. Consider now three points $[u], [v], [w] \in P$; our goal is to verify the triangle inequality

$$d([u],[w]) \leqslant d([u],[v]) + d([v],[w]).$$

We choose unit vectors $u, v, w$ representing these points so that

$$0 \leqslant \alpha = \angle(u,v) \leqslant \frac{\pi}{2}, \quad 0 \leqslant \beta = \angle(v,w) \leqslant \frac{\pi}{2}.$$

Then,

$$\gamma = \angle(u,w) \leqslant \alpha + \beta$$

and the triangle inequality for the metric $d$ is equivalent to the inequality

$$\sin(\gamma) \leqslant \sin(\alpha) + \sin(\beta).$$

We leave verification of the last inequality as an exercise to the reader. Thus, we obtain

THEOREM 1.79. *Chordal metric is a metric on $P$ in both archimedean and nonarchimedean cases.*

EXERCISE 1.80. Suppose that $F$ is a normed field (either nonarchimedean or archimedean).

1. Verify that metric $d$ determines the topology on $P$ which is the quotient topology induced from $V \setminus \{0\}$.
2. Assuming that $F$ is local, verify that $P$ is compact.
3. If the norm on $F$ is complete, show that the metric space $(P, d)$ is complete.
4. If $H$ is a hyperplane in $V = F^{n+1}$, given as $\operatorname{Ker} f$, where $f : V \to F$ is a linear function, show that

$$\operatorname{dist}([v], [H]) = \frac{|f(v)|}{\|v\| \, \|f\|} .$$

### 1.9. Kernels and distance functions

A *kernel* on a set $X$ is a symmetric map $\psi : X \times X \to \mathbb{R}_+$ such that $\psi(x, x) = 0$. Fix $p \in X$ and define the associated *Gromov kernel*

$$k(x, y) := \frac{1}{2} \left(\psi(x, p) + \psi(p, y) - \psi(x, y)\right) .$$

If $X$ were a metric space and $\psi(x, y) = \operatorname{dist}^2(x, y)$, then this quantity is just the Gromov product in $X$ where distances are replaced by their squares (see Section 9.3 for the definition of Gromov product in metric spaces). Clearly,

$$\forall x \in X, \quad k(x, x) = \psi(x, p).$$

DEFINITION 1.81. 1. A kernel $\psi$ is *positive semidefinite* if for every natural number $n$, every subset $\{x_1, \ldots, x_n\} \subset X$ and every vector $\lambda \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \psi(x_i, x_j) \geqslant 0 . \tag{1.10}$$

2. A kernel $\psi$ is *conditionally negative semidefinite* if for every $n \in \mathbb{N}$, every subset $\{x_1, \ldots, x_n\} \subset X$ and every vector $\lambda \in \mathbb{R}^n$ with $\sum_{i=1}^{n} \lambda_i = 0$, the following holds:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \psi(x_i, x_j) \leqslant 0 . \tag{1.11}$$

This is not a particularly transparent definition. A better way to think about this definition is in terms of the vector space $V = V(X)$ of consisting of functions with finite support $X \to \mathbb{R}$. Then each kernel $\psi$ on $X$ defines a symmetric bilinear form on $V$ (denoted $\Psi$):

$$\Psi(f, g) = \sum_{x, y \in X} \psi(x, y) f(x) g(y).$$

With this notation, the left hand side of (1.10) becomes simply $\Psi(f, f)$, where

$$\lambda_i := f(x_i), \quad \operatorname{Supp}(f) \subset \{x_1, \ldots, x_n\} \subset X.$$

Thus, a kernel is positive semidefinite if and only if $\Psi$ is a positive semidefinite bilinear form. Similarly, $\psi$ is conditionally negative semidefinite if and only if the restriction of $-\Psi$ to the subspace $V_0$ consisting of functions with zero average, is a positive semidefinite bilinear form.

NOTATION 1.82. We will use the lower case letters to denote kernels and the corresponding upper case letters to denote the associated bilinear forms on $V$.

Below is yet another interpretation of the conditionally negative semidefinite kernels. For a subset $\{x_1, \ldots, x_n\} \subset X$ define the symmetric matrix $M$ with the entries

$$m_{ij} = -\psi(x_i, x_j), \quad 1 \leqslant i, j \leqslant n.$$

For $\lambda = (\lambda_1, \ldots, \lambda_n)$, the left hand-side of the inequality (1.11) equals

$$q(\lambda) = \lambda^T M \lambda,$$

a symmetric bilinear form on $\mathbb{R}^n$. Then, the condition (1.11) means that $q$ is positive semi-definite on the hyperplane

$$\sum_{i=1}^{n} \lambda_i = 0$$

in $\mathbb{R}^n$. Suppose, for a moment, that this form is actually positive-definite, Since $\psi(x_i, x_j) \geqslant 0$, it follows that the form $q$ on $\mathbb{R}^n$ has signature $(n-1, 1)$. The standard basis vectors $e_1, \ldots, e_n$ in $\mathbb{R}^n$ are null-vectors for $q$; the condition $m_{ij} \leqslant 0$ amounts to the requirement that these vectors belong to the same, say, positive, light cone.

The following theorem gives yet another interpretation of conditionally negative semidefinite kernels in terms of embedding in Hilbert spaces. It was first proven by J. Schoenberg in [**Sch38**] in the case of finite sets, but the same proof works for infinite sets as well.

THEOREM 1.83. *A kernel $\psi$ on $X$ is conditionally negative definite if and only if there exists a map $F : X \to \mathcal{H}$ to a Hilbert space so that*

$$\psi(x, y) = \|F(x) - F(y)\|^2.$$

PROOF. 1. Suppose that the map $F$ exists. Then, for every $p = x_0 \in X$, the associated Gromov kernel $k(x, y)$ equals

$$k(x, y) = \langle F(x), F(y) \rangle,$$

and, hence, for every finite subset $\{x_0, x_1, \ldots, x_n\} \subset X$, the corresponding matrix with the entries $k(x_i, x_j)$ is the Gramm matrix of the set

$$\{y_i := F(x_i) - F(x_0) : i = 1, \ldots, n\} \subset \mathcal{H}.$$

Hence, this matrix is positive semidefinite. Accordingly, Gromov kernel determines a positive semidefinite bilinear form on the vector space $V = V(X)$.

We will verify that $\psi$ is conditionally negative semidefinite by considering subsets $X_0$ in $X$ of the form $\{x_0, x_1, \ldots, x_n\}$. (Since the point $x_0$ was arbitrary, this will suffice.)

Let $f : X_0 \to \mathbb{R}$ be such that

$$\sum_{i=0}^{n} f(x_i) = 0. \tag{1.12}$$

Thus,

$$f(x_0) := -\sum_{i=1}^{n} f(x_i).$$

Set $y_i := F(x_i), i = 0, \ldots, n$. Since the kernel $K$ is positive semidefinite, we have

$$\sum_{i,j=1}^{n} \left(|y_0 - y_i|^2 + |y_0 - y_j|^2 - |y_i - y_j|^2\right) f(x_i) f(x_j) = \tag{1.13}$$

$$2\sum_{i,j=1}^{n} k(x_i,x_j)f(x_i)f(x_j) \geqslant 0.$$

The left hand side of this equation equals

$$2\left(\sum_{i=1}^{n} f(x_i)\right)\cdot\left(\sum_{j=1}^{n}|y_0-y_j|^2 f(x_j)\right) -$$

$$\sum_{i,j=1}^{n}|y_i-y_j|^2 f(x_i)f(x_j).$$

Since $f(x_0) := -\sum_{i=1}^{n} f(x_i)$, we can rewrite this expression as

$$-f(x_0)^2|y_0-y_0|^2 - 2\left(\sum_{j=1}^{n}|y_0-y_j|^2 f(x_0)f(x_j)\right) - \sum_{i,j=1}^{n}|y_i-y_j|^2 f(x_i)f(x_j) =$$

$$\sum_{i,j=0}^{n}|y_i-y_j|^2 f(x_i)f(x_j) = \sum_{i,j=0}^{n}\psi(x_i,x_j)f(x_i)f(x_j).$$

Taking into account the inequality (1.13), we conclude that

$$\sum_{i,j=0}^{n}\psi(x_i,x_j)f(x_i)f(x_j) \leqslant 0. \tag{1.14}$$

In other words, the kernel $\psi$ on $X$ is conditionally negative semidefinite.

2. Suppose that $\psi$ is conditionally negative definite. Fix $p \in X$ and define the Gromov kernel

$$k(x,y) := (x,y)_p := \frac{1}{2}\left(\psi(x,p)+\psi(p,y)-\psi(x,y)\right).$$

The key to the proof is:

Lemma 1.84. *$k$ is a positive semidefinite kernel on $X$.*

Proof. Consider a subset $X_0 = \{x_1,\ldots,x_n\} \subset X$ and a function $f : X_0 \to \mathbb{R}$.

a. We first consider the case when $p \notin X_0$. Then we set $x_0 := p$ and extend the function $f$ to $p$ by

$$f(x_0) := -\sum_{i=1}^{n} f(x_i).$$

The resulting function $f : \{x_0,\ldots,x_n\} \to \mathbb{R}$ satisfies (1.12) and, hence,

$$\sum_{i,j=0}^{n}\psi(x_i,x_j)f(x_i)f(x_j) \leqslant 0.$$

The same argument as in the first part of the proof of Theorem 1.83 (run in the reverse) then shows that

$$\sum_{i,j=1}^{n} k(x_i,x_j)f(x_i)f(x_j) \geqslant 0.$$

Thus, $k$ is positive semidefinite on functions whose support is disjoint from $\{p\}$.

b. Suppose that $p \in X_0$, $f(p) = c \neq 0$. We define a new function $g(x) := f(x) - c\delta_p$. Here $\delta_p$ is the characteristic function of the subset $\{p\} \subset X$. Then $p \notin \mathrm{Supp}(g)$ and, hence, by the Case (a),

$$K(g, g) \geqslant 0.$$

On the other hand,

$$K(f, f) = F(g, g) + 2cK(g, \delta_p) + c^2 K(\delta_p, \delta_p) = F(g, g),$$

since the other two terms vanish (as $k(x, p) = 0$ for every $x \in X$). Thus, $K$ is positive semidefinite. □

Now, consider the vector space $V = V(X)$ equipped with the positive semi-definite bilinear form $\langle f, g\rangle = K(f, g)$. Define the Hilbert space $\mathcal{H}$ as the metric completion of

$$V/\{f \in V : \langle f, f\rangle = 0\}.$$

Then we have a natural map $F : X \to \mathcal{H}$ which sends $x \in X$ to the projection of the $\delta$-function $\delta_x$; we obtain:

$$\langle F(x), F(y)\rangle = k(x, y).$$

Let us verify now that

$$\langle F(x) - F(y), F(x) - F(y)\rangle = \psi(x, y). \tag{1.15}$$

The left hand side of this expression equals

$$\langle F(x), F(x)\rangle + \langle F(y), F(y)\rangle - 2k(x, y) = \psi(x, p) + \psi(y, p) - 2k(x, y).$$

Then, the equality (1.15) follows from the definition of the Gromov kernel $k$. □

According to **[Sch38]**, for every conditionally negative definite kernel $\psi : X \times X \to \mathbb{R}_+$ and every $0 < \alpha \leqslant 1$, the power $\psi^\alpha$ is also a conditionally negative definite kernel.

CHAPTER 2

# Geometric preliminaries

## 2.1. Differential and Riemannian geometry

In this book we will use some elementary Differential and Riemannian geometry, basics of which are reviewed in this section. All the manifolds that we consider are second countable.

**2.1.1. Smooth manifolds.** We expect the reader to know basics of differential topology, that can be found, for instance, in [**GP10**], [**Hir76**], [**War83**]. Below is only a brief review.

Recall that, given a smooth $n$–dimensional manifold $M$, a *$k$–dimensional submanifold* is a closed subset $N \subset M$ with the property that every point $p \in N$ is contained in the domain $U$ of a chart $\varphi : U \to \mathbb{R}^n$ such that $\varphi(U \cap N) = \varphi(U) \cap \mathbb{R}^k$.

If $k = n$ then, by the inverse function theorem, $N$ is an open subset in $M$; in this case $N$ is also called an *open submanifold* in $M$. (The same is true in the topological category, but the proof is harder and requires Brouwer's Invariance of Domain Theorem, see e.g. [**Hat02**], Theorem 2B.3.)

Suppose that $U \subset \mathbb{R}^n$ is an open subset. A *piecewise-smooth function* $f : U \to \mathbb{R}^m$ is a continuous function such that for every $x \in U$ there exists a neighborhood $V$ of $x$ in $U$, a diffeomorphism $\phi : V \to V' \subset \mathbb{R}^n$, a triangulation $T$ of $V'$, so that the composition

$$f \circ \phi^{-1} : (V', T) \to \mathbb{R}^m$$

is smooth on each simplex. Note that composition $g \circ f$ is again piecewise-smooth, provided that $g$ is smooth; however, composition of piecewise-smooth maps need not be piecewise-smooth.

One then defines *piecewise smooth $k$–dimensional submanifolds* $N$ of a smooth manifold $M$. Such $N$ is a topological submanifold which is locally the image of $\mathbb{R}^k$ in $\mathbb{R}^n$ under a piecewise-smooth homeomorphism $\mathbb{R}^n \to \mathbb{R}^n$. We refer the reader to [**Thu97**] for the detailed discussion of piecewise-smooth manifolds.

If $k = n - 1$ we also sometimes call a submanifold a (*piecewise smooth*) *hypersurface*.

Below we review two alternative ways of defining submanifolds. Consider a smooth map $f : M \to N$ of a $m$-dimensional manifold $M = M^m$ to an $n$-dimensional manifold $N = N^n$. The map $f : M \to N$ is called an *immersion* if for every $p \in M$, the linear map $df_p : T_pM \to T_{f(p)}N$ is injective. If, moreover, $f$ defines a homeomorphism from $M$ to $f(M)$ with the subspace topology, then $f$ is called a *smooth embedding.*

EXERCISE 2.1. Construct an injective immersion $\mathbb{R} \to \mathbb{R}^2$ which is not a smooth embedding.

If $N$ is a submanifold in $M$ then the inclusion map $i : N \to M$ is a smooth embedding. This, in fact, provides an alternative definition for $k$-dimensional submanifolds: They are images of smooth embeddings with $k$–dimensional manifolds (see Corollary 2.4). Images of immersions provide a large class of subsets, called *immersed submanifolds.*

A smooth map $f : M^k \to N^n$ is called a *submersion* if for every $p \in M$, the linear map $df_p$ is surjective. The following theorem can be found for instance, in **[GP10]**, **[Hir76]**, **[War83]**.

THEOREM 2.2. (1) *If $f : M^m \to N^n$ is an immersion, then for every $p \in M$ and $q = f(p)$ there exists a chart $\varphi : U \to \mathbb{R}^m$ of $M$ with $p \in U$, and a chart $\psi : V \to \mathbb{R}^n$ of $N$ with $q \in V$ such that $\overline{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$ is of the form*

$$\overline{f}(x_1, \dots, x_m) = (x_1, \dots, x_m, \underbrace{0, \dots, 0}_{n-m \text{ times}}).$$

(2) *If $f : M^m \to N^n$ is a submersion, then for every $p \in M$ and $q = f(p)$ there exists a chart $\varphi : U \to \mathbb{R}^m$ of $M$ with $p \in U$, and a chart $\psi : V \to \mathbb{R}^n$ of $N$ with $q \in V$ such that $\overline{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$ is of the form*

$$\overline{f}(x_1, \dots, x_n, \dots, x_m) = (x_1, \dots, x_n).$$

EXERCISE 2.3. Prove Theorem 2.2.

*Hint.* Use the Inverse Function Theorem and the Implicit Function Theorem from Vector Calculus.

COROLLARY 2.4. (1) *If $f : M^m \to N^n$ is a smooth embedding then $f(M^m)$ is a m-dimensional submanifold of $N^n$.*

(2) *If $f : M^m \to N^n$ is a submersion then for every $x \in N^n$ the fiber $f^{-1}(x)$ is a submanifold of dimension $m - n$.*

EXERCISE 2.5. Every submersion $f : M \to N$ is an *open map*, i.e., the image of an open subset in $M$ is an open subset in $N$.

Let $f : M^m \to N^n$ be a smooth map and $y \in N$ is a point such that for some $x \in f^{-1}(y)$, the map $df_x : T_xM \to T_yN, y = f(x)$, is not surjective. Then the point $y \in N$ is called a *singular value* of $f$. A point $y \in N$ which is not a singular value of $f$ is called a *regular value* of $f$. Thus, for every regular value $y \in N$ of $f$, the preimage $f^{-1}(y)$ is either empty or a smooth submanifold of dimension $m - n$.

THEOREM 2.6 (Sard's theorem). *Almost every point $y \in N$ is a regular value of $f$.*

Sard's theorem has an important quantitative improvement due to Y. Yomdin which we will describe below. Let $B$ be the closed unit ball in $\mathbb{R}^{n-1}$. Consider a $C^n$-smooth function $f : B \to \mathbb{R}$. For every multi-index $i = (i_1, \dots, i_k)|$ set $|i| := k$, and for $k \leqslant n$ let

$$\partial^i f := \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}$$

be the *i-th mixed partial derivative of $f$*. Let

$$\|\partial^i f\| := \max_x |\partial^i f(x)|.$$

Define the $C^n$-*norm* of $f$ as

$$\|f\|_{C^n} := \max_{i, 0 \leqslant |i| \leqslant n+1} \|\partial^i f\|.$$

Given $\epsilon > 0$ let $E_\epsilon \subset \mathbb{R}$ denote the set

$$\{y \in \mathbb{R} : \exists x \in f^{-1}(y), |\nabla f(x)| < \epsilon\}.$$

Thus, the set $E_\epsilon$ consists of "almost" critical values of $f$. Yomdin's theorem informally says that for small $\epsilon$ the set $E_\epsilon$ is small. Below is the precise statement.

THEOREM 2.7 (Y. Yomdin, [**Yom83**]). *There exists a constant $c = c(n, \|f\|_{C^n})$ so that for every $C^n$-smooth function $f : B \to \mathbb{R}$, and every $\epsilon \in (0, 1)$ the set $E_\epsilon$ can be covered by at most $c/\epsilon$ intervals of length $\epsilon^{n/(n-1)}$. In particular:*

*1. Lebesgue measure of $E_\epsilon$ is at most*

$$c\epsilon^{\frac{1}{n-1}}.$$

*2. Whenever an interval $J \subset \mathbb{R}$ has length $\ell > c\epsilon^{1/(n-1)}$, there exists a subinterval $J' \subset J \setminus E_\epsilon$, so that $J'$ has length at least*

$$\frac{c}{\epsilon}\left(\ell - c\epsilon^{1/(n-1)}\right).$$

#### 2.1.2. Smooth partition of unity.

DEFINITION 2.8. Let $M$ be a smooth manifold and $\mathcal{U} = \{B_i : i \in I\}$ a locally finite covering of $M$ by open subsets diffeomorphic to Euclidean balls. A collection of smooth functions $\{\eta_i : i \in I\}$ on $M$ is called a *smooth partition of unity* for the cover $\mathcal{U}$ if the following conditions hold:

(1) $\sum_i \eta_i \equiv 1$.
(2) $0 \leqslant \eta_i \leqslant 1, \quad \forall i \in I$.
(3) $\mathrm{Supp}(\eta_i) \subset \overline{B_i}, \quad \forall i \in I$.

LEMMA 2.9. *Every open cover $\mathcal{U}$ as above admits a smooth partition of unity.*

**2.1.3. Riemannian metrics.** A *Riemannian metric* on a smooth $n$-dimensional manifold $M$, is a positive definite inner product $\langle \cdot, \cdot \rangle_p$ defined on the tangent spaces $T_pM$ of $M$; this inner product is required to depend smoothly on the point $p \in M$. We will suppress the subscript $p$ in this notation; we let $\| \cdot \|$ denote the norm on $T_pM$ determined by the Riemannian metric. The Riemannian metric is usually denoted $g = g_x = g(x), x \in M$ or $ds^2$. We will use the notation $|dx|^2$ to denote the Euclidean Riemannian metric on $\mathbb{R}^n$:

$$dx^2 := dx_1^2 + \ldots + dx_n^2.$$

Here and in what follows we use the convention that for tangent vectors $u, v$,

$$dx_i dx_j(u, v) = u_i v_j$$

and $dx_i^2$ stands for $dx_i dx_i$.

A *Riemannian manifold* is a smooth manifold equipped with a Riemannian metric.

Two Riemannian metrics $g, h$ on a manifold $M$ are said to be *conformal* to each other, if $h_x = \lambda(x) g_x$, where $\lambda(x)$ is a smooth positive function on $M$, called *conformal factor*. In matrix notation, we just multiply the matrix $A_x$ of $g_x$ by a scalar function. Such modification of Riemannian metrics does not change the angles between tangent vectors. A Riemannian metric $g_x$ on a domain $U$ in $\mathbb{R}^n$ is called *conformally-Euclidean* if it is conformal to $|dx|^2$, i.e., it is given by

$$\lambda(x)|dx|^2 = \lambda(x)(dx_1^2 + \ldots + dx_n^2).$$

Thus, the square of the norm of a vector $v \in T_xU$ with respect to $g_x$ is given by

$$\lambda(x)\sum_{i=1}^{n} v_i^2.$$

Given an immersion $f : M^m \to N^n$ and a Riemannian metric $g$ on $N$, one defines the *pull-back* Riemannian metric $f^*(g)$ by

$$\langle v, w\rangle_p = \langle df(v), df(w)\rangle_q, p \in M, q = f(p) \in N,$$

where the right-hand side we use the inner product defined by $g$ and in the left-hand side the one defined by $f^*(g)$. It is useful to rewrite this definition in terms of symmetric matrices, when $M, N$ are open subsets of $\mathbb{R}^n$. Let $A_y$ be the matrix-function defining $g$. Then $f^*(g)$ is given by the matrix-function $B_x$, where

$$y = f(x), \quad B_x = (D_xf)A_y\ (D_xf)^T$$

and $D_xf$ is the Jacobian matrix of $f$ at the point $x$.

Let us compute how pull-back works in "calculus terms" (this is useful for explicit computation of the pull-back metric $f^*(g)$), when $g(y)$ is a Riemannian metric on an open subset $U$ in $\mathbb{R}^n$. Suppose that

$$g(y) = \sum_{i,j} g_{ij}(y)dy_i dy_j$$

and $f = (f_1, \ldots, f_n)$ is a diffeomorphism $V \subset \mathbb{R}^n \to U$. Then

$$f^*(g) = h,$$

$$h(x) = \sum_{i,j} g_{ij}(f(x))df_i df_j.$$

Here for a function $\phi : \mathbb{R}^n \to \mathbb{R}$, e.g., $\phi(x) = f_i(x)$,

$$d\phi = \sum_{k=1}^{n} d_k\phi = \sum_{k=1}^{n} \frac{\partial \phi}{\partial x_k} dx_k,$$

and, thus,

$$df_i df_j = \sum_{k,l=1}^{n} \frac{\partial f_i}{\partial x_k}\frac{\partial f_j}{\partial x_l} dx_k dx_l.$$

A particular case of the above is when $N$ is a submanifold in a Riemannian manifold $M$. One can define a Riemannian metric on $N$ either by using the inclusion map and the pull-back metric, or by considering, for every $p \in N$, the subspace $T_pN$ of $T_pM$, and restricting the inner product $\langle\cdot,\cdot\rangle_p$ to it. Both procedures define the same Riemannian metric on $N$.

**Measurable Riemannian metrics.** The same definition makes sense if the inner product depends only measurably on the point $p \in M$, equivalently, the matrix-function $A_x$ is only measurable. This generalization of Riemannian metrics will be used in our discussion of quasi-conformal groups, Chapter 21, section 21.7.

**Length and distance.** Given a Riemannian metric on $M$, one defines the *length* of a path $\mathfrak{p} : [a, b] \to M$ by

$$\text{length}(\mathfrak{p}) = \int_a^b \|\mathfrak{p}'(t)\| dt. \tag{2.1}$$

By abusing the notation, we will frequently denote length($\mathfrak{p}$) by length($\mathfrak{p}([a, b])$).

Then, provided that $M$ is connected, one defines the Riemannian *distance function*

$$\text{dist}(p, q) = \inf_{\mathfrak{p}} \text{length}(\mathfrak{p}),$$

where the infimum is taken over all paths in $M$ connecting $p$ to $q$.

A smooth map $f : (M, g) \to (N, h)$ of Riemannian manifolds is called a *Riemannian isometry* if $f^*(h) = g$. In most cases, such maps do not preserve the Riemannian distances. This leads to a somewhat unfortunate terminological confusion, since the same name *isometry* is used to define maps between metric spaces which preserve the distance functions. Of course, if a Riemannian isometry $f : (M, g) \to (N, h)$ is also diffeomorphism, then it preserves the Riemannian distance function.

A *Riemannian geodesic segment* is a path $\mathfrak{p} : [a, b] \subset \mathbb{R} \to M$ which is a local length-minimizer, i.e.:

There exists $c \geqslant 0$ so that for all $t_1, t_2$ in $J$ sufficiently close to each other,

$$\text{dist}(\mathfrak{p}(t_1), \mathfrak{p}(t_2)) = \text{length}(\mathfrak{p}([t_1, t_2])) = c|t_1 - t_2|.$$

If $c = 1$, we say that $\mathfrak{p}$ has *unit speed*. Thus, a unit speed geodesic is a locally-distance preserving map from an interval to $(M, g)$. This definition extends to infinite geodesics in $M$, which are maps $\mathfrak{p} : J \to M$, defined on intervals $J \subset M$, whose restrictions to each finite interval are finite geodesics.

A smooth map $f : (M, g) \to (N, h)$ is called *totally-geodesic* if it maps geodesics in $(M, g)$ to geodesics in $(N, h)$. If, in addition, $f^*(h) = g$, then such $f$ is locally distance-preserving.

**Injectivity and convexity radii.** For every complete Riemannian manifold $M$ and a point $p \in M$, there exists the *exponential map*

$$\exp_p : T_pM \to M$$

which sends every vector $v \in T_pM$ to the point $\gamma_v(1)$, where $\gamma_v(t)$ is the unique geodesic in $M$ with $\gamma(0) = p$ and $\gamma'(0) = v$. The *injectivity radius* $InjRad(p)$ is the supremum of the numbers $r$ so that $\exp_p |B(0, r)$ is a diffeomorphism to its image. The *radius of convexity* $ConRad(p)$ is the supremum of $r$'s so that $r \leqslant InRad(p)$ and $C = \exp_p(B(0, r))$ is a convex subset of $M$, i.e., every $x, y \in C$ are connected by a (distance–realizing) geodesic segment entirely contained in $C$. It is a basic fact of Riemannian geometry that for every $p \in M$,

$$ConRad(p) > 0,$$

see e.g. **[dC92]**.

**2.1.4. Riemannian volume.** For every $n$-dimensional Riemannian manifold $(M, g)$ one defines the *volume element* (or *volume density*) denoted $dV$ (or $dA$ if $M$ is 2-dimensional). Given $n$ vectors $v_1, \ldots, v_n \in T_pM$, $dV(v_1 \wedge \ldots \wedge v_n)$ is the volume of the parallelepiped in $T_pM$ spanned by these vectors, this volume is nothing but $\sqrt{|\det(G(v_1, \ldots, v_n))|}$, where $G(v_1, \ldots, v_n)$ is the Gramm matrix with the entries

$\langle v_i, v_j \rangle$. If $ds^2 = \rho^2(x)|dx|^2$, is a conformally-Euclidean metric, then its volume density is given by

$$\rho^n(x)dx_1 \dots dx_n.$$

Thus, every Riemannian manifold has a canonical measure, given by the integral of its volume form

$$mes(E) = \int_A dV.$$

THEOREM 2.10 (Generalized Rademacher's theorem). *Let $f : M \to N$ be a Lipschitz map of Riemannian manifolds. Then $f$ is differentiable almost everywhere.*

EXERCISE 2.11. Deduce Theorem 2.10 from Theorem 1.40 and the fact that $M$ is second countable.

We now define *volumes of maps and submanifolds.* The simplest and the most familiar notion of volume comes from the vector calculus. Let $\Omega$ be a bounded region in $\mathbb{R}^n$ and $f : \Omega \to \mathbb{R}^n$ be a smooth map. Then the *geometric* volume of $f$ is defined as

$$Vol(f) := \int_\Omega |J_f(x)| dx_1 \dots dx_n \tag{2.2}$$

where $J_f$ is the Jacobian determinant of $f$. Note that we are integrating here a non-negative quantity, so geometric volume of a map is always non-negative. If $f$ were 1-1 and $J_f(x) > 0$ for every $x$, then, of course,

$$Vol(f) = \int_\Omega J_f(x) dx_1 \dots dx_n = Vol(f(\Omega)).$$

More generally, if $f : \Omega \to \mathbb{R}^m$ (now, $m$ need not be equal to $n$), then

$$Vol(f) = \int_\Omega \sqrt{|\det(G_f)|}$$

where $G_f$ is the Gramm matrix with the entries $\left\langle \frac{\partial f}{\partial x_i}, \frac{\partial f}{\partial x_j} \right\rangle$, where brackets denote the usual inner product in $\mathbb{R}^m$. In case $f$ is 1-1, the reader will recognize in this formula the familiar expression for the volume of an immersed submanifold $\Sigma = f(\Omega)$ in $\mathbb{R}^m$,

$$Vol(f) = \int_\Sigma dS.$$

The Gramm matrix above makes sense also for maps whose target is an $m$-dimensional Riemannian manifold $(M, g)$, with partial derivatives replaced with vectors $df(X_i)$ in $M$, where $X_i$ are coordinate vector fields in $\Omega$:

$$X_i = \frac{\partial}{\partial x_i}, i = 1, \dots, n.$$

Furthermore, one can take the domain of the map $f$ to be an arbitrary smooth manifold $N$ (possibly with boundary). Definition still makes sense and is independent of the choice of local charts on $N$ used to define the integral: this independence is a corollary of the change of variables formula in the integral in a domain in $\mathbb{R}^n$.

More precisely, consider charts $\varphi_\alpha : U_\alpha \to V_\alpha \subset N$, so that $\{V_\alpha\}_{\alpha\in J}$ is a locally-finite open covering of $N$. Let $\{\eta_\alpha\}$ be a partition of unity on $N$ corresponding to this covering. Then for $\zeta_\alpha = \eta_\alpha \circ \varphi_\alpha$, $f_\alpha = f \circ \varphi_\alpha$,

$$Vol(f) = \sum_{\alpha\in J} \int_{U_\alpha} \zeta_\alpha \sqrt{|\det(G_{f_\alpha})|}dx_1 \ldots dx_n$$

In particular, if $f$ is 1-1 and $\Sigma = f(N)$, then

$$Vol(f) = Vol(\Sigma).$$

REMARK 2.12. The formula for $Vol(f)$ makes sense when $f : N \to M$ is merely Lipschitz, in view of Theorem 2.10.

Thus, one can define the volume of an immersed submanifold, as well as that of a piecewise smooth submanifold; in the latter case we subdivide a piecewise-smooth submanifold in a union of images of simplices under smooth maps.

By abuse of language, sometimes, when we consider an open submanifold $N$ in $M$, so that boundary $\partial N$ of $N$ a submanifold of codimension 1, while we denote the volume of $N$ by $Vol(N)$, we shall call the volume of $\partial N$ the *area*, and denote it by $Area\,(\partial N)$.

EXERCISE 2.13. (1) Suppose that $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ is a smooth map so that $|d_x f(u)| \leqslant 1$ for every unit vector $u$ and every $x \in \Omega$. Show that $|J_f(x)| \leqslant 1$ for every $x$ and, in particular,

$$Vol(f(\Omega)) = |\int_\Omega J_f dx_1 \ldots dx_n| \leqslant Vol(f) \leqslant Vol(\Omega).$$

Hint: Use that under the linear map $A = d_x f$, the image of every $r$-ball is contained in $r$-ball.

(2) Prove the same thing if the map $f$ is merely 1-Lipschitz.

More general versions of the above exercises are the following.

EXERCISE 2.14. Let $(M, g)$ and $(N, h)$ be $n$-dimensional Riemannian manifolds.

(1) Let $f : M \to N$ be a smooth map such that for every $x \in M$, the norm of the linear map

$$df_x : \left(T_x M, \langle\cdot,\cdot\rangle_g\right) \to (T_f(x)N, \langle\cdot,\cdot\rangle_h)$$

is at most $L$.

Prove that $|J_f(x)| \leqslant L^n$ for every $x$ and that for every open subset $U$ of $M$

$$Vol(f(\Omega)) \leqslant L^n Vol(\Omega).$$

(2) Prove the same statement for an $L$–Lipschitz map $f : M \to N$.

A consequence of Theorem 2.2 is the following.

THEOREM 2.15. *Consider a compact Riemannian manifold $M^m$, a submersion $f : M^m \to N^n$, and a point $p \in N$. For every $x \in N$ set $M_x := f^{-1}(x)$. Then, for every $p \in N$ and every $\epsilon > 0$ there exists an open neighborhood $W$ of $p$ such that for every $x \in W$,*

$$1 - \epsilon \leqslant \frac{Vol(M_x)}{Vol(M_p)} \leqslant 1 + \epsilon.$$

Proof. First note that, by compactness of $M_p$, for every neighborhood $U$ of $M_p$ there exists a neighborhood $W$ of $p$ such that $f^{-1}(W) \subset U$.

According to Theorem 2.2, (2), for every $x \in M_p$ there exists a chart of $M$, $\varphi_x : U_x \to \overline{U}_x$, with $U_x$ containing $x$, and a chart of $N$, $\psi_x : V_x \to \overline{V}_x$ with $V_x$ containing $p$, such that $\psi_x \circ f \circ \varphi_x^{-1}$ is a restriction of the projection to the first $n$ coordinates. Without loss of generality we may assume that $\overline{U}_x$ is an open cube in $\mathbb{R}^m$. Therefore, $\overline{V}_x$ is also a cube in $\mathbb{R}^n$, and $\overline{U}_x = \overline{V}_x \times \overline{Z}_x$ , where $\overline{Z}_x$ is an open subset in $\mathbb{R}^{m-n}$ .

Since $M_p$ is compact, it can be covered by finitely many such domains of charts $U_1, \dots, U_k$. Let $V_1, \dots, V_k$ be the corresponding domains of charts containing $p$. For the open neighborhood $U = \bigcup_{i=1}^k U_i$ of $M_p$ consider an open neighborhood $W$ of $p$, contained in $\bigcap_{i=1}^k V_i$, such that $f^{-1}(W) \subseteq U$.

For every $x \in W$, $M_x = \bigcup_{l=1}^k (U_i \cap M_x)$. Fix $l \in \{1, \dots, k\}$. Let $(g_{ij}(y))_{1 \leqslant i,j \leqslant n}$ be the matrix-valued function on $\overline{U}_l$, defining the pull-back by $\varphi_l$ of the Riemannian metric on $M$ .

Since $g_{ij}$ is continuous, there exists a neighborhood $\overline{W}_l$ of $\bar{p} = \psi_l(p)$ such that for every $\bar{x} \in W_l$ and for every $\bar{t} \in \overline{Z}_l$ we have,

$$(1-\epsilon)^2 \leqslant \frac{\det\left[g_{ij}(\bar{x},\bar{t})\right]_{n+1\leqslant i,j\leqslant m}}{\det\left[g_{ij}(\bar{p},\bar{t})\right]_{n+1\leqslant i,j\leqslant m}} \leqslant (1+\epsilon)^2 \,.$$

Recall that the volumes of $M_x \cap U_i$ and of $M_p \cap U_l$ are obtained by integrating respectively $(\det\left[g_{ij}(\bar{x},\bar{t})\right]_{n+1\leqslant i,j\leqslant k})^{1/2}$ and $(\det\left[g_{ij}(\bar{p},\bar{t})\right]_{n+1\leqslant i,j\leqslant k})^{1/2}$ on $Z_l$ . The volumes of $M_x$ and $M_p$ are obtained by combining this with a partition of unity.

It follows that for $x \in \bigcap_{i=1}^k \psi_i^{-1}(\overline{W}_l)$,

$$1-\epsilon \leqslant \frac{Vol(M_x)}{Vol(M_p)} \leqslant 1+\epsilon \,.$$

□

Finally, we recall an important formula for volume computations:

Theorem 2.16 (Coarea formula, see e.g. Theorem 6.3 **[Cha06]** ). *Let $U$ be an open connected subset with compact closure $\overline{U}$ in a Riemannian manifold $M$ and let $f : U \to (0,\infty)$ be a smooth submersion with a continuous extension to $\overline{U}$ such that $f$ restricted to $\overline{U} \setminus U$ is constant. For every $t \in (0,\infty)$ let $\mathcal{H}_t$ denote the level set $f^{-1}(t)$, and let $\mathrm{d}A_t$ be the Riemannian area density induced on $\mathcal{H}_t$.*

*Then, for every function $g \in L^1(U)$,*

$$\int_U g\,|\mathrm{grad} f| \mathrm{d}V = \int_0^\infty \mathrm{d}t \int_{\mathcal{H}_t} g\, \mathrm{d}A_t$$

*where $\mathrm{d}V$ is the Riemannian volume density of $M$*

**2.1.5. Growth function and Cheeger constant.** In this section we present two basic notions initially introduced in Riemannian geometry and later adapted and used in group theory and in combinatorics.

Given a Riemannian manifold $(M,g)$ and a point $x_0 \in M$, we define the *growth function*

$$\mathfrak{G}_{M,x_0}(r) := Vol\, B(x_0, r),$$

the volume of the metric ball of radius $r$ and center at $x$ in $(M,g)$

Remarks 2.17. (1) For two different points $x_0, y_0$, we have

$$\mathfrak{G}_{M,x_0}(r) \leqslant \mathfrak{G}_{M,y_0}(r+d), \text{ where } d = \text{dist}(x_0, y_0).$$

(2) Suppose that the action of the group of isometries of $M$ is cobounded, i.e., there exists $\kappa$ such that the Isom($M$)-orbit of $B(x_0, \kappa)$ equals $M$. In this case, for every two basepoints $x_0, y_0$

$$\mathfrak{G}_{M,x_0}(r) \leqslant \mathfrak{G}_{M,y_0}(r+\kappa).$$

Thus, in this case the growth rate of the function $\mathfrak{G}$ does not depend on the choice of the basepoint.

We refer the reader to Section 12.1 for the detailed discussion of volume growth and its relation to group growth.

Exercise 2.18. Assume again that the action Isom($M$) $\curvearrowright M$ is cobounded and that $(M, g)$ is complete.

(1) Prove that the growth function is *almost sub-multiplicative*, that is:

$$\mathfrak{G}_{M,x_0}\left((r+t)\kappa\right) \leqslant \mathfrak{G}_{M,x_0}(r\kappa)\mathfrak{G}_{M,x_0}((t+1)\kappa).$$

(2) Prove that the growth function of $M$ is at most exponential, that is there exists $a > 1$ such that

$$\mathfrak{G}_{M,x_0}(x) \leqslant a^x, \text{ for every } x \geqslant 0.$$

Definition 2.19. An *isoperimetric inequality* in a manifold $M$ is an inequality satisfied by all open submanifolds $\Omega$ with compact closure and smooth boundary, of the form

$$Vol(\Omega) \leq f(\Omega) g\left(Area \partial\Omega\right),$$

where $f$ and $g$ are real-valued functions, $g$ defined on $\mathbb{R}_+$.

Definition 2.20. The *Cheeger (isoperimetric) constant* $h(M)$ (or *isoperimetric ratio*) of $M$ is the infimum of the ratios

$$\frac{Area(\partial\Omega)}{\min\left[Vol(\Omega)\,,\, Vol(M\setminus\Omega)\right]},$$

where $\Omega$ varies over all open submanifolds with compact closure and smooth boundary.

If in particular $h(M) \geq \kappa > 0$ then the following isoperimetric inequality holds in $M$:

$$Vol(\Omega) \leqslant \frac{1}{\kappa} Area(\partial\Omega) \text{ for every } \Omega.$$

This notion was defined by Cheeger for compact manifolds in [**Che70**]. Further details can be found for instance in P. Buser's book [**Bus10**]. Note that when $M$ is a Riemannian manifold of infinite volume, one may replace the denominator in the ratio defining the Cheeger constant by $Vol(\Omega)$.

Assume now that $M$ is the universal cover of a compact Riemannian manifold $N$. A natural question to ask is to what extent the growth function and the Cheeger constant of $M$ depend on the choice of the Riemannian metric on $N$. The first question, in a way, was one of the origins of the geometric group theory.

V.A. Efremovich [**Efr53**] noted that two growth functions corresponding to two different choices of metrics on $N$ increase at the same rate, and, moreover, that their behavior is essentially determined by the fundamental group only. See Proposition 12.12 for a slightly more general statement.

A similar phenomenon occurs with the Cheeger constant: Positivity of $h(M)$ does not depend on the metric on $N$, it depends only on a certain property of $\pi_1(N)$, namely, the non-amenability, see Remark 16.12. This was proved much later by R. Brooks [**Bro81a, Bro82a**]. Brooks' argument has a global analytic flavor, as it uses the connection established by Cheeger [**Che70**] between positivity of the isoperimetric constant and positivity of spectrum of the Laplace-Beltrami operator on $M$. Note that even though in the quoted paper Cheeger only considers compact manifolds, the same argument works for universal covers of compact manifolds. This result was highly influential in global analysis on manifolds and harmonic analysis on graphs and manifolds.

**2.1.6. Curvature.** Instead of defining the Riemannian curvature tensor, we will only describe some properties of Riemannian curvature. First, if $(M, g)$ is a 2-dimensional Riemannian manifold, one defines *Gaussian curvature* of $(M, g)$, which is a smooth function $K : M \to \mathbb{R}$, whose values are denoted $K(p)$ and $K_p$.

More generally, for an $n$-dimensional Riemannian manifold $(M, g)$, one defines the *sectional curvature*, which is a function $\Lambda^2 M \to \mathbb{R}$, denoted $K_p(u, v) = K_{p,g}(u, v)$:

$$K_p(u, v) = \frac{\langle R(u, v)u, v\rangle}{|u \wedge v|^2},$$

provided that $u, v \in T_pM$ are linearly independent. Here $R$ is the Riemannian curvature tensor and $|u\wedge v|$ is the area of the parallelogram in $T_pM$ spanned by the vectors $u, v$. Sectional curvature depends only on the 2-plane $P$ in $T_pM$ spanned by $u$ and $v$. The curvature tensor $R(u, v)w$ does not change if we replace the metric $g$ with a conformal metric $h = ag$, where $a > 0$ is a constant. Thus,

$$K_{p,h}(u, v) = a^{-2}K_{p,g}(u, v).$$

Totally geodesic Riemannian isometric immersions $f : (M, g) \to (N, h)$ preserve sectional curvature:

$$K_p(u, v) = K_q(df(u), df(v)), \quad q = f(p).$$

In particular, sectional curvature is invariant under Riemannian isometries of equidimensional Riemannian manifolds. In the case when $M$ is 2-dimensional, $K_p(u, v) = K_p$, is the Gaussian curvature of $M$.

**Gauss-Bonnet formula.** Our next goal is to connect areas of triangles to curvature.

THEOREM 2.21 (Gauss-Bonnet formula). *Let $(M, g)$ be a Riemannian surface with the Gaussian curvature $K(p), p \in M$ and the area form $dA$. Then for every 2-dimensional triangle $\blacktriangle \subset M$ with geodesic edges and vertex angles $\alpha, \beta, \gamma$,*

$$\int_{\blacktriangle} K(p)dA = (\alpha + \beta + \gamma) - \pi.$$

*In particular, if $K(p)$ is constant equal $\kappa$, we get*

$$-\kappa Area(\blacktriangle) = \pi - (\alpha + \beta + \gamma).$$

The quantity $\pi - (\alpha + \beta + \gamma)$ is called the *angle deficit* of the triangle $\Delta$.

**Manifolds of bounded geometry.** A (complete) Riemannian manifold $M$ is said to have *bounded geometry* if there are constants $a, b$ and $\epsilon > 0$ so that:

1. Sectional curvature of $M$ varies in the interval $[a, b]$.
2. Injectivity radius of $M$ is $\geqslant \epsilon$.

The numbers $a, b, \epsilon$ are called *geometric bounds* on $M$. For instance, every compact Riemannian manifold $M$ has bounded geometry, every covering space of $M$ (with pull-back Riemannian metric) also has bounded geometry.

THEOREM 2.22 (See e.g. Theorem 1.14, [**Att94**]). *Let $M$ be a Riemannian manifold of bounded geometry with geometric bounds $a, b, \epsilon$. Then for every $x \in M$ and $0 < r < \epsilon/2$, the exponential map*

$$\exp_x : B(0, r) \to B(x, r) \subset M$$

*is an $L$-bi-Lipschitz diffeomorphism, where $L = L(a, b, \epsilon)$.*

This theorem also allows one to refine the notion of partition of unity in the context of Riemannian manifolds of bounded geometry:

LEMMA 2.23. *Let $M$ be a Riemannian manifold of bounded geometry and let $\mathcal{U} = \{B_i = B(x_i, r_i) : i \in I\}$ a locally finite covering of $M$ by metric balls so that $InjRad_M(x_i) > 2r_i$ for every $i$ and*

$$B\left(x_i, \frac{3}{4} r_i\right) \cap B\left(x_j, \frac{3}{4} r_j\right) = \emptyset, \ \forall i \neq j.$$

*Then $\mathcal{U}$ admits a smooth partition of unity $\{\eta_i : i \in I\}$ which, in addition, satisfies the following properties:*

*1. $\eta_i \equiv 1$ on every ball $B(x_i, \frac{r_i}{2})$.*
*2. Every smooth functions $\eta_i$ is $L$–Lipschitz for some $L$ independent of $i$.*

**Curvature and volume.**

Below we describe without proof certain consequences of uniform lower and upper bounds on the sectional curvature on the growth of volumes of balls, that will be used in the sequel. The references for the result below are [**BC01**, Section 11.10], [**CGT82**], [**Gro86**], [**G$\ddot{\text{6}}$0**]. See also [**GHL04**], Theorem 3.101, p. 140.

Below we will use the following notation: For $\kappa \in \mathbb{R}$, $A_\kappa(r)$ and $V_\kappa(r)$ will denote the area of the sphere, respectively the volume of the ball of radius $r$, in the $n$–dimensional space of constant sectional curvature $\kappa$. We will also denote by $A(x, r)$ the area of the geodesic sphere of radius $r$ and center $x$ in a Riemannian manifold $M$. Likewise, $V(x, r)$ will denote the volume of the geodesic ball centered at $x$ and of radius $r$ in $M$.

THEOREM 2.24 (Bishop–Gromov–Günther). *Let $M$ be a complete $n$–dimensional Riemannian manifold.*

(1) *Assume that the sectional curvature on $M$ is at least $a$. Then, for every point $x \in M$:*
   - *$A(x, r) \leqslant A_a(r)$ and $V(x, r) \leqslant V_a(r)$.*
   - *The functions $r \mapsto \frac{A(x,r)}{A_a(r)}$ and $r \mapsto \frac{V(x,r)}{V_a(r)}$ are non-increasing.*

(2) *Assume that the sectional curvature on $M$ is at most $b$. The, for every $x \in M$ with injectivity radius $\rho_x = InjRad_M(x)$:*
   - *For all $r \in (0, \rho_x)$, we have $A(x,r) \geqslant A_b(r)$ and $V(x,r) \geqslant V_b(r)$.*
   - *The functions $r \mapsto \frac{A(x,r)}{A_b(r)}$ and $r \mapsto \frac{V(x,r)}{V_b(r)}$ are non-decreasing on $(0, \rho_x)$.*

The results (1) in the theorem above are also true if the Ricci curvature of $M$ is at least $(n-1)a$.

Theorem 2.24 follows from infinitesimal versions of the above inequalities (see Theorems 3.6 and 3.8 in [**Cha06**]). A consequence of the infinitesimal version of Theorem 2.24, (1), is the following theorem which will be useful in the proof of quasi-isometric invariance of positivity of the Cheeger constant:

THEOREM 2.25 (Buser's inequality [**Bus82**], [**Cha06**], Theorem 6.8). *Let $M$ be a complete $n$–dimensional manifold with sectional curvature at least $a$. Then there exists a positive constant $\lambda$ depending on $n, a$ and $r > 0$, such that the following holds. Given a hypersurface $\mathcal{H} \subset M$ and a ball $B(x,r) \subset M$ such that $B(x,r) \setminus \mathcal{H}$ is the union of two open subsets $\mathcal{O}_1 \mathcal{O}_2$ separated by $\mathcal{H}$, we have:*

$$\min\left[Vol(\mathcal{O}_1)\,,\, Vol(\mathcal{O}_2)\right] \leqslant \lambda Area\left[\mathcal{H} \cap B(x,r)\right].$$

**2.1.7. Harmonic functions.** For the detailed discussion of the material in this section we refer the reader to [**Li04**] and [**SY94**].

Let $M$ be a Riemannian manifold. Given a smooth function $f : M \to \mathbb{R}$, we define the *energy* of $f$ as the integral

$$E(f) = \int_M |df|^2 dV = \int_M |\nabla f|^2 dV.$$

Here the gradient vector field $\nabla f$ is obtained by dualizing the differential 1-form $df$ using the Riemannian metric on $M$. Note that energy is defined even if $f$ only belongs to the Sobolev space $W^{1,2}_{loc}(M)$ of functions differentiable a.e. on $M$ with locally square-integrable partial derivatives.

THEOREM 2.26 (Lower semicontinuity of the energy functional). *Let $(f_i)$ be a sequence of functions in $W^{1,2}_{loc}(M)$ which converges (in $W^{1,2}_{loc}(M)$) to a function $f$. Then*

$$E(f) \leqslant \liminf_{i \to \infty} E(f_i).$$

DEFINITION 2.27. A function $h \in W^{1,2}_{loc}$ is called *harmonic* if it is *locally energy-minimizing*: For every point $p \in M$ and a small metric ball $B = B(p,r) \subset M$,

$$E(h|B) \leqslant E(u), \quad \forall u : \bar{B} \to \mathbb{R}, u|_{\partial B} = h|_{\partial B}.$$

Equivalently, for every relatively compact open subset $\Omega \subset M$ with smooth boundary

$$E(h|B) \leqslant E(u), \quad \forall u : \bar{\Omega} \to \mathbb{R}, u|_{\partial \Omega} = h|_{\partial \Omega}.$$

It turns out that harmonic functions $h$ on $M$ are automatically smooth and, moreover, satisfy the equation $\Delta h = 0$, where $\Delta$ is the *Laplace–Beltrami operator* on $M$:

$$\Delta u = \operatorname{div} \nabla u$$

Here for a vector field $X$ on $M$, the *divergence* $\operatorname{div} X$ is a function on $M$ satisfying

$$\operatorname{div} X dV = L_X dV,$$

where $L_X$ is the *Lie derivative* along the vector field $X$:

$$L_X : \Omega^k(M) \to \Omega^k(M),$$

$$L_X(\omega) = i_X d\omega + d(i_X \omega),$$

$$i_X : \Omega^{\ell+1}(M) \to \Omega^\ell(M), \quad i_X(\omega)(X_1, \ldots, X_\ell) = \omega(X, X_1, \ldots, X_\ell).$$

In local coordinates (assuming that $M$ is $n$-dimensional):

$$\operatorname{div} X = \sum_{i=1}^n \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x_i} \left( \sqrt{|g|} X_i \right)$$

where

$$|g| = \det((g_{ij})),$$

and

$$(\nabla u)^i = \sum_{j=1}^n g^{ij} \frac{\partial u}{\partial x_j}$$

and $(g^{ij}) = (g_{ij})^{-1}$, the inverse matrix of the metric tensor. Thus,

$$\Delta u = \sum_{i,j=1}^n \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x_i} \left( g^{ij} \sqrt{|g|} \frac{\partial u}{\partial x_j} \right).$$

In terms of the Levi–Civita connection $\nabla$ on $M$,

$$\Delta(u) = Trace(H(u)), \quad H(u)(X_i, X_j) = \nabla_{X_i} \nabla_{X_j}(u) - \nabla_{\nabla_{X_i} X_j}(u),$$

$$Trace(H) = \sum_{i,j=1}^n g^{ij} H_{ij},$$

where $X_i, X_j$ are vector fields on $M$.

If $M = \mathbb{R}^n$ with the flat metric, then $\Delta$ is the usual Laplace operator:

$$\Delta u = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} u.$$

THEOREM 2.28 (Yau's gradient estimate). *Suppose that $M^n$ is a complete $n$-dimensional Riemannian manifold with Ricci curvature $\geqslant a$. Then for every harmonic function $h$ on $M$, every $x \in M$ with $InjRad(x) \geqslant \epsilon$,*

$$|\nabla h(x)| \leqslant h(x) C(\epsilon, n).$$

THEOREM 2.29 (Compactness property). *Suppose that $(f_i)$ is a sequence of harmonic functions on $M$ so that there exists $p \in M$ for which the sequence $(f_i(p))$ is bounded. Then the family of functions $(f_i)$ is precompact in $W^{1,2}_{loc}(M)$. Furthermore, every limit of a subsequence in $(f_i)$ is a harmonic function.*

THEOREM 2.30 (Maximum Principle). *Let $\Omega \subset M$ be a relatively compact domain with smooth boundary and $h : M \to \mathbb{R}$ be a harmonic function. Then $h|\overline{\Omega}$ attains maximum on the boundary of $\Omega$ and, moreover, if $h|\Omega$ attains its maximum at a point of $\Omega$, then $h$ is constant.*

**2.1.8. Alexandrov curvature and $CAT(\kappa)$ spaces.** In the more general setting of metric spaces it is still possible to define a notion of (upper and lower bound for the) sectional curvature, which moreover coincide with the standard ones for Riemannian manifolds. This is done by comparing geodesic triangles in a metric space to geodesic triangles in a *model space of constant curvature*. In what follows, we only discuss the metric definition of upper bound for the sectional curvature, the lower bound case is similar but less used.

For a given $\kappa \in \mathbb{R}$, we denote by $X_\kappa$ the *model surface of constant curvature* $\kappa$. If $\kappa = 0$ then $X_\kappa$ is the Euclidean plane, if $\kappa < 0$ then $X_\kappa$ will be discussed in detail in Chapter 8, it is the *upper half-plane with the rescaled hyperbolic metric:*

$$X_\kappa = \left( U^2, |\kappa|^{-1} \frac{dx^2 + dy^2}{y^2} \right).$$

If $\kappa > 0$ then $X_\kappa$ is the 2–dimensional sphere $S\left(0, \frac{1}{\sqrt{\kappa}}\right)$ in $\mathbb{R}^3$ with the Riemannian metric induced from $\mathbb{R}^3$.

Let $X$ be a geodesic metric space, and let $\Delta$ be a geodesic triangle in $X$. Given $\kappa > 0$ we say that $\Delta$ is $\kappa$*–compatible* if its perimeter is at most $\frac{2\pi}{\sqrt{\kappa}}$. By default, every triangle is $\kappa$–compatible for $\kappa \leqslant 0$.

We will prove later on (see §8.10) the following:

LEMMA 2.31. *Let $\kappa \in \mathbb{R}$ and let $a \leqslant b \leqslant c$ be three numbers such that $c \leqslant a + b$ and $a + b + c < \frac{2\pi}{\sqrt{\kappa}}$ if $\kappa > 0$. Then there exists a geodesic triangle in $X_\kappa$ with lengths of edges $a, b$ and $c$, and it is unique up to congruence.*

Therefore, for every $\kappa \in \mathbb{R}$ and every $\kappa$–compatible triangle $\Delta = \Delta(A, B, C) \subset X$ with vertices $A, B, C \in X$ and lengths $a, b, c$ of the opposite sides, there exists a triangle (unique, up to congruence)

$$\tilde{\Delta}(\tilde{A}, \tilde{B}, \tilde{C}) \subset X_\kappa$$

with the side-lengths $a, b, c$. The triangle $\tilde{\Delta}(\tilde{A}, \tilde{B}, \tilde{C})$ is called the $\kappa$*–comparison triangle* or a $\kappa$*–Alexandrov triangle.*

For every point $P$ on, say, the side $[AB]$ of $\Delta$, we define the $\kappa$*–comparison point* $\tilde{P} \in [\tilde{A}, \tilde{B}]$, so that

$$d(A, P) = d(\tilde{A}, \tilde{P}).$$

Thus, for $P \in [A, B], Q \in [B, C]$ we define $\kappa$–comparison points $\tilde{P}, \tilde{Q} \in \tilde{\Delta}$.

DEFINITION 2.32. We say that *the triangle $\Delta$ is $CAT(\kappa)$* if it is $\kappa$–compatible and for every pair of points $P$ and $Q$ on the triangle, their $\kappa$–comparison points $\tilde{P}, \tilde{Q}$ satisfy

$$\mathrm{dist}_{X_\kappa}\left(\tilde{P}, \tilde{Q}\right) \geqslant \mathrm{dist}_X\left(P, Q\right).$$

DEFINITION 2.33.
(1) A $CAT(\kappa)$*–domain* in $X$ is an open convex set $U \subset X$, and such that all the geodesic triangles entirely contained in $U$ are $CAT(\kappa)$.
(2) We say that $X$ has *Alexandrov curvature at most* $\kappa$ if it is covered by $CAT(\kappa)$–domains.

Note that a $CAT(\kappa)$–domain $U$ for $\kappa > 0$ must have diameter strictly less than $\frac{\pi}{\sqrt{\kappa}}$. Otherwise, one can construct geodesic triangles in $U$ with two equal edges and the third reduced to a point, with perimeter $\geqslant \frac{2\pi}{\sqrt{\kappa}}$.

The point of Definition 2.33 is that it applies to non-Riemannian metric spaces where such notions as tangent vectors, Riemannian metric, curvature tensor cannot be defined, while one can still talk about curvature being bounded from above by $\kappa$.

PROPOSITION 2.34. *Let $X$ be a Riemannian manifold. Its Alexandrov curvature is at most $\kappa$ if and only if its sectional curvature in every point is $\leqslant \kappa$.*

PROOF. The "if" implication follows from the Rauch-Toponogov comparison theorem (see [**dC92**, Proposition 2.5]). For the "only if" implication we refer to [**Rin61**] or to [**GHL04**, Chapter III]. □

DEFINITION 2.35. A metric space $X$ is called a $CAT(\kappa)$*-space* if the entire $X$ is a $CAT(\kappa)$-domain. We will use the definition only for $\kappa \leqslant 0$. A metric space $X$ is said to be a $CAT(-\infty)$*-space* if $X$ is a $CAT(\kappa)$-space for every $\kappa$.

Note that for the moment we do not assume $X$ to be metrically complete. This is because there are naturally occurring incomplete $CAT(0)$ spaces, called *Euclidean buildings*, which, nevertheless, are *geodesically complete* (every geodesic segment is contained in a complete geodesic). On the other hand, Hilbert spaces provide natural examples of complete CAT(0) metric spaces.

EXERCISE 2.36. Let $X$ be a simplicial tree with a path-metric $d$. Show that $(X, d)$ is $CAT(-\infty)$.

In the case of non-positive curvature there exists a local-to-global result.

THEOREM 2.37 (Cartan-Hadamard Theorem). *If $X$ is a simply connected complete metric space with Alexandrov curvature at most $\kappa$ for some $\kappa \leqslant 0$, then $X$ is a $CAT(\kappa)$–space.*

We refer the reader to [**Bal95**] and [**BH99**] for proofs of this theorem, and a detailed discussion of $CAT(\kappa)$–spaces, with $\kappa \leqslant 0$.

DEFINITION 2.38. Simply-connected complete Riemannian manifolds of sectional curvature $\leqslant 0$ are called *Hadamard manifolds*. Thus, every Hadamard manifold is a $CAT(0)$ space.

An important property of $CAT(0)$-spaces is *convexity of the distance function.* Suppose that $X$ is a geodesic metric space. We say that a function $F : X \times X \to \mathbb{R}$ is *convex* if for every pair of geodesics $\alpha(s), \beta(s)$ in $X$ (which are parameterized with constant, but not necessarily unit, speed), the function

$$f(s) = F(\alpha(s), \beta(s))$$

is a convex function of one variable. Thus, the distance function dist of $X$ is convex, whenever for every pair of geodesics $[a_0, a_1]$ and $[b_0, b_1]$ in $X$, the points $a_s \in [a_0, a_1]$ and $b_s \in [b_0, b_1]$ such that $\text{dist}(a_0, a_s) = s\text{dist}(a_0, a_1)$ and $\text{dist}(b_0, b_s) = s\text{dist}(b_0, b_1)$ satisfy

$$\text{dist}(a_s, b_s) \leq (1 - s)\text{dist}(a_0, b_0) + s\text{dist}(a_1, b_1)\,. \tag{2.3}$$

Note that in the case of a normed vector space $X$, a function $f : X \times X \to \mathbb{R}$ is convex if and only if the sup-graph

$$\{(x, y, t) \in X^2 \times \mathbb{R} : f(x, y) \geqslant t\}$$

is convex.

PROPOSITION 2.39. *A geodesic metric space $X$ is $CAT(0)$ if and only if the distance on $X$ is convex.*

PROOF. Consider two geodesics $[a_0, b_0]$ and $[a_1, b_1]$ in $X$. On the geodesic $[a_0, b_1]$ consider the point $c_s$ such that $\text{dist}(a_0, c_s) = s\text{dist}(a_0, b_1)$ . The fact that the triangle with edges $[a_0, a_1]$, $[a_0, b_1]$ and $[a_1, b_1]$ is $CAT(0)$ and the Thales theorem in $\mathbb{R}^2$, imply that $\text{dist}(a_s, c_s) \leq s\text{dist}(a_1, b_1)$. The same argument applied to the triangle with edges $[a_0, b_1], [a_0, b_0], [b_0, b_1]$, implies that $\text{dist}(c_s, b_s) \leq (1 - s)\text{dist}(a_0, b_0)$. The inequality (2.3) follows from

$$\text{dist}(a_s, b_s) \leqslant \text{dist}(a_s, c_s) + \text{dist}(c_s, b_s) .$$

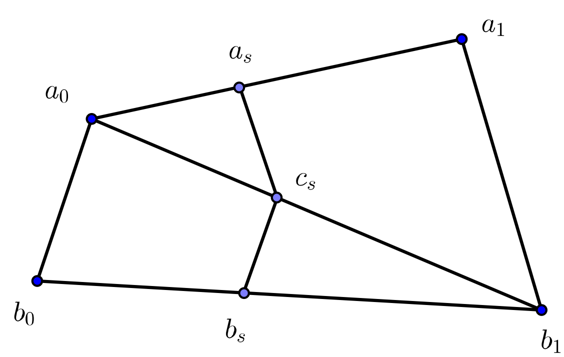



FIGURE 2.1. Argument for convexity of the distance.

Conversely, assume that (2.3) is satisfied.

In the special case when $a_0 = a_1$, this implies the comparison property in Definition 2.32 when one of the two points $P, Q$ is a vertex of the triangle. When $a_0 = b_0$, (2.3) again implies the comparison property when $\frac{\text{dist}(A,P)}{\text{dist}(A,B)} = \frac{\text{dist}(B,Q)}{\text{dist}(B,C)}$ .

We now consider the general case of two points $P \in [A, B]$ and $Q \in [B, C]$ such that $\frac{\text{dist}(A,P)}{\text{dist}(A,B)} = s$ and $\frac{\text{dist}(B,Q)}{\text{dist}(B,C)} = t$ with $s < t$ . Consider $B' \in [A, B]$ such that $\text{dist}[A, B'] = \frac{s}{t}\text{dist}[A, B]$ . Then, according to the above, $\text{dist}(B', C) \leqslant \text{dist}(\widetilde{B'}, \widetilde{C})$, and $\text{dist}(P, Q) \leqslant t\text{dist}(B', C) \leqslant t\text{dist}(\widetilde{B'}, \widetilde{C}) = \text{dist}(\widetilde{P}, \widetilde{Q})$ . $\square$

COROLLARY 2.40. *Every $CAT(0)$-space $X$ is uniquely geodesic.*

Proof. It suffices to apply the inequality (2.3) to any geodesic bigon, that is, in the special case when $a_0 = b_0$ and $a_1 = b_1$. □

**2.1.9. Cartan's fixed point theorem.** Let $X$ be a metric space and $A \subset X$ be a subset. Define the function

$$\rho(x) = \rho_A(x) = \sup_{a \in A} d^2(x, a).$$

Proposition 2.41. *Let $X$ be a complete CAT(0) space. Then for every bounded subset $A \subset X$, the function $\rho = \rho_A$ attains unique minimum in $X$.*

Proof. Consider a sequence $(x_n)$ in $X$ such that

$$\lim_{n \to \infty} \rho(x_n) = r = \inf_{x \in X} \rho(x).$$

We claim that the sequence $(x_n)$ is Cauchy. Given $\epsilon > 0$ let $x = x_i, x' = x_j$ be points in this sequence such that

$$r \leqslant \rho(x) < r + \epsilon, \quad r \leqslant \rho(x') < r + \epsilon.$$

Let $p$ be the midpoint of $[x, x'] \subset X$; hence, $r \leqslant \rho(p)$. Let $a \in A$ be such that

$$\rho(p) - \epsilon < d^2(p, a).$$

Consider the Euclidean comparison triangle $\tilde{T} = T(\tilde{x}, \tilde{x}', \tilde{a})$ for the triangle $T(x, x', a)$. In the Euclidean plane we have (by the parallelogram identity (1.2)):

$$d^2(\tilde{x}, \tilde{x}') + 4\, d^2(\tilde{a}, \tilde{p}) = 2\left(d^2(\tilde{a}, \tilde{x}) + d^2(\tilde{a}, \tilde{x}')\right).$$

Applying the comparison inequality for the triangles $T$ and $\tilde{T}$, we obtain:

$$d(a, p) \leqslant d(\tilde{a}, \tilde{p}).$$

Thus:

$$d(x, x')^2 + 4(r - \epsilon) < d^2(x, x') + 4d^2(a, p) \leqslant 2\left(d^2(a, x) + d^2(a, x')\right) <$$

$$2(\rho(x) + \rho(x')) < 4r + 4\epsilon.$$

It follows that

$$d(x, x')^2 < 8\epsilon$$

and, therefore, the sequence $(x_n)$ is Cauchy. By completeness of $X$, the function $\rho$ attains minimum in $X$; the same Cauchy argument implies that the point of minimum is unique. □

As a corollary, we obtain a fixed-point theorem for isometric group actions on complete $CAT(0)$ spaces, which was first proven by Cartan in the context of Riemannian manifolds of nonpositive curvature:

Theorem 2.42. *Let $X$ be a complete $CAT(0)$ metric space and $G \curvearrowright X$ be a group acting isometrically with bounded orbits. Then $G$ fixes a point in $X$.*

Proof. Let $A$ denote a bounded orbit of $G$ in $X$ and let $\rho_A$ be the corresponding function on $X$. Then, by uniqueness of the minimum point $m$ of $\rho_A$, the group $G$ will fix $m$. □

Corollary 2.43. *1. Every finite group action on a complete $CAT(0)$ space has a fixed point. For instance, every action of a finite group on a tree or on a Hilbert space fixes a point.*

*2. If $G$ is a compact group acting isometrically and continuously on a Hilbert space $\mathcal{H}$, then $G$ fixes a point in $\mathcal{H}$.*

**2.1.10. Ideal boundary, horoballs and horospheres.** In this section we define the ideal boundary of a metric space. This is a particularly significant object when the metric space is $CAT(0)$, and it generalizes the concept introduced for non-positively curved simply connected Riemannian manifolds by P. Eberlein and B. O'Neill in [**EO73**, Section 1].

Let $X$ be a geodesic metric space. Two geodesic rays $\rho_1$ and $\rho_2$ in $X$ are called *asymptotic* if they are at finite Hausdorff distance; equivalently if the function $t \mapsto \text{dist}(\rho_1(t), \rho_2(t))$ is bounded on $[0, \infty)$.

Clearly, being asymptotic is an equivalence relation on the set of geodesic rays in $X$.

Definition 2.44. The *ideal boundary* of a metric space $X$ is the collection of equivalence classes of geodesic rays. It is usually denoted either by $\partial_\infty X$ or by $X(\infty)$.

An equivalence class $\alpha \in \partial_\infty X$ is called an *ideal point* or *point at infinity* of $X$, and the fact that a geodesic ray $\rho$ is contained in this class is sometimes expressed by the equality $\rho(\infty) = \alpha$.

The space of geodesic rays in $X$ has a natural compact-open topology, or, equivalently, topology of uniform convergence on compacts (recall that we regard geodesic rays as maps from $[0, \infty)$ to $X$). Thus, we topologize $\partial_\infty X$ by giving it the quotient topology $\tau$.

Exercise 2.45. Every isometry $g : X \to X$ induces a homeomorphism $g_\infty : \partial_\infty X \to \partial_\infty X$.

This exercise explains why we consider rays emanating from different points of $X$: otherwise most isometries of $X$ would not act on $\partial_\infty X$.

*Convention.* From now on, in this section, we assume that $X$ is a complete $CAT(0)$ metric space.

Lemma 2.46. *If $X$ is locally compact then for every point $x \in X$ and every point $\alpha \in \partial_\infty X$ there exists a unique geodesic ray $\rho$ with $\rho(0) = x$ and $\rho(\infty) = \alpha$. We will also use the notation $[x, \alpha)$ for the ray $\rho$.*

Proof. Let $r : [0, \infty) \to X$ be a geodesic ray with $r(\infty) = \alpha$. For every $n \in \mathbb{N}$, according to Corollary 2.40, there exists a unique geodesic $\mathfrak{g}_n$ joining $x$ and $r(n)$. The convexity of the distance function implies that every $\mathfrak{g}_n$ is at Hausdorff distance $\text{dist}(x, r(0))$ from the segment of $r$ between $r(0)$ and $r(n)$.

By the Arzela-Ascoli Theorem, a subsequence $\mathfrak{g}_{n_k}$ of geodesic segments converges in the compact-open topology to a geodesic ray $\rho$ with $\rho(0) = x$. Moreover, $\rho$ is at Hausdorff distance $\text{dist}(x, r(0))$ from $r$.

Assume that $\rho_1$ and $\rho_2$ are two asymptotic geodesic rays with $\rho_1(0) = \rho_2(0) = x$. Let $M$ be such that $\text{dist}(\rho_1(t), \rho_2(t)) \leqslant M$, for every $t \geqslant 0$. Consider $t \in [0, \infty)$, and $\varepsilon > 0$ arbitrarily small. Convexity of the distance function implies that

$$\text{dist}(\rho_1(t), \rho_2(t)) \leqslant \varepsilon \text{dist}(\rho_1(t/\varepsilon), \rho_2(t/\varepsilon)) \leqslant \varepsilon M .$$

It follows that $\mathrm{dist}(\rho_1(t), \rho_2(t)) = 0$ and, hence, $\rho_1 = \rho_2$. □

In particular, for a fixed point $p \in X$ one can identify the set $\bar{X} := X \sqcup \partial_\infty X$ with the set of geodesic segments and rays with initial point $p$. In what follows, we will equip $\bar{X}$ with the topology induced from the compact-open topology on the space of these segments and rays.

EXERCISE 2.47.
(1) Prove that the embedding $X \hookrightarrow \bar{X}$ is a homeomorphism to its image.
(2) Prove that the topology on $\bar{X}$ is independent of the chosen basepoint $p$. In other words, given $p$ and $q$ two points in $X$, the map $[p, x] \mapsto [q, x]$ (with $x \in \bar{X}$) is a homeomorphism.
(3) In the special case when $X$ is a Hadamard manifold, show that for every point $p \in X$, the ideal boundary $\partial_\infty X$ is homeomorphic to the unit sphere $S$ in the tangent space $T_pM$ *via* the map $v \in S \subset T_pM \to \exp_p(\mathbb{R}_+ v) \in \partial_\infty X$.

An immediate consequence of the Arzela–Ascoli Theorem is that $\bar{X}$ is compact.

Consider a geodesic ray $r : [0, \infty) \to X$, and an arbitrary point $x \in X$. The function $t \mapsto \mathrm{dist}(x, r(t)) - t$ is decreasing (due to the triangle inequality) and bounded from below by $-\mathrm{dist}(x, r(0))$. Therefore, there exists a limit

$$f_r(x) := \lim_{t\to\infty} [\mathrm{dist}(x, r(t)) - t] \,. \tag{2.4}$$

DEFINITION 2.48. The function $f_r : X \to \mathbb{R}$ thus defined, is called the *Busemann function for the ray $r$*.

For the proof of the next lemma see e.g. **[Bal95]**, Chapter 2, Proposition 2.5.

LEMMA 2.49. *If $r_1$ and $r_2$ are two asymptotic rays then $f_{r_1} - f_{r_2}$ is a constant function.*

In particular, it follows that the collections of sublevel sets and the level sets of a Busemann function do not depend on the ray $r$, but only on the point at infinity that $r$ represents.

EXERCISE 2.50. Show that $f_r$ is linear with slope $-1$ along the ray $r$. In particular,

$$\lim_{t\to\infty} f_r(t) = -\infty.$$

DEFINITION 2.51. A sublevel set of a Busemann function, $f_r^{-1}(-\infty, a]$ is called a *(closed) horoball with center (or footpoint)* $\alpha = r(\infty)$; we sometime denote such a set $\overline{B}(\alpha)$. A level set $f_r^{-1}(a)$ of a Busemann function is called a *horosphere with footpoint* $\alpha$, it is denoted $H(\alpha)$. Lastly, an open sublevel set $f_r^{-1}(-\infty, a)$ is called an open horoball with footpoint $\alpha = r(\infty)$, and denoted $B(\alpha)$.

LEMMA 2.52. *Let $r$ be a geodesic ray and let $B$ be the open horoball $f_r^{-1}(-\infty, 0)$. Then $B = \bigcup_{t\geqslant 0} B(r(t), t)$.*

Proof. Indeed, if for a point $x$, $f_r(x) = \lim_{t\to\infty}[\mathrm{dist}(x, r(t)) - t] < 0$, then for some sufficiently large $t$, $\mathrm{dist}(x, r(t)) - t < 0$. Whence $x \in B(r(t), t)$.

Conversely, suppose that $x \in X$ is such that for some $s \geqslant 0$, $\mathrm{dist}(x, r(s)) - s = \delta_s < 0$. Then, because the function $t \mapsto \mathrm{dist}(x, r(t)) - t$ is decreasing, it follows that for every $t \geqslant s$,

$$\mathrm{dist}(x, r(t)) - t \leqslant \delta_s\,.$$

Whence, $f_r(x) \leqslant \delta_s < 0$. □

Lemma 2.53. *Let $X$ be a $CAT(0)$ space. Then every Busemann function on $X$ is convex and 1-Lipschitz.*

Proof. Note that distance function on any metric space is 1-Lipschitz (by the triangle inequality). Since Busemann functions are limits of normalized distance functions, it follows that Busemann functions are 1-Lipschitz as well. (This part does not require $CAT(0)$ assumption.) Similarly, since distance function is convex, Busemann functions are also convex as limits of normalized distance functions. □

Furthermore, if $X$ is a Hadamard manifold, then every Busemann function $f_r$ is smooth, with gradient of constant norm 1, see [**BGS85**].

Lemma 2.54. *Assume that $X$ is a complete $CAT(0)$ space. Then:*

- *Open and closed horoballs in $X$ are convex sets.*
- *A closed horoball is the closure of an open horoball.*

Proof. The first property follows immediately from the convexity of Busemann functions. Let $f = f_r$ be a Busemann function. Consider the closed horoball

$$\bar{B} = \{x : f(x) \leqslant t\}.$$

Since this horoball is a closed subset of $X$, it contains the closure of the open horoball

$$B = \{x : f(x) < t\}.$$

Suppose now that $f(x) = t$. Since $\lim_{s\to\infty} f(s) = -\infty$, there exists $s$ such that $f(r(s)) < t$. Convexity of $f$ implies that

$$f(y) < f(x) = t, \quad \forall y \in [x, r(s)] \setminus \{x\}.$$

Therefore, $x$ belongs to the closure of the open horoball $B$, which implies that $\bar{B}$ is the closure of $B$. □

Exercise 2.55. 1. Suppose that $X$ is the Euclidean space $\mathbb{R}^n$, $r$ is the geodesic ray in $X$ with $r(0) = 0$ and $r'(0) = u$, where $u$ is a unit vector. Show that

$$f_r(x) = -x \cdot u.$$

In particular, closed (resp. open) horoballs in $X$ are closed (resp. open) half-spaces, while horospheres are hyperplanes.

2. Construct an example of a proper $CAT(0)$ space and an open horoball $B \subset X$, $B \neq X$, so that $B$ is not equal to the interior of the closed horoball $\bar{B}$. Can this happen in the case of Hadamard manifolds?

## 2.2. Bounded geometry

In this section we review several notions of bounded geometry for metric spaces.

### 2.2.1. Riemannian manifolds of bounded geometry.

Definition 2.56. We say a Riemannian manifold $M$ has *bounded geometry* if it is connected, it has uniform upper and lower bounds for the sectional curvature, and a uniform lower bound for the injectivity radius $InjRad(x)$ (see Section 2.1.3).

Probably the correct terminology should be "uniformly locally bounded geometry", but we prefer shortness to an accurate description.

A connected Riemannian manifold without boundary, so that the isometry group of $M$ acts cocompactly on $M$ (see section 3.1.1), has bounded geometry.

Remark 2.57. In the literature the condition of bounded geometry on a Riemannian manifold is usually weaker, e.g. that there exists $L \geqslant 1$ and $R > 0$ such that every ball of radius $R$ in $M$ is $L$-bi-Lipschitz equivalent to the ball of radius $R$ in $\mathbb{R}^n$ (**[Gro93]**, §0.5.$A_3$) or that the Ricci curvature has a uniform lower bound (**[Cha06]**, **[Cha01]**).

For the purposes of this book, the restricted condition in Definition 2.56 suffices.

In what follows we keep the notation $V_\kappa(r)$ from Theorem 2.24 to designate the volume of a ball of radius $r$ in the $n$–dimensional space of constant sectional curvature $\kappa$.

Lemma 2.58. *Let $M$ be complete $n$–dimensional Riemannian manifold with bounded geometry, let $a \leqslant b$ and $\rho > 0$ be such that the sectional curvature is at least $a$ and at most $b$, and that at every point the injectivity radius is larger than $\rho$.*

(1) *For every $\delta > 0$, every $\delta$–separated set in $M$ is $\phi$-uniformly discrete, with $\phi(r) = \frac{V_a(r+\lambda)}{V_b(\lambda)}$, where $\lambda$ is the minimum of $\frac{\delta}{2}$ and $\rho$.*

(2) *For every $2\rho > \delta > 0$ and every maximal $\delta$–separated set $N$ in $M$, the multiplicity of the covering $\{B(x, \delta) \mid x \in N\}$ is at most $\frac{V_a\left(\frac{3\delta}{2}\right)}{V_b\left(\frac{\delta}{2}\right)}$.*

Proof. (1) Let $S$ be a $\delta$–separated subset in $M$.

According to Theorem 2.24, for every point $x \in S$ and radius $r > 0$ we have:

$$V_a(r + \lambda) \geqslant Vol\,[B_M(x, r + \lambda)] \geqslant \text{card}\,\left[\overline{B}(x, r) \cap S\right]\, V_b(\lambda)\,.$$

This implies that card $\left[\overline{B}(x, r) \cap S\right] \leqslant \frac{V_a(r+\lambda)}{V_b(\lambda)}$, whence $S$ with the induced metric is $\phi$-uniformly discrete, with the required $\phi$.

(2) Let $F$ be a subset in $N$ such that $\bigcap_{x \in F} B(x, \delta)$ is non-empty. Let $y$ be a point in this intersection. Then the ball $B\left(y, \frac{3\delta}{2}\right)$ contains the disjoint union $\bigsqcup_{x \in F} B\left(x, \frac{\delta}{2}\right)$, whence

$$V_a\left(\frac{3\delta}{2}\right) \geqslant Vol\left[B_M\left(y, \frac{3\delta}{2}\right)\right] \geqslant \text{card}\, F\, V_b\left(\frac{\delta}{2}\right)\,.$$

□

**2.2.2. Metric simplicial complexes of bounded geometry.** Let $X$ be a simplicial complex and $d$ a path-metric on $X$. Then $(X, d)$ is said to be a *metric simplicial complex* if the restriction of $d$ to each simplex is isometric to a Euclidean simplex. The main example of a metric simplicial complex is a generalization of a graph with the standard metric described below.

Let $X$ be a connected simplicial complex. As usual, we will often conflate $X$ and its geometric realization. Metrize each $k$-simplex of $X$ to be isometric to the standard $k$-simplex $\Delta^k$ in the Euclidean space:

$$\Delta^k = (\mathbb{R}_+)^{k+1} \cap \{x_0 + \ldots + x_n = 1\}.$$

Thus, for each $m$-simplex $\sigma^m$ and its face $\sigma^k$, the inclusion $\sigma^k \to \sigma^m$ is an isometric embedding. This allows us to define a length-metric on $X$ so that each simplex is isometrically embedded in $X$, similarly to the definition of the standard metric on a graph. Namely, a *piecewise-linear (PL) path* $\mathfrak{p}$ in $X$ is a path $\mathfrak{p} : [a, b] \to X$, whose domain can be subdivided in finitely many intervals $[a_i, a_{i+1}]$ so that $\mathfrak{p}|[a_i, a_{i+1}]$ is a piecewise-linear path whose image is contained in a single simplex of $X$. Lengths of such paths are defined using metric on simplices of $X$. Then

$$d(x, y) = \inf_{\mathfrak{p}} \operatorname{length}(\mathfrak{p})$$

where the infimum is taken over all PL paths in $X$ connecting $x$ to $y$. The metric $d$ is then a path-metric; we call this metric the *standard metric* on $X$.

Exercise 2.59. Verify that the standard metric is complete and that $X$ is a geodesic metric space.

Definition 2.60. A metric simplicial complex $X$ has *bounded geometry* if it is connected and if there exist $L \geqslant 1$ and $N < \infty$ so that:

- every vertex of $X$ is incident to at most $N$ edges;
- the length of every edge is in the interval $[L^{-1}, L]$.

In particular, the set of vertices of $X$ with the induced metric is a uniformly discrete metric space.

Thus, a metric simplicial complex of bounded geometry is necessarily finite-dimensional.

Examples 2.61.

- If $Y$ is a finite connected metric simplicial complex, then its universal cover (with the pull-back path metric) has bounded geometry.
- A connected simplicial complex has bounded geometry if and only if there is a uniform bound on the valency of the vertices in its 1-skeleton.

Metric simplicial complexes of bounded geometry appear naturally in the context of Riemannian manifolds with bounded geometry. Given a simplicial complex $X$, we will equip it with the *standard metric*, where each simplex is isometric to a Euclidean simplex with unit edges.

Theorem 2.62 (See Theorem 1.14, [**Att94**]). *Let $M$ be an $n$-dimensional Riemannian manifold of bounded geometry with geometric bounds $a, b, \epsilon$. Then $M$ admits a triangulation $X$ of bounded geometry (whose geometric bounds depend only on $n, a, b, \epsilon$) and an $L$-bi-Lipschitz homeomorphism $f : X \to M$, where $L = L(n, a, b, \epsilon)$.*

Another procedure of approximation of Riemannian manifolds by simplicial complexes will be described in Section 5.3.

CHAPTER 3

# Algebraic preliminaries

## 3.1. Geometry of group actions

**3.1.1. Group actions.** Let $G$ be a group or a semigroup and $E$ be a set. An *action of $G$ on $E$ on the left* is a map

$$\mu : G \times E \to E, \quad \mu(g, a) = g(a),$$

so that

(1) $\mu(1, x) = x$ ;
(2) $\mu(g_1 g_2, x) = \mu(g_1, \mu(g_2, x))$ for all $g_1, g_2 \in G$ and $x \in E$ .

REMARK 3.1. If, in addition, $G$ is a group, then the two properties above imply that

$$\mu(g, \mu(g^{-1}, x)) = x$$

for all $g \in G$ and $x \in E$ .

An *action of $G$ on $E$ on the right* is a map

$$\mu : E \times G \to E, \quad \mu(a, g) = (a)g,$$

so that

(1) $\mu(x, 1) = x$ ;
(2) $\mu(x, g_1 g_2) = \mu(\mu(x, g_1), g_2)$ for all $g_1, g_2 \in G$ and $x \in E$ .

Note that the difference between an action on the left and an action on the right is the order in which the elements of a product act.

If not specified, an action of a group $G$ on a set $E$ is always to the left, and it is often denoted $G \curvearrowright E$.

If $E$ is a metric space, an *isometric action* is an action so that $\mu(g, \cdot)$ is an isometry of $E$ for each $g \in G$.

A group action $G \curvearrowright X$ is called *free* if for every $x \in X$, the *stabilizer of $x$ in $G$*,

$$G_x = \{g \in G : g(x) = x\}$$

is $\{1\}$.

Given an action $\mu : G \curvearrowright X$, a map $f : X \to Y$ is called *$G$–invariant* if

$$f\left(\mu(g, x)\right) = f(x), \quad \forall g \in G, x \in X.$$

Given two actions $\mu : G \curvearrowright X$ and $\nu : G \curvearrowright Y$, a map $f : X \to Y$ is called *$G$–equivariant* if

$$f\left(\mu(g, x)\right) = \nu(g, f(x)), \quad \forall g \in G, x \in X.$$

In other words, for each $g \in G$ we have a commutative diagram,

$$\begin{array}{ccc} X & \xrightarrow{g} & X \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ Y & \xrightarrow{g} & Y \end{array}$$

A *topological group* is a group $G$ equipped with the structure of a topological space, so that the group operations (multiplication and inversion) are continuous maps. If $G$ is a group without specified topology, we will always assume that $G$ is *discrete*, i.e., is given the discrete topology.

If $G$ is a topological group and $E$ is a topological space, a *continuous action* of $G$ on $E$ is a continuous map $\mu$ satisfying the above *action* axioms.

A topological group action $\mu : G \curvearrowright X$ is called *proper* if for every compact subsets $K_1, K_2 \subset X$, the set

$$G_{K_1,K_2} = \{g \in G : g(K_1) \cap K_2 \neq \emptyset\} \subset G$$

is compact. If $G$ has discrete topology, a proper action is called *properly discontinuous* action, as $G_{K_1,K_2}$ is finite.

Exercise 3.2. Suppose that $X$ is locally compact and $G \curvearrowright X$ is proper. Show that the quotient $X/G$ is Hausdorff.

A topological action $G \curvearrowright X$ is called *cocompact* if there exists a compact $C \subset X$ so that

$$G \cdot C := \bigcup_{g \in G} gC = X.$$

Exercise 3.3. If $G \curvearrowright X$ is cocompact then $X/G$ (equipped with the quotient topology) is compact.

The following is a converse to the above exercise:

Lemma 3.4. *Suppose that $X$ is locally compact and $G \curvearrowright X$ is such that $X/G$ is compact. Then $G$ acts cocompactly on $X$.*

Proof. Let $p : X \to Y = X/G$ be the quotient. For every $x \in X$ choose a relatively compact (open) neighborhood $U_x \subset X$ of $x$. Then the collection

$$\{p(U_x)\}_{x \in X}$$

is an open covering of $Y$. Since $Y$ is compact, this open covering has a finite subcovering

$$\{p(U_{x_i} : i = 1, \ldots, n\}$$

The union

$$C := \bigcup_{i=1}^{n} cl(U_{x_i})$$

is compact in $X$ and projects onto $Y$. Hence, $G \cdot C = X$. □

In the context of non-proper metric space the concept of cocompact group action is replaced with the one of *cobounded action.* An isometric action $G \curvearrowright X$ is called *cobounded* if there exists $D < \infty$ such that for some point $x \in X$,

$$\bigcup_{g \in G} g(B(x, D)) = X.$$

Equivalently, given any pair of points $x, y \in X$, there exists $g \in G$ such that $\text{dist}(g(x), y) \leqslant 2D$. Clearly, if $X$ is proper, the action $G \curvearrowright X$ is cobounded if and only if it is cocompact. We call a metric space $X$ *quasi-homogeneous* if the action $\text{Isom}(X) \curvearrowright X$ is cobounded.

Similarly, we have to modify the notion of a properly discontinuous action: An isometric action $G \curvearrowright X$ on a metric space is called *properly discontinuous* if for every bounded subset $B \subset X$, the set

$$G_{B,B} = \{g \in G : g(B) \cap B \neq \emptyset\}$$

is finite. Assigning two different meaning to the same notation of course, creates ambiguity, the way out of this conundrum is to think of the concept of proper discontinuity applied to different categories of actions: Topological and isometric. In the former case we use compact subsets, in the latter case we use bounded subsets. For proper metric spaces, both definitions, of course, are equivalent.

**3.1.2. Lie groups.** References for this section are [**Hel01, OV90, War83**].

A *Lie group* is a group $G$ which has structure of a smooth manifold, so that the binary operation $G \times G \to G$ and inversion $g \mapsto g^{-1}, G \to G$ are smooth. Actually, every Lie group $G$ can be made into a real analytic manifold with real analytic group operations. We will assume that $G$ is a real $n$-dimensional manifold, although one can also consider *complex Lie groups*.

EXAMPLE 3.5. Groups $GL(n, \mathbb{R}), SL(n, \mathbb{R})$, $O(n)$, $O(p, q)$ are (real) Lie groups. Every countable discrete group (a group with discrete topology) is a Lie group.

Here $O(p, q)$ is the group of linear isometries of the quadratic form

$$x_1^2 + \ldots x_p^2 - x_{p+1}^2 - \ldots - x_{p+q}^2$$

of signature $(p, q)$. The most important, for us, case is $O(n, 1) \cong O(1, n)$. The group $PO(n, 1) = O(n, 1)/\pm I$ is the group of isometries of the hyperbolic $n$-space.

EXERCISE 3.6. Show that the group $PO(n, 1)$ embeds in $O(n, 1)$ as the subgroup stabilizing the *future light cone*

$$x_1^2 + \ldots + x_n^2 - x_{n+1}^2 > 0, \quad x_{n+1} > 0.$$

The tangent space $V = T_e G$ of a Lie group $G$ at the identity element $e \in G$ has structure of a Lie algebra, called the *Lie algebra* $\mathfrak{g}$ of the group $G$.

EXAMPLE 3.7. 1. The Lie algebra $\mathfrak{sl}(n, \mathbb{C})$ of $SL(n, \mathbb{C})$ consists of trace-free $n \times n$ complex matrices. The Lie bracket operation on $\mathfrak{sl}(n, \mathbb{C})$ is given by

$$[A, B] = AB - BA.$$

2. The Lie algebra of the unitary subgroup $U(n) < GL(n, \mathbb{C})$ equals the space of skew-hermitian matrices

$$\mathfrak{u}(n) = \{A \in Mat_n(\mathbb{C}) : A = -A^*\}.$$

3. The Lie algebra of the orthogonal subgroup $O(n) < GL(n, \mathbb{R})$ equals the space of skew-symmetric matrices

$$\mathfrak{o}(n) = \{A \in Mat_n(\mathbb{R}) : A = -A^T\}.$$

EXERCISE 3.8. $\mathfrak{u}(n) \oplus i\mathfrak{u}(n) = Mat_n(\mathbb{C})$, is the Lie algebra of the group $GL(n, \mathbb{C})$.

Theorem 3.9. *For every finite-dimensional real Lie algebra $\mathfrak{g}$ there exists unique, up to isomorphism, simply-connected Lie group $G$ whose Lie algebra is isomorphic to $\mathfrak{g}$.*

Every Lie group $G$ has a left-invariant Riemannian metric. Indeed, pick a positive-definite inner product $\langle \cdot, \cdot \rangle_e$ on $V = T_e G$. For every $g \in G$ we consider the left multiplication $L_g : G \to G, L_g(x) = gx$. Then $L_g : G \to G$ is a smooth map and the action of $G$ on itself *via* left multiplication is simply-transitive. We define the inner product $\langle \cdot, \cdot \rangle_g$ on $T_g G$ as the image of $\langle \cdot, \cdot \rangle_e$ under the derivative $Dg : T_e G \to T_g G$.

Every Lie group $G$ acts on itself *via* inner automorphisms

$$\rho(g)(x) = gxg^{-1}.$$

This action is smooth and the identity element $e \in G$ is fixed by the entire group $G$. Therefore $G$ acts linearly on the tangent space $V = T_e G$ at the identity $e \in G$. The action of $G$ on $V$ is called *the adjoint representation of the group $G$* and denoted by Ad. Therefore we have the homomorphism

$$Ad : G \to GL(V).$$

Lemma 3.10. *For every connected Lie group $G$ the kernel of $Ad : G \to GL(V)$ is contained in the center of $G$.*

Proof. There is a local diffeomorphism

$$\exp : V \to G$$

called *the exponential map* of the group $G$, sending $0 \in V$ to $e \in G$. In the case when $G = GL(n, \mathbb{R})$ this map is the ordinary matrix exponential map. The map exp satisfies the identity

$$g \exp(v) g^{-1} = \exp(\mathrm{Ad}(g)v), \quad \forall v \in V, g \in G.$$

Thus, if $\mathrm{Ad}(g) = Id$ then $g$ commutes with every element of $G$ of the form $\exp(v), v \in V$. The set of such elements is open in $G$. Now, if we are willing to use a real analytic structure on $G$ then it would immediately follow that $g$ belongs to the center of $G$. Below is an alternative argument. Let $g \in \mathrm{Ker}(Ad)$. The centralizer $Z(g)$ of $g$ in $G$ is given by the equation

$$Z(g) = \{h \in G : [h, g] = 1\}.$$

Since the commutator is a continuous map, $Z(g)$ is a closed subgroup of $G$. Moreover, as we observed above, this subgroup has nonempty interior in $G$ (containing $e$). Since $Z(g)$ acts transitively on itself by, say, left multiplication, $Z(g)$ is open in $G$. As $G$ is connected, we conclude that $Z(g) = G$. Therefore kernel of $Ad$ is contained in the center of $G$. □

Theorem 3.11 (E. Cartan). *Every closed subgroup $H$ of a Lie group $G$ has structure of a Lie group so that the inclusion $H \hookrightarrow G$ is an embedding of smooth manifolds.*

A Lie group $G$ is called *simple* if $G$ contains no connected proper normal subgroups. Equivalently, a Lie group $G$ is simple if its Lie algebra $\mathfrak{g}$ is simple, i.e., $\mathfrak{g}$ is nonabelian and contains no ideals.

Example 3.12. The group $SL(2,\mathbb{R})$ is simple, but its center is isomorphic to $\mathbb{Z}_2$.

Thus, a simple Lie group need not be simple as an abstract group. A Lie group $G$ is *semisimple* if its Lie algebra splits as a direct sum

$$\mathfrak{g} = \oplus_{i=1}^{k} \mathfrak{g}_i,$$

where each $\mathfrak{g}_i$ is a simple Lie algebra.

### 3.1.3. Haar measure and lattices.

Definition 3.13. A (left) Haar measure on a topological group $G$ is a countably additive, nontrivial measure $\mu$ on Borel subsets of G satisfying:

(1) $\mu(gE) = \mu(E)$ for every $g \in G$ and every Borel subset $E \subset G$.
(2) $\mu(K)$ is finite for every compact $K \subset G$.
(3) Every Borel subset $E \subset G$ is *outer regular*:

$$\mu(E) = \inf\{\mu(U) : E \subset U,\ U \text{ is open in } G\}$$

(4) Every open set $E \subset G$ is *inner regular*:

$$\mu(E) = \sup\{\mu(U) : U \subset E,\ U \text{ is open in } G\}$$

By Haar's Theorem, see [**Bou63**], every locally compact topological group $G$ admits a Haar measure and this measure is unique up to scaling. Similarly, one defines right-invariant Haar measures. In general, left and right Haar measures are not the same, but they are for some important classes of groups:

Definition 3.14. A locally compact group $G$ is *unimodular* if left and right Haar measures are constant multiples of each other.

Important examples of Haar measures come from Riemannian geometry. Let $X$ be a homogeneous Riemannian manifold, $G$ is the isometry group. Then $X$ has a natural measure $\omega$ defined by the volume form of the Riemannian metric on $X$. We have the natural surjective map $G \to X$ given by the orbit map $g \mapsto g(o)$, where $o \in X$ is a base-point. The fibers of this map are stabilizers $G_x$ of points $x \in X$. Arzela-Ascoli theorem implies that each subgroup $G_x$ is compact. Transitivity of the action $G \curvearrowright X$ implies that all the subgroups $G_x$ are conjugate. Setting $K = G_o$, we obtain the identification $X = G/K$. Now, let $\mu$ be the pull-back of $\omega$ under the projection map $G \to X$. By construction, $\mu$ is left-invariant (since the metric on $X$ is $G$–invariant).

Definition 3.15. Let $G$ be a topological group with finitely many connected components and $\mu$ a Haar measure on $G$. A *lattice* in $G$ is a discrete subgroup $\Gamma < G$ so that the quotient $Q = \Gamma\backslash G$ admits a finite $G$–invariant measure (for the action to the right of $G$ on $Q$) induced by the Haar measure. A lattice $\Gamma$ is called *uniform* if the quotient $Q$ is compact.

If $G$ is a Lie group then the measure above can also be obtained by taking a Riemannian metric on $G$ which is left-invariant under $G$ and right-invariant under $K$, the maximal compact subgroup of $G$. Note that when $G$ is unimodular, the volume form thus obtained is also right-invariant under $G$.

Thus if one considers the quotient $X := G/K$, then $X$ has a Riemannian metric which is (left) invariant under $G$. Hence, $\Gamma$ is a lattice if and only if $\Gamma$ acts on $X$

properly discontinuously so that $vol(\Gamma \setminus X)$ is finite. Note that the action of $\Gamma$ on $X$ is *a priori* not free.

THEOREM 3.16. *A locally compact second countable group $G$ is unimodular provided that it contains a lattice.*

PROOF. For arbitrary $g \in G$ consider the push-forward $\nu = R_g(\mu)$ of the (left) Haar measure $\mu$ on $G$; here $R_g$ is the right multiplication by $g$:

$$\nu(E) = \mu(Eg).$$

Then $\nu$ is also a left Haar measure on $G$. By the uniqueness of Haar measure, $\nu = c\mu$ for some constant $c > 0$.

LEMMA 3.17. *Every discrete subgroup $\Gamma < G$ admits a* measurable fundamental set*, i.e., a measurable subset of $D \subset G$ such that*

$$\bigcup_{\gamma \in \Gamma} \gamma D = G, \quad \mu(\gamma D \cap D) = 0, \quad \forall \gamma \in \Gamma \setminus 1.$$

PROOF. Since $\Gamma < G$ is discrete, there exists an open neighborhood $V$ of $1 \in G$ such that $\Gamma \cap V = \{1\}$. Let $U \subset V$ be another open neighborhood of $1 \in G$ such that $UU^{-1} \subset V$. Then for $\gamma \in \Gamma$ we have

$$\gamma u = u', u \in U, u' \in U \Rightarrow \gamma = u'u^{-1} \in U \Rightarrow \gamma = 1.$$

In other words, $\Gamma$-images of $U$ are pairwise disjoint. Since $G$ is a second countable, there exists a countable subset

$$E = \{g_i \in G : i \in \mathbb{N}\}$$

so that

$$G = \bigcup_i U g_i.$$

Clearly, each set

$$W_n := U g_n \setminus \bigcup_{i<n} \Gamma U g_i$$

is measurable, and so is their union

$$D = \bigcup_{n=1}^{\infty} W_n.$$

Let us verify that $D$ is a measurable fundamental set. First, note that for every $x \in G$ there exists the least $n$ such that $x \in U g_n$. Therefore,

$$G = \bigcup_{n=1}^{\infty} \left( U g_n \setminus \bigcup_{i<n} U g_i \right).$$

Next,

$$\Gamma \cdot D = \bigcup_{n=1}^{\infty} \left( \Gamma U g_n \setminus \bigcup_{i<n} \Gamma U g_i \right) =$$

$$\Gamma \cdot \bigcup_{n=1}^{\infty} \left( U g_n \setminus \bigcup_{i<n} U g_i \right) \supset \bigcup_{n=1}^{\infty} \left( U g_n \setminus \bigcup_{i<n} U g_i \right) = G.$$

Therefore, $\Gamma \cdot D = G$. Next, suppose that

$$x \in \gamma D \cap D.$$

Then, for some $n$, $m$

$$x \in W_n \cap \gamma W_m.$$

If $m < n$ then

$$\gamma W_m \subset \Gamma \bigcup_{i<n} U g_i$$

which is disjoint from $W_n$, a contradiction. Thus, $W_n \cap \gamma W_m = \emptyset$ for $m < n$ and all $\gamma \in \Gamma$. If $n < m$ then

$$W_n \cap \gamma W_m = \gamma^{-1} (\gamma W_n \cap W_m) = \emptyset.$$

Thus, $n = m$, which implies that

$$U g_n \cap \gamma U g_n \neq \emptyset \Rightarrow U \cap \gamma U \neq \emptyset \Rightarrow \gamma = 1.$$

Thus, for all $\gamma \in \Gamma \setminus \{1\}$, $\gamma D \cap D = \emptyset$. ☐

Let $D \subset G$ be a measurable fundamental set for a lattice $\Gamma < G$. Then

$$0 < \mu(D) = \mu(\Gamma \backslash G) < \infty$$

since $\Gamma$ is a lattice. For every $g \in G$, $Dg$ is again a fundamental set for $\Gamma$ and, thus, $\mu(D) = \mu(Dg)$. Hence, $\mu(D) = \mu(Dg) = c\nu(D)$. It follows that $c = 1$. Thus, $\mu$ is also a right Haar measure. ☐

**3.1.4. Geometric actions.** Suppose now that $X$ is a metric space. We will equip the group of isometries $\mathrm{Isom}(X)$ of $X$ with the *compact-open topology*, equivalent to the topology of uniform convergence on compact sets. A subgroup $G \subset \mathrm{Isom}(X)$ is called *discrete* if it is discrete with respect to the subset topology.

EXERCISE 3.18. Suppose that $X$ is proper. Show that the following are equivalent for a subgroup $G \subset \mathrm{Isom}(X)$:

a. $G$ is discrete.

b. The action $G \curvearrowright X$ is properly discontinuous.

c. For every $x \in X$ and an infinite sequence $g_i \in G$, $\lim_{i\to\infty} d(x, g_i(x)) = \infty$.

Hint: Use Arzela–Ascoli theorem.

DEFINITION 3.19. A *geometric action* of a group $G$ on a metric space $X$ is an isometric properly discontinuous cobounded action $G \curvearrowright X$.

For instance, if $X$ is a homogeneous Riemannian manifold with the isometry group $G$ and $\Gamma < G$ is a uniform lattice, then $\Gamma$ acts geometrically on $X$. Note that every geometric action on a proper metric space is cocompact.

LEMMA 3.20. *Suppose that a group $G$ acts geometrically on a proper metric space $X$. Then $G\backslash X$ has a metric defined by*

$$\mathrm{dist}(\bar{a}, \bar{b}) = \inf\{\mathrm{dist}(p, q)\ ;\ p \in Ga\,,\ q \in Gb\} = \inf\{\mathrm{dist}(a, q)\ ;\ q \in Gb\}\,, \tag{3.1}$$

*where $\bar{a} = Ga$ and $\bar{b} = Gb$.*

*Moreover, this metric induces the quotient topology of $G\backslash X$.*

Proof. The infimum in (3.1) is attained, i.e. there exists $g \in G$ such that

$$\mathrm{dist}(\bar{a}, \bar{b}) = \mathrm{dist}(a, gb).$$

Indeed, take $g_0 \in G$ arbitrary, and let $R = \mathrm{dist}(a, gb)$. Then

$$\mathrm{dist}(\bar{a}, \bar{b}) = \inf\{\mathrm{dist}(a, q) \ ; \ q \in Gb \cap \overline{B}(a, R)\}.$$

Now, for every $gb \in \overline{B}(a, R)$,

$$gg_0^{-1}\overline{B}(a, R) \cap \overline{B}(a, R) \neq \emptyset.$$

Since $G$ acts properly discontinuously on $X$, this implies that the set $Gb \cap \overline{B}(a, R)$ is finite, hence the last infimum is over a finite set, and it is attained. We leave it to the reader to verify that dist is the Hausdorff distance between the orbits $G \cdot a$ and $G \cdot b$. Clearly the projection $X \to G\backslash X$ is a contraction. One can easily check that the topology induced by the metric dist on $G\backslash X$ coincides with the quotient topology. □

## 3.2. Complexes and group actions

**3.2.1. Simplicial complexes.** As we expect the reader to be familiar with basics of algebraic topology, we will discuss simplicial complexes and (in the next section) cell complexes only very briefly.

We will use the notation $X^{(i)}$ to denote the $i$-th skeleton of the simplicial complex $X$. A *gallery* in an $n$-dimensional simplicial complex $X$ is a chain of $n$-simplices $\sigma_1, \ldots, \sigma_k$ so that $\sigma_i \cap \sigma_{i+1}$ is an $n-1$-simplex for every $i = 1, \ldots, k-1$.

Let $\sigma, \tau$ be simplices of dimensions $m$ and $n$ respectively with the vertex sets

$$\sigma^{(0)} = \{v_0, \ldots, v_m\}, \quad \tau^{(0)} = \{w_0, \ldots, w_n\}$$

The product $\sigma \times \tau$, of course, is not a simplex (unless $nm = 0$), but it admits a standard triangulation, whose vertex set is

$$\sigma^{(0)} \times \tau^{(0)}.$$

This triangulation is defined as follows. Pairs $u_{ij} = (v_i, w_j)$ are the vertices of $\sigma \times \tau$. Distinct vertices

$$(u_{i_0, j_0}, \ldots, u_{i_k, j_k})$$

span a $k$-simplex in $\sigma \times \tau$ if and only if $j_0 \leqslant \ldots \leqslant j_k$.

A homotopy between simplicial maps $f_0, f_1 : X \to Y$ is a simplicial map $F : X \times I \to Y$ which restricts to $f_i$ on $X \times \{i\}, i = 0, 1$. The *tracks* of the homotopy $F$ are the paths $\mathfrak{p}(t) = F(x, t), x \in X$.

Let $X$ be a simplicial complex. Recall that besides usual cohomology groups $H^*(X; A)$ (with coefficients in a ring $A$ that the reader can assume to be $\mathbb{Z}$ or $\mathbb{Z}_2$), we also have *cohomology with compact support* $H_c^*(X, A)$ which are defined as follows. Consider the usual cochain complex $C^*(X; A)$. We say that a cochain $\sigma \in C^*(X; A)$ has *compact support* if it vanishes outside of a finite subcomplex in $X$. Thus, in each chain group $C^k(X; A)$ we have the subgroup $C_c^k(X; A)$ consisting of compactly supported cochains. Then the usual coboundary operator $\delta$ satisfies

$$\delta : C_c^k(X; A) \to C_c^{k+1}(X; A).$$

The cohomology of the new cochain complex $(C_c^*(X; A), \delta)$ is denoted $H_c^*(X; A)$ and is called *cohomology of $X$ with compact support*. Maps of simplicial complexes no longer induce homomorphisms of $H_c^*(X;A)$ since they do not preserve the compact support property of cochains; however, *proper* maps of simplicial complexes do induce natural maps on $H_c^*$. Similarly, maps which are *properly* homotopic induce equal homomorphisms of $H_c^*$ and *proper homotopy equivalences* induce isomorphisms of $H_c^*$. In other words, $H_c^*$ satisfies the functoriality property of the usual cohomology groups as long as we restrict to the category of proper maps.

**3.2.2. Cell complexes.** A cell complex (or CW complex) $X$ is defined as the increasing union of subspaces denoted $X^{(n)}$ (or $X^n$), called *n-skeleta* of $X$. The 0-skeleton $X^{(0)}$ of $X$ is a set with discrete topology. Assume that $X^{(n-1)}$ is defined. Let

$$U_n := \sqcup_{j \in J} D_j^n,$$

a (possibly empty) disjoint union of closed $n$-balls $D_j^n$. Suppose that for each $D_j^n$ we have a continuous *attaching map* $e_j : \partial D_j^n \to X^{(n-1)}$. This defines a map $e = e^n : \partial U_n \to X^{(n-1)}$ and an equivalence relation $x \equiv y = e(x)$, $x \in U, y \in X^{(n-1)}$. We then declare $X^{(n)}$ to be the quotient space of $X^{(n-1)} \sqcup U_n$ with the quotient topology with respect to the above equivalence relation. We will use the notation $D_j^n/e_j$ the image of $D^n$ in $X^n$, i.e., the quotient $D_j^n/\equiv$. We then equip

$$X := \bigcup_{n \in \mathbb{N}} X_n$$

with the *weak topology*, where a subset $C \subset X$ is closed if and only if the intersection of $C$ with each skeleton is closed (equivalently, intersection of $C$ with the image of each $D^n$ in $X$ is closed). By abuse of terminology, both the balls $D_j^n$ and their projections to $X$ are called *n-cells* in $X$. Similarly, we will conflate $X$ and its underlying topological space.

Exercise 3.21. A subset $K \subset X$ is compact if and only if is closed and contained in a finite union of cells.

**Regular and almost regular cell complexes.** A cell complex $X$ is said to be *regular* if every attaching map $e_j$ is 1-1. For instance, every simplicial complex is a regular cell complex. A regular cell complex is called *triangular* if every cell is naturally isomorphic to a simplex. (Note that $X$ itself need not be simplicial since intersections of cells could be unions of simplices.) A cell complex $X$ is *almost regular* if the boundary $S^{n-1}$ of every cell $D_j^n$ is given structure of a regular cell complex $K_j$ so that the attaching map $e_j$ is 1-1 on every cell in $S^{n-1}$. Almost regular 2-dimensional cell complexes (with a single vertex) appear naturally in the context of group presentations, see Definition 4.79.

**Barycentric subdivision of an almost regular cell complex.** Our goal is to (canonically) subdivide an almost regular cell complex $X$ so that the result is a triangular regular cell complex $X' = Y$ where every cell is a simplex. We define $Y$ as an increasing union of regular subcomplexes $Y_n$ (where $Y_n \subset Y^{(n)}$ but, in general, is smaller).

First, set $Y_0 := X^{(0)}$. Suppose that $Y_{n-1} \subset Y^{(n-1)}$ is defined, so that $|Y_{n-1}| = X^{(n-1)}$. Consider attaching maps $e_j : \partial D_j^n \to X^{(n-1)}$. We take the preimage of the regular cell complex structure of $Y_{n-1}$ under $e_j$ to be a *refinement* $L_j$ of the regular cell complex structure $K_j$ on $S^{n-1}$. We then define a regular cell complex

$M_j$ on $D^n_j$ by conning off every cell in $L_j$ from the origin $o \in D^n_j$. Then cells in $M_j$ are the cones $Cone_{o_j}(s)$, where $s$'s are cells in $L_j$.

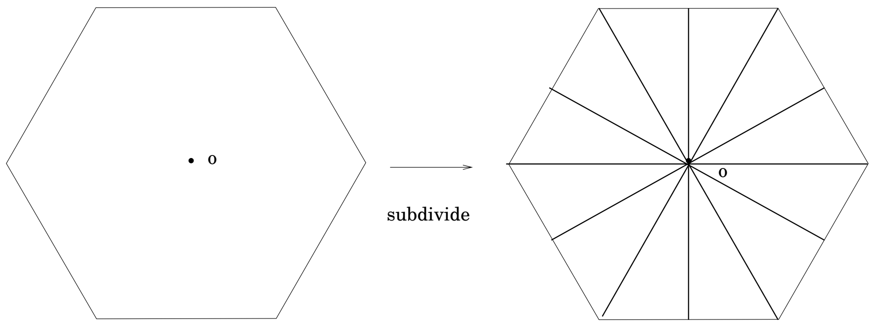



FIGURE 3.1. *B*arycentric subdivision of a 2-cell.

Since, by the induction assumption, every cell in $Y_{n-1}$ is a simplex, its preimage $s$ in $S^{n-1}$ is also a simplex, this $Cone_o(s)$ is a simplex as well. We then attach each cell $D^n_j$ to $Y_n$ by the original attaching map $e_j$. It is clear that the new cells $Cone_{o_j}(s)$ are embedded in $Y_n$ and each is naturally isomorphic to a simplex. Lastly, we set

$$Y := \bigcup_n Y_n.$$

**Second barycentric subdivision.** Note that the complex $X'$ constructed above may not be a simplicial complex. The problem is that if $x, y$ are distinct vertices of $L_j$, their images under the attaching map $e_j$ could be the same (a point $z$). Thus the edges $[o_j, x], [o_j, y]$ in $Y_{n+1}$ will intersect in the set $\{o_j, z\}$. However, if the complex $X$ was regular, this problem does not arise and $X'$ is a simplicial complex. Thus in order to promote $X$ to a simplicial complex (whose geometric realization is homeomorphic to $|X|$), we take the *second barycentric subdivision* $X''$ of $X$: Since $X'$ is a regular cell complex, the complex $X''$ is naturally isomorphic to a simplicial complex.

***G*-cell complexes.** Let $X$ be a cell complex and $G$ be a group. We say that $X$ is a *$G$-cell complex* (or that we have a *cellular* action $G \curvearrowright X$) if $G$ acts on $X$ by homeomorphisms and for every $n$ we have a $G$-action $G \curvearrowright U_n$ so that the attaching map $e^n$ is $G$-equivariant.

DEFINITION 3.22. A cellular action $G \curvearrowright X$ is said to be *without inversions* if whenever $g \in G$ preserves a cell $s$ in $X$, it fixes this cell pointwise.

An action $G \curvearrowright X$ on a simplicial complex is called *simplicial* if it sends simplices to simplices and is linear on each simplex.

Assuming that $X$ is naturally isomorphic to a simplicial complex and $G \curvearrowright X$ is without inversions, without loss of generality we may assume that $G \curvearrowright X$ is linear on every simplex in $X$.

The following is immediate from the definition of $X''$, since barycentric subdivisions are canonical:

Lemma 3.23. *Let $X$ be an almost regular cell complex and $G \curvearrowright X$ be an action without inversions. Then $G \curvearrowright X$ induces a simplicial action without inversions $G \curvearrowright X''$.*

Lemma 3.24. *Let $X$ be a simplicial complex and $G \curvearrowright X$ be a free simplicial action. Then this action is properly discontinuous on $X$ (in the weak topology).*

Proof. Let $K$ be a compact in $X$. Then $K$ is contained in a finite union of simplices $\sigma_1, \dots, \sigma_k$ in $X$. Let $F \subset G$ be the subset consisting of elements $g \in G$ so that $gK \cap K \neq \emptyset$. Then, assuming that $F$ is infinite, it contains distinct elements $g, h$ such that $g(\sigma) = h(\sigma)$ for some $\sigma \in \{\sigma_1, \dots, \sigma_n\}$. Then $f := h^{-1}g(\sigma) = \sigma$. Since the action $G \curvearrowright X$ is linear on each simplex, $f$ fixes a point in $\sigma$. This contradicts the assumption that the action of $G$ on $X$ is free. $\square$

**3.2.3. Borel construction.** Recall that every group $G$ admits a *classifying space* $E(G)$, which is a contractible cell complex admitting a free cellular action $G \curvearrowright E(G)$. The space $E(G)$ is far from being unique, we will use the one obtained by *Milnor's Construction*, see for instance [**Hat02**, Section 1.B]. A benefit of this construction is that $E(G)$ is a simplicial complex and the construction of $G \curvearrowright E(G)$ is canonical. Simplices in $E(G)$ are ordered tuples of elements of $g$: $[g_0, \dots, g_n]$ is an $n$-simplex with the obvious inclusions. To verify contractibility of $E = E(G)$, note that each $i+1$-skeleton $E^{i+1}$ contains the cone over the $i$-skeleton $E^i$, consisting of simplices of the form

$$[1, g_0, \dots, g_n], g_0, \dots, g_n \in G.$$

(The point $[1, \dots, 1] \in E^{i+1}$ is the tip of this cone.) Therefore, the straight-line homotopy to $[1, \dots, 1]$ gives the required contraction.

The group $G$ acts on $E(G)$ by the left multiplication

$$g \times [g_0, \dots, g_n] \to [gg_0, \dots, gg_n].$$

Clearly, this action is free and, moreover, each simplex has trivial stabilizer. The action $G \curvearrowright E(G)$ has two obvious properties that we will be using:

1. If $G$ is finite then each skeleton $E(G)^i$ is compact.
2. If $G_1 < G_2$ then there exists an equivariant embedding $E(G_1) \hookrightarrow E(G_2)$.

We will use only these properties and not the actual construction of $E(G)$ and the action $G \curvearrowright E(G)$.

Suppose now that $X$ is a cell complex and $G \curvearrowright X$ is a cellular action without inversions. Our next goal is to replace $X$ with a new cell complex $\widehat{X}$ which admits a homotopy-equivalence $p : \widehat{X} \to X$ so that the action $G \curvearrowright X$ lifts (*via* $p$) to a free cellular action $G \curvearrowright \widehat{X}$. The construction of $G \curvearrowright \widehat{X}$ is called the *Borel Construction*. We first explain the construction in the case when $X$ is a simplicial complex since the idea is much clearer in this case.

For each simplex $\sigma \in X$ consider its (pointwise) stabilizer $G_\sigma \leqslant G$. Clearly, if $\sigma_1 \subset \sigma_2$ then

$$G_{\sigma_2} \leqslant G_{\sigma_1}.$$

For each simplex $\sigma$ define $\widehat{X}_\sigma := \sigma \times E(G_\sigma)$. The group $G_\sigma$ acts naturally on $\widehat{X}_\sigma$. Whenever $\sigma_1 \subset \mathrm{Supp}(\sigma_2)$ we have the natural embedding $E(G_{\sigma_1}) \hookrightarrow E(G_{\sigma_2})$ and hence embeddings

$$\widehat{X}_{\sigma_1} = \sigma_1 \times E(G_{\sigma_1}) \supset \sigma_1 \times E(G_{\sigma_2}) \subset \widehat{X}_{\sigma_2}.$$

Henceforth, we glue $\widehat{X}_{\sigma_2}$ to $\widehat{X}_{\sigma_1}$ by identifying the two copies of the product subcomplex $\sigma_1 \times E(G_{\sigma_2})$. Let $\widehat{X}$ denote the regular cell complex resulting from these identifications.

For general cell complexes we have to modify the above construction. Define the *support* $\operatorname{Supp}(\sigma)$ of an $n$-cell $\sigma$ in $X$ to be the smallest subcomplex in $X$ whose underlying space contains the image of $S^{n-1}$ under the attaching map of $\sigma$. Since $G$ acts on $X$ without inversions, for every $\sigma_1 \subset \operatorname{Supp}(\sigma_2)$,

$$G_{\sigma_2} \leqslant G_{\sigma_1}$$

where $G_\sigma$ is the stabilizer of $\sigma$ in $G$. As before, for each $n$-dimensional cell $\sigma$ define $\widehat{X}_\sigma := D^n \times E(G_\sigma)$. The group $G_\sigma$ acts on $\widehat{X}_\sigma$ preserving the product structure and fixing $D^n$ pointwise. Whenever $\sigma_1 \subset \operatorname{Supp}(\sigma_2)$ we have the natural embedding $E(G_{\sigma_1}) \hookrightarrow E(G_{\sigma_2})$ and hence embeddings

$$\widehat{X}_{\sigma_1} = \sigma_1 \times E(G_{\sigma_1}) \supset \sigma_1 \times E(G_{\sigma_2}) \subset \operatorname{Supp}(\sigma_2) \times E(G_{\sigma_2}).$$

At the same time, we have the attaching map $e_{\sigma_2} : \partial D^n \to \operatorname{Supp}(\sigma_2)$ and, thus the attaching map

$$\widehat{e}_{\sigma_2} := e_{\sigma_2} \times Id : \partial D^n \times E(G_{\sigma_2}) \to \operatorname{Supp}(\sigma_2) \times E(G_{\sigma_2})$$

Here $n$ is the dimension of the cell $\sigma_2$. We now define $\widehat{X}$ by induction on skeleta of $X$. We begin with $\widehat{X}_0$ obtained by replacing each 0-cell $\sigma$ in $X$ with $\widehat{X}_\sigma$. Assume that $\widehat{X}_{n-1}$ is constructed by gluing spaces $\widehat{X}_\tau$, where $\tau$'s are cells in $X^{(n-1)}$. For each $n$-cell $\sigma$ the attaching map $\widehat{e}_\sigma$ defined above will yield an attaching map

$$\partial D^n \times E(G_\sigma) \to \hat{X}_{n-1}.$$

We then glue the spaces $\widehat{X}_\sigma$ to $\widehat{X}_{n-1}$ *via* these attaching maps. We have a natural projection $p : \widehat{X} \to X$ which corresponds to the projection

$$\widehat{X}_\sigma := D^n \times E(G_\sigma) \to D^n$$

for each $n$-cell $\sigma$ in $X$. Since each $D^n$ is contractible, it follows that $p$ restricts to a homotopy-equivalence

$$\widehat{X}_n \to X^{(n)}$$

for every $n$. Naturality of the construction ensures that the action $G \curvearrowright X$ lifts to an action $G \curvearrowright \widehat{X}$; it is clear from the construction that for each cell $\sigma$, the stabilizer of $\widehat{X}_\sigma$ in $G$ is $G_\sigma$. Since $G_\sigma$ acts freely on $E(G_\sigma)$, it follows that the action $G \curvearrowright \widehat{X}$ is free. Suppose now that $G \curvearrowright X$ is properly discontinuous. Then, $G_\sigma$ is finite for each $\sigma$ and, thus $\widehat{X}_\sigma$ has finite $i$-skeleton for each $i$. Moreover, if $X/G$ were compact, then the action of $G$ on each $i$-skeleton of $\widehat{X}$ is compact as well.

The construction of the complex $\widehat{X}$ and the action $G \curvearrowright \widehat{X}$ is called the *Borel construction.* One application of the Borel construction is the following

LEMMA 3.25. *Suppose that $G \curvearrowright X$ is a cocompact properly discontinuous action. Then there exists a properly discontinuous, cellular, free action $G \curvearrowright \widehat{X}$ which is cocompact on each skeleton and so that $X$ is homotopy-equivalent to $\widehat{X}$.*

**3.2.4. Groups of finite type.** If $G$ is a group admitting a free properly discontinuous cocompact action on a graph $\Gamma$, then $G$ is finitely generated, as, by the covering theory, $G \cong \pi_1(\Gamma/G)/p_*(\pi_1(\Gamma))$, where $p : \Gamma \to \Gamma/G$ is the covering map. Groups of *finite type* $\mathbf{F}_n$ are higher-dimensional generalizations of this example.

DEFINITION 3.26. A group $G$ is said to have *type* $\mathbf{F}_n$, $1 \leqslant n \leqslant \infty$, if it admits a free properly discontinuous cellular action on an $n-1$-connected $n$-dimensional cell complex $Y$, which is cocompact on each skeleton.

Note that we allow the complex $Y$ to be infinite-dimensional.

EXERCISE 3.27. A group $G$ is finitely-presented if and only if it has type $\mathbf{F}_2$.

In view of Lemma 3.25, we obtain:

COROLLARY 3.28. *A group $G$ has type $\mathbf{F}_n$ if and only if it admits a properly discontinuous cocompact cellular action on an $n-1$-connected $n$-dimensional cell complex $X$, which is cocompact on each skeleton.*

PROOF. One direction is obvious. Suppose, therefore, that we have an action $G \curvearrowright X$ as above. We apply Borel construction to this action and obtain a free properly discontinuous action $G \curvearrowright \widehat{X}$ which is cocompact on each skeleton of $\widehat{X}$. If $n = \infty$, we let $Y := \widehat{X}$. Otherwise, we let $Y$ denote the $n$-skeleton of $\widehat{X}$. Recall that the inclusion $Y \hookrightarrow \widehat{X}$ induces monomorphisms of all homotopy groups $\pi_j$, $j \leqslant n-1$. Since $X$ is $n-1$–connected, the same holds for $\widehat{X}$ and hence $Y$. □

COROLLARY 3.29. *Every finite group has type $\mathbf{F}_\infty$.*

PROOF. Start with the action of $G$ on a complex $X$ which is a point and then apply the above corollary. □

## 3.3. Subgroups

Given two subgroups $H, K$ in a group $G$ we denote by $HK$ the subset

$$\{hk \,;\; h \in H,\, k \in K\} \subset G.$$

Recall that a *normal subgroup* $K$ in $G$ is a subgroup such that for every $g \in G$, $gKg^{-1} = K$ (equivalently $gK = Kg$). We use the notation $K \lhd G$ to denote that $K$ is a normal subgroup in $G$. When either $H$ or $K$ is a normal subgroup, the set $HK$ defined above becomes a subgroup of $G$.

A subgroup $K$ of a group $G$ is called *characteristic* if for every automorphism $\phi : G \to G$, $\phi(K) = K$. Note that every characteristic subgroup is normal (since conjugation is an automorphism). But not every normal subgroup is characteristic.

EXAMPLE 3.30. Let $G$ be the group $(\mathbb{Z}^2, +)$. Since $G$ is abelian, every subgroup is normal. But, for instance, the subgroup $\mathbb{Z} \times \{0\}$ is not invariant under the automorphism $\phi : \mathbb{Z}^2 \to \mathbb{Z}^2$ , $\phi(m, n) = (n, m)$.

DEFINITION 3.31. A *subnormal descending series*indexsubnormal descending series in a group $G$ is a series

$$G = N_0 \rhd N_1 \rhd \cdots \rhd N_n \rhd \cdots$$

such that $N_{i+1}$ is a normal subgroup in $N_i$ for every $i \geqslant 0$.

If all $N_i$ are normal subgroups of $G$ then the series is called *normal.*

A subnormal series of a group is called a *refinement* of another subnormal series if the terms of the latter series all occur as terms in the former series.

The following is a basic result in group theory:

LEMMA 3.32. *If $G$ is a group, $N \lhd G$, and $A \lhd B \leqslant G$, then $BN/AN$ is isomorphic to $B/A(B\cap N)$.*

DEFINITION 3.33. Two subnormal series

$$G = A_0 \rhd A_1 \rhd \dots \rhd A_n = \{1\} \text{ and } G = B_0 \rhd B_1 \rhd \dots \rhd B_m = \{1\}$$

are called *isomorphic* if $n = m$ and there exists a bijection between the sets of partial quotients $\{A_i/A_{i+1} \mid i = 1, \dots, n-1\}$ and $\{B_i/B_{i+1} \mid i = 1, \dots, n-1\}$ such that the corresponding quotients are isomorphic.

LEMMA 3.34. *Any two finite subnormal series*

$$G = H_0 \geqslant H_1 \geqslant \dots \geqslant H_n = \{1\} \text{ and } G = K_0 \geqslant K_1 \geqslant \dots \geqslant K_m = \{1\}$$

*possess isomorphic refinements.*

PROOF. Define $H_{ij} = (K_j \cap H_i)H_{i+1}$. The following is a subnormal series

$$H_{i0} = H_i \geqslant H_{i1} \geqslant \dots \geqslant H_{im} = H_{i+1}\,.$$

When inserting all these in the series of $H_i$ one obtains the required refinement.

Likewise, define $K_{rs} = (H_s \cap K_r)K_{r+1}$ and by inserting the series

$$K_{r0} = K_r \geqslant K_{r1} \geqslant \dots \geqslant K_{rn} = K_r$$

in the series of $K_r$, we define its refinement.

According to Lemma 3.32

$$H_{ij}/H_{ij+1} = (K_j \cap H_i)H_{i+1}/(K_{j+1}\cap H_i)H_{i+1} \simeq K_j \cap H_i/(K_{j+1}\cap H_i)(K_j \cap H_{i+1})\,.$$

Similarly, one proves that $K_{ji}/K_{ji+1} \simeq K_j \cap H_i/(K_{j+1}\cap H_i)(K_j \cap H_{i+1})$. □

DEFINITION 3.35. The *center* $Z(G)$ of a group $G$ is defined as the subgroup consisting of elements $h \in G$ so that $[h, g] = 1$ for each $g \in G$.

It is easy to see that the center is a characteristic subgroup of $G$.

DEFINITION 3.36. A group $G$ is a *torsion group* if all its elements have finite order.

A group $G$ is said to be *without torsion* (or *torsion-free*) if all its non-trivial elements have infinite order.

Note that the subset $\mathrm{Tor}\, G = \{g \in G \mid g \text{ of finite order}\}$ of the group $G$, sometimes called the *torsion* of $G$, is in general not a subgroup.

DEFINITION 3.37. A group $G$ is said to have property * *virtually* if a finite index subgroup $H$ of $G$ has the property *.

The following properties of finite index subgroups will be useful.

LEMMA 3.38. *If $N \lhd H$ and $H \lhd G$, $N$ of finite index in $H$ and $H$ finitely generated, then $N$ contains a finite index subgroup $K$ which is normal in $G$.*

Proof. By hypothesis, the quotient group $F = H/N$ is finite. For an arbitrary $g \in G$ the conjugation by $g$ is an automorphism of $H$, hence $H/gNg^{-1}$ is isomorphic to $F$. A homomorphism $H \to F$ is completely determined by the images in $F$ of elements of a finite generating set of $H$. Therefore there are finitely many such homomorphisms, and finitely many possible kernels of them. Thus, the set of subgroups $gNg^{-1}$, $g \in G$, forms a finite list $N, N_1, .., N_k$. The subgroup $K = \bigcap_{g\in G} gNg^{-1} = N \cap N_1 \cap \cdots \cap N_k$ is normal in $G$ and has finite index in $N$, since each of the subgroups $N_1, \ldots, N_k$ has finite index in $H$. □

Proposition 3.39. *Let $G$ be a finitely generated group. Then:*

(1) *For every $n \in \mathbb{N}$ there exist finitely many subgroups of index $n$ in $G$.*

(2) *Every finite index subgroup $H$ in $G$ contains a subgroup $K$ which is finite index and characteristic in $G$.*

Proof. (1) Let $H \leqslant G$ be a subgroup of index $n$. We list the left cosets of $H$:

$$H = g_1 \cdot H, g_2 \cdot H, \ldots, g_n \cdot H,$$

and label these cosets by the numbers $\{1, \ldots, n\}$. The action by left multiplication of $G$ on the set of left cosets of $H$ defines a homomorphism $\phi : G \to S_n$ such that $\phi(G)$ acts transitively on $\{1, 2, \ldots, n\}$ and $H$ is the inverse image under $\phi$ of the stabilizer of 1 in $S_n$. Note that there are $(n-1)!$ ways of labeling the left cosets, each defining a different homomorphism with these properties.

Conversely, if $\phi : G \to S_n$ is such that $\phi(G)$ acts transitively on $\{1, 2, \ldots, n\}$ then $G/\phi^{-1}(\mathrm{Stab}\,(1))$ has cardinality $n$.

Since the group $G$ is finitely generated, a homomorphism $\phi : G \to S_n$ is determined by the image of a generating finite set of $G$, hence there are finitely many distinct such homomorphisms. The number of subgroups of index $n$ in $H$ is equal to the number $\eta_n$ of homomorphisms $\phi : G \to S_n$ such that $\phi(G)$ acts transitively on $\{1, 2, \ldots, n\}$, divided by $(n-1)!$.

(2) Let $H$ be a subgroup of index $n$. For every automorphism $\varphi : G \to G$, $\varphi(H)$ is a subgroup of index $n$. According to (1) the set $\{\varphi(H) \mid \varphi \in \mathrm{Aut}\,(G)\}$ is finite, equal $\{H, H_1, \ldots, H_k\}$. It follows that

$$K = \bigcap_{\varphi\in \mathrm{Aut}\,(G)} \varphi(H) = H \cap H_1 \cap \ldots \cap H_k.$$

Then $K$ is a characteristic subgroup of finite index in $H$ hence in $G$. □

Let $S$ be a subset in a group $G$, and let $H \leqslant G$ be a subgroup. The following are equivalent:

(1) $H$ is the smallest subgroup of $G$ containing $S$ ;

(2) $H = \bigcap_{S\subset G_1 \leqslant G} G_1$ ;

(3) $H = \left\{ s_1 s_2 \cdots s_n \; ; \; n \in \mathbb{N}, s_i \in S \text{ or } s_i^{-1} \in S \text{ for every } i \in \{1, 2, \ldots, n\} \right\}.$

The subgroup $H$ satisfying any of the above is denoted $H = \langle S \rangle$ and is said to be *generated by* $S$. The subset $S \subset H$ is called a *generating set* of $H$. The elements in $S$ are called *generators* of $H$.

When $S$ consists of a single element $x$, $\langle S \rangle$ is usually written as $\langle x \rangle$; it is the cyclic subgroup consisting of powers of $x$.

We say that a normal subgroup $K \lhd G$ is *normally generated* by a set $R \subset K$ if $K$ is the smallest normal subgroup of $G$ which contains $R$, i.e.

$$K = \bigcap_{R \subset N \lhd G} N\,.$$

We will use the notation

$$K = \langle\langle R \rangle\rangle$$

for this subgroup.

### 3.4. Equivalence relations between groups

Definition 3.40.
(1) Two groups $G_1$ and $G_2$ are called *co-embeddable* if there exist injective group homomorphisms $G_1 \to G_2$ and $G_2 \to G_1$.

(2) The groups $G_1$ and $G_2$ are *commensurable* if there exist finite index subgroups $H_i \leqslant G_i$, $i = 1, 2$, such that $H_1$ is isomorphic to $H_2$.

An isomorphism $\varphi : H_1 \to H_2$ is called an *abstract commensurator* of $G_1$ and $G_2$.

(3) We say that two groups $G_1$ and $G_2$ are *virtually isomorphic* (abbreviated as VI) if there exist finite index subgroups $H_i \subset G_i$ and finite normal subgroups $F_i \lhd H_i$, $i = 1, 2$, so that the quotients $H_1/F_1$ and $H_2/F_2$ are isomorphic.

An isomorphism $\varphi : H_1/F_1 \to H_2/F_2$ is called a *virtual isomorphism* of $G_1$ and $G_2$. When $G_1 = G_2$, $\varphi$ is called *virtual automorphism.*

Example 3.41. All countable free groups are co-embeddable. However, a free group of infinite rank is not virtually isomorphic to a free group of infinite rank.

Proposition 3.42. *All the relations in Definition 3.40 are equivalence relation between groups.*

Proof. The fact that weak commensurability is an equivalence relation is immediate. It suffices to prove that virtual isomorphism is am equivalence relation. The only non-obvious property is transitivity. We need

Lemma 3.43. *Let $F_1, F_2$ be normal finite subgroups of a group $G$. Then their normal closure $F = \langle\langle F_1, F_2 \rangle\rangle$ (i.e., the smallest normal subgroup of $G$ containing $F_1$ and $F_2$) is again finite.*

Proof. Let $f_1 : G \to G_1 = G/F_1$, $f_2 : G_1 \to G_1/f_1(F_2)$ be the quotient maps. Since the kernel of each $f_1, f_2$ is finite, it follows that the kernel of $f = f_2 \circ f_1$ is finite as well. On the other hand, the kernel of $f$ is clearly the subgroup $F$. □

Suppose now that $G_1$ is VI to $G_2$ and $G_2$ is VI to $G_3$. Then we have

$$F_i \lhd H_i < G_i, |G_i : H_i| < \infty, |F_i| < \infty, \quad i = 1, 2, 3,$$

and

$$F_2' \lhd H_2' < G_2, |G_2 : H_2'| < \infty, |F_2'| < \infty,$$

so that

$$H_1/F_1 \cong H_2/F_2, \quad H_2'/F_2' \cong H_3/F_3.$$

The subgroup $H_2'' := H_2 \cap H_2'$ has finite index in $G_2$. By the above lemma, the normal closure in $H_2''$

$$K_2 := \langle\langle F_2 \cap H_2'', F_2' \cap H_2'' \rangle\rangle$$

is finite. We have quotient maps

$$f_i : H_2'' \to C_i = f_i(H_2'') \leqslant H_i/F_i, i = 1, 3,$$

with finite kernels and cokernels. The subgroups $E_i := f_i(K_2)$, are finite and normal in $C_i$, $i = 1, 3$. We let $H_i', F_i' \subset H_i$ denote the preimages of $C_i$ and $E_i$ under the quotient maps $H_i \to H_i/F_i$, $i = 1, 3$. Then $|F_i'| < \infty, |G_i : H_i'| < \infty, i = 1, 3$. Lastly,

$$H_i'/F_i' \cong C_i/E_i \cong H_2''/K_2, i = 1, 3.$$

Therefore, $G_1, G_3$ are virtually isomorphic. □

Given a group $G$, we define $VI(G)$ as the set of equivalence classes of virtual automorphisms of $G$ with respect to the following equivalence relation. Two virtual automorphisms of $G$, $\varphi : H_1/F_1 \to H_2/F_2$ and $\psi : H_1'/F_1' \to H_2'/F_2'$, are *equivalent* if for $i = 1, 2$, there exist $\widetilde{H}_i$, a finite index subgroup of $H_i \cap H_i'$, and $\widetilde{F}_i$, a normal subgroup in $\widetilde{H}_i$ containing the intersections $\widetilde{H}_i \cap F_i$ and $\widetilde{H}_i \cap F_i'$, such that $\varphi$ and $\psi$ induce the same automorphism from $\widetilde{H}_1/\widetilde{F}_1$ to $\widetilde{H}_2/\widetilde{F}_2$.

Lemma 3.43 implies that the composition induces a binary operation on $VI(G)$, and that $VI(G)$ with this operation becomes a group, called *the group of virtual automorphisms of* $G$.

Let $Comm(G)$ be the set of equivalence classes of abstract commensurators of $G$ with respect to an equivalence relation defined as above, with the normal subgroups $F_i$ and $F_i'$ trivial. As in the case of $VI(G)$, the set $Comm(G)$, endowed with the binary operation defined by the composition, becomes a group, called the *abstract commensurator of the group* $G$.

Let $\Gamma$ be a subgroup of a group $G$. The *commensurator of* $\Gamma$ *in* $G$, denoted by $\mathrm{Comm}_G(\Gamma)$, is the set of elements $g$ in $G$ such that the conjugation by $g$ defines an abstract commensurator of $\Gamma$: $g\Gamma g^{-1} \cap \Gamma$ has finite index in both $\Gamma$ and $g\Gamma g^{-1}$.

Exercise 3.44. Show that $Comm_G(\Gamma)$ is a subgroup of $G$.

Exercise 3.45. Show that for $G = SL(n, \mathbb{R})$ and $\Gamma = SL(n, \mathbb{Z})$, $Comm_G(\Gamma)$ contains $SL(n, \mathbb{Q})$.

### 3.5. Residual finiteness

Even though, studying infinite groups is our primary focus, questions in group theory can be, sometimes, reduced to questions about finite groups. *Residual finiteness* is the concept that (sometimes) allows such reduction.

Definition 3.46. A group $G$ is said to be *residually finite* if

$$\bigcap_{i \in I} G_i = \{1\},$$

where $\{G_i : i \in I\}$ is the set of all finite-index subgroups in $G$.

Clearly, subgroups of residually finite groups are also residually finite. In contrast, if $G$ is an infinite simple group, then $G$ cannot be residually-finite.

Lemma 3.47. *A finitely generated group $G$ is residually finite if and only if for every $g \in G \setminus \{1\}$, there exists a finite group $\Phi$ and a homomorphism $\varphi : G \to \Phi$, so that $\varphi(g) \neq 1$.*

Proof. Suppose that $G$ is residually finite. Then, for every $g \in G \setminus \{1\}$ there exists a finite-index subgroup $G_i \leqslant G$ so that $g \notin G_i$. Since $G$ is finitely generated, it contains a normal subgroup of finite index $N_i \lhd G$, so that $N_i \leqslant G_i$. Indeed, we can take

$$N_i := \bigcap_{x \in S} G_i^x$$

where $S$ is a finite generating set of $G$ and $G_i^x$ denotes the subgroup $xG_ix^{-1}$. Then $N_i$ is invariant under all inner automorphisms of $G$ and, hence, is normal in $G$. Clearly, $g \notin N_i$ and $|G : N_i| < \infty$. Now, setting $\Phi := G/N_i$, we obtain the required homomorphism $\varphi : G \to \Phi$.

Conversely, suppose that for every $g \neq 1$ we have a homomorphism $\varphi_g : G \to \Phi_g$, where $\Phi_g$ is a finite group, so that $\varphi_g(g) \neq 1$. Setting $N_g := \mathrm{Ker}(\varphi_g)$, we get

$$\bigcap_{g \in G} N_g = \{1\}.$$

The above intersection, of course, contains the intersection of all finite index subgroups in $G$. □

Example 3.48. The group $G = GL(n, \mathbb{Z})$ is residually finite. Indeed, we take subgroups $G_p \leqslant G$, $G_p = \mathrm{Ker}(\varphi_p)$, $\varphi_p : G \to GL(n, \mathbb{Z}_p)$). If $g \in G$ is a nontrivial element, we consider its nonzero off-diagonal entry $g_{ij} \neq 0$. Then $g_{ij} \neq 0 \mod p$, whenever $p > |g_{ij}|$. Thus, $\varphi_p(g) \neq 1$ and $G$ is residually finite.

Corollary 3.49. *Free group of rank 2 $F_2$ is residually finite. Every free group of (at most) countable rank is residually finite.*

Proof. We will see in Example 4.38 that $F_2$ embeds in $SL(2, \mathbb{Z})$. Furthermore, every free group of (at most) countable rank embeds in $F_2$. Now, the assertion follows from the above example. □

The simple argument for $GL(n, \mathbb{Z})$ is a model for a proof of a harder theorem:

Theorem 3.50 (A. I. Mal'cev [**Mal40**]). *Let $G$ be a finitely generated subgroup of $GL(n, R)$, where $R$ is a commutative ring with unity. Then $G$ is residually finite.*

Mal'cev's theorem is complemented by the following result, known as *Selberg Lemma* [**Sel60**]:

Theorem 3.51 (Selberg Lemma). *Let $G$ be a finitely generated subgroup of $GL(n, F)$, where $F$ is a field of characteristic zero. Then $G$ contains a torsion-free subgroup of finite index.*

We refer the reader to [**Rat94**, §7.5] and [**Nic**] for the proofs. Note that Selberg Lemma fails for fields of positive characteristic, see e.g. [**Nic**].

## 3.6. Commutators, commutator subgroup

Definition 3.52. The *commutator of two elements $h, k$* in a group $G$ is

$$[h, k] = hkh^{-1}k^{-1}.$$

Note that:

- two elements $h, k$ commute (i.e., $hk = kh$) if and only if $[h, k] = 1$.
- $hk = [h, k]kh$ ;

Thus, the commutator $[h, k]$ 'measures de degree of non-commutativity' of the elements $h$ and $k$. In Lemma 10.25 we will prove some further properties of commutators.

Let $H, K$ be two subgroups of $G$. We denote by $[H, K]$ the subgroup of $G$ generated by all commutators $[h, k]$ with $h \in H$, $k \in K$.

DEFINITION 3.53. The *commutator subgroup (or* derived subgroup*) of* $G$ is the subgroup $G' = [G, G]$. As above, we may say that the commutator subgroup $G'$ of $G$ 'measures the degree of non-commutativity' of the group $G$.

A group $G$ is *abelian* if every two elements of $G$ *commute*, i.e., $ab = ba$ for all $a, b \in G$.

EXERCISE 3.54. Suppose that $S$ is a generating set of $G$. Then $G$ is abelian if and only if $[a, b] = 1$ for all $a, b \in S$.

PROPOSITION 3.55. (1) *$G'$ is a characteristic subgroup of $G$;*

(2) *$G$ is* abelian *if and only if $G' = \{1\}$;*

(3) *$G_{ab} = G/G'$ is an abelian group (called the* abelianization *of $G$);*

(4) *if $\varphi : G \to A$ is a homomorphism to an abelian group $A$, then $\varphi$ factors through the abelianization: Given the quotient map $p : G \to G_{ab}$, there exists a homomorphism $\overline{\varphi} : G_{ab} \to A$ such that $\varphi = \overline{\varphi} \circ p$.*

PROOF. (1) The set $S = \{[x, y] \mid x, y \in G\}$ is a generating set of $G'$ and for every automorphism $\psi : G \to G$, $\psi(S) = S$.

(2) follows from the equivalence $xy = yx \Leftrightarrow [x, y] = 1$ , and (3) is an immediate consequence of (2).

(4) follows from the fact that $\varphi(S) = \{1\}$. □

Recall that the *finite dihedral group* of order $2n$, denoted by $D_{2n}$ or $I_2(n)$, is the group of symmetries of the regular Euclidean $n$-gon, i.e. the group of isometries of the unit circle $S^1 \subset \mathbb{C}$ generated by the rotation $r(z) = e^{\frac{2\pi i}{n}} z$ and the reflection $s(z) = \bar{z}$. Likewise, the *infinite dihedral group* $D_\infty$ is the group of isometries of $\mathbb{Z}$ (with the metric induced from $\mathbb{R}$); the group $D_\infty$ is generated by the translation $t(x) = x + 1$ and the symmetry $s(x) = -x$.

EXERCISE 3.56. Find the commutator subgroup and the abelianization for the finite dihedral group $D_{2n}$ and for the infinite dihedral group $D_\infty$.

EXERCISE 3.57. Let $S_n$ (the symmetric group on $n$ symbols) be the group of permutations of the set $\{1, 2, \ldots, n\}$, and $A_n \subset S_n$ be the alternating subgroup, consisting of even permutations.

(1) Prove that for every $n \notin \{2, 4\}$ the group $A_n$ is generated by the set of cycles of length 3.

(2) Prove that if $n \geqslant 3$, then for every cycle $\sigma$ of length 3 there exists $\rho \in S_n$ such that $\sigma^2 = \rho\sigma\rho^{-1}$.

(3) Use (1) and (2) to find the commutator subgroup and the abelianization for $A_n$ and for $S_n$.

(4) Find the commutator subgroup and the abelianization for the group $H$ of permutations of $\mathbb{Z}$ defined in Example 4.7.

Note that it is not necessarily true that the commutator subgroup $G'$ of $G$ consists entirely of commutators $\{[x, y] : x, y \in G\}$ (see **[Vav]** for some finite group examples). However, occasionally, every element of the derived subgroup is indeed a single commutator. For instance, every element of the alternating group $A_n < S_n$ is the commutator in $S_n$, see **[Ore51]**.

This leads to an interesting invariant (of geometric flavor) called the *commutator norm* (or *commutator length*) $\ell_c(g)$ of $g \in G'$, which is the least number $k$ so that $g$ can be expressed as a product

$$g = [x_1, y_1] \cdots [x_k, y_k],$$

as well as the *stable commutator norm* of $g$:

$$\limsup_{n\to\infty} \frac{\ell_c(g^n)}{n}.$$

See **[Bav91, Cal08]** for further details. For instance, if $G$ is the free group on two generators (see Definition 4.16), then every nontrivial element of $G'$ has stable commutator norm greater than 1.

### 3.7. Semi-direct products and short exact sequences

Let $G_i, i \in I$, be a collection of groups. The *direct product* of these groups, denoted

$$G = \prod_{i\in I} G_i$$

is the Cartesian product of sets $G_i$ with the group operation given by

$$(a_i) \cdot (b_i) = (a_i b_i).$$

Note that each group $G_i$ is the quotient of $G$ by the (normal) subgroup

$$\prod_{j\in I\setminus\{i\}} G_j.$$

A group $G$ is said to *spit* as a direct product of its normal subgroups $N_i \lhd G, i = 1, \ldots, k$, if one of the following equivalent statements holds:

- $G = N_1 \cdots N_k$ and $N_i \cap N_j = \{1\}$ for all $i \neq j$;
- for every element $g$ of $G$ there exists a unique $k$-tuple $(n_1, \ldots, n_k), n_i \in N_i, i = 1, \ldots, k$ such that $g = n_1 \cdots n_k$.

Then, $G$ is isomorphic to the direct product $N_1 \times \ldots \times N_k$. Thus, finite direct products $G$ can be defined either *extrinsically*, using groups $N_i$ as quotients of $G$, or *intrinsically*, using normal subgroups $N_i$ of $G$.

Similarly, one defines *semidirect products* of two groups, by taking the above *intrinsic* definition and relaxing the normality assumption:

DEFINITION 3.58. (1) (*with the ambient group as given data*) A group $G$ is said to *split as a semidirect product of two subgroups* $N$ and $H$, which is denoted by $G = N \rtimes H$ if and only if $N$ is a *normal subgroup* of $G$, $H$ is a *subgroup* of $G$, and one of the following equivalent statements holds:

- $G = NH$ and $N \cap H = \{1\}$;
- $G = HN$ and $N \cap H = \{1\}$;
- for every element $g$ of $G$ there exists a unique $n \in N$ and $h \in H$ such that $g = nh$;
- for every element $g$ of $G$ there exists a unique $n \in N$ and $h \in H$ such that $g = hn$;
- there exists *a retraction* $G \to H$, i.e., a homomorphism which restricts to the identity on $H$, and whose kernel is $N$.

Observe that the map $\varphi : H \to \mathrm{Aut}\,(N)$ defined by $\varphi(h)(n) = hnh^{-1}$, is a group homomorphism.

(2) (*with the quotient groups as given data*) Given any two groups $N$ and $H$ (not necessarily subgroups of the same group) and a group homomorphism $\varphi : H \to \mathrm{Aut}\,(N)$, one can define a new group $G = N \rtimes_\varphi H$ which is a semidirect product of a copy of $N$ and a copy of $H$ in the above sense, defined as follows. As a set, $N \rtimes_\varphi H$ is defined as the cartesian product $N \times H$. The binary operation $*$ on $G$ is defined by

$$(n_1, h_1) * (n_2, h_2) = (n_1\varphi(h_1)(n_2), h_1h_2)\,,\ \forall n_1, n_2 \in N \text{ and } h_1, h_2 \in H\,.$$

The group $G = N \rtimes_\varphi H$ is called the *semidirect product of* $N$ *and* $H$ *with respect to* $\varphi$.

REMARKS 3.59. (1) If a group $G$ is the semidirect product of a normal subgroup $N$ with a subgroup $H$ in the sense of (1) then $G$ is isomorphic to $N \rtimes_\varphi H$ defined as in (2), where

$$\varphi(h)(n) = hnh^{-1}\,.$$

(2) The group $N \rtimes_\varphi H$ defined in (2) is a semidirect product of the normal subgroup $N_1 = N \times \{1\}$ and the subgroup $H = \{1\} \times H$ in the sense of (1).

(3) If both $N$ and $H$ are normal subgroups in (1) then $G$ is a direct product of $N$ and $H$.

If $\varphi$ is the trivial homomorphism, sending every element of $H$ to the identity automorphism of $N$, then $N \rtimes_\phi H$ is the direct product $N \times H$.

Here is yet another way to define semidirect products. An *exact sequence* is a sequence of groups and group homomorphisms

$$\dots G_{n-1} \stackrel{\varphi_{n-1}}{\longrightarrow} G_n \stackrel{\varphi_n}{\longrightarrow} G_{n+1} \dots$$

such that $\mathrm{Im}\,\varphi_{n-1} = \mathrm{Ker}\,\varphi_n$ for every $n$. A *short exact sequence* is an exact sequence of the form:

$$\{1\} \longrightarrow N \stackrel{\varphi}{\longrightarrow} G \stackrel{\psi}{\longrightarrow} H \longrightarrow \{1\}\,. \tag{3.2}$$

In other words, $\varphi$ is an isomorphism from $N$ to a normal subgroup $N' \lhd G$ and $\psi$ descends to an isomorphism $G/N' \simeq H$.

DEFINITION 3.60. A short exact sequence *splits* if there exists a homomorphism $\sigma : H \to G$ (called a *section*) such that

$$\psi \circ \sigma = Id.$$

When the sequence splits we shall sometimes write it as

$$1 \to N \to G \overset{\curvearrowleft}{\to} H \to 1.$$

Then, every split exact sequence determines a decomposition of $G$ as the semidirect product $\varphi(N) \rtimes \sigma(H)$. Conversely, every semidirect product decomposition $G = N \rtimes H$ defines a split exact sequence, where $\varphi$ is the identity embedding and $\psi : G \to H$ is the retraction.

EXAMPLES 3.61.
(1) The dihedral group $D_{2n}$ is isomorphic to $\mathbb{Z}_n \rtimes_\varphi \mathbb{Z}_2$, where $\varphi(1)(k) = n - k$.
(2) The infinite dihedral group $D_\infty$ is isomorphic to $\mathbb{Z} \rtimes_\varphi \mathbb{Z}_2$, where $\varphi(1)(k) = -k$.
(3) The permutation group $S_n$ is the semidirect product of $A_n$ and $\mathbb{Z}_2 = \{id, (12)\}$.
(4) The group $(\mathrm{Aff}(\mathbb{R}), \circ)$ of affine maps $f : \mathbb{R} \to \mathbb{R}, f(x) = ax + b$, with $a \in \mathbb{R}^*$ and $b \in \mathbb{R}$ is a semidirect product $\mathbb{R} \rtimes_\varphi \mathbb{R}^*$, where $\varphi(a)(x) = ax$.

PROPOSITION 3.62.
(1) *Every isometry $\phi$ of $\mathbb{R}^n$ is of the form $\phi(x) = Ax + b$, where $b \in \mathbb{R}^n$ and $A \in O(n)$.*
(2) *The group $\mathrm{Isom}(\mathbb{R}^n)$ splits as the semidirect product $\mathbb{R}^n \rtimes O(n)$, with the obvious action of the orthogonal $O(n)$ on $\mathbb{R}^n$.*

*Sketch of proof of* (1). For every vector $a \in \mathbb{R}^n$ we denote by $T_a$ the translation of vector $a$, $x \mapsto x + a$.

If $\phi(0) = b$ then the isometry $\psi = T_{-b} \circ \phi$ fixes the origin 0. Thus it suffices to prove that an isometry fixing the origin is a linear map in $O(n)$. Indeed:

- an isometry of $\mathbb{R}^n$ preserves straight lines, because these are bi-infinite geodesics;
- an isometry is a homogeneous map, i.e. $\psi(\lambda v) = \lambda \psi(v)$; this is due to the fact that (for $0 < \lambda \leqslant 1$) $w = \lambda v$ is the unique point in $\mathbb{R}^n$ satisfying

$$d(0, w) + d(w, v) = d(0, v).$$

- an isometry map is an additive map, i.e. $\psi(a + b) = \psi(a) + \psi(b)$ because an isometry preserves parallelograms.

Thus, $\psi$ is a linear transformation of $\mathbb{R}^n$, $\psi(x) = Ax$ for some matrix $A$. Orthogonality of the matrix $A$ follows from the fact that the image of an orthonormal basis under $\psi$ is again an orthonormal basis. □

EXERCISE 3.63. Prove statement (2) of Proposition 3.62. Note that $\mathbb{R}^n$ is identified to the group of translations of the $n$-dimensional affine space *via* the map $b \mapsto T_b$.

In sections 3.11 and 3.12 we discuss semidirect products and short exact sequences in more detail.

## 3.8. Direct sums and wreath products

Let $X$ be a non-empty set, and let $\mathcal{G} = \{G_x \mid x \in X\}$ be a collection of groups indexed by $X$. Consider the set of maps $Map_f(X, \mathcal{G})$ with finite support, i.e.,

$$Map_f(X, \mathcal{G}) := \{f : X \to \bigsqcup_{x \in X} G_x \; ; \; f(x) \in G_x \, , \; f(x) \neq 1_{G_x}$$

$$\text{for only finitely many } x \in X\} \, .$$

Definition 3.64. The *direct sum* $\bigoplus_{x \in X} G_x$ is defined as $Map_f(X, \mathcal{G})$, endowed with the pointwise multiplication of functions:

$$(f \cdot g)(x) = f(x) \cdot g(x) \, , \; \forall x \in X.$$

Clearly, if $A_x$ are abelian groups then $\bigoplus_{x \in X} A_x$ is abelian.

When $G_x = G$ is the same group for all $x \in X$, the direct sum is the set of maps

$$Map_f(X, G) := \{f : X \to G \mid f(x) \neq 1_G \text{ for only finitely many } x \in X\} \; ,$$

and we denote it either by $\bigoplus_{x \in X} G$ or by $G^{\oplus X}$.

If, in this latter case, the set $X$ is itself a group $H$, then there is a natural action of $H$ on the direct sum, defined by

$$\varphi : H \to \mathrm{Aut}\left( \bigoplus_{h \in H} G \right) \, , \; \varphi(h)f(x) = f(h^{-1}x) \, , \; \forall x \in H \, .$$

Thus, we define the semi-direct product

$$\left( \bigoplus_{h \in H} G \right) \rtimes_{\varphi} H \, .$$

Definition 3.65. The semidirect product $(\bigoplus_{h \in H} G) \rtimes_{\varphi} H$ is called *the wreath product of* $G$ *with* $H$, and it is denoted by $G \wr H$. The wreath product $G = \mathbb{Z}_2 \wr \mathbb{Z}$ is called the *lamplighter group*.

## 3.9. Group cohomology

The purpose of this section is to introduce cohomology of groups and to give explicit formulae for cocycles and coboundaries in small degrees. We refer the reader to [**Bro82b**, Chapter III, Section 1] for the more thorough discussion.

Let $G$ be a group and let $M, N$ be left $G$-modules; then $Hom_G(M, N)$ denotes the subspace of $G$-invariants in the $G$-module $Hom(M, N)$, where $G$ acts on homomorphisms $u : M \to N$ by the formula:

$$(gu)(m) = g \cdot u(g^{-1}m).$$

If $C_*$ is a chain complex and $A$ is a $G$-module, then $Hom_G(C_*, A)$ is the chain complex formed by subspaces $Hom_G(C_k, A)$ in $Hom(C_k, A)$. The *standard chain complex* $C_* = C_*(G)$ of $G$ with coefficients in $A$ is defined as follows:

$C_k(G) = \mathbb{Z} \otimes \prod_{i=0}^{k} G$, is the $G$-module freely generated by $(k+1)$-tuples $(g_0, \ldots, g_k)$ of elements of $G$ with the $G$-action given by

$$g \cdot (g_0, \ldots, g_k) = (gg_0, \ldots, gg_k).$$

The reader should think of each tuple as spanning a $k$-simplex. The boundary operator on this chain complex is the natural one:

$$\partial_k(g_0, \dots, g_k) = \sum_{i=0}^{k} (-1)^i (g_0, \dots, \hat{g}_i, \dots g_k),$$

where $\hat{g}_i$ means that we omit this entry in the tuple. Then $C_* = C_*(G)$ is the simplicial chain complex of the simplicial complex defining the Milnor's classifying space $EG$ of the group $G$ (see Section 3.2.3). The dual cochain complex $C^*$ is defined by:

$$C^k = Hom(C_k, A), \quad \delta_k(f)((g_0, \dots, g_{k+1})) = f(\partial_{k+1}(g_0, \dots, g_{k+1})), f \in C^k.$$

Suppose for a moment that $A$ is a trivial $G$-module. Then, for $BG = (EG)/G$, the simplicial cochain complex $C^*(BG, A)$ is naturally isomorphic to the subcomplex of $G$-invariant cochains in $C^*(G, A)$, i.e., the subcomplex $(C^*(G, A))^G = Hom_G(C_*, A)$. If $A$ is a nontrivial $G$-module then the $Hom_G(C_*, A)$ is still isomorphic to a certain natural cochain complex based on the simplicial complex $C_*(BG)$ (cochain complex with twisted coefficients, or coefficients in a certain sheaf), but the definition is more involved and we will omit it.

Definition 3.66. The cohomology groups of $G$ with coefficients in the $G$-module $A$ are defined as $H^*(G, A) := H_*(Hom_G(C_*, A))$. In other words,

$$H^*(G, A) = \mathrm{Ker}(\delta_k)/\mathrm{Im}(\delta_{k-1}), \quad H^i(G, A) = Z^i(G, A)/B^i(G, A).$$

In particular, if $A$ is a trivial $G$-module, then $H^*(G, A) = H^*(BG, A)$.

So far, all definitions looked very natural. Our next step is to reduce the number of variables in the definition of cochains by one using the fact that cochains in $Hom_G(C_k, A)$ are $G$-invariant. The drawback of this reduction, as we will see, will be lack of naturality, but the advantage will be new formulae for cohomology groups which are useful in some applications.

By $G$-invariance, for $f \in Hom_G(C_k, A)$ we have:

$$f(g_0, \dots, g_k) = g_0 \cdot f(1, g_0^{-1} g_1, \dots, g_0^{-1} g_k)$$

In other words, it suffices to restrict cochains to the set of $(k+1)$-tuples where the first entry is $1 \in G$. Every such tuple has the form

$$(1, g_1, g_1 g_2, \dots, g_1 \cdots g_k)$$

(we will see below why). The latter is commonly denoted

$$[g_1 | g_2 | \dots | g_k].$$

Note that computing the value of the coboundary,

$$\delta_{k-1} f(1, g_1, g_1 g_2, \dots, g_1 \cdots g_k) = \delta_{k-1} f([g_1 | g_2 | \dots | g_k])$$

we get

$$\delta_{k-1} f(1, g_1, g_1 g_2, \dots, g_1 \cdots g_k) =$$

$$f(g_1, \dots, g_1 \cdots g_k) - f(1, g_1 g_2, \dots, g_1 \cdots g_k) + f(1, g_1, g_1 g_2 g_3, \dots, g_1 \cdots g_k) - \dots =$$

$$g_1 \cdot f(1, g_2, \dots, g_2 \cdots g_k) - f([g_1 g_2 | g_3 | \dots | g_k]) + f([g_1 | g_2 g_3 | g_4 | \dots | g_k]) - \dots =$$

$$g_1 \cdot f([g_2 | \dots | g_k]) - f([g_1 g_2 | g_3 | \dots | g_k]) + f([g_1 | g_2 g_3 | g_4 | \dots | g_k]) - \dots$$

Thus,

$$\delta_{k-1} f([g_1|g_2|\dots|g_k]) = g_1 \cdot f([g_2|\dots|g_k]) - f([g_1g_2|g_3|\dots|g_k]) +$$

$$f([g_1|g_2g_3|g_4|\dots|g_k]) - \dots$$

Then, we let $\bar{C}^k$ ($k \geq 1$) denote the abelian group of functions $f$ sending $k$-tuples $[g_1|\dots|g_k]$ of elements of $G$ to elements of $A$; we equip these groups with the above coboundary homomorphisms $\delta_k$. For $k = 0$, we have to use the empty symbol $[\ ]$, $f([\ ]) = a \in A$, so that such functions $f$ are identified with elements of $A$. Thus, $\bar{C}_0 = A$ and the above formula for $\delta_0$ reads as:

$$\delta_0 : a \mapsto c_a, \quad c_a([g]) = g \cdot a - a.$$

The resulting chain complex $(\bar{C}_*, \delta_*)$ is called the *inhomogeneous bar complex* of $G$ with coefficients in $A$. We now compute the coboundary maps $\delta_k$ for this complex for small values of $k$:

(1) $\delta_0 : a \mapsto f_a, \quad f_a([g]) = g \cdot a - a$.
(2) $\delta_1(f)([g_1, g_2]) = g_1 \cdot f([g_2]) - f([g_1g_2]) + f([g_1])$.
(3) $\delta_2(f)([g_1|g_2|g_3]) = g_1 \cdot f([g_2|g_3]) - f([g_1g_2|g_3]) + f([g_1|g_2g_3]) - f([g_1|g_2])$.

Therefore, spaces of coboundaries and cocycles for $(\bar{C}_*, \delta_*)$ in small degrees are (we now drop the bar notation for simplicity):

(1) $B^1(G, A) = \{f_a : G \to A, \forall a \in A | f_a(g) = g \cdot a - a\}$.
(2) $Z^1(G, A) = \{f : G \to A | f(g_1g_2) = f(g_1) + g_1 \cdot f(g_2)\}$.
(3) $B^2(G, A) = \{h : G \times G \to A | \exists f : G \to A, h(g_1, g_2) = f(g_1) - f(g_1g_2) + g_1 \cdot f(g_2)\}$.
(4) $Z^2(G, A) = \{f : G \times G \to A | g_1 \cdot f(g_2, g_3) - f(g_1, g_2) = f(g_1g_2, g_3) - f(g_1, g_2g_3)\}$.

Let us look at the definition of $Z^1(G, A)$ more closely. In addition to the left action of $G$ on $A$, we define a *trivial right action* of $G$ on $A$: $a \cdot g = a$. Then a function $f : G \to A$ is a 1-cocycle if and only if

$$f(g_1g_2) = f(g_1) \cdot g_2 + g_1 \cdot f(g_2).$$

The reader will immediately recognize here the Leibnitz formula for the derivative of the product. Hence, 1-cocycles $f \in Z^1(G, A)$ are called *derivations* of $G$ with values in $A$. The 1-coboundaries are called *principal derivations*. If $A$ is trivial as a left $G$-module, then, of course, all principal derivations are zero and derivations are just homomorphisms $G \to A$.

**Nonabelian derivations.** The notions of derivation and principal derivation can be extended to the case when the target group is nonabelian; we will use the notation $N$ for the target group with the binary operation $\star$ and $g \cdot n$ for the action of $G$ on $N$ by automorphisms, i.e.,

$$g \cdot n = \varphi(g)(n), \quad \text{where } \varphi : G \to Aut(N) \text{ is a homomorphism.}$$

DEFINITION 3.67. A function $d : G \to N$ is called a *derivation* if

$$d(g_1g_2) = d(g_1) \star g_1 \cdot d(g_2), \quad \forall g_1, g_2 \in G.$$

A derivation is called *principal* if it is of the form $d = d_n$, where

$$d_n(g) = n^{-1} \star (g \cdot n).$$

The space of derivations is denoted $Der(G, N)$ and the subspace of principal derivations is denoted $Prin(G, N)$ or, simply, $P(G, N)$.

EXERCISE 3.68. Verify that every principal derivation is indeed a derivation.

EXERCISE 3.69. Verify that every derivation $d$ satisfies

- $d(1) = 1$;
- $d(g^{-1}) = g^{-1} \cdot [d(g)]^{-1}$.

We will use derivations in the context of free solvable groups in Section 11.2. In section (§3.11) we will discuss derivations in the context of semidirect products, while in §3.12 we explain how 2nd cohomology group $H^2(G, A)$ can be used to describe central co-extensions.

**Nonabelian cohomology.** We would like to define the 1-st cohomology $H^1(G, N)$, where the group $N$ is nonabelian and we have an action of $G$ on $N$. The problem is that neither $Der(G, N)$ nor $Prin(G, N)$ is a group, so taking quotient $Der(G, N)/Prin(G, N)$ makes no sense. Nevertheless, we can think of the formula

$$f \mapsto f + d_a, a \in A,$$

in the abelian case (defining action of $Prin(G, A)$ on $Der(G, A)$) as the *left* action of the group $A$ on $Der(G, A)$:

$$a(f) = f', \quad f'(g) = -a + f(g) + (g \cdot a).$$

The latter generalizes in the nonabelian case, the group $N$ acts to the left on $Der(G, N)$ by

$$n(f) = f', \quad f'(g) = n^{-1} \star f(g) \star (g \cdot n).$$

Then, one defines $H^1(G, N)$ as the quotient

$$N \backslash Der(G, N).$$

EXAMPLE 3.70. 1. Suppose that $G$-action on $N$ is trivial. Then $Der(G, N) = Hom(G, N)$ and $N$ acts on homomorphisms $f : G \to N$ by postcomposition with inner automorphisms. Thus, $H^1(G, N)$ in this case is

$$N \backslash Hom(G, N),$$

the set of conjugacy classes of homomorphisms $G \to N$.

2. Suppose that $G \cong \mathbb{Z} = \langle 1 \rangle$ and the action $\varphi$ of $\mathbb{Z}$ on $N$ is arbitrary. We have $\eta := \varphi(1) \in Aut(N)$. Then $H^1(G, N)$ is the set of *twisted conjugacy classes of elements of* $N$: Two elements $m_1, m_2 \in N$ are said to be in the same $\eta$-twisted conjugacy class if there exists $n \in N$ so that

$$m_2 = n^{-1} \star m_1 \star \eta(n).$$

Indeed, every derivation $d \in Der(\mathbb{Z}, N)$ is determined by the image $m = d(1) \in N$. Then two derivations $d_i$ so that $m_i = d_i(1)$ $(i = 1, 2)$ are in the same $N$-orbit if $m_1, m_2$ are in the same $\eta$-twisted conjugacy class.

## 3.10. Ring derivations

Our next goal is to extend the notion of derivation in the context of (noncommutative) rings. Typical rings that the reader should have in mind are *integer group rings*.

**Group rings.** The *(integer) group ring* $\mathbb{Z}G$ of a group $G$ is the set of formal sums $\sum_{g\in G} m_g\, g$, where $m_g$ are integers which are equal to zero for all but finitely many values of $g$. Then $\mathbb{Z}G$ is a ring when endowed with the two operations:

- addition:

$$\sum_{g\in G} m_g\, g + \sum_{g\in G} n_g\, g = \sum_{g\in G}(m_g + n_g)\, g$$

- multiplication defined by the convolution of maps to $\mathbb{Z}$, that is

$$\sum_{a\in G} m_a\, a + \sum_{b\in G} n_b\, b = \sum_{g\in G}\left(\sum_{ab=g} m_a n_b\right) g\,.$$

According to a Theorem of G. Higman [**Hig40**], every group ring is an integral domain. Both $\mathbb{Z}$ and $G$ embed as subsets of $\mathbb{Z}G$ by identifying every $m\in\mathbb{Z}$ with $m1_G$ and every $g\in G$ with $1g$. Every homomorphism between groups $\varphi : G\to H$ induces a homomorphism between group rings, which by abuse of notation we shall denote also by $\varphi$. In particular, the trivial homomorphism $o : G\to\{1\}$ induces a retraction $o : \mathbb{Z}G\to\mathbb{Z}$, called the *augmentation*. If the homomorphism $\varphi : G\to H$ is an isomorphism then so is the homomorphism between group rings. This implies that an action of a group $G$ on another group $H$ (by automorphisms) extends to an action of $G$ on the group ring $\mathbb{Z}H$ (by automorphisms).

Let $L$ be a ring and $M$ be an abelian group. We say that $M$ is a (left) $L$-module if we are given a map

$$(\ell, m)\mapsto \ell\cdot m, L\times M\to M,$$

which is additive in both variables and so that

$$(\ell_1 \star \ell_2)\cdot m = \ell_1\cdot(\ell_2\cdot m), \tag{3.3}$$

where $\star$ denotes the multiplication operation in $L$.

Similarly, the ring $M$ is the *right* $L$-module if we are given an additive in both variables map

$$(m, \ell)\mapsto m\cdot \ell, M\times L\to M,$$

so that

$$m\cdot(\ell_1\star\ell_2) = (m\cdot\ell_1)\cdot\ell_2. \tag{3.4}$$

Lastly, $M$ is an $L$-*bimodule* if $M$ has structure of both left and right $L$-module.

Definition 3.71. Let $M$ be an $L$-bimodule. A *derivation* (with respect to this bimodule structure) is a map $d : L\to M$ so that:

(1) $d(\ell_1+\ell_2) = d(\ell_1) + d(\ell_2)$,
(2) $d(\ell_1\star\ell_2) = d(\ell_1)\cdot\ell_2 + \ell_1\cdot d(\ell_2)$.

The space of derivations is an abelian group, which will be denoted $Der(L, M)$.

Below is the key example of a bimodule that we will be using in the context of derivations. Let $G, H$ be groups, $\varphi : G \to Bij(H)$ is an action of $G$ on $H$ by set-theoretic automorphisms. We let $L := \mathbb{Z}G, M := \mathbb{Z}H$, where we regard the ring $M$ as an abelian group and ignore its multiplicative structure.

Every action $\varphi : G \curvearrowright H$ determines the left $L$-module structure on $M$ by:

$$(\sum_i a_i g_i) \cdot (\sum_j b_i h_i) := \sum_{i,j} a_i b_i g_i \cdot h_i, \quad a_i \in \mathbb{Z}, b_j \in \mathbb{Z},$$

where $g \cdot h = \varphi(g)(h)$ for $g \in G, h \in H$. We define the structure of right $L$-module on $M$ by:

$$(m, \ell) \mapsto mo(\ell) = o(\ell)m, \quad o(\ell) \in \mathbb{Z}$$

where $o : L \to \mathbb{Z}$ is the augmentation of $\mathbb{Z}G = L$.

Derivations with respect for the above group ring bimodules will be called *group ring derivations*.

EXERCISE 3.72. Verify the following properties of group ring derivations:

$(P_1)$ $d(1_G) = 0$, whence $d(m) = 0$ for every $m \in \mathbb{Z}$;

$(P_2)$ $d(g^{-1}) = -g^{-1} \cdot d(g)$;

$(P_3)$ $d(g_1 \cdots g_m) = \sum_{i=1}^m (g_1 \cdots g_{i-1}) \cdot d(g_i) o(g_{i+1} \cdots g_m)$.

$(P_4)$ Every derivation $d \in Der(\mathbb{Z}G, \mathbb{Z}H)$ is uniquely determined by its values $d(x)$ on generators $x$ of $G$.

**Fox Calculus.** We now consider the special case when $G = H = F_X$, is the free group on the generating set $X$. In this context, theory of derivations was developed in [**Fox53**].

LEMMA 3.73. *Every map $d : X \to M = \mathbb{Z}G$ extends to a group ring derivation $d \in Der(\mathbb{Z}G, M)$.*

PROOF. We set

$$d(x^{-1}) = -x^{-1} \cdot d(x), \quad \forall x \in X$$

and $d(1) = 0$. We then extend $d$ inductively to the free group $G$ by

$$d(yu) = d(y) + y \cdot d(u),$$

where $y = x \in X$ or $y = x^{-1}$ and $yu$ is a reduced word in the alphabet $X \cup X^{-1}$. We then extend $d$ by additivity to the rest of the ring $L = \mathbb{Z}G$. In order to verify that $d$ is a derivation, we need to check only that

$$d(uv) = d(u) + u \cdot d(v),$$

where $u, v \in F_X$. The verification is a straightforward induction on the length of the reduced word $u$ and is left to the reader. $\square$

NOTATION 3.74. To each generator $x_i \in X$ we associate a derivation $\partial_i$, called *Fox derivative*, defined by $\partial_i x_j = \delta_{ij} \in \mathbb{Z} \subset \mathbb{Z}G$. In particular,

$$\partial_i(x_i^{-1}) = -x_i^{-1}.$$

PROPOSITION 3.75. *Suppose that $G = F_r$ is free group of rank $r < \infty$. Then every derivation $d \in Der(\mathbb{Z}G, \mathbb{Z}G)$ can be written as a sum*

$$d = \sum_{i=1}^{r} k_i \partial_i \,, \quad \text{where } k_i = d(x_i) \in \mathbb{Z}\,.$$

*Furthermore, $Der(\mathbb{Z}G, \mathbb{Z}G)$ is a free abelian group with the basis $\partial_i, i = 1, \ldots, r$.*

PROOF. The first assertion immediately follows from Exercise 3.72 (part $(P_4)$), and from the fact that both sides of the equation evaluated on $x_j$ equal $k_j$. Thus, the derivations $\partial_i, i = 1, \ldots, k$ generate $Der(\mathbb{Z}G, \mathbb{Z}G)$. Independence of these generators follows from $\partial_i x_j = \delta_{ij}$. □

## 3.11. Derivations and split extensions

**Components of homomorphisms to semidirect products.**

DEFINITION 3.76. Let $G$ and $L$ be two groups and let $N, H$ be subgroups in $G$.

(1) Assume that $G = N \times H$. Every group homomorphism $F : L \to G$ splits as a product of two homomorphisms $F = (f_1, f_2)$, $f_1 : L \to N$ and $f_2 : L \to H$, called the *components* of $F$.

(2) Assume now that $G$ is a semidirect product $N \rtimes H$. Then every homomorphism $F : L \to G$ is determines (and is determined by) a pair $(d, f)$, where
- $f : L \to H$ is a homomorphism (the composition of $F$ and the retraction $G \to H$);
- a map $d = d_F : L \to N$, called *derivation* associated with $F$. The derivation $d$ is determined by the formula
$$F(\ell) = d(\ell) f(\ell).$$

EXERCISE 3.77. Show that $d$ is indeed a derivation.

EXERCISE 3.78. Verify that for every derivation $d$ and a homomorphism $f : L \to H$ there exists a homomorphism $F : L \to G$ with the components $d, f$.

**Extensions and co-extensions.**

DEFINITION 3.79. Given a short exact sequence

$$\{1\} \longrightarrow N \longrightarrow G \longrightarrow H \longrightarrow \{1\},$$

we call the group $G$ an *extension of $N$ by $H$* or a *co-extension of $H$ by $N$*.

Given two classes of groups $\mathcal{A}$ and $\mathcal{B}$, the groups that can be obtained as extensions of $N$ by $H$ with $N \in \mathcal{A}$ and $H \in \mathcal{B}$, are called $\mathcal{A}$-by-$\mathcal{B}$ *groups* (e.g. abelian-by-finite, nilpotent-by-free etc.).

Two extensions defined by the short exact sequences

$$\{1\} \longrightarrow N_i \xrightarrow{\varphi_i} G_i \xrightarrow{\psi_i} H_i \longrightarrow \{1\}$$

$(i = 1, 2)$ are *equivalent* if there exist isomorphisms

$$f_1 : N_1 \to N_2, \quad f_2 : G_1 \to G_2, \quad f_3 : H_1 \to H_2$$

that determine a commutative diagram:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & N_1 & \longrightarrow & G_1 & \longrightarrow & H_1 & \longrightarrow & 1 \\ & & \Big\downarrow f_1 & & \Big\downarrow f_2 & & \Big\downarrow f_3 & & \\ 1 & \longrightarrow & N_2 & \longrightarrow & G_2 & \longrightarrow & H_2 & \longrightarrow & 1 \end{array}$$

We now use the notion of isomorphism of exact sequences to reinterpret the notion of split extension.

PROPOSITION 3.80. *Consider a short exact sequence*

$$1 \to N \overset{\iota}{\to} G \overset{\pi}{\to} Q \to 1\,. \tag{3.5}$$

*The following are equivalent:*

(1) *the sequence splits;*

(2) *there exists a subgroup $H$ in $G$ such that the projection $\pi$ restricted to $H$ becomes an isomorphism.*

(3) *the extension $G$ is equivalent to an extension corresponding to a semidirect product $N \rtimes Q$;*

(4) *there exists a subgroup $H$ in $G$ such $G = N \rtimes H$.*

PROOF. It is clear that (2) $\Rightarrow$ (1).

(1) $\Rightarrow$ (2): Let $\sigma : Q \to \sigma(H) \subset G$ be a section. The equality $\pi \circ \sigma = \mathrm{id}_Q$ implies that $\pi$ restricted to $H$ is both surjective and injective.

The implication (2) $\Rightarrow$ (3) is obvious.

(3) $\Rightarrow$ (2): Assume that there exists $H$ such that $\pi|_H$ is an isomorphism. The fact that it is surjective implies that $G = NH$. The fact that it is injective implies that $H \cap N = \{1\}$.

(2) $\Rightarrow$ (4): Since $\pi$ restricted to $H$ is surjective, it follows that for every $g \in G$ there exists $h \in H$ such that $\pi(g) = \pi(h)$, hence $gh^{-1} \in \operatorname{Ker} \pi = \operatorname{Im} \iota$.

Assume that $g \in G$ can be written as $g = \iota(n_1)h_1 = \iota(n_2)h_2$, with $n_1, n_2 \in N$ and $h_1, h_2 \in H$. Then $\pi(h_1) = \pi(h_2)$, which, by the hypothesis that $\pi$ restricted to $H$ is an isomorphism, implies $h_1 = h_2$, whence $\iota(n_1) = \iota(n_2)$ and $n_1 = n_2$ by the injectivity of $\iota$.

(4) $\Rightarrow$ (2): The existence of the decomposition for every $g \in G$ implies that $\pi$ restricted to $H$ is surjective.

The uniqueness of the decomposition implies that $H \cap \operatorname{Im} \iota = \{1\}$, whence $\pi$ restricted to $H$ is injective. □

REMARK 3.81. Every sequence with free nonabelian group $Q$ splits: Construct a section $\sigma : Q \to G$ by sending each free generator $x_i$ of $Q$ to an element $\tilde{x}_i \in G$ so that $\pi(\tilde{x}_i) = x_i$. In particular, every group which admits an epimorphism to a free nonabelian group $F$, also contains a subgroup isomorphic to $F$.

EXAMPLES 3.82. (1) The short exact sequence

$$1 \longrightarrow (2\mathbb{Z})^n \longrightarrow \mathbb{Z}^n \longrightarrow \mathbb{Z}_2^n \longrightarrow 1$$

does not split.

(2) Let $F_n$ be a free group of rank $n$ (see Definition 4.16) and let $F_n'$ be its commutator subgroup (see Definition 3.53). Note that the abelianization of $F_n$ as defined in Proposition 3.55, (3), is $\mathbb{Z}^n$. The short exact sequence

$$1 \longrightarrow F_n' \longrightarrow F_n \longrightarrow \mathbb{Z}^n \longrightarrow 1$$

does not split.

From now on, we restrict to the case of exact sequences

$$1 \to A \xrightarrow{\iota} G \xrightarrow{\pi} Q \to 1, \tag{3.6}$$

where $A$ is an abelian group. Recall that the set of derivations $Der(Q, A)$ has natural structure of an abelian group.

REMARKS 3.83. (1) The short exact sequence (3.6) uniquely defines an action of $Q$ in $A$. Indeed $G$ acts on $A$ by conjugation and, since the kernel of this action contains $A$, it defines an action of $Q$ on $A$. In what follows we shall denote this action by $(q, a) \mapsto q \cdot a$, and by $\varphi$ the homomorphism $Q \to \mathrm{Aut}(A)$ defined by this action.

(2) If the short exact sequence (3.6) splits, the group $G$ is isomorphic to $A \rtimes_{\varphi} Q$.

**Classification of splittings.**
Below we discuss classification of all splittings of short exact sequences (3.6) which do split. We use the additive notation for the binary operation on $A$. We begin with few observations. From now on, we fix a section $\sigma_0$ and, hence, a semidirect product decomposition $G = A \rtimes Q$. Note that every splitting of a short exact sequence (3.6), is determined by a section $\sigma : Q \to G$. Furthermore, every section $\sigma : Q \to G$ is determined by its components $(d_\sigma, \pi)$ with respect to the semidirect product decomposition given by $\sigma_0$ (see Remark 3.76). Since $\pi$ is fixed, a section $\sigma$ is uniquely determined by its derivation $d_\sigma$. Conversely, every derivation $d \in Der(Q, A)$ determines a section $\sigma$, so that $d = d_\sigma$. Thus, the set of sections of (3.6) is in bijective correspondence with the abelian group of derivations $Der(Q, A)$.

Our next goal is to discuss the equivalence relation between different sections (and derivations). We say that an automorphism $\alpha \in Aut(G)$ is a *shearing* (with respect to the semidirect product decomposition $G = A \rtimes Q$) if $\alpha(A) = A, \alpha|A = Id$ and $\alpha$ projects to the identity on $Q$. Examples of shearing automorphisms are *principal shearing automorphisms*, which are given by conjugations by elements $a \in A$. It is clear that shearing automorphisms act on splittings of the short exact sequence (3.6).

EXERCISE 3.84. The group of shearing automorphisms of $G$ is isomorphic to the abelian group $Der(Q, A)$: Every derivation $d \in Der(Q, A)$ determines a shearing automorphism $\alpha = \alpha_d$ of $G$ by the formula

$$\alpha(a \star q) = (a + d(q)) \star q$$

which gives the bijective correspondence.

In view of this exercise, the classification of splittings modulo shearing automorphisms yields a very boring answer: All sections are equivalent under the group of shearing transformations. A finer classification of splittings is given by the following definition. We say that two splittings $\sigma_1, \sigma_2$ are *A-conjugate* if they differ by a principal shearing automorphism:

$$\sigma_2(q) = a\sigma_1(q)a^{-1}, \forall q \in Q,$$

where $a \in A$. If $d_1, d_2$ are the derivations corresponding to the sections $\sigma_1, \sigma_2$, then

$$(d_2(q), q) = (a, 1)(d_1(q), q)(-a, 1) \Leftrightarrow d_2(q) = d_1(q) - [q \cdot a - a] .$$

In other words, $d_1, d_2$ differ by the principal derivation corresponding to $a \in A$.

Thus, we proved the following

PROPOSITION 3.85. *A-conjugacy classes of splittings of the short exact sequence* (3.6) *are in bijective correspondence with the quotient*

$$Der(Q, A)/Prin(Q, A),$$

*where $Prin(Q, A)$ is the subgroup of principal derivations.*

Note that $Der(Q, A) \cong Z^1(Q, A)$, $Prin(Q, A) = B^1(Q, A)$ and the quotient $Der(Q, A)/Prin(Q, A)$ is $H^1(Q, A)$, the first cohomology group of $Q$ with coefficients in the $\mathbb{Z}Q$–module $A$.

Below is another application of $H^1(Q, A)$. Let $L$ be a group and $F : L \to G = A \rtimes Q$ be a homomorphism. The group $G$, of course, acts on the homomorphisms $F$ by postcomposition with inner automorphisms. Two homomorphisms are said to be *conjugate* if they belong to the same orbit of this $G$-action.

LEMMA 3.86. *1. A homomorphism $F : L \to G$ is conjugate to a homomorphism with the image in $Q$ if and only if the derivation $d_F$ of $F$ is principal.*

*2. Furthermore, suppose that $F_i : L \to G$ are homomorphisms with components $(d_i, \pi), i = 1, 2$. Then $F_1$ and $F_2$ are A-conjugate if and only if $[d_1] = [d_2] \in H^1(L, A)$.*

PROOF. Let $g = qa \in G, a \in A, q \in Q$. If $(qa)F(\ell)(qa)^{-1} \in Q$, then $aF(\ell)a^{-1} \in Q$. Thus, for (1) it suffices to consider $A$-conjugation of homomorphisms $F : L \to G$. Hence, (2) $\Rightarrow$ (1). To prove (2) we note that the composition of $F$ with an inner automorphism defined by $a \in A$ has the derivation equal to $d_F - d_a$, where $d_a$ is the principal derivation determined by $a$. □

### 3.12. Central co-extensions and 2-nd cohomology

We restrict ourselves to the case of central co-extensions (a similar result holds for general extensions with abelian kernels, see e.g. [**Bro82b**]). In this case, $A$ is trivial as a $G$-module and, hence, $H^*(G, A) \cong H^k(K(G, 1), A)$. This cohomology group can be also computed as $H^k(Y, A)$, where $G = \pi_1(Y)$ and $Y$ is $k+1$-connected cell complex.

Let $G$ be a group and $A$ an abelian group. A *central co-extension* of $G$ by $A$ is a short exact sequence

$$1 \to A \xrightarrow{\iota} \tilde{G} \xrightarrow{r} G \to 1$$

where $\iota(A)$ is contained in the center of $\tilde{G}$. Choose a set-theoretic section $s : G \to \tilde{G}, s(1) = 1, r \circ s = Id$. Then, the group $\tilde{G}$ is be identified (as a set) with the direct product $A \times G$. With this identification, the group operation on $\tilde{G}$ has the form

$$(a, g) \cdot (b, h) = (a + b + f(g, h), gh),$$

where $f(1,1) = 0 \in A$. Here the function $f : G \times G \to A$ measures the failure of $s$ to be a homomorphism:

$$f(g,h) = s(g)s(h)\left(s(gh)\right)^{-1}.$$

Not every function $f : G \times G \to A$ corresponds to a central extension: A function $f$ gives rise to a central co-extension if and only if it satisfies the *cocycle identity*:

$$f(g,h) + f(gh,k) = f(h,k) + f(g,hk).$$

In other words, the set of such functions is the abelian group of cocycles $Z^2(G,A)$, see §3.9. We will refer to $f$ simply as a *cocycle*.

Two central co-extensions are said to be equivalent if there exist an isomorphism $\tau$ making the following diagram commutative:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & A & \longrightarrow & \tilde{G}_1 & \longrightarrow & G & \longrightarrow & 1\\
 & & \downarrow{\scriptstyle id} & & \downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle id} & & \\
1 & \longrightarrow & A & \longrightarrow & \tilde{G}_2 & \longrightarrow & G & \longrightarrow & 1
\end{array}$$

For instance, a co-extension is trivial, meaning equivalent to the product $A \times G$, if and only if the central co-extension splits. We will use the notation $\mathbb{E}(G,A)$ to denote the set of equivalence classes of co-extensions. In the language of cocycles, $r_1 \sim r_2$ if and only if

$$f_1 - f_2 = \delta c,$$

where $c : G \to A$, and

$$\delta c(g,h) = c(g) + c(h) - c(gh)$$

is the coboundary, $c \in B^2(G,A)$. Recall that $H^2(G,A) = Z^2(G,A)/B^2(G,A)$ is the 2-nd cohomology group of $G$ with coefficients in $A$.

The set $\mathbb{E}(G,A)$ has natural structure of an abelian group, where the sum of two co-extensions

$$A \to G_i \xrightarrow{r_i} G$$

is defined by

$$G_3 = \{(g_1,g_2) \in G_1 \times G_2 | r_1(g_1) = r_2(g_2)\} \xrightarrow{r} G,$$

$r(g_1,g_2) = r_1(g_1) = r_2(g_2)$. The kernel of this co-extension is the subgroup $A$ embedded diagonally in $G_1 \times G_2$. In the language of cocycles $f : G \times G \to A$, the sum of co-extensions corresponds to the sum of cocycles and the trivial element is represented by the cocycle $f = 0$.

To summarize:

THEOREM 3.87 (See Chapter IV in **[Bro82b]**.). *There exists an isomorphism of abelian groups*

$$H^2(K(G,1),A) \cong H^2(G,A) \to \mathbb{E}(G,A).$$

**Co-extensions and group presentations.** Below we describe the isomorphism in Theorem 3.87 in terms of generators and relators, which will require familiarity with some of the material in Chapter 4.

Start with a presentation $\langle \mathcal{X} | \mathcal{R} \rangle$ of the group $G$ and let $Y^2$ denote the corresponding presentation complex (see Definition 4.80). Embed $Y^2$ in a 3-connected cell complex $Y$ by attaching appropriate 3-cells to $Y^2$. Then $H^2(Y,A) \cong H^2(G,A)$. Each cohomology class $[\zeta] \in H^2(G,A)$ is realized by a cocycle $\zeta \in Z^2(Y,A)$, which will assigns elements of $A$ to each 2-cell in $Y$. The 2-cells $c_i$ of $Y$ are indexed by the defining relators $R_i, i \in I$, of $G$. By abusing the notation, we set $\zeta(R_i) := \zeta(c_i)$, so that $\zeta(R_i^{-1}) = -\zeta(c_i)$. Given such $\zeta$, define the group $\tilde{G} = \tilde{G}_\zeta$ by the presentation

$$\tilde{G} = \left\langle \tilde{\mathcal{X}} = \mathcal{X} \cup A | [a,x] = 1, \forall a \in A, \forall x \in \tilde{\mathcal{X}}; R_i(\zeta(R_i))^{-1} = 1, i \in I \right\rangle.$$

In particular, if $w$ is a word in the alphabet $\mathcal{X}$, which is a product of conjugates of the relators $R_{i_j}^{t_j}, t_j = \pm 1$, then

$$w \cdot \left( \sum_j t_j \zeta(c_{i_j}) \right) = 1 \tag{3.7}$$

in $\tilde{G}$.

Clearly, we have the epimorphism $r : \tilde{G} \to G$ which sends every $a \in A \subset \tilde{\mathcal{X}}$ to $1 \in G$. We need to identify the kernel $r$. We have a homomorphism $\iota : A \to \tilde{G}$, defined by $a \to a \in A \subset \tilde{\mathcal{X}}, a \in A$. Furthermore, $\iota(A)$ is a central subgroup of $\tilde{G}$, hence, $Ker(r) = \iota(A)$, since the homomorphism $r$ amounts to dividing $\tilde{G}$ by $\tilde{A}$.

We next show that $\iota$ is injective. Let $\tilde{Y}$ denote the presentation complex $\tilde{Y}$ for $\tilde{G}$; the homomorphism $r : \tilde{G} \to G$ is induced by the map $F : \tilde{Y} \to Y$ which collapses each loop corresponding to $a \in A$ to the vertex of $Y$ and sends 2-cells corresponding to the relators $[x,a], x \in X$, to the base-point in $Y$. So far we did not use the assumption that $\zeta$ is a cocycle, i.e., that $\zeta(\sigma) = 0$ whenever $\sigma$ is the boundary o a 3-cycle in $Y$. Suppose that $\iota(a) = 1 \in \tilde{G}, a \in A$. Then the loop $\alpha$ in $\tilde{Y}$ corresponding to $a$ bounds a 2-disk $\tilde{\sigma}$ in $\tilde{Y}$. The image of this disk under $f$ is a spherical 2-cycle $\sigma$ in $Y$ since $F$ is constant on $\alpha$. The spherical cycle $\sigma$ is null-homologous since $Y$ is 2-connected, $\sigma = \partial \xi$, $\xi \in C^3(Y,A)$. Since $\zeta$ is a cocycle, $0 = \zeta(\partial \xi) = \zeta(\sigma)$. Thus, equation (3.7), implies that $a = \zeta(\sigma) = 0$ in $A$. This means that $\iota$ is injective.

Suppose the cocycle $\zeta$ is a coboundary, $\zeta = \delta \eta$, where $\eta \in C^1(Y^1, A)$, i.e., $\eta$ yields a homomorphism $\eta' : G \to A, \eta'(x_k) = a_k$. We then define a map $s : G \to \tilde{G}$ by $s(x_k) = x_k a_k$. Then relations $R_i = \zeta(R_i)$ imply that $s(R_i) = 1$ in $\tilde{G}$, so the co-extension defined by $\zeta$ splits and, hence, is trivial.

We, thus, have a map from $H^2(Y,A)$ to the set $\mathbb{E}(G,A)$.

If, $\zeta \in Z^2(Y,A)$ maps to a trivial co-extension $\tilde{G} \to G$ of $G$ by $A$, this means that we have a section $s : G \to \tilde{G}$. Then, for every generator $x_k \in X$ of the group $G$, we have $s(x_k) = x_k a_k$, for some $a_k \in A$. Thus, we define a 1-cochain $\eta \in C^1(Y^1, A)$ by $\eta(x_k) = a_k$, where we identify $x_k$ with a 1-cell in $Y^1$. Then the same arguments as above, run in the reverse, imply that $\zeta = \delta \eta$ and, hence $[\zeta] = 0 \in H^2(Y,A)$.

EXAMPLE 3.88. Let $G$ be the fundamental group of a genus $p \geqslant 1$ closed oriented surface $S$. Take the standard presentation of $G$, so that $S$ is the (aspherical) presentation complex. Let $A = \mathbb{Z}$ and take $[\zeta] \in H^2(G,\mathbb{Z}) \cong H^2(S,\mathbb{Z})$ be the class Poincaré dual to the fundamental class of $S$. Then for the unique 2-cell $c$ in $S$ corresponding to the relator

$$R = [a_1, b_1] \cdots [a_p, b_p],$$

we have $\zeta(c) = -1 \in \mathbb{Z}$. The corresponding group $\tilde{G}$ has the presentation

$$\langle a_1, b_1, \ldots, a_p, b_p, t | \quad [a_1, b_1] \cdots [a_p, b_p] t, [a_i, t], [b_i, t], i = 1, \ldots, p \rangle .$$

The conclusion, thus, is that a group $G$ with nontrivial 2-nd cohomology group $H^2(G, A)$ admits nontrivial central co-extensions with the kernel $A$. How does one construct groups with nontrivial $H^2(G, A)$? Suppose that $G$ admits an aspherical presentation complex $Y$ so that $\chi(G) = \chi(Y) \geqslant 2$. Then for $A \cong \mathbb{Z}$, we have

$$\chi(G) = 1 - b_1(Y) + b_2(Y) \geqslant 2 \Rightarrow b_2(Y) > 0.$$

The universal coefficients theorem then shows that if $A$ is an abelian group which admits an epimorphism to $\mathbb{Z}$, then $H^2(G, A) \neq 0$ provided that $\chi(Y) \geqslant 2$ as before.

CHAPTER 4

# Finitely generated and finitely presented groups

## 4.1. Finitely generated groups

A group which has a finite generating set is called *finitely generated.*

REMARK 4.1. In French, the terminology for finitely generated groups is *groupe de type fini.* On the other hand, in English, *group of finite type* is a much stronger requirement than finite generation (typically, this means that the group has type $\mathbf{F}_\infty$).

EXERCISE 4.2. Show that every finitely generated group is countable.

EXAMPLES 4.3.
(1) The group $(\mathbb{Z}, +)$ is finitely generated by both $\{1\}$ and $\{-1\}$. Also, any set $\{p, q\}$ of coprime integers generates $\mathbb{Z}$.
(2) The group $(\mathbb{Q}, +)$ is not finitely generated.

EXERCISE 4.4. Prove that the transposition $(12)$ and the cycle $(12\dots n)$ generate the permutation group $S_n$.

REMARKS 4.5.
(1) Every quotient $\bar{G}$ of a finitely generated group $G$ is finitely generated; we can take as generators of $\bar{G}$ the images of the generators of $G$.
(2) If $N$ is a normal subgroup of $G$, and both $N$ and $G/N$ are finitely generated, then $G$ is finitely generated. Indeed, take a finite generating set $\{n_1, .., n_k\}$ for $N$, and a finite generating set $\{g_1N, ..g_mN\}$ for $G/N$. Then
$$\{g_i, n_j \ : \ 1 \leqslant i \leqslant m\}, 1 \leqslant j \leqslant k\}\}$$
is a finite generating set for $G$.

REMARK 4.6. If $N$ is a normal subgroup in a group $G$ and $G$ is finitely generated, it *does not* necessarily follow that $N$ is finitely generated (not even if $G$ is a semidirect product of $N$ and $G/N$).

EXAMPLE 4.7. Let $H$ be the group of permutations of $\mathbb{Z}$ generated by the transposition $t = (01)$ and the translation map $s(i) = i+1$. Let $H_i$ be the group of permutations of $\mathbb{Z}$ supported on $[-i, i] = \{-i, -i+1, \dots, 0, 1, \dots, i-1, i\}$, and let $H_\omega$ be the group of finitely supported permutations of $\mathbb{Z}$ (i.e. the group of bijections $f : \mathbb{Z} \to \mathbb{Z}$ such that $f$ is the identity outside a finite subset of $\mathbb{Z}$),
$$H_\omega = \bigcup_{i=0}^{\infty} H_i \,.$$
Then $H_\omega$ is a normal subgroup in $H$ and $H/H_\omega \simeq \mathbb{Z}$, while $H_\omega$ is not finitely generated.

Indeed from the relation $s^k t s^{-k} = (k\, k+1)$, $k \in \mathbb{Z}$, it immediately follows that $H_\omega$ is a subgroup in $H$. It is likewise easy to see that $s^k H_i s^{-k} \subset H_{i+k}$, whence $s^k H_\omega s^{-k} \subset H_\omega$ for every $k \in \mathbb{Z}$.

If $g_1, \ldots, g_k$ is a finite set generating $H_\omega$, then there exists an $i \in \mathbb{N}$ so that all $g_j$'s are in $H_i$, hence $H_\omega = H_i$. On the other hand, clearly, $H_i$ is a proper subgroup of $H_\omega$.

EXERCISE 4.8. 1. Let $F$ be a non-abelian free group (see Definition 4.16). Let $\varphi : F \to \mathbb{Z}$ be any non-trivial homomorphism. Prove that the kernel of $\varphi$ is not finitely generated.

2. Let $F$ be a free group of finite rank with free generators $x_1, \ldots, x_n$; set $G := F \times F$. Then $G$ has the generating set

$$\{(x_i, 1), (1, x_j) : 1 \le i, j \le n\}.$$

Define homomorphism $\phi : G \to \mathbb{Z}$ sending every generator of $G$ to $1 \in \mathbb{Z}$. Show that the kernel $K$ of $\phi$ is finitely generated. Hint: Use the elements $(x_i, x_j^{-1})$, $(x_i x_j^{-1}, 1)$, $(1, x_i x_j^{-1})$, $1 \le i, j \le n$, of the subgroup $K$.

We will see later that a *finite index* subgroup of a finitely generated group is finitely generated (Lemma 4.75 or Theorem 5.29).

Below we describe a finite generating set for the group $GL(n, \mathbb{Z})$. In the proof we use *elementary matrices* $N_{i,j} = I_n + E_{i,j}$ $(i \ne j)$; here $I_n$ is the identity $n \times n$ matrix and the matrix $E_{i,j}$ has a unique non-zero entry 1 in the intersection of the $i$–th row and the $j$–th column.

PROPOSITION 4.9. *The group $GL(n, \mathbb{Z})$ is generated by*

$$s_1 = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \quad s_2 = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

$$s_3 = \begin{pmatrix} 1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \quad s_4 = \begin{pmatrix} -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

PROOF. *Step 1.* The permutation group $S_n$ acts (effectively) on $\mathbb{Z}^n$ by permuting the basis vectors; we, thus, obtain a monomorphism $\varphi : S_n \to GL(n, \mathbb{Z})$, so that $\varphi(12\ldots n) = s_1$, $\varphi(12) = s_2$. Consider now the corresponding action of $S_n$ on $n \times n$ matrices. Multiplication of a matrix by $s_1$ on the left permutes rows cyclically, multiplication to the right does the same with columns. Multiplication by $s_2$ on the left swaps the first two rows, multiplication to the right does the same with columns. Therefore, by multiplying an elementary matrix $A$ by appropriate products of $s_1, s_1^{-1}$ and $s_2$ on the left and on the right, we obtain the matrix $s_3$. In view of Exercise 4.4, the permutation $(12\ldots n)$ and the transposition $(12)$ generate the permutation group $S_n$. Thus, every elementary matrix $N_{ij}$ is a product of $s_1, s_1^{-1}, s_2$ and $s_3$.

Let $d_j$ denote the diagonal matrix with the diagonal entries $(1, \dots, 1, -1, 1, \dots 1)$, where $-1$ occurs in $j$-th place. Thus, $d_1 = s_4$. The same argument as above, shows that for every $d_j$ and $s = (1j) \in S_n$, $sd_js = d_1$. Thus, all diagonal matrices $d_j$ belong to the subgroup generated by $s_1, s_2$ and $s_4$.

*Step 2.* Now, let $g$ be an arbitrary element in $GL(n, \mathbb{Z})$. Let $a_1, \dots, a_n$ be the entries of the first column of $g$. We will prove that there exists an element $p$ in $\langle s_1 \, , \, \dots \, , \, s_4 \rangle \subset GL(n, \mathbb{Z})$, such that $pg$ has the entries $1, 0, \dots, 0$ in its first column. We argue by induction on $k = C_1(g) = |a_1| + \dots + |a_n|$. Note that $k \geqslant 1$. If $k = 1$, then $(a_1, \dots, a_n)$ is a permutation of $(\pm 1, 0, \dots, 0)$; hence, it suffices to take $p$ in $\langle s_1 \, , \, s_2, s_4 \rangle$ permuting the rows so as to obtain $1, 0, \dots, 0$ in the first column.

Assume that the statement is true for all integers $1 \leqslant i < k$; we will prove it for $k$. After to permuting rows and multiplying by $d_1 = s_4$ and $d_2$, we may assume that $a_1 > a_2 > 0$. Then $N_{1,2}d_2g$ has the following entries in the first column: $a_1 - a_2, -a_2, a_3, \dots a_n$. Therefore, $C_1\left(N_{1,2}d_2g\right) < C_1(g)$. By the induction assumption, there exists an element $p$ of $\langle s_1 \, , \, \dots \, , \, s_4 \rangle$ such that $pN_{1,2}d_2g$ has the entries of its first column equal to $1, 0, \dots, 0$. This proves the claim.

*Step 3.* We leave it to the reader to check that for every pair of matrices $A, B \in GL(n-1, \mathbb{R})$ and row vectors $L = (l_1, \dots, l_{n-1})$ and $M = (m_1, \dots, m_{n-1})$

$$\begin{pmatrix} 1 & L \\ 0 & A \end{pmatrix} \cdot \begin{pmatrix} 1 & M \\ 0 & B \end{pmatrix} = \begin{pmatrix} 1 & M + LB \\ 0 & AB \end{pmatrix}.$$

Therefore, the set of matrices

$$\left\{ \begin{pmatrix} 1 & L \\ 0 & A \end{pmatrix} \; ; \; A \in GL(n-1, \mathbb{Z}) \, , \, L \in \mathbb{Z}^{n-1} \right\}$$

is a subgroup of $GL(n, \mathbb{Z})$ isomorphic to $\mathbb{Z}^{n-1} \rtimes GL(n-1, \mathbb{Z})$.

Using this, an induction on $n$ and Step 2, one shows that there exists an element $p$ in $\langle s_1, \dots, s_4 \rangle$ such that $pg$ is upper triangular and with entries on the diagonal equal to 1. It, therefore, suffices to prove that every integer upper triangular matrix as above is in $\langle s_1, \dots s_4 \rangle$. This can be done for instance by repeating the argument in Step 2 with multiplications on the right. □

The wreath product (see Definition 3.65) is a useful construction of a finitely generated group from two finitely generated groups:

EXERCISE 4.10. Let $G$ and $H$ be groups, and $S$ and $X$ be their respective generating sets. Prove that $G \wr H$ is generated by

$$\{(f_s, 1_H) \mid s \in S\} \cup \{(f_1, x) \mid x \in X\},$$

where $f_s : H \to G$ is defined by $f_s(1_H) = s$, $f_s(h) = 1_G$, $\forall h \neq 1_H$.

In particular, if $G$ and $H$ are finitely generated then so is $G \wr H$.

EXERCISE 4.11. Let $G$ be a finitely generated group and let $S$ be an infinite set of generators of $G$. Show that there exists a finite subset $F$ of $S$ so that $G$ is generated by $F$.

EXERCISE 4.12. An element $g$ of the group $G$ is a *non-generator* if for every generating set $S$ of $G$, the complement $S \setminus \{g\}$ is still a generating set of $G$.

(a) Prove that the set of non-generators forms a subgroup of $G$. This subgroup is called the *Frattini subgroup*.

(b) Compute the Frattini subgroup of $(\mathbb{Z}, +)$.
(c) Compute the Frattini subgroup of $(\mathbb{Z}^n, +)$. (*Hint:* You may use the fact that $Aut(\mathbb{Z}^n)$ is $GL(n, \mathbb{Z})$, and that the $GL(n, \mathbb{Z})$–orbit of $e_1$ is the set of vectors $(k_1, \ldots, k_n)$ in $\mathbb{Z}^n$ such that $\gcd(k_1, \ldots, k_n) = 1$.)

DEFINITION 4.13. *A group $G$ is said to have* bounded generation property *(or is* boundedly generated*) if there exists a finite subset $\{t_1, \ldots, t_m\} \subset G$ such that every $g \in G$ can be written as $g = t_1^{k_1} t_2^{k_2} \cdots .t_m^{k_m}$ , where $k_1, k_2, \ldots, k_m$ are integers.*

Clearly, all finitely generated abelian groups have the bounded generation property, and so are all the finite groups. On the other hand, the nonabelian free f groups, which we will introduce in the next section, obviously, do not have the bounded generation property. For other examples of boundedly generated groups see Proposition 11.3.

## 4.2. Free groups

Let $X$ be a set. Its elements are called *letters* or *symbols*. We define the set of *inverse letters* (or *inverse symbols*) $X^{-1} = \{a^{-1} \mid a \in X\}$. We will think of $X \cup X^{-1}$ as an *alphabet*.

A *word* in $X \cup X^{-1}$ is a finite (possibly empty) string of letters in $X \cup X^{-1}$, i.e. an expression of the form

$$a_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \cdots a_{i_k}^{\epsilon_k}$$

where $a_i \in X, \epsilon_i = \pm 1$; here $x^1 = x$ for every $x \in X$. We will use the notation 1 for the *empty word* (the one which has no letters).

Denote by $X^*$ the set of words in the alphabet $X \cup X^{-1}$, where the empty word, denoted by 1, is included. For instance,

$$a_1 a_2 a_1^{-1} a_2 a_2 a_1 \in X^*.$$

The *length* of a word $w$ is the number of letters in this word. The length of the empty word is 0.

A word $w \in X^*$ is *reduced* if it contains no pair of consecutive letters of the form $aa^{-1}$ or $a^{-1}a$. The *reduction* of a word $w \in X^*$ is the deletion of all pairs of consecutive letters of the form $aa^{-1}$ or $a^{-1}a$.

For instance,

$$1, a_2 a_1, a_1 a_2 a_1^{-1}$$

are reduced, while

$$a_2 a_1 a_1^{-1} a_3$$

is not reduced.

More generally, a word $w$ is *cyclically reduced* if it is reduced and, in addition, the first and the last letters of $w$ are not inverses of each other.

We define an equivalence relation on $X^*$ by $w \sim w'$ if $w$ can be obtained from $w'$ by a finite sequence of reductions and their inverses, i.e., the relation $\sim$ on $X^*$ is generated by

$$u a_i a_i^{-1} v \sim uv, \quad u a_i^{-1} a_i v \sim uv$$

where $u, v \in X^*$.

PROPOSITION 4.14. *Any word $w \in X^*$ is equivalent to a unique reduced word.*

Proof. *Existence.* We prove the statement by induction on the length of a word. For words of length 0 and 1 the statement is clearly true. Assume that it is true for words of length $n$ and consider a word of length $n+1$, $w = a_1 \cdots a_n a_{n+1}$, where $a_i \in X \cup X^{-1}$. According to the induction hypothesis there exists a reduced word $u = b_1 \cdots b_k$ with $b_j \in X \cup X^{-1}$ such that $a_2 \cdots a_{n+1} \sim u$. Then $w \sim a_1 u$. If $a_1 \neq b_1^{-1}$ then $a_1 u$ is reduced. If $a_1 = b_1^{-1}$ then $a_1 u \sim b_2 \cdots b_k$ and the latter word is reduced.

*Uniqueness.* Let $F(X)$ be the set of reduced words in $X \cup X^{-1}$. For every $a \in X \cup X^{-1}$ we define a map $L_a : F(X) \to F(X)$ by

$$L_a(b_1 \cdots b_k) = \begin{cases} ab_1 \cdots b_k & \text{if} \quad a \neq b_1^{-1}, \\ b_2 \cdots b_k & \text{if} \quad a = b_1^{-1}. \end{cases}$$

For every word $w = a_1 \cdots a_n$ define $L_w = L_{a_1} \circ \cdots \circ L_{a_n}$. For the empty word 1 define $L_1 = \mathrm{id}$. It is easy to check that $L_a \circ L_{a^{-1}} = \mathrm{id}$ for every $a \in X \cup X^{-1}$, and to deduce from it that $v \sim w$ implies $L_v = L_w$.

We prove by induction on the length that if $w$ is reduced then $w = L_w(1)$. The statement clearly holds for $w$ of length 0 and 1. Assume that it is true for reduced words of length $n$ and let $w$ be a reduced word of length $n+1$. Then $w = au$, where $a \in X \cup X^{-1}$ and $u$ is a reduced word that does not begin with $a^{-1}$, i.e. such that $L_a(u) = au$. Then $L_w(1) = L_a \circ L_u(1) = L_a(u) = au = w$.

In order to prove uniqueness it suffices to prove that if $v \sim w$ and $v, w$ are reduced then $v = w$. Since $v \sim w$ it follows that $L_v = L_w$, hence $L_v(1) = L_w(1)$, that is $v = w$. □

Exercise 4.15. Give a geometric proof of this proposition using identification of $w \in X^*$ with the set of edge-paths $\mathfrak{p}_w$ in a regular tree $T$ of valence $2|X|$, which start at a fixed vertex $e$. The reduced path $\mathfrak{p}^*$ in $T$ corresponding to the reduction $w^*$ of $w$ is the unique geodesic in $T$ connecting $e$ to the terminal point of $\mathfrak{p}$. Uniqueness of $w^*$ then translates to the fact that a tree contains no circuits.

Let $F(X)$ be the set of reduced words in $X \cup X^{-1}$. Proposition 4.14 implies that $X^*/\sim$ can be identified with $F(X)$.

Definition 4.16. The *free group over* $X$ is the set $F(X)$ endowed with the product defined by: $w * w'$ is the unique reduced word equivalent to the word $ww'$. The unit is the empty word.

The cardinality of $X$ is called the *rank* of the free group $F(X)$.

The set $F(X)$ with the product defined in Definition 4.16 is indeed a group. The inverse of a reduced word

$$w = a_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \cdots a_{i_k}^{\epsilon_k}$$

by

$$w^{-1} = a_{i_k}^{-\epsilon_k} a_{i_{k-1}}^{-\epsilon_{k-1}} \cdots a_{i_1}^{-\epsilon_1}.$$

It is clear that $ww^{-1}$ project to the empty word 1 in $F$.

Remark 4.17. A free group of rank at least two is not abelian. Thus *free non-abelian* means free of rank at least two.

The *free semigroup* $F^s(X)$ *with the generating set* $X$ is defined in the fashion similar to $F(X)$, except that we only allow the words in the alphabet $X$ (and not in $X^{-1}$), in particular the reduction is not needed.

Proposition 4.18 (Universal property of free groups). *A map $\varphi : X \to G$ from the set $X$ to a group $G$ can be extended to a homomorphism $\Phi : F(X) \to G$ and this extension is unique.*

Proof. *Existence.* The map $\varphi$ can be extended to a map on $X \cup X^{-1}$ (which we denote also $\varphi$) by $\varphi(a^{-1}) = \varphi(a)^{-1}$.

For every reduced word $w = a_1 \cdots a_n$ in $F(X)$ define

$$\Phi(a_1 \cdots a_n) = \varphi(a_1) \cdots \varphi(a_n).$$

Set $\Phi(e) := 1$, the identity element of $G$. We leave it to the reader to check that $\Phi$ is a homomorphism.

*Uniqueness.* Let $\Psi : F(X) \to G$ be a homomorphism such that $\Psi(x) = \varphi(x)$ for every $x \in X$. Then for every reduced word $w = a_1 \cdots a_n$ in $F(X)$, $\Psi(w) = \Psi(a_1) \cdots \Psi(a_n) = \varphi(a_1) \cdots \varphi(a_n) = \Phi(w)$. □

Corollary 4.19. *Every group is the quotient of a free group.*

Proof. Apply Proposition 4.18 to the group $G$ and the set $X = G$. □

Lemma 4.20. *A short exact sequence $1 \to N \to G \xrightarrow{r} F(X) \to 1$ always splits. In particular, $G$ contains a subgroup isomorphic to $F(X)$.*

Proof. Indeed, for each $x \in X$ consider choose an element $t_x \in G$ projecting to $x$; the map $x \mapsto t_x$ extends to a group homomorphism $s : F(X) \to G$. Composition $r \circ s$ is the identity homomorphism $F(X) \to F(X)$ (since it is the identity on the generating set $X$). Therefore, the homomorphism $s$ is a splitting of the exact sequence. Since $r \circ s = Id$, $s$ a monomorphism. □

Corollary 4.21. *Every short exact sequence $1 \to N \to G \to \mathbb{Z} \to 1$ splits.*

### 4.3. Presentations of groups

Let $G$ be a group and $S$ a generating set of $G$. According to Proposition 4.18, the inclusion map $i : S \to G$ extends uniquely to an epimorphism $\pi_S : F(S) \to G$. The elements of $\operatorname{Ker} \pi_S$ are called *relators* (or *relations*) of the group $G$ with the generating set $S$.

*N.B.* In the above by an abuse of language we used the symbol $s$ to designate two different objects: $s$ is a letter in $F(S)$, as well as an element in the group $G$.

If $R = \{r_i \mid i \in I\} \subset F(S)$ is such that $\operatorname{Ker} \pi_S$ is normally generated by $R$ (i.e. $\langle\langle R \rangle\rangle = \operatorname{Ker} \pi_S$) then we say that the ordered pair $(S, R)$, usually denoted $\langle S | R \rangle$, is *a presentation of* $G$. The elements $r \in R$ are called *defining relators* (or *defining relations*) of the presentation $\langle S | R \rangle$.

By abuse of language we also say that the generators $s \in S$ and the *relations* $r = 1$, $r \in R$, constitute a presentation of the group $G$. Sometimes we will write presentations in the form

$$\langle s_i, i \in I | r_j = 1, j \in J \rangle$$

where

$$S = \{x_i\}_{i \in I}, \quad R = \{r_j\}_{j \in J}.$$

If both $S$ and $R$ are finite then the pair $S, R$ is called *a finite presentation of* $G$. A group $G$ is called *finitely presented* if it admits a finite presentation. Sometimes it is difficult, and even algorithmically impossible, to find a finite presentation of a finitely presented group, see [**BW11**].

Conversely, given an alphabet $S$ and a set $R$ of (reduced) words in the alphabet $S$ we can form the quotient

$$G := F(S)/\langle\langle R\rangle\rangle\,.$$

Then $\langle S|R\rangle$ is a presentation of $G$. By abusing notation, we will often write

$$G = \langle S|R\rangle$$

if $G$ is a group with the presentation $\langle S|R\rangle$. If $w$ is a word in the generating set $S$, we will use $[w]$ to denote its projection to the group $G$. An alternative notation for the equality

$$[v] = [w]$$

is

$$v \equiv_G w.$$

Note that the significance of a presentation of a group is the following:

- every element in $G$ can be written as a finite product $x_1 \cdots x_n$ with $x_i \in S \cup S^{-1} = \{s^{\pm 1} : s \in S\}$, i.e., as a word in the alphabet $S \cup S^{-1}$;
- a word $w = x_1 \cdots x_n$ in the alphabet $S \cup S^{-1}$ is *equal to the identity in* $G$, $w \equiv_G 1$, *if and only if* in $F(S)$ the word $w$ is the product of finitely many conjugates of the words $r_i \in R$, i.e.,

$$w = \prod_{i=1}^{m} r_i^{u_i}$$

for some $m \in \mathbb{N}$, $u_i \in F(S)$ and $r_i \in R$.

Below are few examples of group presentations:

EXAMPLES 4.22. (1) $\langle a_1, \ldots, a_n \mid [a_i, a_j], 1 \leqslant i, j \leqslant n\rangle$ is a finite presentation of $\mathbb{Z}^n$;

(2) $\left\langle x, y \mid x^n, y^2, yxyx\right\rangle$ is a presentation of the finite dihedral group $D_{2n}$;

(3) $\left\langle x, y \mid x^2, y^3, [x, y]\right\rangle$ is a presentation of the cyclic group $\mathbb{Z}_6$.

Let $\langle X|R\rangle$ be a presentation of a group $G$. Let $H$ be a group and $\psi : X \to H$ be a map which "preserves the relators", i.e., $\psi(r) = 1$ for every $r \in R$. Then:

LEMMA 4.23. *The map $\psi$ extends to a group homomorphism $\psi : G \to H$.*

PROOF. By the universal property of free groups, the map $\psi$ extends to a homomorphism $\tilde{\psi} : F(X) \to H$. We need to show that $\langle\langle R\rangle\rangle$ is contained in $\mathrm{Ker}(\tilde{\psi})$. However, $\langle\langle R\rangle\rangle$ consists of products of elements of the form $grg^{-1}$, where $g \in F, r \in R$. Since $\tilde{\psi}(grg^{-1}) = 1$, the claim follows. □

EXERCISE 4.24. The group $\bigoplus_{x\in X} \mathbb{Z}_2$ has presentation

$$\left\langle x \in X | x^2, [x, y], \forall x, y \in X\right\rangle.$$

Proposition 4.25 (Finite presentability is independent of the generating set). *Assume that a group $G$ has finite presentation $\langle S \mid R \rangle$, and let $\langle X \mid T \rangle$ be an arbitrary presentation of $G$, so that $X$ is finite. Then there exists a finite subset $T_0 \subset T$ such that $\langle X \mid T_0 \rangle$ is a presentation of $G$.*

Proof. Every element $s \in S$ can be written as a word $a_s(X)$ in $X$. The map $i_{SX} : S \to F(X)$, $i_{SX}(s) = a_s(X)$ extends to a unique homomorphism $p : F(S) \to F(X)$. Moreover, since $\pi_X \circ i_{SX}$ is an inclusion map of $S$ to $F(X)$, and both $\pi_S$ and $\pi_X \circ p$ are homomorphisms from $F(S)$ to $G$ extending the map $S \to G$, by the uniqueness of the extension we have that $\pi_S = \pi_X \circ p$. This implies that $\operatorname{Ker} \pi_X$ contains $p(r)$ for every $r \in R$.

Likewise, every $x \in X$ can be written as a word $b_x(S)$ in $S$, and this defines a map $i_{XS} : X \to F(S)$, $i_{XS}(x) = b_x(S)$, which extends to a homomorphism $q : F(X) \to F(S)$. A similar argument shows that $\pi_S \circ q = \pi_X$.

For every $x \in X$, $\pi_X(p(q(x))) = \pi_S(q(x)) = \pi_X(x)$. This implies that for every $x \in X$, $x^{-1}p(q(x))$ is in $\operatorname{Ker} \pi_X$.

Let $N$ be the normal subgroup of $F(X)$ normally generated by

$$\{p(r) \mid r \in R\} \cup \{x^{-1}p(q(x)) \mid x \in X\}\,.$$

We have that $N \leqslant \operatorname{Ker} \pi_X$. Therefore, there is a natural projection

$$\operatorname{proj} : F(X)/N \to F(X)/\operatorname{Ker} \pi_X\,.$$

Let $\bar{p} : F(S) \to F(X)/N$ be the homomorphism induced by $p$. Since $\bar{p}(r) = 1$ for all $r \in R$, it follows that $\bar{p}(\operatorname{Ker} \pi_S) = 1$, hence $\bar{p}$ induces a homomorphism $\varphi : F(S)/\operatorname{Ker} \pi_S \to F(X)/N$.

The homomorphism $\varphi$ is onto. Indeed, $F(X)/N$ is generated by elements of the form $xN = p(q(x))N$, and the latter is the image under $\varphi$ of $q(x) \operatorname{Ker} \pi_S$.

Consider the homomorphism $\operatorname{proj} \circ \varphi : F(S)/\operatorname{Ker} \pi_S \to F(X)/\operatorname{Ker} \pi_X$. Both the domain and the target groups are isomorphic to $G$. Each element $x$ of the generating set $X$ is sent by the isomorphism $G \to F(S)/\operatorname{Ker} \pi_S$ to $q(x) \operatorname{Ker} \pi_S$. The same element $x$ is sent by the isomorphism $G \to F(X)/\operatorname{Ker} \pi_X$ to $x \operatorname{Ker} \pi_X$. Note that

$$\operatorname{proj} \circ \varphi\,(q(x) \operatorname{Ker} \pi_S) = \operatorname{proj}(xN) = x \operatorname{Ker} \pi_X.$$

This means that modulo the two isomorphisms mentioned above, the map $\operatorname{proj} \circ \varphi$ is $\operatorname{id}_G$. This implies that $\varphi$ is injective, hence, a bijection. Therefore, proj is also a bijection. This happens if and only if $N = \operatorname{Ker} \pi_X$. In particular, $\operatorname{Ker} \pi_X$ is normally generated by the finite set of relators

$$\Re = \{p(r) \mid r \in R\} \cup \{x^{-1}p(q(x)) \mid x \in X\}.$$

Since $\Re = \langle\langle T \rangle\rangle$, every relator $\rho \in \Re$ can be written as a product

$$\prod_{i \in I_\rho} t_i^{v_i}$$

with $v_i \in F(X), t_i \in T$ and $I_\rho$ finite. It follows that $\operatorname{Ker} \pi_X$ is normally generated by the finite subset

$$T_0 = \bigcup_{\rho \in \Re} \{t_i \mid i \in I_\rho\}$$

of $T$. $\square$

Proposition 4.25 can be reformulated as follows: if $G$ is finitely presented, $X$ is finite and

$$1 \to N \to F(X) \to G \to 1$$

is a short exact sequence, then $N$ is normally generated by finitely many elements $n_1, \ldots, n_k$. This can be generalized to an arbitrary short exact sequence:

Lemma 4.26. *Consider a short exact sequence*

$$1 \to N \to K \stackrel{\pi}{\to} G \to 1\,, \quad \text{with } K \text{ finitely generated.} \tag{4.1}$$

*If $G$ is finitely presented, then $N$ is normally generated by finitely many elements $n_1, \ldots, n_k \in N$.*

Proof. Let $S$ be a finite generating set of $K$; then $\overline{S} = \pi(S)$ is a finite generating set of $G$. Since $G$ is finitely presented, by Proposition 4.25 there exist finitely many words $r_1, \ldots, r_k$ in $S$ such that

$$\left\langle \overline{S} \mid r_1(\overline{S}), \ldots, r_k(\overline{S}) \right\rangle$$

is a presentation of $G$.

Consider $n_j = r_j(S)$, an element of $N$ by the assumption.

Let $n$ be an arbitrary element in $N$ and $w(S)$ a word in $S$ such that $n = w(S)$ in $K$. Then $w(\overline{S}) = \pi(n) = 1$, whence in $F(S)$ the word $w(S)$ is a product of finitely many conjugates of $r_1, \ldots, r_k$. When projecting such a relation *via* $F(S) \to K$ we obtain that $n$ is a product of finitely many conjugates of $n_1, \ldots, n_k$. □

Proposition 4.27. *Suppose that $N$ a normal subgroup of a group $G$. If both $N$ and $G/N$ are finitely presented then $G$ is also finitely presented.*

Proof. Let $X$ be a finite generating set of $N$ and let $Y$ be a finite subset of $G$ such that $\bar{Y} = \{yN \mid y \in Y\}$ is a generating set of $G/N$. Let $\langle X \mid r_1, \ldots, r_k \rangle$ be a finite presentation of $N$ and let $\left\langle \bar{Y} \mid \rho_1, \ldots, \rho_m \right\rangle$ be a finite presentation of $G/N$. The group $G$ is generated by $S = X \cup Y$ and this set of generators satisfies a list of relations of the following form

$$r_i(X) = 1\,,\; 1 \leqslant i \leqslant k\,,\; \rho_j(Y) = u_j(X)\,,\; 1 \leqslant j \leqslant m\,, \tag{4.2}$$

$$x^y = v_{xy}(X)\,,\; x^{y^{-1}} = w_{xy}(X) \tag{4.3}$$

for some words $u_j, v_{xy}, w_{xy}$ in $S$.

We claim that this is a complete set of defining relators of $G$.

All the relations above can be rewritten as $t(X, Y) = 1$ for a finite set $T$ of words $t$ in $S$. Let $K$ be the normal subgroup of $F(S)$ normally generated by $T$.

The epimorphism $\pi_S : F(S) \to G$ defines an epimorphism $\varphi : F(S)/K \to G$. Let $wK$ be an element in $\operatorname{Ker}\varphi$, where $w$ is a word in $S$. Due to the set of relations (4.3), there exist a word $w_1(X)$ in $X$ and a word $w_2(Y)$ in $Y$, such that $wK = w_1(X)w_2(Y)K$.

Applying the projection $\pi : G \to G/N$, we see that $\pi(\varphi(wK)) = 1$, i.e., $\pi(\varphi(w_2(Y)K)) = 1$. This implies that $w_2(Y)$ is a product of finitely many conjugates of $\rho_i(Y)$, hence $w_2(Y)K$ is a product of finitely many conjugates of $u_j(X)K$, by the second set of relations in (4.2). This and the relations (4.3) imply that $w_1(X)w_2(Y)K = v(X)K$ for some word $v(X)$ in $X$. Then the image $\varphi(wK) = \varphi(v(X)K)$ is in $N$; therefore, $v(X)$ is a product of finitely many conjugates of relators $r_i(X)$. This implies that $v(X)K = K$.

We have thus obtained that $\operatorname{Ker}\varphi$ is trivial, hence $\varphi$ is an isomorphism, equivalently that $K = \operatorname{Ker}\pi_S$. This implies that $\operatorname{Ker}\pi_S$ is normally generated by the finite set of relators listed in (4.2) and (4.3). □

We continue with a list of finite presentations of some important groups:

EXAMPLES 4.28. (1) *Surface groups*:

$$G = \langle a_1, b_1, \ldots, a_n, b_n | [a_1, b_1] \cdots [a_n, b_n] \rangle ,$$

is the fundamental group of the closed connected oriented surface of genus $n$, see e.g. **[Mas91]**.

(2) *Right–angled Artin groups (RAAGs).* Let $\mathcal{G}$ be a finite graph with the vertex set $V = \{x_1, \ldots, x_n\}$ and the edge set $E$ consisting of the edges $\{[x_i, x_j]\}_{i,j}$. Define the *right–angled Artin group* by

$$A_{\mathcal{G}} := \langle V | [x_i, x_j], \text{whenever } [x_i, x_j] \in E \rangle .$$

Here we commit a useful abuse of notation: In the first instance $[x_i, x_j]$ denotes the commutator and in the second instance it denotes the edge of $\mathcal{G}$ connecting $x_i$ to $x_j$.

EXERCISE 4.29. a. If $\mathcal{G}$ contains no edges then $A_{\mathcal{G}}$ is a free group on $n$ generators.

b. If $\mathcal{G}$ is the complete graph on $n$ vertices then

$$A_{\mathcal{G}} \cong \mathbb{Z}^n.$$

(3) *Coxeter groups.* Let $\mathcal{G}$ be a finite simple graph. Let $V$ and $E$ denote be the vertex and the edge set of $\mathcal{G}$ respectively. Put a label $m(e) \in \mathbb{N} \setminus \{1\}$ on each edge $e = [x_i, x_j]$ of $\mathcal{G}$. Call the pair

$$\Gamma := (\mathcal{G}, m : E \to \mathbb{N} \setminus \{1\})$$

a *Coxeter graph.* Then $\Gamma$ defines the *Coxeter group* $C_\Gamma$:

$$C_\Gamma := \left\langle x_i \in V | x_i^2, (x_i x_j)^{m(e)}, \text{ whenever there exists an edge } e = [x_i, x_j] \right\rangle .$$

See **[Dav08]** for the detailed discussion of Coxeter groups.

(4) *Artin groups.* Let $\Gamma$ be a Coxeter graph. Define

$$A_\Gamma := \left\langle x_i \in V | \underbrace{x_i x_j \cdots}_{m(e) \text{ terms}} = \underbrace{x_j x_i \cdots}_{m(e) \text{ terms}}, \text{ whenever } e = [x_i, x_j] \in E \right\rangle .$$

Then $A_\Gamma$ is a right-angled Artin group if and only if $m(e) = 2$ for every $e \in E$. In general, $C_\Gamma$ is the quotient of $A_\Gamma$ by the subgroup normally generated by the set

$$\{x_i^2 : x_i \in V\}.$$

(5) *Shephard groups:* Let $\Gamma$ be a Coxeter graph. Label vertices of $\Gamma$ with natural numbers $n_x, x \in V(\Gamma)$. Now, take a group, a Shepherd group, $S_\Gamma$ to be generated by vertices $x \in V(\Gamma)$, subject to Artin relators and, in addition, relators

$$x^{n_x}, \quad x \in V(\Gamma).$$

Note that, in the case $n_x = 2$ for all $x \in V(\Gamma)$, the group which we obtain is the Coxeter group $C_\Gamma$. Shephard groups (and von Dyck groups below) are *complex analogues* of Coxeter groups.

(6) *Generalized von Dyck groups:* Let $\Gamma$ be a labeled graph as in the previous example. Define a group $D_\Gamma$ to be generated by vertices $x \in V(\Gamma)$, subject to the relators

$$x^{n_x}, \quad x \in V(\Gamma);$$

$$(xy)^{m(e)}, e = [x, y] \in E(\Gamma).$$

If $\Gamma$ consists of a single edge, then $D_\Gamma$ is called a *von Dyck group*. Every von Dyck group $D_\Gamma$ is an index 2 subgroup in the Coxeter group $C_\Delta$, where $\Delta$ is the triangle with edge-labels $p, q, r$, which are the vertex-edge labels of $\Gamma$.

(7) *Integer Heisenberg group:*

$$H_{2n+1}(\mathbb{Z}) := \langle x_1, \ldots, x_n, y_1, \ldots, y_n, z \mid$$

$$[x_i, z] = 1, [y_j, z] = 1, [x_i, x_j] = 1, [y_i, y_j] = 1, [x_i, y_j] = z^{\delta_{ij}}, 1 \leqslant i, j \leqslant n \rangle.$$

(8) *Baumslag–Solitar groups:*

$$BS(p, q) = \langle a, b | ab^p a^{-1} = b^q \rangle.$$

Exercise 4.30. Show that $H_{2n+1}(\mathbb{Z})$ is isomorphic to the group appearing in Example 10.29, (3).

Open problem 4.31. It is known that all (finitely generated) Coxeter groups are linear, see e.g. [**Bou02**]. Is the same true for all Artin groups, Shephard groups, generalized von Dyck groups? Note that even linearity of Artin Braid groups was unknown prior to [**Big01**]. Is it at least true that all these groups are residually finite?

An important feature of finitely presented groups is provided by the following theorem, see e.g. [**Hat02**]:

Theorem 4.32. *Every finitely generated group is the fundamental group of a smooth compact manifold of dimension* 4.

Presentations $G = \langle X | R \rangle$ provide a 'compact' form for defining the group $G$. They were introduced by Max Dehn in the early 20-th century. The main problem of the combinatorial group theory is to derive algebraic information about $G$ from its presentation.

*Algorithmic problems in the combinatorial group theory.*

**Word Problem.** Let $G = \langle X|R\rangle$ be a finitely-presented group. Construct a Turing machine (or prove its non-existence) that, given a word $w$ in the generating set $X$ as its input, would determine if $w$ represents the trivial element of $G$, i.e., if

$$w \in \langle\langle R\rangle\rangle .$$

**Conjugacy Problem.** Let $G = \langle X|R\rangle$ be a finitely-presented group. Construct a Turing machine (or prove its non-existence) that, given a pair of word $v, w$ in the generating set $X$, would determine if $v$ and $w$ represent conjugate elements of $G$, i.e., if there exists $g \in G$ so that

$$[w] = g^{-1}[v]g.$$

To simplify the language, we will state such problems below as: Given a finite presentation of $G$, determine if two elements of $G$ are conjugate.

**Simultaneous Conjugacy Problem.** Given $n$-tuples pair of words

$$(v_1, \ldots, v_n), \quad (w_1, \ldots, w_n)$$

in the generating set $X$ and a (finite) presentation $G = \langle X|R\rangle$, determine if there exists $g \in G$ so that

$$[w_i] = g^{-1}[v_i]g, i = 1, \ldots, n.$$

**Triviality Problem.** Given a (finite) presentation $G = \langle X|R\rangle$ as an input, determine if $G$ is trivial, i.e., equals $\{1\}$.

**Isomorphism Problem.** Given two (finite) presentations $G_i = \langle X_i|R_i\rangle, i = 1, 2$ as an input, determine if $G_1$ is isomorphic to $G_2$.

**Embedding Problem.** Given two (finite) presentations $G_i = \langle X_i|R_i\rangle, i = 1, 2$ as an input, determine if $G_1$ is isomorphic to a subgroup of $G_2$.

**Membership Problem.** Let $G$ be a finitely-presented group, $h_1, \ldots, h_k \in G$ and $H$, the subgroup of $G$ generated by the elements $h_i$. Given an element $g \in G$, determine if $g$ belongs to $H$.

Note that a group with solvable conjugacy or membership problem, also has solvable word problem. It was discovered in the 1950-s in the work of Novikov, Boone and Rabin [**Nov58, Boo57, Rab58**] that all of the above problems are *algorithmically unsolvable*. For instance, in the case of the word problem, given a finite presentation $G = \langle X|R\rangle$, there is no algorithm whose input would be a (reduced) word $w$ and the output YES is $w \equiv_G 1$ and NO if not. Fridman [**Fri60**] proved that certain groups have solvable word problem and unsolvable conjugacy problem. We will later see examples of groups with solvable word and conjugacy problems but unsolvable membership problem (Corollary 9.143). Furthermore, there are examples [**BH05**] of finitely-presented groups with solvable conjugacy problem but unsolvable simultaneous conjugacy problem for every $n \geq 2$.

Nevertheless, the main message of the geometric group theory is that *under various geometric assumptions on groups (and their subgroups), all of the above algorithmic problems are solvable.* Incidentally, the idea that geometry can help solving algorithmic problems also goes back to Max Dehn. Here are two simple examples of solvability of word problem:

Proposition 4.33. *Free group $F$ of finite rank has solvable word problem.*

PROOF. Given a word $w$ in free generators $x_i$ (and their inverses) of $F$ we cancel recursively all possible pairs $x_i x_i^{-1}$, $x_i^{-1} x_i$ in $w$. Eventually, this results in a reduced word $w'$. If $w'$ is nonempty, then $w$ represents a nontrivial element of $F$, if $w'$ is empty, then $w \equiv 1$ in $F$. □

PROPOSITION 4.34. *Every finitely-presented residually-finite group has solvable word problem.*

PROOF. First, note that if $\Phi$ is a finite group, then it has solvable word problem (using the multiplication table in $\Phi$ we can "compute" every product of generators as an element of $\Phi$ and decide if this element is trivial or not). Given a residually finite group $G$ with finite presentation $\langle X|R\rangle$ we will run two Turing machines $T_1, T_2$ simultaneously:

The machine $T_1$ will look for homomorphism $\varphi : G \to S_n$, where $S_n$ is the symmetric group on $n$ letters ($n \in \mathbb{N}$): The machine will try to send generators $x_1, \ldots, x_m$ of $G$ to elements of $S_m$ and then check if the images of the relators in $G$ under this map are trivial or not. For every such homomorphism, $T_1$ will check if $\varphi(g) = 1$ or not. If $T_1$ finds $\varphi$ so that $\varphi(g) \neq 1$, then $g \in G$ is nontrivial and the process stops.

The machine $T_2$ will list all the elements of the kernel $N$ of the quotient homomorphism $F_m \to G$: It will multiply conjugates of the relators $r_j \in R$ by products of the generators $x_i \in X$ (and their inverses) and transforms the product to a reduced word. Every element of $N$ is such a product, of course. We first write $g \in G$ as a reduced word $w$ in generators $x_i$ and their inverses. If $T_2$ finds that $w$ equals one of the elements of $N$, then it stops and concludes that $g = 1$ in $G$.

The point of residual finiteness is that, eventually, one of the machines stops and we conclude that $g$ is trivial or not. □

**Laws in groups.**

DEFINITION 4.35. An *identity* (or *law*) is a non-trivial reduced word $w = w(x_1, \ldots, x_n)$ in $n$ letters $x_1, \ldots, x_n$ and their inverses. A group $G$ is said to *satisfy the identity (law)* $w(x_1, \ldots, x_n) = 1$ if the equality is satisfied in $G$ whenever $x_1, \ldots, x_n$ are replaced by arbitrary elements in $G$.

EXAMPLES 4.36 (groups satisfying a law).

(1) Abelian groups. Here the law is
$$w(x_1, x_2) = x_1 x_2 x_1^{-1} x_2^{-1} .$$

(2) Solvable groups, see (11.2).

(3) Free Burnside groups. The *free Burnside group*
$$B(n,m) = \left\langle x_1, \ldots, x_n \mid w^n \text{ for every word } w \text{ in } x_1^{\pm 1}, \ldots, x_n^{\pm 1} \right\rangle .$$
It is known that these groups are infinite for sufficiently large $m$ (see [**Ady79**], [**Ol'91a**], [**Iva94**], [**Lys96**], [**DG**] and references therein).

Note that free nonabelian groups (and, hence, groups containing them) do not satisfy any law.

## 4.4. Ping-pong lemma. Examples of free groups

Lemma 4.37 (Ping–pong, or Table–tennis, lemma). *Let $X$ be a set, and let $g : X \to X$ and $h : X \to X$ be two bijections. If $A, B$ are two non-empty subsets of $X$, such that $A \not\subset B$ and*

$$g^n(A) \subset B \text{ for every } n \in \mathbb{Z} \setminus \{0\},$$

$$h^m(B) \subset A \text{ for every } m \in \mathbb{Z} \setminus \{0\},$$

*then $g, h$ generate a free subgroup of rank 2 in the group $\mathrm{Bij}(X)$ with the binary operation given by composition $\circ$.*

Proof. *Step 1.* Let $w$ be a non-empty reduced word in $\{g, g^{-1}, h, h^{-1}\}$. We want to prove that $w$ is not equal to the identity in $\mathrm{Bij}(X)$. We begin by noting that it is enough to prove this when

$$w = g^{n_1} h^{n_2} g^{n_3} h^{n_4} \dots g^{n_k}, \text{ with } n_j \in \mathbb{Z} \setminus \{0\} \, \forall j \in \{1, \dots, k\}. \tag{4.4}$$

Indeed:

- If $w = h^{n_1} g^{n_2} h^{n_3} \dots g^{n_k} h^{n_{k+1}}$, then $gwg^{-1}$ is as in (4.4), and $gwg^{-1} \neq \mathrm{id} \Rightarrow w \neq \mathrm{id}$.
- If $w = g^{n_1} h^{n_2} g^{n_3} h^{n_4} \dots g^{n_k} h^{n_{k+1}}$, then for any $m \neq -n_1$, $g^m w g^{-m}$ is as in (4.4).
- If $w = h^{n_1} g^{n_2} h^{n_3} \dots g^{n_k}$, then for any $m \neq n_k$, $g^m w g^{-m} \neq \mathrm{id}$ is as in (4.4).

*Step 2.* If $w$ is as in (4.4) then

$$w(A) \subset g^{n_1} h^{n_2} g^{n_3} h^{n_4} \dots g^{n_{k-2}} h^{n_{k-1}}(B) \subset g^{n_1} h^{n_2} g^{n_3} h^{n_4} \dots g^{n_{k-1}}(A) \subset \dots \subset$$

$$g^{n_1}(A) \subset B.$$

If $w = \mathrm{id}$, then it would follow that $A \subset B$, a contradiction. □

Example 4.38. For any integer $k \geqslant 2$ the matrices

$$g = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \text{ and } h = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$$

generate a free subgroup of $SL(2, \mathbb{Z})$.

*1st proof.* The group $SL(2, \mathbb{Z})$ acts on the upper half plane $\mathbb{H}^2 = \{z \in \mathbb{C} \mid \Im(z) > 0\}$ by linear fractional transformations $z \mapsto \frac{az+b}{cz+d}$. The matrix $g$ acts as a horizontal translation $z \mapsto z + k$, while

$$h = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Therefore $h$ acts as represented in Figure 4.1, where $h$ sends the interior of the disk bounded by $C$ to the exterior of the disk bounded by $C'$. We apply Lemma 4.37 to $g, h$ and the subsets $A$ and $B$ represented below, i.e. $A$ is the strip

$$\{z \in \mathbb{H}^2 \ : \ -\frac{k}{2} < \mathrm{Re}\, z < \frac{k}{2}\}$$

and $B$ is the complement of its closure, that is

$$B = \{z \in \mathbb{H}^2 \ : \ \mathrm{Re}\, z < -\frac{k}{2} \text{ or } \mathrm{Re}\, z > \frac{k}{2}\}.$$

Hence $g^n(A) \subset B$ and $h^n(B) \subset A$ for all $n \neq 0$. Therefore, the claim follows from lemma 4.37.

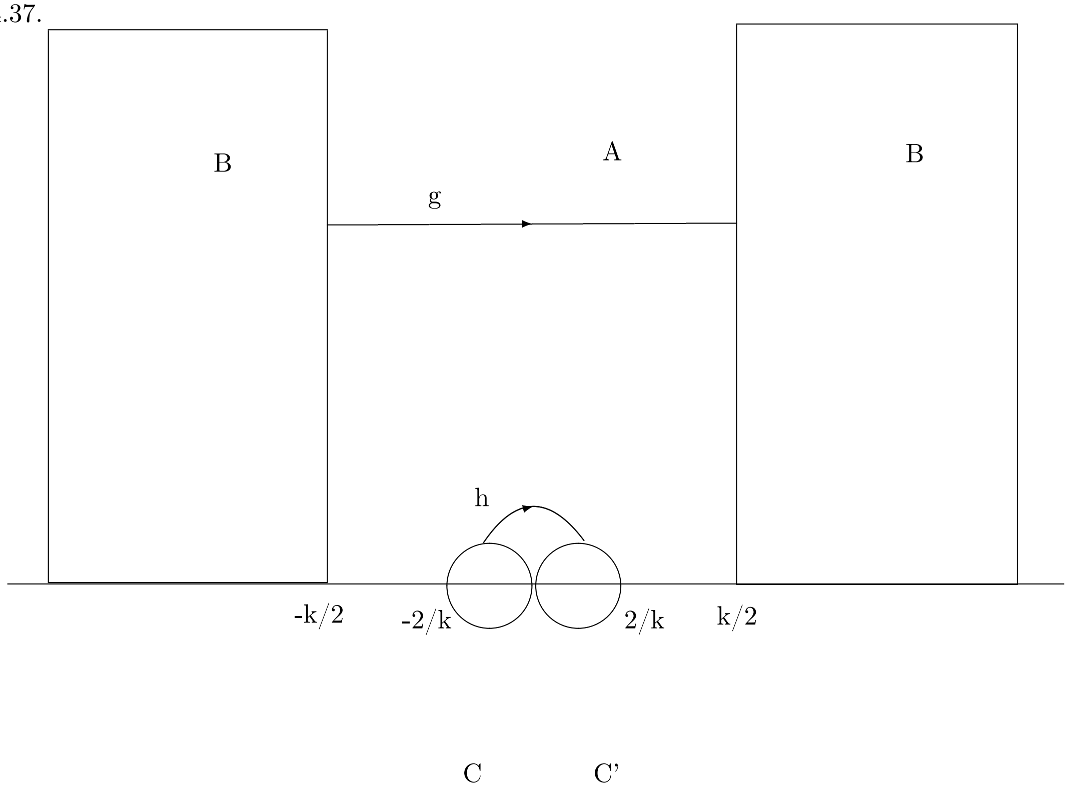



FIGURE 4.1. Example of ping-pong.

*2nd proof.* The group $SL(2,\mathbb{Z})$ also acts linearly on $\mathbb{R}^2$, and we can apply Lemma 4.37 to $g, h$ and the following subsets of $\mathbb{R}^2$

$$A = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \ : \ |x| < |y| \right\} \text{ and } B = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \ : \ |x| > |y| \right\}.$$

□

REMARK 4.39. The statement in the Example above no longer holds for $k = 1$. Indeed, in this case we have

$$g^{-1}hg^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Thus, $(g^{-1}hg^{-1})^2 = I_2$, and, hence, the group generated by $g, h$ is not free.

Lemma 4.37 extends to the case of several bijections as follows.

Lemma 4.40 (The generalized Ping-pong lemma). *Let $X$ be a set, and let $g_i : X \to X$, $i \in \{1,2,\dots,k\}$, be bijections. Suppose that $A_1,\dots,A_k$ are non-empty subsets of $X$, such that $\bigcup_{i=2}^k A_i \not\subset A_1$ and that for every $i \in \{1,2,\dots,k\}$*

$$g_i^n\left(\bigcup_{j\neq i} A_j\right) \subset A_i \text{ for every } n \in \mathbb{Z}\setminus\{0\}\,.$$

*Then $g_1,\dots,g_k$ generate a free subgroup of rank $k$ in the group of bijections* $\mathrm{Bij}(X)$.

Proof. Consider a non-trivial reduced word $w$ in $\left\{g_1^{\pm 1},\dots,g_k^{\pm 1}\right\}$. As in the proof of Lemma 4.37, without loss of generality we may assume that the word $w$ begins with $g_1^a$ and ends with $g_1^b$, where $a,b \in \mathbb{Z}\setminus\{0\}$. We apply $w$ to $\bigcup_{i=2}^k A_i$, and obtain that the image is contained in $A_1$. If $w = \mathrm{id}$ in $Bij(X)$, it would that $\bigcup_{i=2}^k A_i \subset A_1$, a contradiction. □

## 4.5. Ping-pong on a projective space

We will frequently use Ping-Pong lemma in the case when $X$ is a projective space. Since this application of the ping-pong argument is the key for the proof of the Tits' Alternative, we explain it here in detail.

Let $V$ be a finite dimensional space over a *normed field* $\mathbb{K}$, which is either $\mathbb{R},\mathbb{C}$ or has discrete norm and uniformizer $\pi$, as in §1.7. We endow the projective space $\mathbb{P}(V)$ with the metric $d$ as in §1.8.

Lemma 4.41. *Every $g \in GL(n,\mathbb{K})$ induces a bi-Lipschitz transformation of $P(\mathbb{K}^n)$ with Lipschitz constant $\leqslant \frac{|a_1|^2}{|a_n|^2}$, where $a_1,\dots,a_n$ are the singular values of $g$ and*

$$|a_1| \geqslant \ldots \geqslant |a_n|.$$

Proof. According to the Cartan decomposition $g = kdk'$ and since all elements in the subgroup $K$ act by isometries on the projective space, it suffices to prove the statement when $g$ is a diagonal matrix $A$ with diagonal entries $a_1,\dots,a_n$ which are arranged in the order as above. We will do the computation in the case $\mathbb{K} = \mathbb{R}$ and leave the other cases to the reader. Given nonzero vectors $x = (x_1,\dots,x_n), y = (y_1,\dots,y_n)$, we obtain:

$$|gx \wedge gy| = |\sum_{i<j} a_i a_j x_i x_j e_i \wedge e_j| \leqslant |a_1|^2 |\sum_{i<j} x_i x_j| = |a_1|^2 |x \wedge y|,$$

$$|gx| = |\sum_i a_i^2 x_i^2|^{1/2} \geqslant |a_n||x|, \quad |gy| \geqslant |a_n||y|$$

and, hence,

$$d(g[x],g[y]) \leqslant \frac{|a_1|^2}{|a_n|^2}\frac{|x\wedge y|}{|x|\cdot|y|} = \frac{|a_1|^2}{|a_n|^2} d([x],[y]).$$

□

Let $g$ be an element in $GL(n,\mathbb{K})$ such that with respect to some ordered basis $\{u_1,\dots,u_n\}$, the matrix of $g$ is diagonal with diagonal entries $\lambda_1,\dots,\lambda_n$ satisfying

$$|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots \geqslant |\lambda_{n-1}| > |\lambda_n| > 0\,.$$

Let us denote by $A(g)$ and by $H(g)$ the projection to $P(\mathbb{K}^n)$ of the span of $\{u_1\}$, respectively of the span of $\{u_2,\dots,u_n\}$. Note that then $A(g^{-1})$ and $H(g^{-1})$

are the respective projections to $P(\mathbb{K}^n)$ of the span of $\{u_n\}$, respectively, of the span of $\{u_1, \dots, u_{n-1}\}$. Obviously, $A(g) \in H(g^{-1})$ and $A(g^{-1}) \in H(g)$.

LEMMA 4.42. *Assume that $g$ and $h$ are two elements in $GL(n, \mathbb{K})$ as above, which are diagonal with respect to bases $\{u_1, \dots, u_n\}$, $\{v_1, \dots, v_n\}$ respectively. Assume also that the points $A(g^{\pm 1})$ are not in $H(h) \cup H(h^{-1})$, and $A(h^{\pm 1})$ are not in $H(g) \cup H(g^{-1})$. Then there exists a positive integer $N$ such that $g^N$ and $h^N$ generate a free non-abelian subgroup of $GL(n, \mathbb{K})$.*

PROOF. We first claim that for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that for every $m \geqslant N$, $g^{\pm m}$ maps the complement of the $\varepsilon$-neighborhood of $H(g^{\pm 1})$ inside the ball of radius $\varepsilon$ and center $A(g^{\pm 1})$.

According to Lemma 4.41, it suffices to prove the statement when $\{u_1, \dots, u_n\}$ is the standard basis $\{e_1, \dots, e_n\}$ of $V$ (since we can conjugate $g$ to a matrix diagonal with respect to the standard basis). In particular, $A(g^{\pm 1})$ is either $[e_1]$ or $[e_n]$. In the former case we take $f(x) = x \cdot e_1$, in the latter case, take $f(x) = x \cdot e_n$, so that $Ker(f) = H = H(g^{\pm 1})$. Then, for a unit vector $v = (x_1, \dots, x_n) \in V$, according to Exercise 1.80, $\mathrm{dist}([v], [H]) = |f(v)|$. To simplify the notation, we will assume that $f(x) = x \cdot e_1$, since the other case is obtained by relabeling. Then,

$$[v] \notin \mathcal{N}_\varepsilon(H(g^{\pm 1})) \iff |x_1| \geqslant \varepsilon.$$

We have

$$|g^m v \wedge e_1| = |\sum_{i>1} \lambda_i^m x_i e_i \wedge e_1| \leqslant \sqrt{n} |\lambda_2|^m |v| = \sqrt{n} |\lambda_2|^m$$

while

$$|g^m v| \geqslant |\lambda_1|^m |x_1|,$$

which implies that

$$d(g^m[v], [e_1]) = \frac{|g^m v \wedge e_1|}{|g^m v|} \leqslant \frac{\sqrt{n}}{|x_1|} \frac{|\lambda_2|^m}{|\lambda_1|^m} \leqslant \frac{\sqrt{n}}{\varepsilon} \left( \frac{|\lambda_2|}{|\lambda_1|} \right)^m$$

The latter quantity converges to zero as $m \to \infty$, since $|\lambda_1| > |\lambda_2|$. Thus, for all large $m$, $d(g^m[v], [e_1]) < \varepsilon$. The same claim holds for $h^{\pm 1}$.

Now consider $\varepsilon > 0$ such that for every $\alpha \in \{g, g^{-1}\}$ and $be \in \{h, h^{-1}\}$ the points $A(\alpha)$ and $A(\beta^{\pm 1})$ are at distance at least $2\varepsilon$ from $H(\alpha)$. Let $N$ be large enough so that $g^{\pm N}$ maps the complement of the $\varepsilon$-neighborhood of $H(g^{\pm 1})$ inside the ball of radius $\varepsilon$ and center $A(g^{\pm 1})$, and $h^{\pm N}$ maps the complement of the $\varepsilon$-neighborhood of $H(h^{\pm 1})$ inside the ball of radius $\varepsilon$ and center $A(h^{\pm 1})$.

Let $A := B(A(g), \varepsilon) \sqcup B(A(g^{-1}), \varepsilon)$ and $B := B(A(h), \varepsilon) \sqcup B(A(h^{-1}), \varepsilon)$. Clearly, $g^{kN}(A) \subseteq B$ and $h^{kN}(B) \subseteq A$ for every $k \in \mathbb{Z}$. Hence by Lemma 4.37, $g^N$ and $h^N$ generate a free group. □

### 4.6. The rank of a free group determines the group. Subgroups

PROPOSITION 4.43. *Two free groups $F(X)$ and $F(Y)$ are isomorphic if and only if $X$ and $Y$ have the same cardinality.*

PROOF. A bijection $\varphi : X \to Y$ extends to an isomorphism $\Phi : F(X) \to F(Y)$ by Proposition 4.18. Therefore, two free groups $F(X)$ and $F(Y)$ are isomorphic if $X$ and $Y$ have the same cardinality.

Conversely, let $\Phi : F(X) \to F(Y)$ be an isomorphism. Take $N(X) \leqslant F(X)$, the subgroup generated by the subset $\{g^2 \,;\, g \in F(X)\}$; clearly, $N$ is normal in $F(X)$.

Then, $\Phi(N(X)) = N(Y)$ is the normal subgroup generated by $\{h^2\,;\, h \in F(Y)\}$. It follows that $\Phi$ induces an isomorphism $\Psi : F(X)/N(X) \to F(Y)/N(Y)$.

LEMMA 4.44. *The quotient* $\bar{F} := F/N$ *is isomorphic to* $A = \mathbb{Z}_2^{\oplus X}$, *where* $F = F(X)$.

PROOF. Recall that $A$ has the presentation

$$\left\langle x \in X | x^2, [x, y], \forall x, y \in X \right\rangle,$$

see Exercise 4.24. We now prove the assertion of the lemma. Consider the map $\eta : F \to A$ sending the generators of $F$ to the obvious generators of $A$. Thus, $\pi(g) = \pi(g^{-1})$ for all $g \in F$. We conclude that for all $g, h \in X$,

$$1 = \pi((hg)^2) = \pi([g, h]),$$

and, therefore, $\bar{F}$ is abelian.

Since $A$ satisfies the law $a^2 = 1$ for all $a \in A$, it is clear that $\eta = \varphi \circ \pi$, where $\pi : F \to \bar{F}$ is the quotient map. We next construct the inverse $\psi$ to $\phi$. We define $\psi$ on the generators $x \in X$ of $A$: $\psi(x) = \bar{x} = \pi(x)$. We need to show that $\psi$ preserves the relators of $A$ (as in Lemma 4.23): Since $\bar{F}$ is abelian, $[\psi(x), \psi(y)] = 1$ for all $x, y \in X$. Moreover, $\psi(x)^2 = 1$ since $\bar{F}$ also satisfies the law $g^2 = 1$. It is clear that $\phi, \psi$ are inverses to each other. □

Thus, $F(X)/N(X)$ is isomorphic to $\mathbb{Z}_2^{\oplus X}$, while $F(Y)/N(Y)$ is isomorphic to $\mathbb{Z}_2^{\oplus Y}$. It follows that $\mathbb{Z}_2^{\oplus X} \cong \mathbb{Z}_2^{\oplus Y}$ as $\mathbb{Z}_2$–vector spaces. Therefore, $X$ and $Y$ have the same cardinality, by uniqueness of the dimension of vector spaces. □

REMARK 4.45. Proposition 4.43 implies that for every cardinal number $n$ there exists, up to isomorphism, exactly one free group of rank $n$. We denote it by $F_n$.

THEOREM 4.46 (Nielsen–Schreier). *Any subgroup of a free group is a free group.*

This theorem will be proven in Corollary 4.70 using topological methods; see also [**LS77**, Proposition 2.11].

PROPOSITION 4.47. *The free group of rank two contains an isomorphic copy of* $F_k$ *for every finite* $k$ *and* $k = \aleph_0$.

PROOF. Let $x, y$ be the two generators of $F_2$. Let $S$ be the subset consisting of all elements of $F_2$ of the form $x_k := y^k x y^{-k}$, for all $k \in \mathbb{N}$. We claim that the subgroup $\langle S \rangle$ generated by $S$ is isomorphic to the free group of rank $\aleph_0$.

Indeed, consider the set $A_k$ of all reduced words with prefix $y^k x$. With the notation of Section 4.2, the transformation $L_{x_k} : F_2 \to F_2$ has the property that $L_{x_k}(A_j) \subset A_k$ for every $j \neq k$. Obviously, the sets $A_k$ , $k \in \mathbb{N}$ , are pairwise disjoint. This and Lemma 4.40 imply that $\{L_{x_k}\,;\, k \in \mathbb{N}\}$ generate a free subgroup in $\mathrm{Bij}(F_2)$, hence so do $\{x_k\,;\, k \in \mathbb{N}\}$ in $F_2$. □

## 4.7. Free constructions: Amalgams of groups and graphs of groups

**4.7.1. Amalgams.** Amalgams (amalgamated free products and HNN extensions) allow one to build more complicated groups starting with a given pair of groups or a group and a pair of its subgroups which are isomorphic to each other.

**Amalgamated free products.** As a warm-up we define the *free product* of groups $G_1 = \langle X_1 | R_1 \rangle, G_2 = \langle X_2 | R_2 \rangle$ by the presentation:

$$G_1 * G_2 = \langle G_1, G_2 | \quad \rangle$$

which is a shorthand for the presentation:

$$\langle X_1 \sqcup X_2 | R_1 \sqcup R_2 \rangle .$$

For instance, the free group of rank 2 is isomorphic to $\mathbb{Z} * \mathbb{Z}$.

More generally, suppose that we are given subgroups $H_i \leqslant G_i$ $(i = 1, 2)$ and an isomorphism

$$\phi : H_1 \to H_2$$

Define the *amalgamated free product*

$$G_1 *_{H_1 \cong H_2} G_2 = \left\langle G_1, G_2 | \phi(h) h^{-1}, h \in H_1 \right\rangle .$$

In other words, in addition to the relators in $G_1, G_2$ we identify $\phi(h)$ with $h$ for each $h \in H_1$. A common shorthand for the amalgamated free product is

$$G_1 *_H G_2$$

where $H \cong H_1 \cong H_2$ (the embeddings of $H$ into $G_1$ and $G_2$ are suppressed in this notation).

**HNN extensions.** This construction is named after G. Higman, B. Neumann and H. Neumann who first introduced it in [**HNN49**]. It is a variation on the amalgamated free product where $G_1 = G_2$. Namely, suppose that we are given a group $G$, its subgroups $H_1, H_2$ and an isomorphism $\phi : H_1 \to H_2$. Then the HNN extension of $G$ *via* $\phi$ is defined as

$$G \star_{H_1 \cong H_2} = \left\langle G, t | t h t^{-1} = \phi(h), \forall h \in H_1 \right\rangle .$$

A common shorthand for the HNN extension is

$$G \star_H$$

where $H \cong H_1 \cong H_2$ (the two embeddings of $H$ into $G$ are suppressed in this notation).

Exercise 4.48. Suppose that $H_1$ and $H_2$ are both trivial subgroups. Then

$$G \star_{H_1 \cong H_2} \cong G * \mathbb{Z}.$$

**4.7.2. Graphs of groups.** In this section, graphs are no longer assumed to be simplicial, but are assumed to connected. The notion of graphs of groups is a very useful generalization of both the amalgamated free product and the HNN extension.

Suppose that $\Gamma$ is a graph. Assign to each vertex $v$ of $\Gamma$ a *vertex group* $G_v$; assign to each edge $e$ of $\Gamma$ an *edge group* $G_e$. We orient each edge $e$ so it has the *initial and the terminal* (possibly equal) vertices $e_-$ and $e_+$. Suppose that for each edge $e$ we are given *monomorphisms*

$$\phi_{e_+} : G_e \to G_{e_+}, \phi_{e_-} : G_e \to G_{e_-}.$$

Remark 4.49. More generally, one can allow non-injective homomorphisms

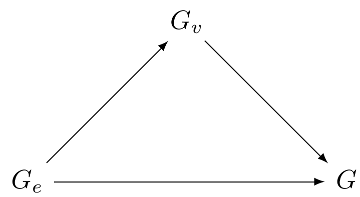
$$G_e \to G_{e_+}, G_e \to G_{e_-},$$

but we will not consider them here.

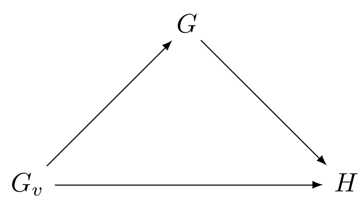
The graph $\Gamma$ together with the collection of vertex and edge groups and the monomorphisms $\phi_{e_\pm}$ is called a *graph of groups* $\mathcal{G}$.

DEFINITION 4.50. The *fundamental group* $\pi(\mathcal{G}) = \pi_1(\mathcal{G})$ of the above graph of groups is a group $G$ satisfying the following:

1. There is a collection of *compatible homomorphisms* $G_v \to G, G_e \to G, v \in V(\Gamma), e \in E(\Gamma)$, so that whenever $v = e_\pm$, we have the commutative diagram




2. The group $G$ is *universal* with respect to the above property, i.e., given any group $H$ and a collection of compatible homomorphisms $G_v \to H, G_e \to H$, there exists a unique homomorphism $G \to H$ so that we have commutative diagrams




for all $v \in V(\Gamma)$.

Note that the above definition easily implies that $\pi(\mathcal{G})$ is unique (up to an isomorphism). For the existence of $\pi(\mathcal{G})$ see **[Ser80]** and discussion below. Whenever $G \cong \pi(\mathcal{G})$, we will say that $\mathcal{G}$ determines a *graph of groups decomposition* of $G$. The decomposition of $G$ is called *trivial* if there is a vertex $v$ so that the natural homomorphism $G_v \to G$ is onto.

EXAMPLE 4.51. 1. Suppose that the graph $\Gamma$ is a single edge $e = [1, 2]$, $\phi_{e_-}(G_e) = H_1 \leqslant G_1$, $\phi_{e_+}(G_e) = H_2 \leqslant G_2$. Then

$$\pi(\mathcal{G}) \cong G_1 \star_{H_1 \cong H_2} G_2.$$

2. Suppose that the graph $\Gamma$ is a single loop $e = [1, 1]$, $\phi_{e_-}(G_e) = H_1 \leqslant G_1$, $\phi_{e_+}(G_e) = H_2 \leqslant G_1$. Then

$$\pi(\mathcal{G}) \cong G_1 \star_{H_1 \cong H_2} .$$

Once this example is understood, one can show that for every graph of groups $\mathcal{G}$, $\pi_1(\mathcal{G})$ exists by describing this group in terms of generators and relators in the manner similar to the definition of the amalgamated free product and HNN extension. In the next section we will see how to construct $\pi_1(\mathcal{G})$ using topology.

**4.7.3. Converting graphs of groups to amalgams.** Suppose that $\mathcal{G}$ is a graph of groups and $G = \pi_1(\mathcal{G})$. Our goal is to convert $\mathcal{G}$ in an amalgam decomposition of $G$. There are two cases to consider:

1. Suppose that the graph $\Gamma$ underlying $\mathcal{G}$ contains a oriented edge $e = [v_1, v_2]$ so that $e$ separates $\Gamma$ in the sense that the graph $\Gamma'$ obtained form $\Gamma$ by removing $e$ (and keeping $v_1, v_2$) is a disjoint union of connected subgraphs $\Gamma_1 \sqcup \Gamma_2$, where

$v_i \in V(\Gamma_i)$. Let $\mathcal{G}_i$ denote the subgraph in the graph of groups $\mathcal{G}$, corresponding to $\Gamma_i, i = 1, 2$. Then set

$$G_i := \pi_1(\mathcal{G}_i), i = 1, 2, \quad G_3 := G_e.$$

We have composition of embeddings $G_e \to G_{v_i} \to G_i \to G$. Then the universal property of $\pi_1(\mathcal{G}_i)$ and $\pi_1(\mathcal{G})$ implies that $G \cong G_1 \star_{G_3} G_2$: One simply verifies that $G$ satisfies the universal property for the amalgam $G_1 \star_{G_3} G_2$.

2. Suppose that $\Gamma$ contains an oriented edge $e = [v_1, v_2]$ so $e$ does not separate $\Gamma$. Let $\Gamma_1 := \Gamma'$, where $\Gamma'$ is obtained from $\Gamma$ by removing the edge $e$ as in Case 1. Set $G_1 := \pi_1(\mathcal{G}_1)$ as before. Then embeddings

$$G_e \to G_{v_i}, i = 1, 2$$

induce embeddings $G_e \to G_i$ with the images $H_1, H_2$ respectively. Similarly to the Case 1, we obtain

$$G \cong G_1 \star_{G_e} = G_1 \star_{H_1 \cong H_2}$$

where the isomorphism $H_1 \to H_2$ is given by the composition

$$H_1 \to G_e \to H_2.$$

Clearly, $\mathcal{G}$ is trivial if and only if the corresponding amalgam $G_1 \star_{G_3} G_2$ or $G_1 \star_{G_e}$ is trivial.

**4.7.4. Topological interpretation of graphs of groups.** Let $\mathcal{G}$ be a graph of groups. Suppose that for all vertices and edges $v \in V(\Gamma)$ and $e \in E(\Gamma)$ we are given connected cell complexes $M_v, M_e$ with the fundamental groups $G_v, G_e$ respectively. For each edge $e = [v, w]$ assume that we are given a continuous map $f_{e_\pm} : M_e \to M_{e_\pm}$ which induces the monomorphism $\phi_{e_\pm}$. This collection of spaces and maps is called a *graph of spaces*

$$\mathcal{G}_M := \{M_v, M_e, f_{e_\pm} : M_e \to M_{e_\pm} : v \in V(\Gamma), e \in E(\Gamma)\}.$$

In order to construct $\mathcal{G}_M$ starting from $\mathcal{G}$, recall that each group $G$ admits a cell complex $K(G, 1)$ whose fundamental group is $G$ and whose universal cover is contractible, see e.g. [**Hat02**]. Given a group homomorphism $\phi : H \to G$, there exists a continuous map, unique up to homotopy,

$$f : K(H, 1) \to K(G, 1)$$

which induces the homomorphism $\phi$. Then one can take $M_v := K(G_v, 1)$, $M_e := K(G_e, 1)$, etc.

To simplify the picture (although this is not the general case), the reader can think of each $M_v$ as a manifold with several boundary components which are homeomorphic to $M_{e_1}, M_{e_2}, \ldots$, where $e_j$ are the edges having $v$ as their *initial* or *final vertex*. Then assume that the maps $f_{e_\pm}$ are homeomorphisms onto the respective boundary components.

For each edge $e$ form the product $M_e \times [0, 1]$ and then form the double mapping cylinders for the maps $f_{e_\pm}$, i.e. identify points of $M_e \times \{0\}$ and $M_e \times \{1\}$ with their images under $f_{e_-}$ and $f_{e_+}$ respectively.

Let $M$ denote the resulting cell complex. It then follows from the Seifert–Van Kampen theorem [**Mas91**] that

THEOREM 4.52. *The group $\pi_1(M)$ is isomorphic to $\pi(\mathcal{G})$.*

This theorem allows one to think of the graphs of groups and their fundamental groups *topologically* rather than *algebraically*. Given the above interpretation, one can easily see that for each vertex $v \in V(\Gamma)$ the canonical homomorphism $G_v \to \pi(\mathcal{G})$ is injective.

EXAMPLE 4.53. The group $F(X)$ is isomorphic to $\pi_1(\vee_{x\in X}\mathbb{S}^1)$.

**4.7.5. Graphs of groups and group actions on trees.** An *action* of a group $G$ on a tree $T$ is an action $G \curvearrowright T$ so that each element of $G$ acts as an automorphism of $T$, i.e., such action is a homomorphism $G \to Aut(T)$. A tree $T$ with the prescribed action $G \curvearrowright T$ is called a $G$–tree. An action $G \curvearrowright T$ is said to be *without inversions* if whenever $g \in G$ preserves an edge $e$ of $T$, it fixes $e$ pointwise. The action is called *trivial* if there is a vertex $v \in T$ fixed by the entire group $G$.

REMARK 4.54. Later on, we will encounter more complicated (non-simplicial) trees and actions.

Our next goal is to explain the relation between the graph of groups decompositions of $G$ and actions of $G$ on simplicial trees without inversions.

Suppose that $G \cong \pi(\mathcal{G})$ is a graph of groups decomposition of $G$. We associate with $\mathcal{G}$ a graph of spaces $M = M_{\mathcal{G}}$ as above. Let $X$ denote the universal cover of the corresponding cell complex $M$. Then $X$ is the disjoint union of the copies of the universal covers $\tilde{M}_v, \tilde{M}_e \times (0,1)$ of the complexes $M_v$ and $M_e \times (0,1)$. We will refer to this partitioning of $X$ as the *tiling* of $X$. In other words, $X$ has the structure of a graph of spaces, where each vertex/edge space is homeomorphic to $\tilde{M}_v, v \in V(\Gamma)$, $\tilde{M}_e \times [0,1], e \in E(\Gamma)$. Let $T$ denote the graph corresponding to $X$: Each copy of $\tilde{M}_v$ determines a vertex in $T$ and each copy of $\tilde{M}_e \times [0,1]$ determines an edge in $T$.

EXAMPLE 4.55. Suppose that $\Gamma$ is a single segment $[1,2]$, $M_1$ and $M_2$ are surfaces of genus 1 with a single boundary component each. Let $M_e$ be the circle. We assume that the maps $f_{e_\pm}$ are homeomorphisms of this circle to the boundary circles of $M_1, M_2$. Then, $M$ is a surface of genus 2. The graph $T$ is sketched in Figure 4.2.

The Mayer–Vietoris theorem, applied to the above tiling of $X$, implies that $0 = H_1(X,\mathbb{Z}) \cong H_1(T,\mathbb{Z})$. Therefore, $T = T(\mathcal{G})$ is a tree. The group $G = \pi_1(M)$ acts on $X$ by deck-transformations, preserving the tiling. Therefore we get the induced action $G \curvearrowright T$. If $g \in G$ preserves some $\tilde{M}_e \times (0,1)$, then it comes from the fundamental group of $M_e$. Therefore such $g$ also preserves the orientation on the segment $[0,1]$. Hence the action $G \curvearrowright T$ is without inversions. Observe that the stabilizer of each $\tilde{M}_v$ in $G$ is conjugate in $G$ to $\pi_1(M_v) = G_v$. Moreover, $T/G = \Gamma$.

EXAMPLE 4.56. Let $G = BS(p,q)$ be the Baumslag-Solitar group described in Example 4.28, (8). The group $G$ clearly has the structure of a graph of groups since it is isomorphic to the HNN extension of $\mathbb{Z}$,

$$\mathbb{Z}\star_{H_1\cong H_2}$$

where the subgroups $H_1, H_2 \subset \mathbb{Z}$ have the indices $p$ and $q$ respectively. In order to construct the cell complex $K(G,1)$ take the circle $S^1 = M_v$, the cylinder $S^1 \times [0,1]$ and attach the ends to this cylinder to $M_v$ by the maps of the degree $p$ and $q$ respectively. Now, consider the associated $G$–tree $T$. Its vertices have valence

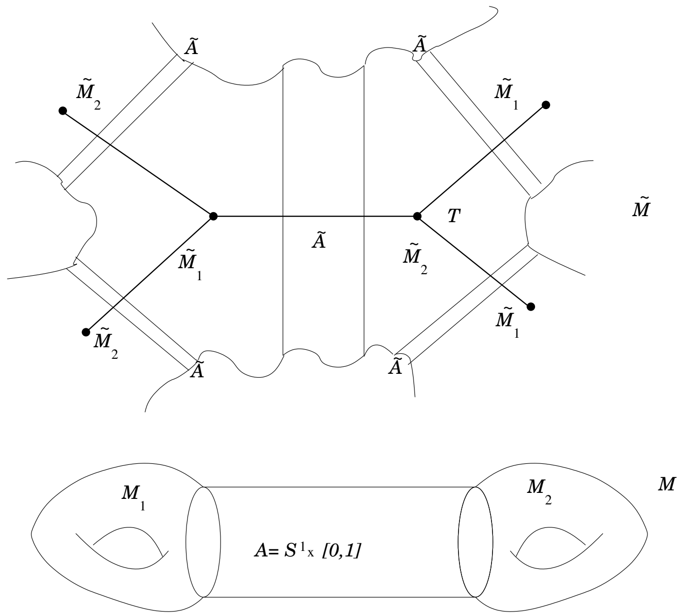



FIGURE 4.2. Universal cover of the genus 2 surface.

$p+q$: Each vertex $v$ has $q$ incoming and $p$ outgoing edges so that for each outgoing edge $e$ we have $v = e_-$ and for each incoming edge we have $v = e_+$. The vertex stabilizer $G_v \cong \mathbb{Z}$ permutes (transitively) incoming and outgoing edges among each other. The stabilizer of each outgoing edge is the subgroup $H_1$ and the stabilizer of each incoming edge is the subgroup $H_2$. Thus the action of $\mathbb{Z}$ on the incoming vertices is *via* the group $\mathbb{Z}/q$ and on the outgoing vertices *via* the group $\mathbb{Z}/p$.

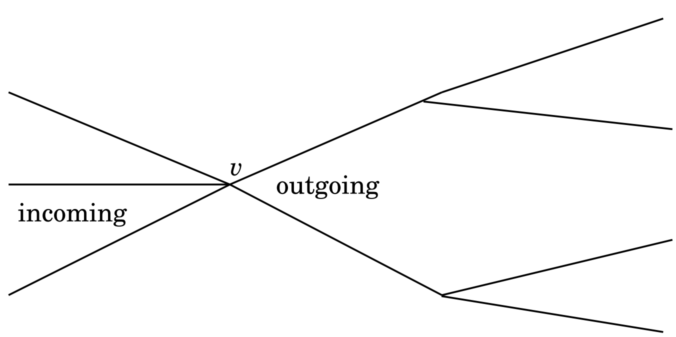



FIGURE 4.3. Tree for the group $BS(2,3)$.

LEMMA 4.57. *$G \curvearrowright T$ is trivial if and only if the graph of groups decomposition of $G$ is trivial.*

PROOF. Suppose that $G$ fixes a vertex $\tilde{v} \in T$. Then $\pi_1(M_v) = G_v = G$, where $v \in \Gamma$ is the projection of $\tilde{v}$. Hence the decomposition of $G$ is trivial. Conversely, suppose that $G_v$ maps onto $G$. Let $\tilde{v} \in T$ be the vertex which projects to $v$. Then

$\pi_1(M_v)$ is the entire $\pi_1(M)$ and hence $G$ preserves $\tilde{M}_{\tilde{v}}$. Therefore, the group $G$ fixes $\tilde{v}$. ☐

Conversely, each action of $G$ on a simplicial tree $T$ yields a realization of $G$ as the fundamental group of a graph of groups $\mathcal{G}$, so that $T = T(\mathcal{G})$. Here is the construction of $\mathcal{G}$. Furthermore, a *nontrivial* action leads to a *nontrivial* graph of groups.

If the action $G \curvearrowright T$ has inversion, we replace $T$ with its barycentric subdivision $T'$. Then the action $G \curvearrowright T'$ is without inversions. If $G \curvearrowright T$ were nontrivial, so is $G \curvearrowright T'$. Thus, from now on, we assume that $G$ acts on $T$ without inversions. Then the quotient $T/G$ is a graph $\Gamma$: $V(\Gamma) = V(T)/G$ and $E(\Gamma) = E(T)/G$. For every vertex $\tilde{v}$ and edge $\tilde{e}$ of $T$ let $G_{\tilde{v}}$ and $G_{\tilde{e}}$ be their respective stabilizes in $G$. Clearly, whenever $\tilde{e} = [\tilde{v}, \tilde{w}]$, we get the embedding

$$G_{\tilde{e}} \to G_{\tilde{v}}.$$

If $g \in G$ maps oriented edge $\tilde{e} = [\tilde{v}, \tilde{w}]$ to an oriented edge $\tilde{e}' = [\tilde{v}', \tilde{w}']$, we obtain isomorphisms

$$G_{\tilde{v}} \to G_{\tilde{v}'}, \quad G_{\tilde{w}} \to G_{\tilde{w}'}, \quad G_{\tilde{e}} \to G_{\tilde{e}'}$$

induced by conjugation *via* $g$ and the following diagram is commutative:

$$\begin{array}{ccc} G_{\tilde{e}} & \longrightarrow & G_{\tilde{v}} \\ \downarrow & & \downarrow \\ G_{\tilde{e}'} & \longrightarrow & G_{\tilde{v}'} \end{array}$$

We then set $G_v := G_{\tilde{v}}, G_e := G_{\tilde{e}}$, where $v$ and $e$ are the projections of $\tilde{v}$ and edge $\tilde{e}$ to $\Gamma$. For every edge $e$ of $\Gamma$ oriented as $e = [v, w]$, we define the monomorphism $G_e \to G_v$ as follows. By applying an appropriate element $g \in G$ as above, we can assume that $\tilde{e} = [\tilde{v}, \tilde{w}]$. Then We define the embedding $G_e \to G_v$ to make the diagram

$$\begin{array}{ccc} G_{\tilde{e}} & \longrightarrow & G_{\tilde{v}} \\ \downarrow & & \downarrow \\ G_{e} & \longrightarrow & G_{v} \end{array}$$

commutative. The result is a graph of groups $\mathcal{G}$. We leave it to the reader to verify that the functor $(G \curvearrowright T) \to \mathcal{G}$ described above is just the reverse of the functor $\mathcal{G} \to (G \curvearrowright T)$ for $\mathcal{G}$ with $G = \pi_1(\mathcal{G})$. In particular, $\mathcal{G}$ is trivial if and only if the action $G \curvearrowright T$ is trivial.

Definition 4.58. $\mathcal{G} \to (G \curvearrowright T) \to \mathcal{G}$ is the *Bass–Serre correspondence* between realizations of groups as fundamental groups of graphs of groups and group actions on trees without inversions.

We refer the reader to **[SW79]** and **[Ser80]** for further details.

## 4.8. Cayley graphs

Finitely generated groups may be turned into geometric object as follows. Given a group $G$ and its generating set $S$, one defines the *Cayley graph* of $G$ with respect to $S$. This is a symmetric directed graph $\mathrm{Cayley}_{\mathrm{dir}}(G, S)$ such that

- its set of vertices is $G$;
- its set of oriented edges is $(g, gs)$, with $s \in S$.

Usually, the underlying non-oriented graph $\mathrm{Cayley}(G, S)$ of $\mathrm{Cayley}_{\mathrm{dir}}(G, S)$, i.e. the graph such that:

- its set of vertices is $G$;
- its set of edges consists of all pairs of elements in $G$, $\{g, h\}$, such that $h = gs$, with $s \in S$,

is also called *Cayley graph of* $G$ *with respect to* $S$.

By abusing notation, we will also use the notation $[g, h] = \overline{gh}$ for the edge $\{g, h\}$.

Since $S$ is a generating set of $G$, it follows that the graph $\mathrm{Cayley}(G, S)$ is connected.

One can attach a *color* (*label*) from $S$ to each oriented edge in $\mathrm{Cayley}_{\mathrm{dir}}(G, S)$: the edge $(g, gs)$ is labeled by $s$.

We endow $\mathrm{Cayley}(G, S)$ with the standard length metric (where every edge has unit length). The restriction of this metric to $G$ is called *the word metric associated to* $S$ and it is denoted by $\mathrm{dist}_S$ or $d_S$.

NOTATION 4.59. For an element $g \in G$ and a generating set $S$ we denote $\mathrm{dist}_S(1, g)$ by $|g|_S$, the *word norm* of $g$. With this notation, $\mathrm{dist}_S(g, h) = |g^{-1}h|_S = |h^{-1}g|_S$.

CONVENTION 4.60. In this book, unless stated otherwise, all Cayley graphs are for finite generating sets $S$.

Much of the discussion in this section though remains valid for arbitrary generating sets, including infinite ones.

REMARK 4.61. 1. Every group acts on itself by left multiplication:

$$G \times G \to G\,,\ (g, h) \mapsto gh\,.$$

This action extends to any Cayley graph: if $[x, xs]$ is an edge of $\mathrm{Cayley}(G, S)$ with the vertices $x, xs$, we extend $g$ to the isometry

$$g : [x, xs] \to [gx, gxs]$$

between the unit intervals. Both actions $G \curvearrowright G$ and $G \curvearrowright \mathrm{Cayley}(G, S)$ are isometric. It is also clear that both actions are free, properly discontinuous and cocompact (provided that $S$ is finite): The quotient $\mathrm{Cayley}(G, S)/G$ is homeomorphic to the bouquet of $n$ circles, where $n$ is the cardinality of $S$.

2. The action of the group on itself by right multiplication defines maps

$$R_g : G \to G\,,\ R_g(h) = hg$$

that are in general not isometries with respect to a word metric, but are at finite distance from the identity map:

$$\mathrm{dist}(\mathrm{id}(h), R_g(h)) = |g|_S\,.$$

Exercise 4.62. Prove that the word metric on a group $G$ associated to a generating set $S$ may also be defined

(1) either as the unique maximal left-invariant metric on $G$ such that

$$\mathrm{dist}(1, s) = \mathrm{dist}(1, s^{-1}) = 1\,,\ \forall s \in S\,;$$

(2) or by the following formula: $\mathrm{dist}(g, h)$ is the length of the shortest word $w$ in the alphabet $S \cup S^{-1}$ such that $w = g^{-1}h$ in $G$.

Below are two simple examples of Cayley graphs.

Example 4.63. Consider $\mathbb{Z}^2$ with set of generators

$$S = \{a = (1, 0), b = (0, 1), a^{-1} = (-1, 0), b^{-1} = (0, -1)\}.$$

The Cayley graph $\mathrm{Cayley}(G, S)$ is the square grid in the Euclidean plane: The vertices are points with integer coordinates, two vertices are connected by an edge if and only if either their first or their second coordinates differ by $\pm 1$. See Figure 4.4

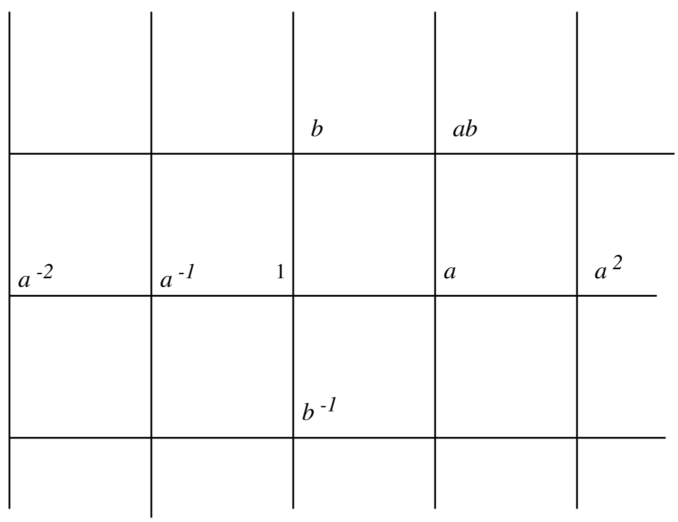



Figure 4.4. Cayley graph of $\mathbb{Z}^2$.

The Cayley graph of $\mathbb{Z}^2$ with respect to the set of generators $\{\pm(1, 0), \pm(1, 1)\}$ has the same set of vertices as the above, but the vertical lines must be replaced by diagonal lines.

Example 4.64. Let $G$ be the free group on two generators $a, b$. Take $S = \{a, b, a^{-1}, b^{-1}\}$. The Cayley graph $\mathrm{Cayley}(G, S)$ is the 4-valent tree (there are four edges incident to each vertex).

See Figure 4.5.

Theorem 4.65. *Fundamental group of every connected graph $\Gamma$ is free.*

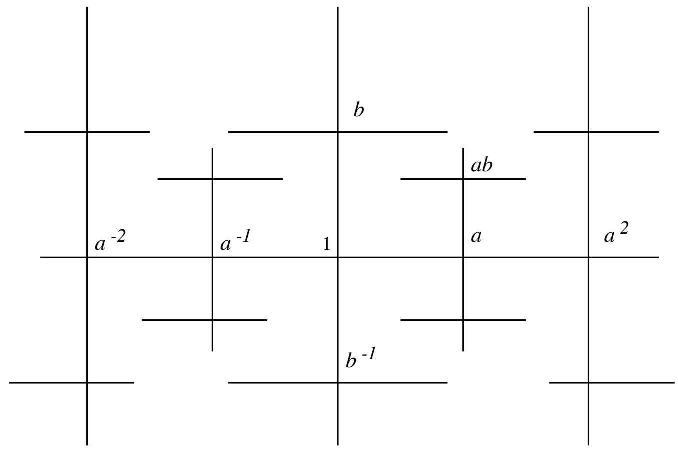



FIGURE 4.5. *F*ree group.

PROOF. By axiom of choice, $\Gamma$ contains a maximal subtree $\Lambda \subset \Gamma$. Let $\Gamma'$ denote the subdivision of $\Gamma$ where very edge $e$ in $\mathcal{E} = E(\Gamma) \setminus E(\Lambda)$ is subdivided in 3 sub-edges. For every such edge $e$ let $e'$ denote the middle 3rd. Now, add to $\Lambda$ all the edges in $E(\Gamma')$ which are not of the form $e'$ ($e \in \mathcal{E}$), and the vertices of such edges, of course, and let $T'$ denote the resulting tree. Thus, we obtain a covering of $\Gamma'$ by the simplicial tree $T'$ and the subgraph $\Gamma_{\mathcal{E}}$ consisting of the pairwise disjoint edges $e'$ ($e \in \mathcal{E}$), and the incident vertices. To this covering we can now apply Seifert–Van Kampen Theorem and conclude that $G = \pi_1(\Gamma)$ is free, with the free generators indexed by the set $\mathcal{E}$. □

COROLLARY 4.66. *A connected graph is simply connected if and only if the graph is a tree.*

COROLLARY 4.67. *1. Every free group $F(X)$ is the fundamental group of the bouquet $B$ of $|X|$ circles. 2. The universal cover of $B$ is a tree $T$, which is isomorphic to the Cayley graph of $F(X)$ with respect to the generating set $X$.*

PROOF. 1. By Theorem 4.65, $G = \pi_1(B)$ is free; furthermore, the proof also shows that the generating set of $G$ is identified with the set of edges of $B$. We now orient every edge of $B$ using this identification. 2. The universal cover $T$ of $B$ is a simply-connected graph, hence, a tree. We lift the orientation of edges of $B$ to orientation of edges of $T$. The group $F(X) = \pi_1(B)$ acts on $T$ by covering transformations, hence, the action on the vertex $V(T)$ set of $T$ is simply-transitive. Therefore, we obtain and identification of $V(T)$ with $G$. Let $v$ be a vertex of $T$. By construction and the standard identification of $\pi_1(B)$ with covering transformations of $T$, every oriented edge $e$ of $B$ lifts to an oriented edge $\tilde{e}$ of $T$ of the form $[v, w]$. Conversely, every oriented edge $[v, w]$ of $T$ projects to an oriented edge of $B$. Thus, we labeled all the oriented edges of $T$ with generators of $F(X)$. Again, by the covering theory, if an oriented edge $[u, w]$ of $T$ is labeled with a generator $x \in F(X)$, then $x$ sends $u$ to $w$. Thus, $T$ is isomorphic to the Cayley graph of $F(X)$. □

COROLLARY 4.68. *A group $G$ is free if and only if it can act freely by automorphisms on a simplicial tree $T$.*

Proof. By the covering theory, $G \cong \pi_1(\Gamma)$ where $\Gamma = T/G$. Now, Theorem 4.65, $G = \pi_1(\Gamma)$ is free. See **[Ser80]** for another proof and more general discussion of group actions on trees. □

Remark 4.69. The concept of a simplicial tree generalizes to the one of a real tree. There are non-free groups acting isometrically and freely on real trees, e.g., surface groups and free abelian groups. Rips proved that every finitely generated group acting freely and isometrically on a real tree is a free product of surface groups and free abelian groups, see e.g. **[Kap01]** for a proof.

Corollary 4.70 (Nielsen–Schreier). *Every subgroup $H$ of a free group $F$ is itself free.*

Proof. Realize the free group $F$ as the fundamental group of a bouquet $B$of circles; the universal cover $T$ of $B$ is a simplicial tree. The subgroup $H \leqslant F$ also acts on $T$ freely. Thus, $H$ is free. □

Exercise 4.71. Let $G$ and $H$ be finitely generated groups, with $S$ and $X$ respective finite generating sets.

Consider the wreath product $G \wr H$ as defined in Definition 3.65, endowed with the finite generating set canonically associated to $S$ and $X$ described in Exercise 4.10. For every function $f : H \to G$ denote by $\operatorname{supp} f$ the set of elements $h \in H$ such that $f(h) \neq \mathbf{1}_G$ .

Let $f$ and $g$ be arbitrary functions from $H$ to $G$ with finite support, and $h, k$ arbitrary elements in $H$. Prove that the word distance in $G \wr H$ from $(f, h)$ to $(g, k)$ with respect to the generating set mentioned above is

$$\operatorname{dist}\left((f,h),(g,k)\right) = \sum_{x \in H} \operatorname{dist}_S(f(x), g(x)) + \operatorname{Length}(\operatorname{supp} g^{-1}f\,; h, k)\,, \tag{4.5}$$

where $\operatorname{Length}(\operatorname{supp} g^{-1}f\,; h, k)$ is the length of the shortest path in $\operatorname{Cayley}(H, X)$ starting in $h$, ending in $k$ and whose image contains the set $\operatorname{supp} g^{-1}f$ .

Thus we succeeded in assigning to every finitely generated group $G$ a metric space $\operatorname{Cayley}(G, S)$. The problem, however, is that this assignment $G \to \operatorname{Cayley}(G, S)$ is far from canonical: different generating sets could yield completely different Cayley graphs. For instance, the trivial group has the presentations:

$$\langle \ \mid \ \rangle\,, \quad \langle a|a\rangle\,, \quad \left\langle a, b|ab, ab^2\right\rangle, \ldots,$$

which give rise to the non-isometric Cayley graphs:

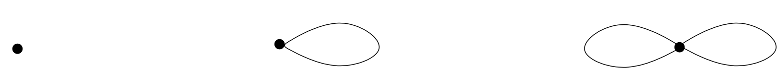


Figure 4.6. *C*ayley graphs of the trivial group.

The same applies to the infinite cyclic group:

In the above examples we did not follow the convention that $S = S^{-1}$.

Note, however, that all Cayley graphs of the trivial group have finite diameter; the same, of course, applies to all finite groups. The Cayley graphs of $\mathbb{Z}$ as above, although they are clearly non-isometric, are within finite distance from each other (when placed in the same Euclidean plane). Therefore, when seen from a (very) large distance (or by a person with a very poor vision), every Cayley graph of a

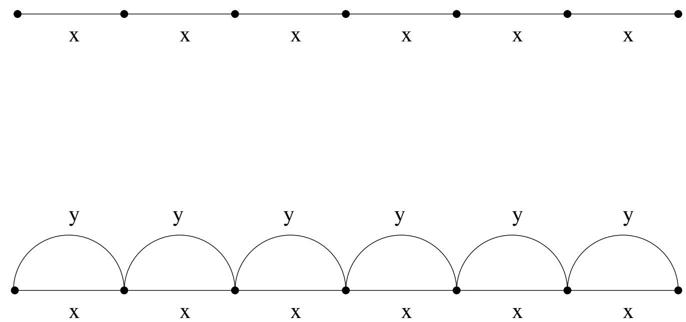



FIGURE 4.7. Cayley graphs of $\mathbb{Z} = \langle x | \rangle$ and $\mathbb{Z} = \left\langle x, y | xy^{-1} \right\rangle$.

finite group looks like a "fuzzy dot"; every Cayley graph of $\mathbb{Z}$ looks like a "fuzzy line," etc. Therefore, although non-isometric, they "look alike".

EXERCISE 4.72.
(1) Prove that if $S$ and $\bar{S}$ are two finite generating sets of $G$ then the word metrics $\mathrm{dist}_S$ and $\mathrm{dist}_{\bar{S}}$ on $G$ are bi-Lipschitz equivalent, i.e. there exists $L > 0$ such that

$$\frac{1}{L}\mathrm{dist}_S(g, g') \leqslant \mathrm{dist}_{\bar{S}}(g, g') \leqslant L\mathrm{dist}_S(g, g')\,, \forall g, g' \in G\,. \tag{4.6}$$

(2) Prove that an isomorphism between two finitely generated groups is a bi-Lipschitz map when the two groups are endowed with word metrics.

CONVENTION 4.73. From now on, unless otherwise stated, by a metric on a finitely generated group we mean a word metric coming from a finite generating set.

EXERCISE 4.74. Show that the Cayley graph of a finitely generated infinite group contains an isometric copy of $\mathbb{R}$, i.e. a bi-infinite geodesic. Hint: Apply Arzela-Ascoli theorem to a sequence of geodesic segments in the Cayley graph.

On the other hand, it is clear that no matter how poor your vision is, the Cayley graphs of, say, $\{1\}$, $\mathbb{Z}$ and $\mathbb{Z}^2$ all look different: They appear to have different "dimension" (0, 1 and 2 respectively).

Telling apart the Cayley graph $\mathrm{Cayley}_1$ of $\mathbb{Z}^2$ from the Cayley graph $\mathrm{Cayley}_2$ of the Coxeter group

$$\Delta := \Delta(4, 4, 4) := \left\langle a, b, c | a^2, b^2, c^2, (ab)^4, (bc)^4, (ca)^4 \right\rangle$$

seems more difficult: They both "appear" 2-dimensional. However, by looking at the larger pieces of $\mathrm{Cayley}_1$ and $\mathrm{Cayley}_2$, the difference becomes more apparent: Within a given ball of radius $R$ in $\mathrm{Cayley}_1$, there seems to be less vertices than in $\mathrm{Cayley}_2$. The former grows quadratically, the latter grows exponentially fast as $R$ goes to infinity.

The goal of the rest of the book is to make sense of this "fuzzy math".

In Section 5.1 we replace the notion of an *isometry* with the notion of a *quasi-isometry*, in order to capture what different Cayley graphs of the same group have in common.

LEMMA 4.75. *A* finite index *subgroup of a finitely generated group is finitely generated.*

PROOF. It follows from Theorem 5.29. We give here another proof, as the set of generators of the subgroup found here will be used in future applications.

Let $G$ be a group and $S$ a finite generating set of $G$, and let $H$ be a finite index subgroup in $G$. Then $G = H \sqcup \bigsqcup_{i=1}^{k} Hg_i$ for some elements $g_i \in G$. Consider

$$R = \max_{1 \leqslant i \leqslant k} |g_i|_S .$$

Then $G = HB(1, R)$. We now prove that $X = H \cap B(1, 2R+1)$ is a generating set of $H$.

Let $h$ be an arbitrary element in $H$ and let $g_0 = 1, g_1, \ldots, g_n = h$ be the consecutive vertices on a geodesic in Cayley$(G, S)$ joining 1 and $h$. In particular, this implies that $\mathrm{dist}_S(1, h) = n$.

For every $1 \leqslant i \leqslant n-1$ there exist $h_i \in H$ such that $\mathrm{dist}_S(g_i, h_i) \leqslant R$. Set $h_0 = 1$ and $h_n = h$. Then $\mathrm{dist}_S(h_i, h_{i+1}) \leqslant 2R+1$, hence $h_{i+1} = h_i x_i$ for some $x_i \in X$, for every $0 \leqslant i \leqslant n-1$. It follows that $h = h_n = x_1 x_2 \cdots x_n$, whence $X$ generates $H$ and $|h|_X \leqslant |h|_S = n$. □

## 4.9. Volumes of maps of cell complexes and Van Kampen diagrams

The goal of this section is to describe several notions of volumes of maps and to relate them to each other and to the word reductions in finitely-presented groups. It turns out that most of these notions are equivalent, but, in few cases, there subtle differences.

Recall that in section 2.1.4 we defined volumes of maps between Riemannian manifolds. More generally, the same definition of volume of a map applies in the context of Lipschitz maps of Euclidean simplicial complexes, i.e., simplicial complexes where each $k$-simplex is equipped with the metric of the Euclidean simplex where every edge has unit length. In order to compute $n$-volume of a map $f$, first compute volumes of restrictions $f|\Delta_i$, for all $n$-dimensional simplices and then add up the results.

**4.9.1. Simplicial and combinatorial volumes of maps.** Suppose that $X, Y$ are simplicial complexes equipped with *standard metrics* and $f: X \to Y$ is a simplicial map, i.e., a map which sends every simplex to simplex so that the restriction is linear. Then the $n$-dimensional simplicial volume $sVol_n(f)$ of $f$ is just the number of $n$-dimensional simplices in the domain $X$. Note that this, somewhat strange, concept, is independent of the map $f$ but is, nevertheless, useful. The more natural concept is the one of the *combinatorial volume* of the map $f$, namely,

$$cVol_n(f) = \sum_{\Delta} \frac{1}{c_n} Vol(f(\Delta))$$

where the sum is taken over all $n$-simplices in $X$ and $c_n$ is the volume of the Euclidean simplex with unit edges. In other words, $cVol_n$ counts the number of $n$-simplices in $X$ which are not mapped by $f$ to simplices of lower dimension.

Both definitions extend in the context of cellular maps of cell-complexes.

Definition 4.76. Let $X, Y$ be $n$-dimensional almost regular cell complexes. A cellular map $f : X \to Y$ is said to be *regular* if for every $n$-cell $\sigma$ in $X$ either:

(a) $f$ *collapses* $\sigma$, i.e., $f(\sigma) \subset Y^{(n-1)}$, or

(b) $f$ maps the interior of $\sigma$ homeomorphically to the interior of an $n$-cell in $Y$.

For instance, simplicial map of simplicial complexes is regular.

We define the *combinatorial* $n$-volume $cVol_n(f)$ of $f$ to be the total number of $n$-cells in $X$ which are not collapsed by $f$. The combinatorial 2-volume is called *area*. Thus, this definition agrees with the notion of combinatorial volume for simplicial maps.

**Geometric volumes of maps.** Similarly, suppose that $X, Y$ are regular $n$-dimensional cell complexes. We define smooth structure on each open $n$-cell in $X$ and $Y$ by using the identification of these cells with the open $n$-dimensional Euclidean balls of unit volume, coming from the regular cell complex structure on $X$ and $Y$.

We say that a cellular map $f : X \to Y$ is *smooth* if for every $y \in Y$ which belongs to an open $n$-cell, $f$ is smooth at every $x \in f^{-1}(y)$. At points $x \in f^{-1}(y)$ for such $y$ we have a continuous function $|J_f(x)|$. We declare $|J_f(x)|$ to be zero at all points $x \in X$ which map to $Y^{(n-1)}$. Then we again define the *geometric volume* $Vol(f)$ by the formula (2.2) where the integral is taken over all open $n$-cells in $X$. We extend this definition to the case where $f$ is not smooth over some open $m$-cells by setting $Vol(f) = \infty$ in this case. In the case when $n = 2$, $Vol(f)$ is called the *area* of $f$ and denoted $Area(f)$.

We now assume that $X$ is an $n$-dimensional finite regular cell complex and $Z \subset X$ is a subcomplex of dimension $n - 1$. The example we will be primarily interested in is when $X$ is the 2-disk and $Z$ is its boundary circle.

Lemma 4.77 (Regular cellular approximation). *After replacing $X$ with its subdivision if necessary, every cellular map $f : X \to Y$ is homotopic, rel. $Z$, to a smooth regular map $h : X \to Y$ so that*

$$Vol(h) = cVol_n(h) \leqslant cVol_n(f)$$

*i.e., the geometric volume equals the combinatorial volume for the map $h$.*

Proof. First, without loss of generality, we may assume that $f$ is smooth. For each open $n$-cell $\sigma^\circ$ in $Y$ we consider components $U$ of $f^{-1}(\sigma^\circ)$. If for some $U$ and $p \in \sigma^\circ$, $f(U) \subset \sigma^\circ \setminus p$, then we compose $f|cl(U)$ with the retraction of $\sigma$ to its boundary from the point $p$. The resulting map $f_1$ is clearly cellular, homotopic to $f$ rel. $Z$ and its $n$-volume is at most the $n$-volume of $f$ (for both geometric and combinatorial volumes). Moreover, for every component $U$ of $f_1^{-1}(\sigma^\circ)$, $f_1(U) = \sigma^\circ$. We let $m(f_1, \sigma)$ denote the number of components of $f^{-1}(\sigma^\circ)$.

Our next goal is to replace $f_1$ with a new (cellular) map $f_2$ so that $f_2$ is 1-1 on each $U$ as above. By Sard's theorem, for every $n$-cell $\sigma$ in $Y$ there exists a point $p = p_\sigma \in \sigma^\circ$ which is a regular value of $f_1$. Let $V = V_\sigma \subset \sigma^\circ$ be a small closed ball whose interior contains $p$ and so that $f_1$ is a covering map over $V$. Let $\rho_\sigma : \sigma \to \sigma$ denote the retraction of $\sigma$ to its boundary which sends $V$ diffeomorphically to $\sigma^\circ$ and which maps $\sigma \setminus V$ to the boundary of $\sigma$. Let $\rho : Y \to Y$ be the map whose restriction to each closed $n$-cell $\sigma$ is $\rho_\sigma$ and whose restriction to $Y^{(n-1)}$ is the identity map. Then we replace $f_1$ with the composition $f_2 := \rho \circ f_1$. It is clear that the new map $f_2$ is cellular and is homotopic to $f_1$ rel. $Z$. Moreover, $f_2$ is a trivial covering over each open $n$-cell in $Y$. By construction, we have:

$$Vol_n(f_2) = \sum_\sigma m(f_1, \sigma) Vol_n(\sigma) = \sum_\sigma m(f_2, \sigma) Vol_n(\sigma) \leqslant Vol_n(f), \tag{4.7}$$

where the sum is taken over all $n$-cells $\sigma$ in $Y$. Furthermore, for each $n$-cell $\sigma$, $f_2^{-1}(\sigma^\circ)$ is a disjoint union of open $n$-balls, each of which is contained in an open $n$-cell in $X$. Moreover, the restriction of $f_2$ to the boundary of each of these balls factors as the composition

$$e_\sigma \circ g$$

where $g$ is a homeomorphism to the Euclidean ball $B^n$ and $e_\sigma : \partial B^n \to Y^{(n-1)}$ is the attaching map of the cell $\sigma$. We then subdivide the cell complex $X$ so that the closure of each $f_2^{-1}(\sigma^\circ)$ is a cell. Then $h := f_2$ is the required regular map. The required equality (and inequality) of volumes is an immediate corollary of the equation (4.7). □

### 4.9.2. Topological interpretation of finite-presentability.

Lemma 4.78. *A group $G$ is isomorphic to the fundamental group of a finite cell complex $Y$ if and only if $G$ is finitely-presented.*

Proof. 1. Suppose that $G$ has a finite presentation

$$\langle X|R\rangle = \langle x_1, \dots, x_n | r_1, \dots, r_m\rangle .$$

We construct a finite 2-dimensional cell-complex $Y$, as follows. The complex $Y$ has unique vertex $v$. The 1-skeleton of $Y$ is the $n$-rose, the bouquet of $n$ circles $\gamma_1, \dots, \gamma_n$ with the common point $v$, the circles are labeled $x_1, \dots, x_n$. Observe that the free group $F_X$ is isomorphic to $\pi_1(Y^1, v)$ where the isomorphism sends each $x_i$ to the circle in $Y^1$ with the label $x_i$. Thus, every word $w$ in $X^*$ determines a based loop $L_w$ in $Y^1$ with the base-point $v$. In particular, each relator $r_i$ determines a loop $\alpha_i := L_{r_i}$. We then attach 2-cells $\sigma_1, \dots, \sigma_m$ to $Y^1$ using the maps $\alpha_i : S^1 \to Y^1$ as the attaching maps. Let $Y$ be the resulting cell complex. It is clear from the construction that $Y$ is *almost regular*.

We obtain a homomorphism $\phi : F_X \to \pi_1(Y^1) \to \pi_1(Y)$. Since each $r_i$ lies in the kernel of this homomorphism, $\phi$ descends to a homomorphism $\psi : G \to \pi_1(Y)$. It follows from the Seifert-Van Kampen theorem that $\psi$ is an isomorphism.

2. Suppose that $Y$ is a finite complex with $G \cong \pi_1(Y)$. Pick a maximal subtree $T \subset Y^1$ and let $X$ be the complex obtained by contracting $T$ to a point. Since $T$ is contractible, the resulting map $Y \to X$ (contracting $T$ to a point $v \in X^0$) is a homotopy-equivalence. The 1-skeleton of $X$ is an $n$-rose with the edges $\gamma_1, \dots, \gamma_n$ which we will label $x_1, \dots, x_n$. It is now again follows from Seifert-Van Kampen theorem that $X$ is a presentation complex for a finite presentation of $G$: The generators $x_i$ are the loops $\gamma_i$ and the relators are the 2-cells (or, rather, their attaching maps $S^1 \to X^1$). □

Definition 4.79. The complex $Y$ constructed in this proof is called the *presentation complex* of $G$ associated with the presentation $\langle X|R\rangle$.

Definition 4.80. The 2-dimensional complex $Y$ constructed in the first part of the above proof is called the *presentation complex* of the presentation

$$\langle x_1, \dots, x_n | r_1, \dots, r_m\rangle .$$

**4.9.3. Van Kampen diagrams and Dehn function. Van Kampen diagrams of relators.** Suppose that $\langle X|R\rangle$ is a (finite) presentation of a group $G$ and $Y$ be the corresponding presentation complex. Suppose that $w \in \langle\langle R\rangle\rangle < F_X$ is a relator in this presentation. Then $w$ corresponds to a null-homotopic loop $\lambda_w$ in the 1-skeleton $Y^{(1)}$ of $Y$. Let $f : D^2 \to Y$ be an extension of $\lambda_w : S^1 \to Y$. By the cellular approximation theorem (see e.g. [**Hat02**]), after subdivision of $D^2$ as a regular cell complex, we can assume that $f$ is cellular. Note, however, that some edges in this cell complex structure on $D^2$ will be mapped to vertices and some 2-cells will be mapped to 1-skeleton. A *Van Kampen diagram* if an convenient (and traditional) way to keep track of these dimension reductions.

DEFINITION 4.81. We say that a contractible finite planar regular cell complex $K$ is a *tree-graded disk* (a *tree of discs* or a *discoid*) provided that every edge of $K$ is contained in the boundary of $K$. In other words, $K$ is obtained from a finite simplicial tree by replacing some vertices with 2-cells, which is why we think of $K$ as a "tree of discs".

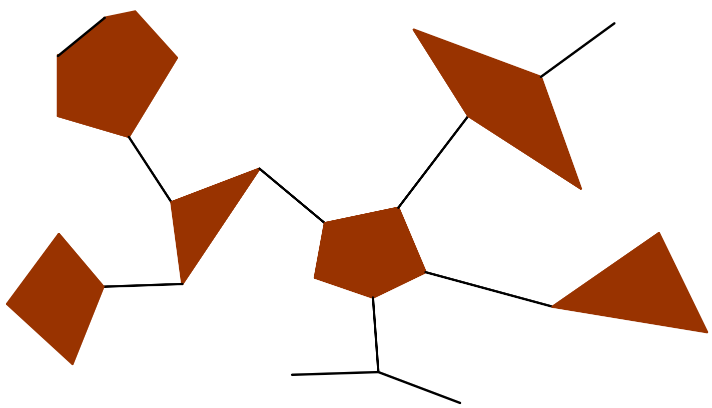


FIGURE 4.8. Example of tree-graded disk.

LEMMA 4.82. *For every $w$ as above, there exists a tree-graded disk $K$, a regular cell complex structure $\tilde{K}$ on $D^2$, a regular cellular map $f : \tilde{K} \to Y$ extending $\lambda_w$ and cellular maps $h : K \to Y, \kappa : \tilde{K} \to K$ so that: $f = h \circ \kappa$.*

PROOF. Write $w$ as a product

$$w = v_1 \cdots v_k, \quad v_i = u_i r_i u_i^{-1}, i = 1, \ldots, k,$$

where each $r_i \in R$ is a defining relator. Then the circle $S^1$ admits a regular cell complex structure so that $\lambda_w$ sends each vertex to the unique vertex $v \in Y$ and for every edge $\alpha_i$, the based loop $f|\alpha_i$ represents the word $v_i \in F_X$. Moreover, the arcs $\alpha_i$ are cyclically ordered on $S^1$ in order of appearance of $v_i$ in $w$. Furthermore, each $\alpha_i$ is subdivided in 3 arcs $\alpha_i^+, \beta_i, \alpha_i^-$ so that the loop $f|\alpha_i^{\pm}$ represents $u_i^{\pm 1}$ and $f|\beta_i$ represents $r_i$. We then construct a collection of pairwise disjoint arcs $\tau_i \subset D^2$ which intersect $S^1$ only at their end-points: For each pair $\alpha_i^+, \alpha_i^-$ we connect the end-points of $\alpha_i^+$ to that of $\alpha_i^-$ by arcs $\epsilon_i^{\pm}$. The result is a cell-complex structure $\tilde{K}$ on $D^2$ where every vertex is in $S^1$. There three types of 2-cells in $\tilde{K}$:

1 Cells $A_i$ bounded by bigons $\gamma_i \cup \epsilon_i^-$,
2 Cells $B_i$ bounded by rectangles $\alpha_i^+ \cup \epsilon^+ \cup \alpha_i^- \cup \epsilon^-$,
3 The rest, not containing any edges in $S^1$.

We now collapse each 2-cell of type (3) to a point, collapse each 2-cell of type (2) to an edge $e_i$ (so that $\alpha_i^{\pm}$ map homeomorphically onto this edge while $\epsilon_i^{\pm}$ map to the end-points of $e_i$). Note that $\alpha_i^{\pm}$ with their orientation inherited from $S^1$ define two opposite orientations on $e_i$.

The result is a tree-graded disk $K$ and a *collapsing map* $\kappa : \tilde{K} \to K$. We define a map $h : K^1 \to Y$ so that $h \circ \kappa|\alpha_i^{\pm} = \lambda_{v_i^{\pm 1}}$ while $h \circ \kappa|\beta_i = \lambda_{r_i}$. Lastly, we extend $h$ to the 2-cells $C_i := \kappa(A_i)$ in $K$: $h : C_i \to Y$ are the 2-cells corresponding to the defining relators $r_i$. □

DEFINITION 4.83. A map $h : K \to Y$ constructed in the above lemma is called a *Van Kampen diagram* of $w$ in $Y$.

The combinatorial area $cArea(h)$ of the Van Kampen diagram $h : K \to Y$ is the number of 2-cells in $K$, i.e., the number $k$ of relators $r_i$ used to describe $w$ as a product of conjugates of defining relators. The (algebraic) *area* of the loop $\lambda_w$ in $Y$, denoted $A(w)$, is

$$\min_{h:K\to Y} cArea(h)$$

where the minimum is taken over all Van Kampen diagrams of $w$ in $Y$. Algebraically, the area $A(w)$ is the least number of defining relators in the representation of $w$ as the product of conjugates of defining relators. This explains the significance of this notion of area: It captures the complexity of the word problem for the presentation $\langle X|R\rangle$ of the group $G$.

We identify all open 2-cells in $Y$ with open 2-disks of unit area. Our next goal is to convert arbitrary disks that bound $L_w$ to Van Kampen diagrams. Let $f : D^2 \to Y$ be a cellular map extending $\lambda_w$, where $D^2$ is given structure of a regular cell complex $W$. By Lemma 4.77, we can replace $f$ with a regular cellular map $f_1 : D^2 \to Y$, which is homotopic to $f$ rel. $Z := \partial D^2$, so that $cArea(f_1) = Area(f_1) \leqslant Area(f)$.

We use the orientation induced from $D^2$ on each 2-cell in $W$. Pick a base-point $x \in \partial D^2$ which is a vertex of $W$. Let $\sigma_1, \dots, \sigma_m$ be the 2-cells in $W$. For each 2-cell $\sigma = \sigma_i$ of $W$ we let $p_\sigma$ denote a path in $W^{(1)}$ connecting $x$ to $\partial\sigma$. Then, by attaching the "tail" $p_\sigma$ to each $\partial\sigma$ (whose orientation is induced from $\sigma$) we get an oriented loop $\tau_\sigma$ based at $x$. By abusing the notation we let $\tau_\sigma$ denote the corresponding elements of $\pi_1(W^{(1)}, x)$. We let $\lambda \in \pi_1(W^{(1)}, x)$ denote the element corresponding to the (oriented) boundary circle of $D^2$. We leave it to the reader to verify that the group $\pi_1(W^{(1)}, x)$ is freely generated by the elements $\tau_\sigma$ and that $\lambda$ is the product

$$\prod_\sigma \tau_\sigma$$

(in some order) of the elements $\tau_\sigma$ where each $\tau_\sigma$ appears exactly once. (This can be shown, for instance, by induction on the number of 2-cells in $W$.) We renumber the 2-cells in $W$ so that the above product has the form

$$\prod_\sigma \tau_\sigma = \tau_{\sigma_1} \dots \tau_{\sigma_m}$$

For each $\sigma_i$ set $\phi_i := \pi_1(f_1)(\tau_{\sigma_i}) \in \pi_1(Y^{(1)}, y)$, $y = f_1(x)$. Then, the element $\pi_1(f_1)(\lambda) \in \pi_1(Y^{(1)}, y)$ (represented by the loop $\lambda_w$) is the product

$$\phi_1 \dots \phi_m \tag{4.8}$$

in the group $\pi_1(Y^{(1)}, y)$. For every 2-cell $\sigma_i$ of $W$ either $\sigma_i$ is collapsed by $f_1$ or not. In the former case, $\phi_i$ represents a trivial element of the free group $\pi_1(Y^{(1)}, y)$. In the latter case, $\phi_i$ has the form

$$u_i r_{j(i)} u_i^{-1}$$

where $r_{j(i)} \in R$ is one of the defining relators of the presentation $\langle X|R\rangle$ and the word $u_i \in F_X$ corresponds to the loop $f_1(p_{\sigma_i})$. Therefore, we can eliminate the elements of the second type from the product (4.8) while preserving the identity

$$w = \phi_{i_1} \cdots \phi_{i_k} \in F_X.$$

This product decomposition, as we observed above, corresponds to a Van Kampen diagram $h : K \to Y$. The number $k$ is nothing but the combinatorial area of the map $f_1$ above. We conclude

Proposition 4.84 (Combinatorial area equals geometric area equals algebraic area).

$$A(w) = \min\{cArea(f) = Area(f) | f : D^2 \to Y\},$$

*where the minimum is taken over all regular cellular maps $f$ extending the map $\lambda_w : S^1 \to Y^{(1)}$.*

Definition 4.85 (Dehn function). Let $G$ be a group with finite presentation $\langle X|R\rangle$ and the corresponding presentation complex $Y$. The *Dehn function of $G$* (with respect to the finite presentation $\langle X|R\rangle$) equals

$$Dehn(n) := \max\{A(w) : |w| \leqslant n\}$$

where $w$'s are elements in $X^*$ representing trivial words in $G$. Geometrically speaking,

$$Dehn(n) = \max_{\lambda, \ell(\lambda) \leqslant n} \min\{cArea(f) | f : D^2 \to Y, f|\partial D^2 = \lambda\}$$

where $\lambda$'s are homotopically trivial regular cellular maps of the triangulated circle to $Y$ and $f$'s are regular cellular maps of the triangulated disk $D^2$ to $Y$.

CHAPTER 5

# Coarse geometry

## 5.1. Quasi-isometry

We now define an important equivalence relation between metric spaces: the quasi-isometry. The quasi-isometry has two equivalent definitions: one which is easy to visualize and one which makes it easier to understand why it is an equivalence relation. We begin with the first definition, continue with the second and prove their equivalence.

DEFINITION 5.1. Two metric spaces $(X, \mathrm{dist}_X)$ and $(Y, \mathrm{dist}_Y)$ are *quasi-isometric* if and only if there exist $A \subset X$ and $B \subset Y$, separated nets, such that $(A, \mathrm{dist}_X)$ and $(B, \mathrm{dist}_Y)$ are bi-Lipschitz equivalent.

EXAMPLES 5.2. (1) A metric space of finite diameter is quasi-isometric to a point.

(2) The space $\mathbb{R}^n$ endowed with a norm is quasi-isometric to $\mathbb{Z}^n$ with the metric induced by that norm.

Historically, quasi-isometry was introduced in order to formalize the relationship between some discrete metric spaces (most of the time, groups) and some "non-discrete" (or continuous) metric spaces like for instance Riemannian manifolds etc. A particular instance of this is the relationship between hyperbolic spaces and certain hyperbolic groups.

When trying to prove that the quasi-isometry relation is an equivalence relation, reflexivity and symmetry are straightforward, but when attempting to prove transitivity, the following question naturally arises:

QUESTION 5.3 ([**Gro93**], p. 23). Can a space contain two separated nets that are not bi-Lipschitz equivalent?

THEOREM 5.4 ([**BK98**]). *There exists a separated net $N$ in $\mathbb{R}^2$ which is not bi-Lipschitz equivalent to $\mathbb{Z}^2$.*

OPEN QUESTION 5.5 ([**BK02**]). When placing a point in the barycenter of each tile of a Penrose tiling, is the resulting separated net bi-Lipschitz equivalent to $\mathbb{Z}^2$?

A more general version of this question: embed $\mathbb{R}^2$ into $\mathbb{R}^n$ as a plane $P$ with irrational slope and take $B$, a bounded subset of $\mathbb{R}^n$ with non-empty interior. Consider all $z \in \mathbb{Z}^n$ such that $z + B$ intersects $P$. The projections of all such $z$ on $P$ compose a separated net. Is such a net bi-Lipschitz equivalent to $\mathbb{Z}^2$?

Fortunately there is a second equivalent way of defining the fact that two metric spaces are quasi-isometric, which is as follows. We begin by loosening up the Lipschitz concept.

DEFINITION 5.6. Let $X, Y$ be metric spaces. A map $f : X \to Y$ is called $(L, C)$–*coarse Lipschitz* if

$$\text{dist}_Y(f(x), f(x')) \leqslant L\text{dist}_X(x, x') + C \tag{5.1}$$

for all $x, x' \in X$. A map $f : X \to Y$ is called an $(L, C)$–*quasi-isometric embedding* if

$$L^{-1}\text{dist}_X(x, x') - C \leqslant \text{dist}_Y(f(x), f(x')) \leqslant L\text{dist}_X(x, x') + C \tag{5.2}$$

for all $x, x' \in X$. Note that a quasi–isometric embedding does not have to be an embedding in the usual sense, however distant points have distinct images.

If $X$ is a finite interval $[a, b]$ then an $(L, C)$–quasi–isometric embedding $\mathfrak{q} : X \to Y$ is called *a quasi-geodesic (segment)*. If $a = -\infty$ or $b = +\infty$ then $\mathfrak{q}$ is called *quasi-geodesic ray*. If both $a = -\infty$ and $b = +\infty$ then $\mathfrak{q}$ is called *quasi-geodesic line*. By abuse of terminology, the same names are used for the image of $\mathfrak{q}$.

An $(L, C)$–quasi-isometric embedding is called an $(L, C)$–**quasi–isometry** if it admits a **quasi–inverse** map $\bar{f} : Y \to X$ which is also an $(L, C)$–quasi–isometric embedding so that:

$$\text{dist}_X(\bar{f}f(x), x) \leqslant C, \quad \text{dist}_Y(f\bar{f}(y), y) \leqslant C \tag{5.3}$$

for all $x \in X, y \in Y$.

Two metric spaces $X, Y$ are *quasi-isometric* if there exists a quasi-isometry $X \to Y$.

We will abbreviate *quasi-isometry*, *quasi–isometric* and *quasi-isometrically* to QI.

EXERCISE 5.7. Let $f_i : X \to X$ be maps so that $f_3$ is $(L_3, A_3)$ coarse Lipschitz and $\text{dist}(f_2, id_X) \leqslant A_2$. Then

$$\text{dist}(f_3 \circ f_1, f_3 \circ f_2, \circ f_1) \leqslant L_3 A_2 + A_3.$$

DEFINITION 5.8. A metric space $X$ is called *quasi-geodesic* if there exist constants $(L, A)$ so that every pair of points in $X$ can be connected by an $(L, A)$–quasi-geodesic.

In most cases the *quasi–isometry constants* $L, C$ do not matter, so we shall use the words *quasi–isometries* and *quasi-isometric embeddings* without specifying constants.

EXERCISE 5.9.
(1) Prove that the composition of two quasi-isometric embeddings is a quasi-isometric embedding, and that the composition of two quasi-isometries is a quasi-isometry.
(2) Prove that quasi-isometry of metric spaces is an equivalence relation.

Some quasi-isometries $X \to X$ are more interesting than others. The *boring* quasi-isometries are the ones which are within finite distance from the identity:

DEFINITION 5.10. Given a metric space $(X, \text{dist})$ we denote by $\mathcal{B}(X)$ the set of maps $f : X \to X$ (not necessarily bijections) which are *bounded perturbations of the identity*, i.e. maps such that

$$\text{dist}(f, id_X) = \sup_{x \in X} \text{dist}(f(x), x) \text{ is finite.}$$

In order to mod out the semigroup of quasi-isometries $X \to X$ by $\mathcal{B}(X)$, one introduces a group $QI(X)$ defined below. Given a metric space $(X, \text{dist})$, consider the set $QI(X)$ of equivalence classes of quasi-isometries $X \to X$, where two quasi-isometries $f, g$ are equivalent if and only if $\text{dist}(f, g)$ is finite. In particular, the set of quasi-isometries equivalent to $id_X$ is $\mathcal{B}(X)$. It is easy to see that the composition defines a binary operation on $QI(X)$, that the quasi-inverse defines an inverse in this group, and that $QI(X)$ is a group when endowed with these operations.

DEFINITION 5.11. The group $(QI(X)\,,\circ)$ is called the *group of quasi-isometries of the metric space* $X$.

There is a natural homomorphism $\text{Isom}(X) \to QI(X)$. In general, this homomorphism is not injective. For instance if $X = \mathbb{R}^n$ then the kernel is the full group of translations $\mathbb{R}^n$. Similarly, the entire group $G = \mathbb{Z}^n \times F$, where $F$ is a finite group, maps trivially to $QI(G)$. In general, kernel $K$ of $G \to QI(G)$ is a subgroup such that for every $k \in K$ the $G$-centralizer of $k$ has finite index in $G$, see Lemma 14.24. Thus, every finitely generated subgroup in $K$ is *virtually central.* In particular, if $G = K$ then $G$ is virtually abelian.

QUESTION 5.12. Is the subgroup $K \leqslant G$ always virtually central? Is it at least true that $K$ is always virtually abelian?

The group $VI(G)$ of virtual automorphisms of $G$ defined in Section 3.4 maps naturally to $QI(G)$ since every virtual isomorphism $\phi$ of $G$ ($\phi : G_1 \overset{\cong}{\to} G_2$, where $G_1, G_2$ are finite-index subgroups of $G$) induces a quasi-isometry $f_\phi : G \to G$. Indeed, $\phi : G_1 \to G_2$ is a quasi-isometry. Since both $G_i \subset G$ are nets, $\phi$ extends to a quasi-isometry $f_\phi : G \to G$.

EXERCISE 5.13. Show that the map $\phi \to f_\phi$ projects to a homomorphism $VI(G) \to QI(G)$.

When $G$ is a finitely generated group, $QI(G)$ is independent of the choice of word metric. More importantly, we will see (Corollary 5.62) that every group quasi-isometric to $G$ admits a natural homomorphism to $QI(G)$.

EXERCISE 5.14. Show that if $f : X \to Y$ is a quasi-isometric embedding such that $f(X)$ is $r$-dense in $Y$ for some $r < \infty$ then $f$ is a quasi-isometry.

*Hint:* Construct a quasi-inverse $\bar{f}$ to the map $f$ by mapping a point $y \in Y$ to $x \in X$ such that

$$\text{dist}_Y(f(x), y) \leqslant r.$$

EXAMPLE 5.15. The cylinder $X = \mathbb{S}^n \times \mathbb{R}$ with a product metric is quasi-isometric to $Y = \mathbb{R}$; the quasi-isometry is the projection to the second factor.

EXAMPLE 5.16. Let $h : \mathbb{R} \to \mathbb{R}$ be an $L$–Lipschitz function. Then the map

$$f : \mathbb{R} \to \mathbb{R}^2, \quad f(x) = (x, h(x))$$

is a QI embedding.

Indeed, $f$ is $\sqrt{1 + L^2}$–Lipschitz. On the other hand, clearly,

$$\text{dist}(x, y) \leqslant \text{dist}(f(x), f(y))$$

for all $x, y \in \mathbb{R}$.

Example 5.17. Let $\varphi : [1, \infty) \to \mathbb{R}_+$ be a differentiable function so that

$$\lim_{r\to\infty} \varphi(r) = \infty,$$

and there exists $C \in \mathbb{R}$ for which $|r\varphi'(r)| \leqslant C$ for all $r$. For instance, take $\varphi(r) = \log(r)$. Define the function $F : \mathbb{R}^2 \setminus B(0,1) \to \mathbb{R}^2 \setminus B(0,1)$ which in the polar coordinates takes the form

$$(r, \theta) \mapsto (r, \theta + \varphi(r)).$$

Hence $F$ maps radial straight lines to spirals. Let us check that $F$ is $L$–bi-Lipschitz for $L = \sqrt{1 + C^2}$. Indeed, the Euclidean metric in the polar coordinates takes the form

$$ds^2 = dr^2 + r^2 d\theta^2.$$

Then

$$F^*(ds^2) = ((r\varphi'(r))^2 + 1)dr^2 + r^2 d\theta^2$$

and the assertion follows. Extend $F$ to the unit disk by the zero map. Therefore, $F : \mathbb{R}^2 \to \mathbb{R}^2$, is a QI embedding. Since $F$ is onto, it is a quasi-isometry $\mathbb{R}^2 \to \mathbb{R}^2$.

Exercise 5.18. If $f, g : X \to Y$ are within finite distance from each other, i.e.

$$\sup \mathrm{dist}(f(x), g(x)) < \infty$$

and $f$ is a quasi-isometry, then $g$ is also a quasi-isometry.

Proposition 5.19. *Two metric spaces $(X, \mathrm{dist}_X)$ and $(Y, \mathrm{dist}_Y)$ are quasi-isometric in the sense of Definition 5.1 if and only if there exists a quasi-isometry $f : X \to Y$.*

Proof. Assume there exists an $(L, C)$–quasi-isometry $f : X \to Y$. Let $\delta = L(C + 1)$ and let $A$ be a $\delta$–separated $\varepsilon$–net in $X$. Then $B = f(A)$ is a 1–separated $(L\varepsilon + 2C)$–net in $Y$. Moreover for any $a, a' \in A$,

$$\mathrm{dist}_Y(f(a), f(a')) \leqslant L\mathrm{dist}_X(a, a') + C \leqslant \left(L + \frac{C}{\delta}\right) \mathrm{dist}_X(a, a')$$

and

$$\mathrm{dist}_Y(f(a), f(a')) \geqslant \frac{1}{L}\mathrm{dist}_X(a, a') - C \geqslant \left(\frac{1}{L} - \frac{C}{\delta}\right) \mathrm{dist}_X(a, a') =$$

$$\frac{1}{L(C+1)}\mathrm{dist}_X(a, a')\,.$$

It follows that $f$ restricted to $A$ and with target $B$ is bi-Lipschitz.

Conversely, assume that $A \subset X$ and $B \subset Y$ are two $\varepsilon$–separated $\delta$–nets, and that there exists a bi-Lipschitz map $g : A \to B$ which is onto. We define a map $f : X \to Y$ as follows: for every $x \in X$ we choose one $a_x \in A$ at distance at most $\delta$ from $x$ and define $f(x) = g(a_x)$.

*N.B.* The axiom of choice makes here yet another important appearance, if we do not count the episodic appearance of Zorn's Lemma, which is equivalent to the axiom of choice. Details on this axiom will be provided later on. Nevertheless, when $X$ is proper (for instance $X$ is a finitely generated group with a word metric) there are finitely many possibilities for $a_x$, so the axiom of choice need not be assumed, in the finite case it follows from the Zermelo–Fraenkel axioms.

Since $f(X) = g(A) = B$ it follows that $Y$ is contained in the $\varepsilon$–tubular neighborhood of $f(X)$. For every $x, y \in X$,

$$\mathrm{dist}_Y(f(x), f(y)) = \mathrm{dist}_Y(g(a_x), g(a_y)) \leqslant L\mathrm{dist}_X(a_x, a_y) \leqslant L(\mathrm{dist}_X(x, y) + 2\varepsilon)\,.$$

Also

$$\mathrm{dist}_Y(f(x), f(y)) = \mathrm{dist}_Y(g(a_x), g(a_y)) \geqslant \frac{1}{L}\mathrm{dist}_X(a_x, a_y) \geqslant \frac{1}{L}(\mathrm{dist}_X(x, y) - 2\varepsilon)\,.$$

Now the proposition follows from Exercise 5.14. □

Below is yet another variation on the definition of quasi-isometry, based on relations.

First, some terminology: Given a relation $R \subset X \times Y$, for $x \in X$ let $R(x)$ denote $\{(x, y) \in X \times Y : (x, y) \in R\}$. Similarly, define $R(y)$ for $y \in Y$. Let $\pi_X, \pi_Y$ denote the projections of $X \times Y$ to $X$ and $Y$ respectively.

DEFINITION 5.20. Let $X$ and $Y$ be metric spaces. A subset $R \subset X \times Y$ is called an $(L, A)$–*quasi-isometric relation* if the following conditions hold:

1. Each $x \in X$ and each $y \in Y$ are within distance $\leqslant A$ from the projection of $R$ to $X$ and $Y$, respectively.
2. For each $x, x' \in \pi_X(R)$

$$\mathrm{dist}_{Haus}(\pi_Y(R(x)), \pi_Y(R(x'))) \leqslant L\mathrm{dist}(x, x') + A.$$

3. Similarly, for each $y, y' \in \pi_Y(R)$

$$\mathrm{dist}_{Haus}(\pi_X(R(y)), \pi_X(R(y'))) \leqslant L\mathrm{dist}(y, y') + A.$$

Observe that for any $(L, A)$–quasi-isometric relation $R$, for all pair of points $x, x' \in X$, and $y \in R(x), y' \in R(x')$ we have

$$\frac{1}{L}\mathrm{dist}(x, x') - \frac{A}{L} \leqslant \mathrm{dist}(y, y') \leqslant L\mathrm{dist}(x, x') + A.$$

The same inequality holds for all pairs of points $y, y' \in Y$, and $x \in R(y), x' \in R(y')$.

In particular, by using the axiom of choice as in the proof of Proposition 5.19, if $R$ is an $(L, A)$–quasi-isometric relation between nonempty metric spaces, then it induces an $(L_1, A_1)$–quasi-isometry $X \to Y$. Conversely, every $(L, A)$–quasi-isometry is an $(L_2, A_2)$–quasi-isometric relation.

In some cases, in order to show that a map $f : X \to Y$ is a quasi-isometry, it suffices to check a weaker version of (5.3). We discuss this weaker version below.

Let $X, Y$ be topological spaces. Recall that a (continuous) map $f : X \to Y$ is called *proper* if the inverse image $f^{-1}(K)$ of each compact in $Y$ is a compact in $X$.

DEFINITION 5.21. A map $f : X \to Y$ between proper metric spaces is called *uniformly proper* if $f$ is coarse Lipschitz and there exists a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\mathrm{diam}(f^{-1}(B(y, R))) \leqslant \psi(R)$ for each $y \in Y, R \in \mathbb{R}_+$. Equivalently, there exists a proper continuous function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\mathrm{dist}(f(x), f(x')) \geqslant \eta(\mathrm{dist}(x, x'))$.

The functions $\psi$ and $\eta$ are called *upper* and *lower distortion function*, respectively.

For instance, the following function is $L$-Lipschitz, proper, but not uniformly proper:

$$f(x) = (|x|, \arctan(x)).$$

EXERCISE 5.22. 1. Composition of uniformly proper maps is again uniformly proper.

2. If $f_1, f_2 : X \to Y$ are such that $\operatorname{dist}(f_1, f_2) < \infty$ and $f_1$ is uniformly proper, then so is $f_2$.

LEMMA 5.23. *Suppose that $Y$ is a geodesic metric space, $f : X \to Y$ is a uniformly proper map whose image is $r$-dense in $Y$ for some $r < \infty$. Then $f$ is a quasi-isometry.*

PROOF. Construct a quasi-inverse to the map $f$. Given a point $y \in Y$ pick a point $\bar{f}(y) := x \in X$ such that $\operatorname{dist}(f(x), y) \leqslant r$. Let us check that $\bar{f}$ is coarse Lipschitz. Since $Y$ is a geodesic metric space it suffices to verify that there is a constant $A$ such that for all $y, y' \in Y$ with $\operatorname{dist}(y, y') \leqslant 1$, one has:

$$\operatorname{dist}(\bar{f}(y), \bar{f}(y')) \leqslant A.$$

Pick $t > 2r + 1$ which is in the image of the lower distortion function $\eta$. Then take $A \in \eta^{-1}(t)$.

It is also clear that $f, \bar{f}$ are quasi-inverse to each other. □

LEMMA 5.24. *Suppose that $G$ is a finitely generated group equipped with word metric and $G \curvearrowright X$ is a properly discontinuous isometric action on a metric space $X$. Then for every $o \in X$ the* orbit map *$f : G \to X$, $f(g) = g \cdot o$, is uniformly proper.*

PROOF. 1. Let $S$ denote the finite generating set of $G$ and set

$$L = \max_{s \in S}(d(s(o), o).$$

Then for every $g \in G$, $d_S(gs, g) = 1$, while

$$d(gs(o), g(o)) = d(s(o), o) \leqslant L.$$

Therefore, $f$ is $L$-Lipschitz.

2. Define the function

$$\eta(n) = \min\{d(go, o) : |g| = n\}.$$

Since the action $G \curvearrowright X$ is properly discontinuous,

$$\lim_{n \to \infty} \eta(n) = \infty.$$

We extend $\eta$ linearly to unit intervals $[n, n+1] \subset \mathbb{R}$ and retain the notation $\eta$ for the extension. Thus, $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and proper. By definition of the function $\eta$, for every $g \in G$,

$$d(f(g), f(1)) = d(go, o) \geqslant \eta(d(g, 1)).$$

Since $G$ acts on itself and on $X$ isometrically, it follows that

$$d(f(g), f(h)) \geqslant \eta(d(g, h)), \quad \forall g, h \in G.$$

Thus, the map $f$ is uniformly proper. □

**Coarse convergence.**

DEFINITION 5.25. Suppose that $X$ is a proper metric space. A sequence $(f_i)$ of maps $X \to Y$ is said to *coarsely uniformly converge to a map $f : X \to Y$ on compacts*, if:

There exists a number $R < \infty$ so that for every compact $K \subset X$, there exits $i_K$ so that for all $i > i_K$,

$$\forall x \in K, \quad d(f_i(x), f(x)) \leqslant R.$$

Proposition 5.26 (Coarse Arzela–Ascoli theorem.). *Fix real numbers $L, A$ and $D$ and let $X, Y$ be proper metric spaces so that $X$ admits a separated $R$-net. Let $f_i : X \to Y$ be a sequence of $(L_1, A_1)$-Lipschitz maps, so that for some points $x_0 \in X, y_0 \in Y$ we have $d(f(x_0), y_0) \leqslant D$. Then there exists a subsequence $(f_{i_k})$, and a $(L_2, A_2)$–Lipschitz map $f : X \to Y$, so that*

$$\lim_{k\to\infty}^{c} f_i = f.$$

*Furthermore, if the maps $f_i$ are $(L_1, A_1)$ quasi-isometries, then $f$ is also an $(L_3, A_3)$ quasi-isometry.*

Proof. Let $N \subset X$ be a separated net. We can assume that $x_0 \in N$. Then the restrictions $f_i|N$ are $L'$-Lipschitz maps and, by the usual Arzela-Ascoli theorem, the sequence $(f_i|N)$ subconverges (uniformly on compacts) to an $L'$ -Lipschitz map $f : N \to Y$. We extend $f$ to $X$ by the rule: For $x \in X$ pick $x' \in N$ so that $d(x, x') \leqslant R$ and set $f(x) := f(x')$. Then $f : X \to Y$ is an $(L_2, A_2)$–Lipschitz. For a metric ball $B(x_0, r) \subset X, r \geqslant R$, there exists $i_r$ so that for all $i \geqslant i_r$ and all $x \in N \cap B(x_0, r)$, we have $d(f_i(x), f(x)) \leqslant 1$. For arbitrary $x \in K$, we find $x' \in N \cap B(x_0, r + R)$ so that $d(x', x) \leqslant R$. Then

$$d(f_i(x), f(x)) \leqslant d(f_i(x'), f(x')) \leqslant L_1(R + 1) + A.$$

This proves coarse convergence. The argument for quasi-isometries is similar. $\square$

## 5.2. Group-theoretic examples of quasi-isometries

We begin by noting that given a finitely generated group $G$ endowed with a word metric the space $\mathcal{B}(G)$ is particularly easy to describe. To begin with it contains all the right translations $R_g : G \to G$, $R_g(x) = xg$ (see Remark 4.61).

Lemma 5.27. *In a finitely generated group $(G, \mathrm{dist}_S)$ endowed with a word metric, the set of maps $\mathcal{B}(G)$ is consisting of* piecewise right translations. *That is, given a map $f \in \mathcal{B}(G)$ there exist finitely many elements $h_1, \ldots, h_n$ in $G$ and a decomposition $G = T_1 \sqcup T_1 \sqcup \ldots \sqcup T_n$ such that $f$ restricted to $T_i$ coincides with $R_{h_i}$.*

Proof. Since $f \in \mathcal{B}(G)$ there exists a constant $R > 0$ such that for every $x \in G$, $\mathrm{dist}(x, f(x)) \leqslant R$. This implies that $x^{-1}f(x) \in B(1, R)$. The ball $B(1, R)$ is a finite set. We enumerate its distinct elements $\{h_1, \ldots, h_n\}$. Thus for every $x \in G$ there exists $h_i$ such that $f(x) = xh_i = R_{h_i}(x)$ for some $i \in \{1, 2, \ldots, n\}$. We define $T_i = \{x \in X \ ; \ f(x) = R_{h_i}(x)\}$. If there exists $x \in T_i \cap T_j$ then $f(x) = xh_i = xh_j$, which implies $h_i = h_j$, a contradiction. $\square$

The main example of quasi-isometry, which partly justifies the interest in such maps, is given by the following result, proved in the context of Riemannian manifolds first by A. Schwarz [**Šva55**] and, 13 years later, by J. Milnor [**Mil68**]. At the time, both were motivated by relating volume growth in universal covers of compact Riemannian manifolds and growth of their fundamental groups. Note that in the literature it is at times this theorem (stating the equivalence between the growth function of the fundamental group of a compact manifold and that of the universal cover of the manifold) that is referred to as the Milnor–Schwarz Theorem, and not Theorem 5.29 below.

In fact, it had been observed already by V.A. Efremovich in [**Efr53**] that two growth functions as above (i.e. of the volume of metric balls in the universal cover of a compact Riemannian manifold, and of the cardinality of balls in the fundamental group with a word metric) increase at the same rate.

REMARK 5.28 (What is in the name?). Schwarz is a German-Jewish name which was translated to Russian (presumably, at some point in the 19-th century) as Шварц. In the 1950-s, the AMS, in its infinite wisdom, decided to translate this name to English as Švarc. A. Schwarz himself eventually moved to the United States and is currently a colleague of the second author at University of California, Davis. See http://www.math.ucdavis.edu/∼schwarz/bion.pdf for his mathematical autobiography. The transformation

$$\text{Schwarz} \to \text{Шварц} \to \text{Švarc}$$

is a good example of a composition of a quasi-isometry and its quasi-inverse.

THEOREM 5.29 (Milnor–Schwarz). *Let* $(X, \mathrm{dist})$ *be a proper geodesic metric space (which is equivalent, by Theorem 1.29, to* $X$ *being a length metric space which is complete and locally compact) and let* $G$ *be a group acting geometrically on* $X$*. Then:*

(1) *the group* $G$ *is finitely generated;*
(2) *for any word metric* $\mathrm{dist}_w$ *on* $G$ *and any point* $x \in X$*, the map* $G \to X$ *given by* $g \mapsto gx$ *is a quasi-isometry.*

PROOF. We denote the orbit of a point $y \in X$ by $Gy$. Given a subset $A$ in $X$ we denote by $GA$ the union of all orbits $Ga$ with $a \in A$.

*Step 1: The generating set.*

As every geometric action, the action $G \curvearrowright X$ is cobounded: There exists a closed ball $\overline{B}$ of radius $D$ such that $G\overline{B} = X$. Since $X$ is proper, $\overline{B}$ is compact. Define

$$S = \{s \in G\,;\ s \neq 1\,,\ s\overline{B} \cap \overline{B} \neq \emptyset\}\,.$$

Note that $S$ is finite because the action of $G$ is proper, and that $S^{-1} = S$ by the definition of $S$.

*Step 2: Outside of the generating set.*

Now consider $\inf\{\mathrm{dist}(\overline{B}, g\overline{B})\ ;\ g \in G \setminus (S \cup \{1\})\}$. For some $g \in G \setminus (S \cup \{1\})$ the distance $\mathrm{dist}(\overline{B}, g\overline{B})$ is a positive constant $R$, by the definition of $S$. The set $H$ of elements $h \in G$ such that $\mathrm{dist}(\overline{B}, h\overline{B}) \leqslant R$ is contained in the set $\{g \in G\ ;\ g\overline{B}(x, D+R) \cap \overline{B}(x, D+R) \neq \emptyset\}$, hence it is finite. Now $\inf\{\mathrm{dist}(\overline{B}, g\overline{B})\ ;\ g \in G \setminus (S \cup \{1\})\} = \inf\{\mathrm{dist}(\overline{B}, g\overline{B})\ ;\ g \in H \setminus (S \cup \{1\})\}$ and the latter infimum is over finitely many positive numbers, therefore there exists $h_0 \in H \setminus (S \cup \{1\})$ such that $\mathrm{dist}(\overline{B}, h_0\overline{B})$ realizes that infimum, which is therefore positive. Let then $2d$ be this infimum. By definition $\mathrm{dist}(\overline{B}, g\overline{B}) < 2d$ implies that $g \in S \cup \{1\}$.

*Step 3: G is finitely generated.*

Consider a geodesic $[x, gx]$ and $k = \left\lfloor \frac{\mathrm{dist}(x,gx)}{d} \right\rfloor$. Then there exists a finite sequence of points on the geodesic $[x, gx]$, $y_0 = x, y_1, \ldots, y_k, y_{k+1} = gx$ such that

$\text{dist}(y_i, y_{i+1}) \leqslant d$ for every $i \in \{0, \ldots, k\}$. For every $i \in \{1, \ldots, k\}$ let $h_i \in G$ be such that $y_i \in h_i\overline{B}$. We take $h_0 = 1$ and $h_{k+1} = g$. As $\text{dist}(\overline{B}, h_i^{-1}h_{i+1}\overline{B}) = \text{dist}(h_i\overline{B}, h_{i+1}\overline{B}) \leqslant \text{dist}(y_i, y_{i+1}) \leqslant d$ it follows that $h_i^{-1}h_{i+1} = s_i \in S$, that is $h_{i+1} = h_i s_i$. Then $g = h_{k+1} = s_0 s_1 \cdots s_k$. We have thus proved that $G$ is generated by $S$, consequently $G$ is finitely generated.

*Step 4: The quasi-isometry.*

Since all word metrics on $G$ are bi-Lipschitz equivalent it suffices to prove (2) for the word metric $\text{dist}_S$, where $S$ is the finite generating set found as above for the chosen arbitrary point $x$. The space $X$ is contained in the $2D$–tubular neighborhood of the image $Gx$ of the map defined in (2). It therefore remains to prove that the map is a quasi-isometric embedding. The previous argument proved that $|g|_S \leqslant k+1 \leqslant \frac{1}{d}\text{dist}(x, gx) + 1$. Now let $|g|_S = m$ and let $w = s_1' \cdots s_m'$ be a word in $S$ such that $w = g$ in $G$. Then, by the triangle inequality,

$$\text{dist}(x, gx) = \text{dist}(x, s_1' \cdots s_m' x) \leqslant \text{dist}(x, s_1' x) + \text{dist}(s_1' x, s_1' s_2' x) + \ldots +$$

$$+\text{dist}(s_1' \cdots s_{m-1}' x, s_1' \cdots s_m' x) = \sum_{i=1}^{m} \text{dist}(x, s_i' x) \leqslant 2Dm = 2D|g|_S \,.$$

We have, thus, obtained that for any $g \in G$,

$$d\text{dist}_S(1, g) - d \leqslant \text{dist}(x, gx) \leqslant 2\text{dist}_S(1, g)\,.$$

Since both the word metric $\text{dist}_S$ and the metric dist on $X$ are left-invariant with respect to the action of $G$, in the above argument, $1 \in G$ can be replaced by any element $h \in G$. □

Corollary 5.30. *Given $M$ a compact connected Riemannian manifold, let $\widetilde{M}$ be its universal covering endowed with the pull-back Riemannian metric, so that the fundamental group $\pi_1(M)$ acts isometrically on $\widetilde{M}$.*

*Then the group $\pi_1(M)$ is finitely generated, and the metric space $\widetilde{M}$ is quasi-isometric to $\pi_1(M)$ with some word metric.*

A natural question to ask is whether two infinite finitely generated groups $G$ and $H$ that are quasi-isometric are also bi-Lipschitz equivalent. In fact, this question was asked in **[Gro93]**, p. 23. We discuss this question in Chapter 23.

Corollary 5.31. *Let $G$ be a finitely generated group.*

(1) *If $G_1$ is a finite index subgroup in $G$ then $G_1$ is also finitely generated; moreover the groups $G$ and $G_1$ are quasi-isometric.*
(2) *Given a finite normal subgroup $N$ in $G$, the groups $G$ and $G/N$ are quasi-isometric.*

Proof. (1) is a particular case of Theorem 5.29, with $G_2 = G$ and $X$ a Cayley graph of $G$.

(2) follows from Theorem 5.29 applied to the action of the group $G$ on a Cayley graph of the group $G/N$. □

LEMMA 5.32. *Let $(X, \mathrm{dist}_i)$, $i = 1, 2$, be proper geodesic metric spaces. Suppose that the action $G \curvearrowright X$ is geometric with respect to both metrics* $\mathrm{dist}_1, \mathrm{dist}_2$. *Then the identity map*

$$id : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$$

*is a quasi-isometry.*

PROOF. The group $G$ is finitely generated by Theorem 5.29, choose a word metric $\mathrm{dist}_G$ on $G$ corresponding to any finite generating set. Pick a point $x_0 \in X$; then the maps

$$f_i : (G, \mathrm{dist}_G) \to (X, \mathrm{dist}_i), \quad f_i(g) = g(x_0)$$

are quasi-isometries, let $\bar{f}_i$ denote their quasi-inverses. Then the map

$$id : (X, \mathrm{dist}_1) \to (X, \mathrm{dist}_2)$$

is within finite distance from the quasi-isometry $f_2 \circ \bar{f}_1$. $\square$

COROLLARY 5.33. *Let* $\mathrm{dist}_1, \mathrm{dist}_2$ *be as in Lemma 5.32. Then any geodesic $\gamma$ with respect to the metric* $\mathrm{dist}_1$ *is a quasi-geodesic with respect to the metric* $\mathrm{dist}_2$.

LEMMA 5.34. *Let $X$ be a proper geodesic metric space, $G \curvearrowright X$ is a geometric action. Suppose, in addition, that we have an isometric properly discontinuous action $G \curvearrowright X'$ on another metric space $X'$ and a $G$-equivariant coarsely Lipschitz map $f : X \to X'$. Then $f$ is uniformly proper.*

PROOF. Pick a point $p \in X$ and set $o := f(p)$. We equip $G$ with a word metric corresponding to a finite generating set $S$ of $G$; then the orbit map $\phi : g \mapsto g(p), \phi : G \to X$ is a quasi-isometry by Milnor–Schwarz theorem. We have the second orbit map $\psi : G \to X'$, $\psi(g) = g(p)$. The map $\psi$ is uniformly proper according to Lemma 5.24. We leave it to the reader to verify that

$$\mathrm{dist}(f \circ \phi, \psi) < \infty.$$

Thus, the map $f \circ \phi$ is uniformly proper as well (see Exercise 5.22). Taking $\bar{\phi} : X \to G$, a quasi-inverse to $\phi$, we see that the composition

$$f \circ \phi \circ \bar{\phi}$$

is uniformly proper too. Since

$$\mathrm{dist}(f \circ \phi \circ \bar{\phi}, f) < \infty,$$

we conclude that $f$ is also uniformly proper. $\square$

Let $G \curvearrowright X, G \curvearrowright X'$ be isometric actions and let $f : X \to X'$ be a quasi-isometric embedding. We say that $f$ is (quasi) equivariant if for every $g \in G$

$$\mathrm{dist}(g \circ f, f \circ g) \leqslant C,$$

where $C < \infty$ is independent of $G$.

LEMMA 5.35. *Suppose that $X, X'$ are proper geodesic metric spaces, $G, G'$ are groups acting geometrically on $X$ and $X'$ respectively and $\rho : G \to G'$ is an isomorphism. Then there exists a $\rho$–equivariant quasi-isometry $f : X \to X'$.*

PROOF. Pick points $x \in X$, $x' \in X'$. According to Theorem 5.29 the maps

$$G \to G \cdot x \hookrightarrow X, \quad G' \to G' \cdot x' \hookrightarrow X'$$

are quasi-isometries; therefore the map

$$f : G \cdot x \to G' \cdot x, \quad f(gx) := \rho(g)x$$

is also a quasi-isometry.

We now define a $G$–equivariant projection $\pi : X \to X$ such that $\pi(X) = G \cdot x$, and $\pi$ is at bounded distance from the identity map on $X$. We start with a closed ball $\overline{B}$ in $X$ such that $G\overline{B} = X$. Using the axiom of choice, pick a subset $\Delta$ of $\overline{B}$ intersecting each orbit of $G$ in exactly one point. For every $y \in X$, there exists a unique $g \in G$ such that $gy \in \Delta$. Define $\pi(y) = g^{-1}x$. Clearly $\mathrm{dist}_X(y, \pi(y)) = \mathrm{dist}(gy, x) \leqslant \mathrm{diam}(\overline{B})$.

Then the map $\tilde{f}$ below is a $\rho$–equivariant quasi-isometry:

$$\tilde{f} : X \to X', \tilde{f} = f \circ \pi,$$

since $\tilde{f}$ is a composition of two equivariant quasi-isometries. □

COROLLARY 5.36. *Two virtually isomorphic (VI) finitely generated groups are quasi-isometric (QI).*

PROOF. Let $G$ be a finitely generated group, $H < G$ a finite index subgroup and $F \lhd H$ a finite normal subgroup. According to Corollary 5.31, $G$ is QI to $H/F$.

Recall now that two groups $G_1, G_2$ are virtually isomorphic if there exist finite index subgroups $H_i < G_i$ and finite normal subgroups $F_i \lhd H_i$, $i = 1, 2$, so that $H_1/F_1 \cong H_2/F_2$. Since $G_i$ is QI to $H_i/F_i$, we conclude that $G_1$ is QI to $G_2$. □

The next example shows that VI is not equivalent to QI.

EXAMPLE 5.37. Let $A$ be a matrix diagonalizable over $\mathbb{R}$ in $SL(2, \mathbb{Z})$ so that $A^2 \neq I$. Thus the eigenvalues $\lambda, \lambda^{-1}$ of $A$ have absolute value $\neq 1$. We will use the notation $Hyp(2, \mathbb{Z})$ for the set of such matrices. Define the action of $\mathbb{Z}$ on $\mathbb{Z}^2$ so that the generator $1 \in \mathbb{Z}$ acts by the automorphism given by $A$. Let $G_A$ denote the associated semidirect product $G_A := \mathbb{Z}^2 \rtimes_A \mathbb{Z}$. We leave it to the reader to verify that $\mathbb{Z}^2$ is a unique maximal normal abelian subgroup in $G_A$. By diagonalizing the matrix $A$, we see that the group $G_A$ embeds as a discrete cocompact subgroup in the Lie group

$$Sol_3 = \mathbb{R}^2 \rtimes_D \mathbb{R}$$

where

$$D(t) = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, t \in \mathbb{R}.$$

In particular, $G_A$ is torsion-free. The group $Sol_3$ has its left-invariant Riemannian metric, so $G_A$ acts isometrically on $Sol_3$ regarded as a metric space. Hence, every group $G_A$ as above is QI to $Sol_3$. We now construct two groups $G_A, G_B$ of the above type which are not VI to each other. Pick two matrices $A, B \in Hyp(2, \mathbb{Z})$ so that for every $n, m \in \mathbb{Z} \setminus \{0\}$, $A^n$ is not conjugate to $B^m$. For instance, take

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}.$$

(The above property of the powers of $A$ and $B$ follows by considering the eigenvalues of $A$ and $B$ and observing that the fields they generate are different quadratic extensions of $\mathbb{Q}$.) The group $G_A$ is QI to $G_B$ since they are both QI to $Sol_3$. Let us check that $G_A$ is not VI to $G_B$. First, since both $G_A, G_B$ are torsion-free, it suffices to show that they are not commensurable, i.e., do not contain isomorphic finite index subgroups. Let $H = H_A$ be a finite index subgroup in $G_A$. Then $H$ intersects the normal rank 2 abelian subgroup of $G_A$ along a rank 2 abelian subgroup $L_A$. The image of $H$ under the quotient homomorphism $G_A \to G_A/\mathbb{Z}^2 = \mathbb{Z}$ has to be an infinite cyclic subgroup, generated by some $n \in \mathbb{N}$. Therefore, $H_A$ is isomorphic to $\mathbb{Z}^2 \rtimes_{A^n} \mathbb{Z}$. For the same reason, $H_B \cong \mathbb{Z}^2 \rtimes_{B^m} \mathbb{Z}$. It is easy to see that an isomorphism $H_A \to H_B$ would have to carry $L_A$ isomorphically to $L_B$. However, this would imply that $A^n$ is conjugate to $B^m$. Contradiction.

Example 5.38. Another example where QI does not imply VI is as follows. Let $S$ be a closed oriented surface of genus $n \geqslant 2$. Let $G_1 = \pi_1(S) \times \mathbb{Z}$. Let $M$ be the total space of the unit tangent bundle $UT(S)$ of $S$. Then the fundamental group $G_2 = \pi_1(M)$ is a nontrivial central extension of $\pi_1(S)$:

$$1 \to \mathbb{Z} \to G_2 \to \pi_1(S) \to 1,$$

$$G_2 = \left\langle a_1, b_1, \ldots, a_n, b_n, t | [a_1, b_1] \cdots [a_n, b_n] t^{2n-2}, [a_i, t], [b_i, t], i = 1, \ldots, n \right\rangle.$$

We leave it to the reader to check that passing to any finite index subgroup in $G_2$ does not make it a trivial central extension of the fundamental group of a hyperbolic surface. On the other hand, since $\pi_1(S)$ is hyperbolic, the groups $G_1$ and $G_2$ are quasi-isometric, see section 9.14.

Another example of quasi-isometry is the following.

Example 5.39. All non-abelian free groups of finite rank are quasi-isometric to each other.

Proof. We present two proofs: One is algebraic and the other is geometric.

1. **Algebraic proof.** We claim that all free groups $F_n, 2 \leqslant n < \infty$ groups are commensurable. Indeed, let $a, b$ denote the generators of $F_2$. Define the epimorphism $\rho_m : F_2 \to \mathbb{Z}_m$ by sending $a$ to 1 and $b$ to 0. Then the kernel $K_m$ of $\rho_m$ has index $m$ in $F_2$. Then $K_m$ is a finitely generated free group $F$. In order to compute the rank of $F$, it is convenient to argue topologically. Let $R$ be a finite graph with the (free) fundamental group $\pi_1(R)$. Then $\chi(R) = 1 - b_1(R) = 1 - \mathrm{rank}\,(\pi_1(R))$. Let $R_2$ be such a graph for $F_2$, then $\chi(R_2) = 1 - 2 = -1$. Let $R \to R_2$ be the $m$-fold covering corresponding to the inclusion $F_n \hookrightarrow F_2$. Then $\chi(R) = m\chi(R_2) = -m$. Hence, $\mathrm{rank}\,(F) = 1 - \chi(R) = 1 + m$. Thus, for every $n = 1 + m \geqslant 2$, we have a finite-index inclusion $F_n \hookrightarrow F_2$. Since commensurability is a transitive relation which implies quasi-isometry, the claim follows.

2. **Geometric proof.** The Cayley graph of $F_n$ with respect to a set of $n$ generators and their inverses is the regular simplicial tree of valence $2n$.

We claim that all regular simplicial trees of valence at least 3 are quasi-isometric. We denote by $\mathcal{T}_k$ the regular simplicial tree of valence $k$ and we show that $\mathcal{T}_3$ is quasi-isometric to $\mathcal{T}_k$ for every $k \geqslant 4$.

We define a piecewise-linear map $\mathfrak{q} : \mathcal{T}_3 \to \mathcal{T}_k$ as in Figure 5.1: Sending all edges drawn in thin lines isometrically onto edges and collapsing each edge-path of length $k - 3$ (drawn in thick lines) to a single vertex. The map $\mathfrak{q}$ thus defined is surjective and it satisfies the inequality

$$\frac{1}{k-2}\,\mathrm{dist}(x, y) - 1 \leqslant \mathrm{dist}(\mathfrak{q}(x), \mathfrak{q}(y)) \leqslant \mathrm{dist}(x, y)\,.$$

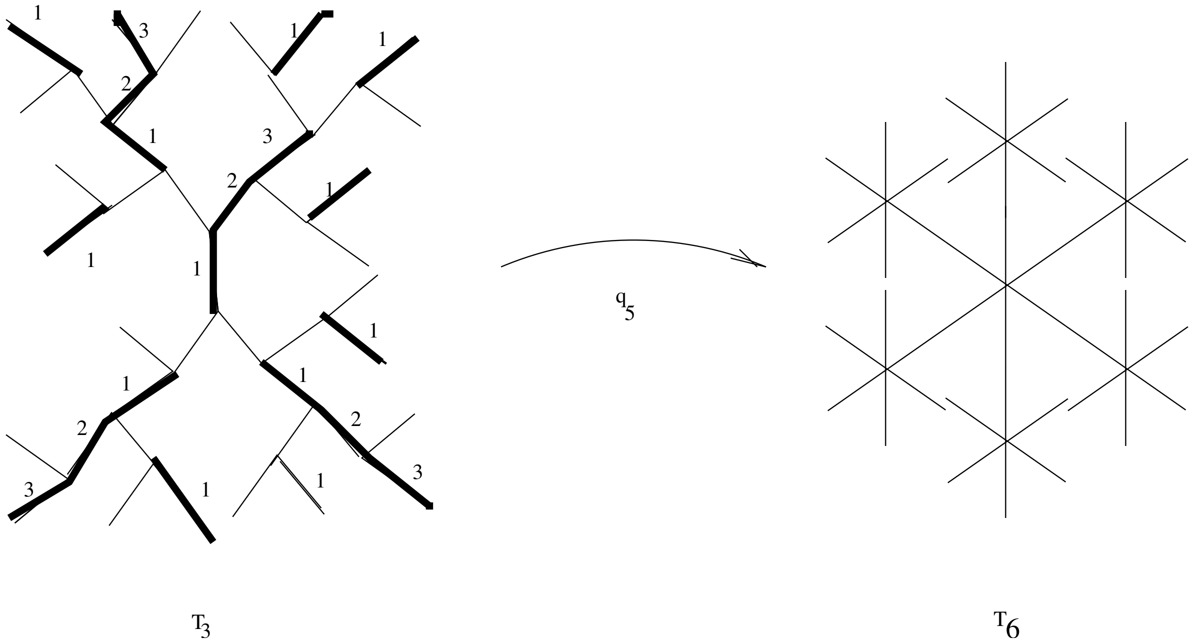



FIGURE 5.1. All regular simplicial trees are quasi-isometric.

□

### 5.3. Metric version of the Milnor–Schwarz Theorem

In the case of a Riemannian manifold, or more generally a metric space, without a geometric action of a group, one can still use a purely metric argument and create a discretization of the manifold/space, that is a simplicial graph quasi-isometric to the manifold. We begin with a few simple observations.

LEMMA 5.40. *Let $X$ and $Y$ be two discrete metric spaces that are bi-Lipschitz equivalent. If $X$ is uniformly discrete then so is $Y$.*

PROOF. Assume $f : X \to Y$ is an $L$–bi-Lipschitz bijection,, where $L \geqslant 1$, and assume that $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a function such that for every $r > 0$ every closed ball $\overline{B}(x, r)$ in $X$ contains at most $\phi(r)$ points. Every closed ball $\overline{B}(y, R)$ in $Y$ is in 1-to-1 correspondence with a subset of $B(f^{-1}(y), LR)$, whence it contains at most $\phi(LR)$ points. □

*Notation*: Let $A$ be a subset in a metric space. We denote by $\mathcal{G}_\kappa(A)$ the simplicial graph with set of vertices $A$ and set of edges

$$\{(a_1, a_2) \mid a_1, a_2 \in A,\, 0 < \mathrm{dist}(a_1, a_2) \leqslant \kappa\}\,.$$

In other words, $\mathcal{G}_\kappa(A)$ is the 1-skeleton of the Rips complex $\mathrm{Rips}_\kappa(A)$.

As usual, we will equip $\mathcal{G}_\kappa(A)$ with the standard metric.

THEOREM 5.41. (1) *Let $(X, \mathrm{dist})$ be a proper geodesic metric space (equivalently a complete, locally compact length metric space, see Theorem 1.29). Let $N$ be an $\varepsilon$–separated $\delta$–net, where $0 < \varepsilon < 2\delta < 1$ and let $\mathcal{G}$ be the metric graph $\mathcal{G}_{8\delta}(N)$. Then the metric space $(X, \mathrm{dist})$ and the graph $\mathcal{G}$ are quasi-isometric. More precisely, for all $x, y \in N$ we have that*

$$\frac{1}{8\delta}\mathrm{dist}_X(x,y) \leqslant \mathrm{dist}_{\mathcal{G}}(x,y) \leqslant \frac{3}{\varepsilon}\mathrm{dist}_X(x,y)\,. \tag{5.4}$$

(2) *If, moreover, $(X, \mathrm{dist})$ is either a complete Riemannian manifold of bounded geometry or a metric simplicial complex of bounded geometry, then $\mathcal{G}$ is a graph of bounded geometry.*

PROOF. (1) Let $x, y$ be two fixed points in $N$. If $\mathrm{dist}_X(x,y) \leqslant 8\delta$ then, by construction, $\mathrm{dist}_{\mathcal{G}}(x,y) = 1$ and both inequalities in (5.4) hold. Let us suppose that $\mathrm{dist}_X(x,y) > 8\delta$.

The distance $\mathrm{dist}_{\mathcal{G}}(x,y)$ is the length $s$ of an edge-path $e_1e_2\dots e_s$, where $x$ is the initial vertex of $e_1$ and $y$ is the terminal vertex of $e_s$. It follows that

$$\mathrm{dist}_{\mathcal{G}}(x,y) = s \geqslant \frac{1}{8\delta}\mathrm{dist}_X(x,y)\,.$$

The distance $\mathrm{dist}_X(x,y)$ is the length of a geodesic $\mathfrak{c}\colon [0, \mathrm{dist}_X(x,y)] \to X$. Let

$$t_0 = 0, t_1, t_2, \dots, t_m = \mathrm{dist}_X(x,y)$$

be a sequence of numbers in $[0, \mathrm{dist}_X(x,y)]$ such that $5\delta \leqslant t_{i+1} - t_i \leqslant 6\delta$, for every $i \in \{0, 1, \dots, m-1\}$.

Let $x_i = \mathfrak{c}(t_i)$, $i \in \{0, 1, 2, \dots, m\}$. For every $i \in \{0, 1, 2, \dots, m\}$ there exists $w_i \in N$ such that $\mathrm{dist}_X(x_i, w_i) \leqslant \delta$. We note that $w_0 = x, w_m = y$. The choice of $t_i$ implies that

$$3\delta \leqslant \mathrm{dist}_X(w_i, w_{i+1}) \leqslant 8\delta\,, \quad \text{for every } i \in \{0, \dots, m-1\}$$

In particular:

- $w_i$ and $w_{i+1}$ are the endpoints of an edge in $\mathcal{G}$, for every $i \in \{0, \dots, m-1\}$;
- $\mathrm{dist}_X(x_i, x_{i+1}) \geqslant \mathrm{dist}(w_i, w_{i+1}) - 2\delta \geqslant \mathrm{dist}(w_i, w_{i+1}) - \frac{2}{3}\mathrm{dist}(w_i, w_{i+1}) = \frac{1}{3}\mathrm{dist}(w_i, w_{i+1})$.

We can then write

$$\mathrm{dist}_X(x,y) = \sum_{i=0}^{m-1}\mathrm{dist}_X(x_i, x_{i+1}) \geqslant \frac{1}{3}\sum_{i=0}^{m-1}\mathrm{dist}(w_i, w_{i+1}) \geqslant \frac{\varepsilon}{3}m \geqslant \frac{\varepsilon}{3}\mathrm{dist}_{\mathcal{G}}(x,y)\,. \tag{5.5}$$

(2) According to the discussion following Definition 2.60, the graph $\mathcal{G}$ has bounded geometry if and only if its set of vertices with the induced simplicial distance is uniformly discrete. Lemma 5.40 implies that it suffices to show that the set of vertices of $\mathcal{G}$ (i.e. the net $N$) with the metric induced from $X$ is uniformly discrete.

When $X$ is a Riemannian manifold, this follows from Lemma 2.58. When $X$ is a simplicial complex this follows from the fact that the set of vertices of $X$ is uniformly discrete. □

Note that one can also discretize a Riemannian manifold $M$ (i.e. of replace $M$ by a quasi-isometric simplicial complex) using Theorem 2.62, which implies:

THEOREM 5.42. *Every Riemannian manifold $M$ of bounded geometry is quasi-isometric to a simplicial complex homeomorphic to $M$.*

## 5.4. Metric filling functions

In this section we define notions of loops, filling disks and minimal filling area in the setting of geodesic metric spaces, following [**Gro93**]. Let $X$ be a geodesic metric space and $\delta > 0$ be a fixed constant. In this present setting of isoperimetric inequalities, by *loops* we always mean Lipschitz maps $\mathfrak{c}$ from the unit circle $\mathbb{S}^1$ to $X$. We will use the notation $\ell_X$ for the length of an arc in $X$.

A *$\delta$-loop* in $X$ is a triangulated circle $S^1$ together with a (Lipschitz) map $\mathfrak{c} : S^1 \to X$, so that for $\ell_X(c(e)) \leqslant \delta$ for every edge $e$ of the triangulation.

A *filling disk of* $\mathfrak{c}$ is a pair consisting of a triangulation $\mathcal{D}$ of the 2-dimensional unit disk $\mathbb{D}^2$ extending the triangulation of its boundary circle $S^1$ and a map

$$\mathfrak{d} : \mathcal{D}^{(0)} \to X$$

extending the map $\mathfrak{c}$ restricted to the set of boundary vertices. Here $\mathcal{D}^{(0)}$ is the set of vertices in $\mathcal{D}$. Sometimes by abuse of language we call the image of the map $\mathfrak{d}$ also *filling disk of* $\mathfrak{c}$.

We next extend the map $\mathfrak{d}$ to the 1-skeleton of $\mathcal{D}$. For every edge $e$ of $\mathcal{D}$ (not contained in the boundary circle) we pick a geodesic connecting the images of the end-points of $e$ under $\mathfrak{d}$. For every boundary edge $e$ of the 2-disk we use the restriction of the map $\mathfrak{d}$ to $e$ in order to connect the images of the vertices. The triangles in $X$ thus obtained are called *bricks*. The *length of a brick* is the sum of the lengths of its edges. The *mesh of a filling disk* is the maximum of the lengths of its bricks. By abusing the notation, we will refer to this extension of $\mathfrak{d}$ to $\mathcal{D}^{(1)}$ as a $\delta$-filling disk as well.

A *$\delta$-filling disk of* $\mathfrak{c}$ is a filling disk with mesh at most $\delta$. The combinatorial area of such a disk is just the number of 2-simplices in the triangulation of $D^2$.

DEFINITION 5.43. The *$\delta$-filling area of* $\mathfrak{c}$ is the minimal combinatorial area of a $\delta$-filling disk of $\mathfrak{c}$. We will use the double notation $\mathrm{Ar}_\delta(\mathfrak{c}) = P(\mathfrak{c}, \delta)$ for the $\delta$-filling area.

Note that $\mathrm{Ar}_\delta$ is a function defined on the set $\Omega$ of loops and taking values in $\mathbb{Z}_+$.

We, likewise, define the *$\delta$-filling radius function* as

$$r_\delta : \Omega \to \mathbb{R}_+,$$

$$r_\delta(\mathfrak{c}) = \inf \left\{ \max_{x \in \mathcal{D}^{(0)}} \mathrm{dist}_X(\mathfrak{d}(x), \mathfrak{c}(\mathbb{S}^1)) \; ; \; \mathfrak{d} \text{ is a } \delta - \text{filling disk of the loop } \mathfrak{c} \right\}.$$

Both functions depend on the parameter $\delta$, and may take infinite values. In order to obtain finite valued functions, we add the hypothesis that there exists a sufficiently large $\mu$ so that for all $\delta \geqslant \mu$, every loop has a $\delta$-filling disk. Such spaces will be called *$\mu$-simply connected*.

EXERCISE 5.44. Show that a geodesic metric space is coarsely simply-connected in the sense of Definition 6.13 if and only if $X$ is $\mu$-simply connected for some $\mu$.

In the sequel we only deal with $\mu$-simply connected metric spaces. We occasionally omit to recall this hypothesis.

We can now define the *$\delta$-filling function* $Ar_\delta : \mathbb{R}_+ \to \mathbb{Z}_+$, $Ar_\delta(\ell) :=$ the maximal area needed to fill a loop of length at most $\ell$. For our convenience, we use in parallel the notation $P(\ell, \delta)$ for this function. We will also use the name *$\delta$-isoperimetric function* for $Ar_\delta(\ell)$.

To get a better feel for the $\delta$-filling function, let us relate $Ar_\delta$ with the usual area function in the case $X = \mathbb{R}^2$. Recall (see **[Fed69]**) that every loop $c$ in $\mathbb{R}^2$ satisfies the *Euclidean isoperimetric inequality*

$$4\pi A(c) \leqslant \ell^2(c), \tag{5.6}$$

where the equality is realized in the case when $c$ is a round circle. Suppose that $\mathfrak{c}$ is a loop in $\mathbb{R}^2$ and $\mathfrak{d} : \mathcal{D}^{(1)} \to X$ is a $\delta$-filling disk for $\mathfrak{c}$. Then $\mathfrak{d}$ extends to a map $\mathfrak{d} : D^2 \to \mathbb{R}^2$, where we extend the restriction of $\mathfrak{d}$ to each 2-simplex $\sigma$ by the least area disk bounded by the loop $\mathfrak{d}|\partial\sigma$. In view of the isoperimetric inequality (5.6) he resulting map $\mathfrak{d}$ will have area

$$Area(\mathfrak{d}) \leqslant \sum_{\sigma} \ell(\mathfrak{d}_{\partial\sigma}) \leqslant Ar_\delta(\mathfrak{d}) \frac{\delta^2}{4\pi}, \tag{5.7}$$

where the sum is taken over all 2-simplices in $\mathcal{D}$. In general, it is impossible to estimate $Ar_\delta$ from above, however, one can do so for carefully chosen maps $\mathfrak{d}$. Namely, we will think of the map $\mathfrak{c}$ as a function $f$ of the angular coordinate $\theta \in [0, 2\pi]$. Suppose that $f$ is $L$-Lipschitz. Choose coordinates in $\mathbb{R}^2$ so that the origin is $\mathfrak{c}(0)$ and define a function

$$F(r, \theta) = r\mathfrak{c}(\theta).$$

Then $F$ is $L' = \sqrt{1 + 4\pi^2} L$-Lipschitz. Subdivide the rectangle $[0, 1] \times [0, 2\pi]$ (the domain of $F$) in subrectangles of width $\epsilon_1$ and height $\epsilon_2$ and draw the diagonal in each rectangle. Then the restriction of $F$ to the boundary of each 2-simplex of the resulting triangulation is a $2L'(\epsilon_1 + \epsilon_2)$-brick. Therefore, in order to ensure that $F$ is a $\delta$-filling of the map $f$, we take:

$$n = \lceil \frac{4L'}{\delta} \rceil, m = \lceil \frac{8\pi L'}{\delta} \rceil.$$

Hence, $Ar_\delta(\mathfrak{c})$ is at most

$$2nm \leqslant \frac{1}{\delta^2} 32(L')^2 = \frac{1 + 4\pi^2}{\delta^2} L^2.$$

In terms of the length $\ell$ of $\mathfrak{c}$,

$$Ar_\delta(\mathfrak{c}) \leqslant \frac{1 + 4\pi^2}{\delta^2 4\pi^2} \ell^2 \leqslant \frac{2}{\delta^2} \ell^2.$$

Likewise, using the radius function we define the *filling radius function* as

$$r : \mathbb{R}_+ \to \mathbb{R}_+,\ r(\ell) = \sup\{r(\mathfrak{c})\ ;\ \mathfrak{c} \text{ loop of length } \leqslant \ell\}\ .$$

Two filling functions corresponding to different $\delta$'s for a metric space, or, more generally, for two quasi-isometric metric spaces, satisfy a certain equivalence relation.

In a geodesic metric space $X$ that is $\mu$-simply connected, if $\mu \leqslant \delta_1 \leqslant \delta_2$ then one can easily see, by considering partitions of bricks of length at most $\delta_2$ into bricks of length at most $\delta_1$ that

$$A_{\delta_1}(\ell) \leqslant A_{\delta_2}(\ell) \leqslant A_{\delta_2}(\delta_1)\, A_{\delta_1}(\ell)$$

and that

$$r_{\delta_1}(\ell) \leqslant r_{\delta_2}(\ell) \leqslant r_{\delta_2}(\delta_1)\, r_{\delta_1}(\ell)\,.$$

EXERCISE 5.45. (1) Prove that if two geodesic metric spaces $X_i$, $i = 1, 2$, are coarsely simply connected and quasi-isometric, then their filling functions, respectively their filling radii, are asymptotically equal. Hint: Suppose that $f : X_1 \to X_2$ is an $(L, A)$-quasi-isometry. Start with a 1-loop $\mathfrak{c}_1 : S^1 \to X_1$, then fill-in $\mathfrak{c}_2 = f \circ \mathfrak{c}$ in $X_2$ using a $\delta_2$-disk $\mathcal{D}_2$, where $\delta_2 = L + A$; then compose $\mathcal{D}_2$ with quasi-inverse to $f$ in order to fill-in the original loop $\mathfrak{c}_1$ using a $\delta_1$-disk $\mathcal{D}_1$, where $\delta_1 = L\delta_2 + A$. Now, argue that $Ar_{\delta_1}(\mathfrak{c}) \leqslant Ar_{\delta_2}(\mathfrak{c}_2)$).

(2) Prove that for a finitely presented group $G$ the metric filling function for an arbitrary Cayley graph $\Gamma_G$ and the Dehn function have the same order. Hint: It is clear that $Dehn(\ell) \leqslant Ar_\mu(\ell)$, where $\mu$ is the length of the longest relators of $G$. Use optimal Van Kampen diagrams for a loop $\mathfrak{c}$ of length $\ell$, to construct $\mu$-filling disks in $\Gamma_G$ whose area is $\leqslant Dehn(\ell) + 4(\ell + 1)$.

Note that one can also define Riemannian filling functions in the context of simply-connected Riemannian manifolds $M$: Given a Lipschitz loop $c$ in $M$ one defines $Area(c)$ to be the least area of a disk in $M$ bounding $c$. Then the *isoperimetric function* $IP_M(\ell)$ of the manifold $M$ is

$$IP_M(\ell) = \sup\{A(c) : length(c) \leqslant \ell\}$$

where $\ell(c)$ is the length of $c$. Then, assuming that $M$ admits a geometric action of a group $G$, we have

$$Ar_\delta(\ell) \approx Dehn(\ell) \approx IP_M(\ell),$$

see [**BT02**].

The order of the filling function of a metric space $X$ is also called *the filling order of* $X$. Besides the fact that it is a quasi-isometry invariant, the interest of the filling order comes from the following result, a proof of which can be found for instance in [**Ger93a**].

PROPOSITION 5.46. *In a finitely presented group $G$ the following statements are equivalent.*

($S_1$) *$G$ has solvable word problem.*
($S_2$) *the Dehn function of $G$ is recursive.*
($S_3$) *the filling radius function of $G$ is recursive.*

If in a metric space $X$ the filling function $Ar(\ell)$ satisfies $Ar(\ell) \prec \ell$ or $\ell^2$ or $e^\ell$, it is said that the space $X$ *satisfies a linear, quadratic or exponential isoperimetric inequality*.

**Filling area in Rips complex.** Suppose that $X$ is $\mu$-connected. Instead of filling closed curves in $X$ by $\delta$-disks, one can fill in polygonal loops in $P = Rips_\delta(X)$ with simplicial disks. Let $\mathfrak{c}$ be a $\delta$-loop in $X$. Then we have a triangulation of the circle $S^1$ so that $\mathrm{diam}(\mathfrak{c}(\partial e)) \leqslant \delta$ for every edge $e$ of the triangulation. Thus, we define a loop $\mathfrak{c}_\delta$ in $P$ by replacing arcs $\mathfrak{c}(\partial e)$ with edges of $P$ connecting the end-points of these arcs. Then

$$\delta c - length(\mathfrak{c}_\delta) = \delta length(\mathfrak{c}_\delta) \geqslant \text{length}(\mathfrak{c})$$

since every edge of $P$ has unit length. It is clear that for $\delta > 0$ the map

$$\{\text{loops in } X \text{ of length} \leqslant \ell\} \to \{\text{loops in } P \text{ of length} \leqslant \frac{\ell}{\delta}\}$$

$$c \mapsto c_\delta$$

is surjective. Furthermore, every $\delta$-disk $\mathcal{D}$ which fills in $\mathfrak{c}$ yields a simplicial map $\mathcal{D}_\delta : D^2 \to P$ which is an extension of $\mathfrak{c}_\delta$: The maps $\mathcal{D}$ and $\mathcal{D}_\delta$ agree on the vertices of the triangulation of $D^2$, and for every 2-simplex $\sigma$ in $D^2$, the map $\mathcal{D}_\delta|\sigma$ is the canonical linear extension of $\mathcal{D}|\sigma^{(0)}$ to the simplex (of dimension $\leqslant 2$) in $P$ spanned by the vertices $\mathcal{D}(\sigma^{(0)})$. Furthermore, area is preserved by this construction:

$$cArea(\mathcal{D}_\delta) = Ar_\delta(\mathcal{D}).$$

This construction produces all simplicial disks in $P$ bounding $\mathfrak{c}_\delta$ and we obtain

$$cArea(\mathfrak{c}_\delta) = Ar_\delta(\mathfrak{c}).$$

Summarizing all this, we obtain

$$A_{Rips_\delta(X)}(\ell) = Ar_\delta(\frac{\ell}{\delta}).$$

The same argument applies to the filling radius and we obtain:

OBSERVATION 5.47. Studying filling area and filling radius functions in $X$ (up to the equivalence relation $\approx$) is equivalent to studying combinatorial filling area and filling radius functions in $Rips_\delta(X)$.

**Besikovitch inequality.** The following proposition relates filling areas of curvilinear quadrilaterals in $X$ to the product among of separation of their sides.

PROPOSITION 5.48 (The quadrangle or Besikovitch inequality). *Let $X$ be a $\mu$–simply connected geodesic metric space and let $\delta \geqslant \mu$.*

*Consider a loop $\mathfrak{c} \in \Omega_X$ and its decomposition $\mathfrak{c}(\mathbb{S}^1) = \alpha_1 \cup \alpha_2 \cup \alpha_3 \cup \alpha_4$ into four consecutive paths. Then, with the notation $d_1 = \text{dist}(\alpha_1, \alpha_3)$ and $d_2 = \text{dist}(\alpha_2, \alpha_4)$ we have that*

$$\text{Ar}_\delta(\mathfrak{c}) \geqslant \frac{2\pi}{\delta^2} d_1 d_2 \,.$$

PROOF. Let $\mathfrak{d} : \mathcal{D}^{(1)} \to X$ be a filling disk of $\mathfrak{c}$ realizing the filling area. Consider a map $\beta : X \to \mathbb{R}^2$ defined by

$$\beta(x) = (\text{dist}(x, \alpha_1)\,,\, \text{dist}(x, \alpha_2))\,.$$

Since each of its components is a 1–Lipschitz map, the map $\beta$ is $\sqrt{2}$–Lipschitz. The image $\beta(\alpha_1)$ is a vertical segment connecting the origin to a point $(0, y_1)$, with $y_1 \geqslant d_2$, while $\beta(\alpha_2)$ is a horizontal segment connecting the origin to a point $(x_2, 0)$, with $x_2 \geqslant d_1$ . Similarly, the image $\beta(\alpha_3)$ is a path to the right of the vertical line $x = d_1$ and $\beta(\alpha_4)$ another path above the horizontal line $y = d_2$. Thus, the rectangle $R$ with the vertices $(0,0), (d_1, 0), (d_1, d_2), (0, d_2)$ is separated from infinity by the curve $\beta\mathfrak{c}(S^1)$ (see Figure 5.2). In particular, the image of any extension $F$ of $\beta \circ \mathfrak{d}$

to $D^2$ contains the rectangle $R$. Thus, $A(F) \geqslant A(R) = d_1 d_2$, hence, by inequality (5.7),

$$d_1 d_2 \leqslant \frac{2\delta^2}{4\pi} Ar_{\sqrt{2}\delta}(\beta \circ c).$$

Furthermore, since $\beta$ is $\sqrt{2}$-Lipschitz,

$$Ar_{\sqrt{2}\delta}(\beta \circ \mathfrak{c}) \leqslant 2Ar_\delta(\mathfrak{c}).$$

Putting this all together, we get

$$Ar_\delta(\mathfrak{c}) \geqslant \frac{\pi}{\delta^2} d_1 d_2$$

as required. □

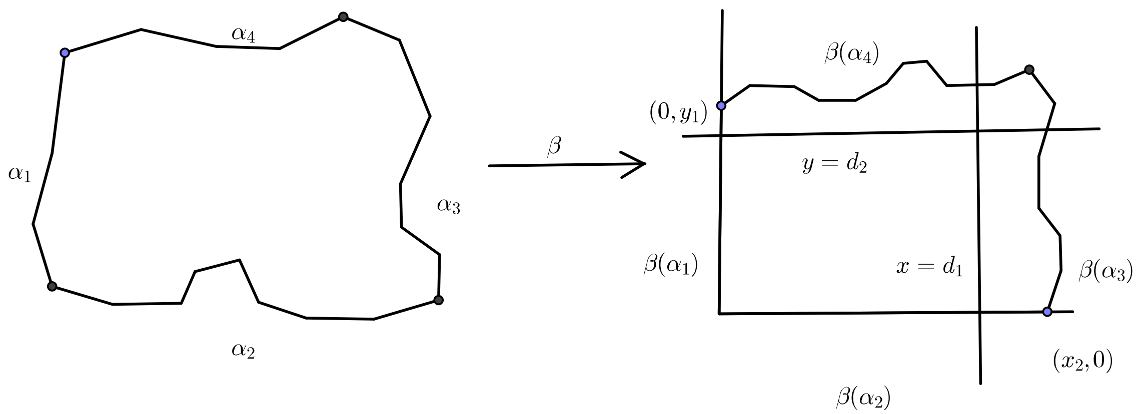



FIGURE 5.2. The map $\beta$.

Besikovitch's inequality generalizes from curvilinear quadrilaterals to curvilinear triangles: This generalization below is has interesting applications to $\delta$-hyperbolic spaces. We first need a definition which would generalize the condition of separation of the opposite edges of a curvilinear quadrilateral.

DEFINITION 5.49. Given a topological triangle $T$, i.e. a loop $\mathfrak{c}$ composed of a concatenation of three paths $\tau_1, \tau_2, \tau_3$, *the minimal size* (minsize) of $T$ is defined as

$$\text{minsize}(T) = \inf\{\text{diam}\{y_1, y_2, y_3\} \; ; \; y_i \in \tau_i, \, i = 1, 2, 3\} \; .$$

PROPOSITION 5.50 (Minsize inequality). *Let $X$ be a $\mu$–simply connected geodesic metric space and let $\delta \geqslant \mu$.*

*Given a topological triangle $T \in \Omega$, we have that*

$$\text{Ar}_\delta(\mathfrak{c}) \geqslant \frac{2\pi}{\delta^2} [\text{minsize}(T)]^2 \, .$$

PROOF. As before, define a $\sqrt{2}$–Lipschitz map $\beta : X \to \mathbb{R}^2$,

$$\beta(x) = (\beta_1(x), \beta_2(x)) = (\text{dist}(x, \tau_1), \quad \text{dist}(x, \tau_2))$$

and note that, as in the proof of Besikovitch's inequality, $\beta$ maps $\tau_1, \tau_2$ to coordinate segments, while the restriction of $\beta$ to $\tau_3$ satisfies:

$$\min(\beta_1(x), \beta_2(x)) \geqslant m,$$

where $m = \text{minsize}(T)$. Therefore, the loop $\beta \circ \mathfrak{c}$ separates from infinity the square $Q$ with the vertices $(0, 0), (m, 0), (m, m), (0, m)$. Then, as before,

$$m^2 \leqslant \frac{\delta^2}{2\pi} \text{Ar}(\mathfrak{c})$$

and claim follows. □

The Dehn function/area filling function can be generalized to higher dimensions and $n$-Dehn functions, which give information about the way to fill topological spheres $\mathbb{S}^n$ with topological balls $\mathbf{B}^{n+1}$ ([**Gro93**, Chapter 5], [**ECH$^+$92**, Chapter 10], [**Pap00**]). The following result was proven by P. Papasoglou:

THEOREM 5.51 (P. Papasoglou, [**Pap00**]). *The second Dehn function of a group of type* $\mathbf{F}_3$ *is bounded by a recursive function.*

The condition $\mathbf{FP}_3$ is a 3-dimensional version of the condition of finite presentability of a group: A group $G$ is of type $\mathbf{F}_3$ if there exists a finite simplicial complex $K$ with $G = \pi_1(K)$ and $\pi_2(K) = 0$. A *basic sphere* in the 2-dimensional skeleton of $K$ is the boundary of an oriented 3-simplex together with a path connecting its vertex to a base-point $v$ in $K$.

This theorem represents a striking contrast with the fact that there are finitely-presented groups with unsolvable word problem and, hence, Dehn function which is not bounded above by any recursive function.

The idea of the proof of Theorem 5.51 is to produce an algorithm which, given $n \in \mathbb{N}$, finds in finite time an upper bound on the number $N$ of basic spheres $\sigma_j$, so that (in $\pi_2(K, v)$)

$$\sum_{i=1}^{N} \sigma_j = \sigma,$$

where $\sigma$ is a spherical 2-cycle in $K$ which consists of at most $n$ 2-dimensional simplices. The algorithm only gives a recursive bound of the second Dehn function, because the filling found by it might be not the smallest possible.

The above algorithm does not work for the ordinary Dehn function since it would require one to recognize which loops in $K$ are homotopically trivial.

## 5.5. Summary of various notions of volume and area

(1) $Vol(f)$ is the Riemannian volume of a map; geometric volume of a smooth map of regular cell-complexes. For $n = 2$, $Vol(f) = Area(f)$.
(2) Combinatorial volume: $cVol_n(f)$, the number of $n$-simplices in the domain not collapsed by $f$. For $n = 2$, $cVol_2(f) = cArea(f)$.
(3) Simplicial volume: $sVol_n(f)$ is the number of $n$-simplices in the domain of $f$.

(4) Combinatorial area: $A(w)$, minimal filling combinatorial area for a trivial word $w$ (algebraic area); algebraically speaking, it equals area of the minimal van Kampen diagram with the given boundary loop $w$.
(5) Coarse area: $\mathrm{A}r_\delta(\mathfrak{c})$, the $\delta$-filling area of a $\delta$-loop $\mathfrak{c}$ in a coarsely simply-connected metric space $X$.
(6) Dehn function: $Dehn_G(n)$, the Dehn function of a presentation complex $Y$ of a group $G$.
(7) Isoperimetric function $IP_M(\ell)$ of a simply-connected Riemannian manifold $M$.

**Summary of relationships between the volume/area concepts**:

(1) Functions $Dehn_G(n)$ and $IP_M(\ell)$ are *approximately equivalent* to each other, provided that $G$ acts geometrically on $M$; both functions are QI invariant, provided that one considers them up to approximate equivalence.
(2) $\mathrm{A}r_\delta(\mathfrak{c}) \asymp Area_P(c)$, where $P = \mathrm{Rips}_\delta(X)$ and $c$ is the loop in $P$ obtained from $\mathfrak{c}$ by connecting "consecutive points" by the edges in $P$.

## 5.6. Topological coupling

We first introduce Gromov's interpretation of quasi-isometry between groups using the language of topological actions.

Given groups $G_1, G_2$, a *topological coupling* of these groups is a metrizable locally compact topological space $X$ together with two commuting cocompact properly discontinuous topological actions $\rho_i : G_i \curvearrowright X, i = 1, 2$. (The actions commute if and only if $\rho_1(g_1)\rho_2(g_2) = \rho_2(g_2)\rho_1(g_1)$ for all $g_i \in G_i$, $i = 1, 2$.) Note that the actions $\rho_i$ are not required to be isometric. The following theorem was first proven by Gromov in [**Gro93**]; see also [**dlH00**, page 98].

THEOREM 5.52. *If $G_1, G_2$ are finitely generated groups, then $G_1$ is QI to $G_2$ if and only if there exists a topological coupling between these groups.*

PROOF. 1. Suppose that $G_1$ is QI to $G_2$. Then there exists an $(L, A)$ quasi-isometry $q : G_1 \to G_2$. Without loss of generality, we may assume that $q$ is $L$–Lipschitz. Consider the space $X$ of such maps $G_1 \to G_2$. We will give $X$ the topology of pointwise convergence. By Arzela–Ascoli theorem, $X$ is locally compact.

The groups $G_1, G_2$ act on $X$ as follows:

$$\rho_1(g_1)(f) := f \circ g_1^{-1}, \quad \rho_2(g_2)(f) := g_2 \circ f, \quad f \in X.$$

It is clear that these actions commute and are topological. For each $f \in X$ there exist $g_1 \in G_1, g_2 \in G_2$ so that

$$g_2 \circ f(1) = 1, f \circ g_1^{-1}(1) \in B(1, A).$$

Therefore, by Arzela–Ascoli theorem, both actions are cocompact. We will check that $\rho_2$ is properly discontinuous as the case of $\rho_1$ is analogous. Let $K \subset X$ be a compact subset. Then there exists $R < \infty$ so that for every $f \in K$, $f(1) \in B(1, R)$. If $g_2 \in G_2$ is such that $g_2 \circ f \in K$ for some $f \in K$, then

$$g_2(B(1, R)) \cap B(1, R) \neq \emptyset. \tag{5.8}$$

Since the action of $G_2$ on itself is free, it follows that the collection of $g_2 \in G_2$ satisfying (5.8) is finite. Hence, $\rho_2$ is properly discontinuous.

Lastly, the space $X$ is metrizable, since it is locally compact, 2nd countable and Hausdorff; more explicitly, one can define distance between functions as the

Gromov–Hausdorff distance between their graphs. Note that this metric is $G_1$–invariant.

2. Suppose that $X$ is a topological coupling of $G_1$ and $G_2$. If $X$ were a geodesic metric space and the actions of $G_1, G_2$ were isometric, we would not need commutation of these action. However, there are examples of QI groups which do not act geometrically on the same geodesic metric space, see Theorem 5.29. Nevertheless, the construction of a quasi-isometry below is pretty much the same as in the proof of Milnor-Schwarz theorem.

Since $G_i \curvearrowright X$ is cocompact, there exists a compact $K \subset X$ so that $G_i \cdot K = X$; pick a point $p \in K$. Then for each $g_i \in G_i$ there exists $\phi_i(g_i) \in G_{i+1}$ so that $g_i(p) \in \phi_i(g_i)(K)$, here and below $i$ is taken mod 2. We have maps $\phi_i : G_i \to G_{i+1}$.

a. Let us check that these maps are Lipschitz. Let $s \in S_i$, a finite generating set of $G_i$, we will use the word metric on $G_i$ with respect to $S_i$, $i = 1, 2$. Define $C$ to be the union

$$\cup_{i=1,2} \bigcup_{s \in S_i} s(K).$$

Since $\rho_i$ are properly discontinuous actions, the sets $G_i^C := \{h \in G_i : h(C) \cap C \neq \emptyset\}$ are finite for $i = 1, 2$. Therefore, the word-lengths of the elements of these sets are bounded by some $L < \infty$. Suppose now that $g_{i+1} = \phi_i(g_i)$, $s \in S_i$. Then $g_i(p) \in g_{i+1}(K), sg_i(p) \in g'_{i+1}(K)$ for some $g'_{i+1} \in G_{i+1}$. Therefore, $sg_{i+1}(K) \cap g'_{i+1}(K) \neq \emptyset$ hence $g_{i+1}^{-1} g'_{i+1}(K) \cap s(K) \neq \emptyset$. (This is where we are using the fact that the actions of $G_1$ and $G_2$ on $X$ commute.) Therefore, $g_{i+1}^{-1} g'_{i+1} \in G_{i+1}^C$, hence $d(g_{i+1}, g'_{i+1}) \leqslant L$. Consequently, $\phi_i$ is $L$–Lipschitz.

b. Let $\phi_i(g_i) = g_{i+1}, \phi_{i+1}(g_{i+1}) = g'_i$. Then $g_i(K) \cap g'_i(K) \neq \emptyset$ hence $g_i^{-1} g'_i \in G_i^C$. Therefore, $\mathrm{dist}(\phi_{i+1} \circ \phi_i, Id_{G_i}) \leqslant L$ and $\phi_i : G_i \to G_{i+1}$ is a quasi-isometry. ☐

The more useful direction of this theorem is, of course, from QI to a topological coupling, see e.g. [**Sha04, Sau06**].

Definition 5.53. Two groups $G_1, G_2$ are said to **have a common geometric model** if there exists a proper quasi-geodesic metric space $X$ such that $G_1, G_2$ both act geometrically on $X$.

In view of Theorem 5.29, if two groups have a common geometric model then they are quasi-isometric. The following theorem shows that the converse is false:

Theorem 5.54 (L. Mosher, M. Sageev, K. Whyte, [**MSW03**]). *Let* $G_1 := \mathbb{Z}_p * \mathbb{Z}_p, G_2 := \mathbb{Z}_q * \mathbb{Z}_q$, *where* $p, q$ *are distinct odd primes. Then the groups* $G_1, G_2$ *are quasi-isometric (since they are virtually isomorphic to the free group on two generators) but do not have a common geometric model.*

This theorem, in particular, implies that in Theorem 5.52 one cannot assume that both group actions are isometric (for the same metric).

### 5.7. Quasi-actions

The notion of an *action* of a group on a space is replaced, in the context of quasi-isometries, by the one of *quasi-action*. Recall that an *action* of a group $G$ on a set $X$ is a homomorphism $\phi : G \to Aut(X)$, where $Aut(X)$ is the group of bijections $X \to X$. Since quasi-isometries are defined only up to "bounded error", the concept of a homomorphism has to be modified when we use quasi-isometries.

Definition 5.55. Let $G$ be a group and $X$ be a metric space. An $(L, A)$-quasi-action of $G$ on $X$ is a map $\phi : G \to Map(X, X)$, so that:

- $\phi(g)$ is an $(L, A)$-quasi-isometry of $X$ for all $g \in G$.
- $d(\phi(1), id_X) \leq A$.
- $d(\phi(g_1 g_2), \phi(g_1)\phi(g_2)) \leq A$ for all $g_1, g_2 \in G$.

Thus, $\phi$ is "almost" a homomorphism with the error $A$.

By abusing notation, we will denote quasi-actions by $\phi : G \curvearrowright X$, even though, what we have is not an action.

Example 5.56. Suppose that $G$ is a group and $\phi : G \to \mathbb{R} \subset \text{Isom}(\mathbb{R})$ is a function. Then $\phi$, of course, satisfies (1), while properties (2) and (3) are equivalent to the single condition:

$$|\phi(g_1 g_2) - \phi(g_1) - \phi(g_2)| \leq A.$$

Such maps $\phi$ are called *quasi-morphisms.* and they appear frequently in geometric group theory, in the context of *2nd bounded cohomology*, see e.g. [**EF97a**]. Many interesting groups do not admit nontrivial homomorphisms of $\mathbb{R}$ but admit unbounded quasi-morphisms. For instance, a hyperbolic Coxeter group $G$ does not admit nontrivial homomorphisms to $\mathbb{R}$. However, unless $G$ is virtually abelian, it has infinite-dimensional space of equivalence classes quasi-morphisms, where

$$\phi_1 \sim \phi_2 \iff \|\phi_1 - \phi_2\| < \infty.$$

See [**EF97a**].

Exercise 5.57. Let $QI(X)$ denote the group of (equivalence classes of) quasi-isometries $X \to X$. Show that every quasi-action determines a homomorphism $\widehat{\phi} : G \to QI(X)$ given by composing $\phi$ with the projection to $QI(X)$.

The *kernel* of the quasi-action $\phi : G \curvearrowright X$ is the kernel of the homomorphism $\widehat{\phi}$.

Exercise 5.58. Construct an example of a geometric quasi-action $G \curvearrowright \mathbb{R}$ whose kernel is the entire group $G$.

We can also define proper discontinuity and cocompactness for quasi-actions by analogy with isometric actions:

Definition 5.59. Let $\phi : G \curvearrowright X$ be a quasi-action.

1. We say that $\phi$ is *properly discontinuous* if for every $x \in X, R \in \mathbb{R}_+$, the set

$$\{g \in G | d(x, \phi(g)(x)) \leq R\}$$

is finite. Note that if $X$ proper and $\phi$ is an isometric action, this definition is equivalent to proper discontinuity of $G \curvearrowright X$.

2. We say that $\phi$ is *cobounded* if there exists $x \in X, R \in \mathbb{R}_+$ so that for every $x' \in X$ there exists $g \in G$ so that $d(x', \phi(g)(x)) \leq R$. Equivalently, there exists $R'$ so that $d(x, \phi(g)(x')) \leq R$.

3. Lastly, we say that quasi-action $\phi$ is *geometric* if it is both properly discontinuous and cobounded.

Below we explain how quasi-actions appear in the context of QI rigidity problems. Suppose that $G_1, G_2$ are groups, $\psi_i : G_i \curvearrowright X_i$ are isometric actions; for instance, $X_i$ could be $G_i$ or its Cayley graph. Suppose that $f : X_1 \to X_2$ is a quasi-isometry with quasi-inverse $\bar{f}$. We then define a *conjugate* quasi-action $\phi = f^*(\psi_2)$ of $G_2$ on $X_1$ by

$$\phi(g) = \bar{f} \circ g \circ f. \tag{5.9}$$

More generally, we say that two quasi-actions $\psi_i : G \curvearrowright X_i$ are *quasi-conjugate* if there exists a quasi-isometry $f : X_1 \to X_2$, so that $\psi_1$ and $f^*(\psi_2)$ project to the same homomorphism

$$G \to QI(X_1).$$

LEMMA 5.60. *1. Under the above assumptions, $\phi = f^*(\psi_2)$ is a quasi-action. 2. If $\psi_2$ is geometric, so is $\phi$.*

PROOF. 1. Suppose that $f$ is an $(L, A)$-quasi-isometry. It is clear that $\phi$ satisfies Parts 1 and 2 of the definition, we only have to verify (3):

$$\mathrm{dist}(\phi(g_1g_2), \phi(g_1)\phi(g_2)) = \mathrm{dist}(\bar{f}g_1g_2f, \bar{f}g_1f\bar{f}g_2f) \le LA + A$$

in view of Exercise 5.7.

2. In order to verify that $\phi$ is geometric, one needs to show proper discontinuity and coboundedness. We will verify the former since the proof of the latter is similar. Pick $x \in X, R \in \mathbb{R}_+$, and consider the set the set

$$G_{x,R} = \{g \in G = G_2 | d(x, \phi(g)(x)) \le R\} \subset G.$$

By definition, $\phi(g)(x) = \bar{f}gf(x)$. Thus, $d(x, g(x)) \le LR + 2A$. Hence, by proper discontinuity of the action $G \curvearrowright X_2$, the set $G_{x,R}$ is finite. □

The same construction of a conjugate quasi-action applies if $G_2 \curvearrowright X_2$ is not an action, but merely a quasi-action.

EXERCISE 5.61. Suppose that $\phi_2 : G \curvearrowright X_2$ is a quasi-action, $f : X_1 \to X_2$ is a quasi-isometry and $\phi_1 : G \curvearrowright X_1$ is the conjugate quasi-action. Then $\phi_2$ is properly discontinuous (respectively, cobounded, or geometric) if and only if $\phi_1$ is properly discontinuous (respectively, cobounded, or geometric).

COROLLARY 5.62. *Let $G_1$ and $G_2$ be finitely generated quasi-isometric groups and let $f : G_1 \to G_2$ be a quasi-isometry. Then:*

*1. The quasi-isometry $f$ induces (by conjugating actions and quasi-actions on $G_2$) an isomorphism $QI(G_2) \to QI(G_1)$ and a homomorphism $f_* : G_2 \to QI(G_1)$*

*2. The kernel of $f_*$ is quasi-finite: For every $K \ge 0$, the set of $g \in G_2$ such that $\mathrm{dist}(f_*(g), \mathrm{id}_{G_1}) \leqslant K$, is finite.*

PROOF. To construct $f_*$ apply Lemma 5.60 to the isometric action $\psi_2 : G_2 \curvearrowright G_2$. Quasifiniteness of the kernel of $f_*$ follows from proper discontinuity of the quasi-action $G_2 \curvearrowright G_1$. The isomorphism $QI(G_2) \to QI(G_1)$ is defined *via* the formula (5.9). The inverse to this homomorphism is defined by switching the roles of $f$ and $\bar{f}$. □

REMARK 5.63. For many groups $G = G_1$, if $h : G \to G$ is an $(L, A)$-quasi-isometry, so that $\mathrm{dist}(f, Id_G) < \infty$, then $\mathrm{dist}(f, Id_G) \le D(L, A)$. For instance, this holds when $G$ is a non-elementary hyperbolic group, see Lemma 9.86. (This is also true for isometry groups of irreducible symmetric spaces and Euclidean buildings and many other spaces, see e.g. [**KKL98**].) In this situation, quasi-finite kernel of $f_*$ above is actually finite.

The following theorem is a weak converse to the construction of a conjugate quasi-action:

THEOREM 5.64 (B. Kleiner, B. Leeb, [**KL09**]). *Suppose that $\phi : G \curvearrowright X_1$ is a quasi-action. Then there exists a metric space $X_2$, a quasi-isometry $f : X_1 \to X_2$ and an isometric action $\psi : G \curvearrowright X_2$, so that $f$ quasi-conjugates $\psi$ to $\phi$.*

Thus, every quasi-action is conjugate to an isometric action, but, *a priori*, on a different metric space. The key issue of the QI rigidity is:

*Can one, under some conditions, take $X_2 = X_1$?*

Most proofs of QI rigidity theorems follow this route:

1. Suppose that groups $G_1, G_2$ are quasi-isometric. Find a "nice space" $X_1$ on which $G_1$ acts geometrically. Take a quasi-isometry $f : X_1 \to X_2 = G_2$, where $\psi : G_2 \curvearrowright G_2$ is the action by left multiplication.

2. Define the conjugate quasi-action $\phi = f^*(\psi)$ of $G_2$ on $X_1$.

3. Show that the quasi-action $\phi$ has finite kernel (or, at least, identify the kernel, prove that it is, say, abelian).

4. Extend, if necessary, the quasi-action $G_2 \curvearrowright X_1$ to a quasi-action $\hat{\phi}$ on a larger space $\hat{X}_1$.

5. Show that $\hat{\phi}$ has the same projection to $QI(\hat{X}_1)$ as a isometric action $\phi' : G_2 \curvearrowright \hat{X}_1$ by verifying, for instance, that $\hat{X}_1$ has very few quasi-isometries, namely, every quasi-isometry of $X$ is within finite distance from an isometry. (Well, maybe no all quasi-isometries of $\hat{X}_1$, but the ones which extend from $X_1$.) Then conclude either that $G_2 \curvearrowright \hat{X}_1$ is geometric, or, that the isometric actions of $G_1, G_2$ are commensurable, i.e., the images of $G_1, G_2$ in $\mathrm{Isom}(\hat{X}_2)$ have a common finite-index subgroup.

We will see how R. Schwarz's proof of QI rigidity for nonuniform lattices follows this line of arguments: $X_1$ will be a truncated hyperbolic space and $\hat{X}_1$ is the hyperbolic space itself. The same is true for QI rigidity of higher rank non-uniform lattices (A. Eskin's theorem [**Esk98**]). This is also true for uniform lattices in the isometry groups of nonpositively curved symmetric spaces other than $\mathbb{H}^n$ and $\mathbb{CH}^n$ (P. Pansu, [**Pan89**], B. Kleiner and B. Leeb [**KL98b**]; A. Eskin and B. Farb [**EF97b**]), except one does not have to enlarge $X_1$. Another example of such argument is the proof by M. Bourdon and H. Pajot [**BP00**] and X. Xie [**Xie06**] of QI rigidity of groups acting geometrically on 2-dimensional hyperbolic buildings.

5'. Part 5 may fail if $X$ has too many quasi-isometries, e.g. if $X_1 = \mathbb{H}^n$ or $X_1 = \mathbb{CH}^n$. Then, instead, one shows that every geometric quasi-action $G_2 \curvearrowright X_1$ is quasi-conjugate to a geometric (isometric!) action. We will see such a proof in the case of Sullivan–Tukia rigidity theorem for uniform lattices in $\mathrm{Isom}(\mathbb{H}^n), n \geq 3$. Similar arguments apply in the case of groups quasi-isometric to the hyperbolic plane.

Not all quasi-isometric rigidity theorems are proven in this fashion. An alternative route is to show QI rigidity of a certain algebraic property (P) is to show that it is equivalent to some geometric property (P'), which is QI invariant. Examples of such proofs are QI rigidity of the class of virtually nilpotent groups and of virtually free groups. The first property is equivalent, by Gromov's theorem, to polynomial growth; the argument in the second case is less direct (see Theorem 18.38), but the key fact is that geometric condition of having infinitely many ends is equivalent to the algebraic condition that a group splits over a finite subgroup.

CHAPTER 6

# Coarse topology

The goal of this section is to provide tools of algebraic topology for studying quasi-isometries and other concepts of the geometric group theory. The class of *metric cell complexes with bounded geometry* provides a class of spaces for which application of algebraic topology is possible.

## 6.1. Ends of spaces

In this section we review the oldest coarse topological notion, the one of *ends* of a topological space. Let $X$ be a connected, locally path-connected topological space which admits an exhaustion by compact subsets, i.e., an increasing family of compact subsets $\{K_i\}_{i\in I}$, where $I$ is an ordered set,

$$K_i \subset K_j, \quad i \leqslant j,$$

so that

$$\bigcup_{i\in I} K_i = X.$$

The key example to consider is when $X$ is a proper metric space, $K_i = \overline{B}(o,i)$, $i \in \mathbb{N}$ and $o \in X$ is a fixed point. We will refer to this as the *standard example.* (An important special case to keep in mind is the Cayley graph of a finitely-generated group, where $o$ is a vertex.) For each $K = K_i$ we let $K^c = X \setminus K$.

We then let $J$ denote the set whose elements are connected components of various $K_i^c$. The set $J$ has the partial order: $C \leqslant C'$ iff $C' \subset C$. Thus, the "larger" $C$'s are the ones which correspond to bigger $K$'s.

Definition 6.1. The set $Ends(X) = \epsilon(X)$ of *ends* of $X$, is the set of unbounded (from above) increasing chains in the poset $J$. Every such chain is called an *end* of $X$.

In the standard example, each end is a sequence of connected nonempty sets

$$C_1 \supset C_2 \supset C_3 \supset \ldots$$

where each $C_i$ is a component of $K_i^c$.

Equivalently, since we assumed that $X$ is locally path-connected, each element of $J$ is an element of the set $\pi_0(K_i^c)$ for some $i$. Thus, we have the inverse system of sets $\{\pi_0(K_i^c)\}$ indexed by $I$, where

$$f_{i,j} : \pi_0(K_j^c) \to \pi_0(K_i^c), i \leqslant j,$$

is the map induced by the inclusion $K_j^c \subset K_i^c$. Then there is a natural bijection between the inverse limit

$$\pi_0^\infty(X) = \varprojlim \pi_0(K_i^c)$$

of this system and the set of ends $\epsilon(X)$: Choosing an element $\sigma$ of $\pi_0(K_j^c)$ is equivalent to choosing the connected component of $K_i^c$ which gives rise to $\sigma$. Note that if $X$ is a Cayley graph, then each $\pi_0(K_i^c)$ is a finite set.

We say that a family of points $(x_i)_{i\in I}$, $x_i \in C_i$, $C_i \subset K_i^c$, *represents* the corresponding end of $X$, since each $x_i$ represents an element of $\pi_0(K_i^c)$. We will use the notation $x_\bullet$ for this end.

We next topologize $\epsilon(X)$. We equip each $\pi_0(K_i^c)$ with the discrete topology (which makes sense in view of the Cayley graph example) and then put the initial topology on the inverse limit as explained in Section 1.1.

Concretely, one describes this topology as follows. Pick some $C \in J$, which is a component of $K_i^c$. Then $C$ defines a subset $\epsilon_C \subset X$, which consists of ends which are represented by those families $(x_j)$ so that, $x_j \in C$ for all $j \geqslant i$. These sets form a basis of the inverse limit topology on $\epsilon(X)$ described above. Since $\epsilon(X)$ is the inverse limit of sets with discrete topology, the space $\epsilon(X)$ is totally disconnected. Furthermore, clearly, $\epsilon(X)$ is Hausdorff.

Exercise 6.2. 1. The above topology on $\epsilon(X)$ defines a compactification $\bar{X} = X \cup \epsilon(X)$ of the topological space $X$.

2. Let $G$ be a group of homeomorphisms of $X$. Then the action of $G$ on $X$ extends to a topological action of $G$ on $\bar{X}$.

Remark 6.3. 1. Some of the sets $\epsilon_C$ could be empty: They correspond to the sets $C$ which are relatively compact. This, of course, means that one should discard such sets $C$ when thinking about the ends of $X$.

2. There is a terminological confusion here coming from the literature in differential geometry and geometric analysis, where $X$ is a smooth manifold: An analyst would call each set $C$ an end of $X$.

Example 6.4. 1. Every compact topological space $X$ has empty set of ends. Conversely, if $\epsilon(X) = \emptyset$, then $X$ is compact.

2. If $X = \mathbb{R}$, then $\epsilon(X)$ is a 2-point set. If $X = \mathbb{R}^n, n \geqslant 2$, then $\epsilon(X)$ is a single point.

3. If $X$ is a binary (i.e., tri-valent) tree then $\epsilon(X)$ is homeomorphic to the Cantor set.

See Figure 6.1 for an example. The space $X$ in this picture has 5 visibly different ends: $\epsilon_1, ..., \epsilon_5$. We have $K_1 \subset K_2 \subset K_3$. The compact $K_1$ separates the ends $\epsilon_1, \epsilon_2$. The next compact $K_2$ separates $\epsilon_3$ from $\epsilon_4$. Finally, the compact $K_3$ separates $\epsilon_4$ from $\epsilon_5$.

Analogously, one defines *higher homotopy groups* $\pi_k^\infty(X, x^\bullet)$ at infinity of $X$, $k \geqslant 1$. We now assume that the set $I$ is the set of natural numbers with the usual order. For each end $x_\bullet \in \epsilon(X)$ pick a representing sequence $(x_i)_{i\in I}$. For each $i \leqslant j$, pick a path $p_{ij}$ in $K_i^c$ connecting $x_i$ to $x_j$. The concatenation of such paths is a proper map $p : \mathbb{R}_+ \to X$. The proper homotopy class of $p$ is denoted $x^\bullet$. Given $p$, we then have the inverse system of group homomorphisms

$$\pi_k(K_j^c, x_j) \to \pi_k(K_i^c, x_i), i \leqslant j,$$

induced by inclusion maps of the components $C_j \hookrightarrow C_i$, where $x_i \in C_i, x_j \in X_j$. Note that the paths $p_{ij}$ are needed here since we are using different base-points for the homotopy groups.

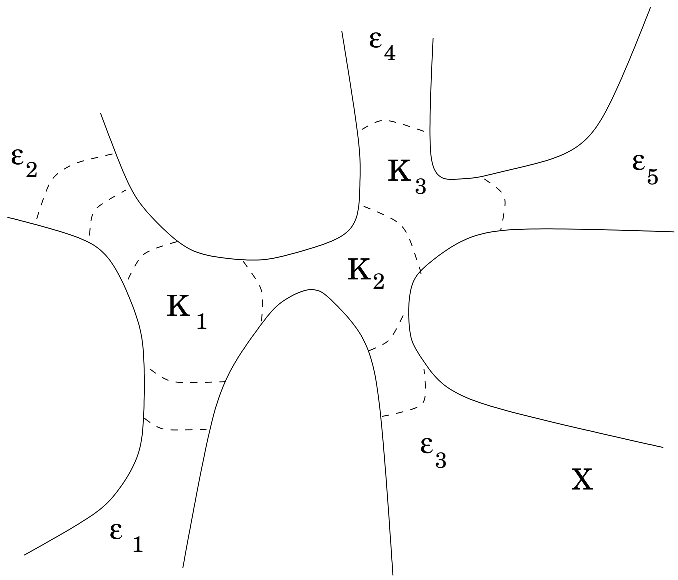



FIGURE 6.1. *E*nds of $X$.

The group $\pi_k^\infty(X, x^\bullet)$ then is the inverse limit

$$\varprojlim \pi_k(K_i^c, x_i).$$

EXERCISE 6.5. Verify that this construction depends only on $x^\bullet$ and not on the paths $p_{ij}$.

For the rest of the book, we will not need $\pi_k^\infty$ for $k > 0$.

PROPOSITION 6.6. *If $f : X \to Y$ is an $(L, A)$–quasi-isometry of proper geodesic metric spaces then $f$ induces a homeomorphism $\epsilon(X) \to \epsilon(Y)$.*

PROOF. For geodesic metric spaces, path-connectedness is equivalent to connectedness. Since $f$ is a quasi-isometry, for each bounded subset $K \subset X$, the image $f(K)$ is again bounded. Note that $f$ need not map connected sets to connected sets since $f$ is not required to be continuous. nevertheless, we have

LEMMA 6.7. *The open $A' = A + 1$-neighborhood $\mathcal{N}_{A'}(f(C))$ is connected for every connected subset $C \subset X$.*

PROOF. For points $x, x' \in C$, and every $\delta > 0$ there exists a *chain* $x_0 = x, x_1, ..., x_n = x'$, so that $x_i \in C$ and $\mathrm{dist}(x_i, x_{i+1}) \leqslant \delta$, $i = 0, ..., n-1$. Then we obtain a chain $y_i = f(x_i)$, $i = 0, ..., n$, so that

$$\mathrm{dist}(y_i, y_{i+1}) \leqslant \delta' = L\delta + A$$

It follows that a geodesic segment $[y_i y_{i+1}]$ is contained in $\mathcal{N}_{\delta'}(f(C))$. Hence, the $\delta'$–neighborhood of $f(C)$ is path-connected for every $\delta > 0$. We conclude that $\mathcal{N}_{A'}(f(C))$ is connected by taking $\delta = 1$. □

Without loss of generality, we may assume that $K_i = \overline{B}(o, i)$ is a closed metric ball in $X$ and $i \in \mathbb{N}$. We define a map $\epsilon(f) : \epsilon(X) \to \epsilon(Y)$ as follows. Set $R := A + 1$. Suppose that $\eta \in \epsilon(X)$ is represented by a nested sequence $(C_i)$, where $C_i$ is a connected component of $X \setminus K_i$, $K_i \subset X$ is compact. By reindexing our system of compacts $K_i$, without loss of generality we may assume that for each $i$, $\mathcal{N}_R(C_i) \subset C_{i-1}$. Thus we get a nested sequence of connected subsets $\mathcal{N}_R(f(C_i)) \subset Y$ each of which is contained in a connected component $V_i$ of the complement to the bounded subset $f(K_{i-1}) \subset Y$. Thus we send $\eta$ to the end $\epsilon(f)(\eta)$ represented by $(V_i)$. By considering the quasi-inverse $\bar{f}$ to $f$, we see that $\epsilon(f)$ has the inverse map $\epsilon(\bar{f})$. It is also clear from the construction that both $\epsilon(f)$ and $\epsilon(\bar{f})$ are continuous. □

If $G$ is a finitely generated group then the space of ends $\epsilon(G)$ is defined to be the set of ends of its Cayley graph. The previous lemma implies that $\epsilon(G)$ does not depend on the choice of a finite generating set and that quasi-isometric groups have homeomorphic sets of ends.

THEOREM 6.8 (Properties of $\epsilon(X)$). *1. The topological space $\epsilon(X)$ is compact, Hausdorff and totally disconnected; $\epsilon(X)$ is empty if and only if $X$ is compact.*

*2. Suppose that $G$ is a finitely-generated group. Then $\epsilon(G)$ consists of 0, 1, 2 points or has cardinality of continuum. In the latter case the set $\epsilon(G)$ is perfect: Each point is a limit point.*

*3. $\epsilon(G)$ is empty iff $G$ is finite. $\epsilon(G)$ consists of 2-points iff $G$ is virtually (infinite) cyclic.*

*4. $|\epsilon(G)| > 1$ iff $G$ splits nontrivially over a finite subgroup.*

COROLLARY 6.9. *1. If $G$ is quasi-isometric to $\mathbb{Z}$ then $G$ contains $\mathbb{Z}$ as a finite index subgroup.*

*2. Suppose that $G$ splits nontrivially as $G_1 \star G_2$ and $G'$ is quasi-isometric to $G$. Then $G'$ splits nontrivially as $G'_1 \star_F G'_2$ (amalgamated product) or as $G'_1 \star_F$ (HNN splitting), where $F$ is a finite group.*

Note that we already know that $\epsilon(X)$ is Hausdorff and totally-disconnected. Compactness of $\epsilon(X)$ follows from the fact that each $K^c$ has only finitely many components which are not relatively compact. Properties 2 and 3 in Theorem 6.8 are also relatively easy, see for instance [**BH99**, Theorem 8.32] for the detailed proofs. The hard part of this theorem is

THEOREM 6.10. *If $|\epsilon(G)| > 1$ then $G$ splits nontrivially over a finite subgroup.*

This theorem is due to Stallings [**Sta68**] (in the torsion-free case) and Bergman [**Ber68**] for groups with torsion. To this day, there is no simple proof of this result. A geometric proof could be found in Niblo's paper [**Nib04**]. For finitely presented groups, there is an easier combinatorial proof due to Dunwoody using minimal tracks, [**Dun85**]; a combinatorial version of this argument could be found in [**DD89**]. In Chapters 18 and 19 we prove Theorem 6.10 first for finitely-presented and then for all finitely-generated groups. We will also prove QI rigidity of the class of virtually free groups.

### 6.2. Rips complexes and coarse homotopy theory

**6.2.1. Rips complexes.** Let $X$ be a uniformly discrete metric space (see Definition 1.19). Recall that the $R$-Rips complex of $X$ is the simplicial complex whose vertices are points of $X$; vertices $x_1, ..., x_n$ span a simplex if and only if

$$\mathrm{dist}(x_i, x_j) \leqslant R, \forall i, j.$$

For each pair $0 \leqslant R_1 \leqslant R_2 < \infty$ we have a natural simplicial embedding

$$\iota_{R_1,R_2} : \mathrm{Rips}_{R_1}(X) \to \mathrm{Rips}_{R_2}(X)$$

and

$$\iota_{R_1,R_2} = \iota_{R_2,R_3} \circ \iota_{R_1,R_2}$$

provided that $R_1 \leqslant R_2 \leqslant R_3$. Thus, the collection of Rips complexes of $X$ forms a direct system $\mathrm{Rips}_\bullet(X)$ of simplicial complexes indexed by positive real numbers.

Following the construction in Section 2.2.2, we metrize (connected) Rips complexes $\mathrm{Rips}_R(X)$ using the *standard length metric* on simplicial complexes. Then, each embedding $\iota_{R_1,R_2}$ is isometric on every simplex and 1-Lipschitz overall. Note that the assumption that $X$ is uniformly discrete implies that $\mathrm{Rips}_R(X)$ is a simplicial complex of bounded geometry (Definition 2.60) for every $R$.

EXERCISE 6.11. Suppose that $X = G$, a finitely-generated group with a word metric. Show that for every $R$, the action of $G$ on itself extends to a simplicial action of $G$ on $\mathrm{Rips}_R(G)$. Show that this action is geometric.

The following simple observation explains why Rips complexes are useful for analyzing quasi-isometries:

LEMMA 6.12. *Let $f : X \to Y$ be an $(L, A)$–coarse Lipschitz map. Then $f$ induces a simplicial map $\mathrm{Rips}_R(X) \to \mathrm{Rips}_{LR+A}(Y)$ for each $R \geqslant 0$.*

PROOF. Consider an $m$-simplex $\sigma$ in $\mathrm{Rips}_R(X)$; the vertices of $\sigma$ are distinct points $x_0, x_1, ..., x_m \in X$ within distance $\leqslant R$ from each other. Since $f$ is $(L, A)$–coarse Lipschitz, the points $f(x_0), ..., f(x_m) \in Y$ are within distance $\leqslant LR + A$ from each other, hence they span a simplex $\sigma'$ of dimension $\leqslant m$ in $\mathrm{Rips}_{LR+A}(Y)$. The map $f$ sends vertices of $\sigma$ to vertices of $\sigma'$; we extend this map linearly to a map $\sigma \to \sigma'$. It is clear that this extension defines a simplicial map of simplicial complexes $\mathrm{Rips}_R(X) \to \mathrm{Rips}_{LR+A}(Y)$. □

The idea behind the next definition is that the "coarse homotopy groups" of a metric space $X$ are the homotopy groups of the Rips complexes $\mathrm{Rips}_R(X)$ of $X$. Literally speaking, this does not make much sense since the above homotopy groups depend on $R$. To eliminate this dependence, we have to take into account the maps $\iota_{r,R}$.

DEFINITION 6.13. 1. A metric space $X$ is *coarsely connected* if $\mathrm{Rips}_r(X)$ is connected for some $r$. (Equivalently, $\mathrm{Rips}_R(X)$ is connected for all sufficiently large $R$.)

2. A metric space $X$ is *coarsely $k$-connected* if for each $r$ there exists $R \geqslant r$ so that the mapping $\mathrm{Rips}_r(X) \to \mathrm{Rips}_R(X)$ induces trivial maps of the $i$-th homotopy groups

$$\pi_i(\mathrm{Rips}_r(X), x) \to \pi_i(\mathrm{Rips}_R(X), x)$$

for all $0 \leqslant i \leqslant k$ and $x \in X$.

In particular, $X$ is *coarsely simply-connected* if it is coarsely 1-connected.

In other words, $X$ is coarsely connected if there exists a number $R$ such that each pair of points $x, y \in X$ can be connected by an $R$-chain of points $x_i \in X$, i.e., a finite sequence of points $x_i$, where $\mathrm{dist}(x_i, x_{i+1}) \leqslant R$ for each $i$.

The definition of coarse $k$-connectedness is not quite satisfactory since it only deals with "vanishing" of coarse homotopy groups without actually defining these groups for general $X$. One way to deal with this issue is to consider *pro-groups* which are direct systems

$$\pi_i(\text{Rips}_r(X)), r \in \mathbb{N}$$

of groups. Given such algebraic objects, one can define their *pro-homomorphisms*, *pro-monomorphisms*, etc., see [**KK05**] where this is done in the category of abelian groups (the homology groups). Alternatively, one can work with the direct limit of the homotopy groups.

### 6.2.2. Direct system of Rips complexes and coarse homotopy.

LEMMA 6.14. *Let $X$ be a metric space. Then for $r, c < \infty$, each simplicial spherical cycle $\sigma$ of diameter $\leqslant c$ in* $\text{Rips}_r(X)$ *bounds a disk of diameter $\leqslant r + c$ within* $\text{Rips}_{r+c}(X)$.

PROOF. Pick a vertex $x \in \sigma$. Then $\text{Rips}_{r+c}(X)$ contains a simplicial cone $\tau(\sigma)$ over $\sigma$ with vertex at $x$. Clearly, $\text{diam}(\tau) \leqslant r + c$. ☐

PROPOSITION 6.15. *Let $f, g : X \to Y$ be maps within distance $\leqslant c$ from each other, which extend to simplicial maps*

$$f, g : \text{Rips}_{r_1}(X) \to \text{Rips}_{r_2}(Y)$$

*Then for $r_3 = r_2 + c$, the maps $f, g : \text{Rips}_{r_1} \to \text{Rips}_{r_3}(Y)$ are homotopic via a 1-Lipschitz homotopy $F : \text{Rips}_{r_1}(X) \times I \to \text{Rips}_{r_3}(Y)$. Furthermore, tracks of this homotopy have length $\leqslant (n+1)$, where $n = \dim(\text{Rips}_{r_1}(X))$.*

PROOF. We give the product $\text{Rips}_{r_1}(X) \times I$ the *standard* structure of a simplicial complex with the vertex set $X \times \{0, 1\}$ (by triangulating the each $k+1$-dimensional prisms $\sigma \times I$, where $\sigma$ are simplices in $X$, this triangulation has in $\leqslant (k + 1)$ top-dimensional simplices); we equip this complex with the standard metric.

The map $F$ of the zero-skeleton of $\text{Rips}_{r_1}(X) \times I$ is, of course, $F(x, 0) = f(x), F(x, 1) = g(x)$. Let $\sigma \subset \text{Rips}_{r_1}(X) \times I$ be an $i$–simplex. Then $\text{diam}(F(\sigma^0)) \leqslant r_3 = r_2 + c$, where $\sigma^0$ is the vertex set of $\sigma$. Therefore, $F$ extends (linearly) from $\sigma^0$ to a (1-Lipschitz) map $F : \sigma \to \text{Rips}_{r_3}(Y)$ whose image is the simplex spanned by $F(\sigma^0)$.

To estimate the lengths of the tracks of the homotopy $F$, we note that for each $x \in \text{Rips}_{r_1}(X)$, the path $F(x, t)$ has length $\leqslant 1$ since the interval $x \times I$ is covered by $\leqslant (n + 1)$ simplices, each of which has unit diameter. ☐

In view of the above lemma, we make the following definition:

DEFINITION 6.16. Maps $f, g : X \to Y$ are *coarsely homotopic* if for all $r_1, r_2$ so that $f$ and $g$ extend to

$$f, g : \text{Rips}_{r_1}(X) \to \text{Rips}_{r_2}(Y),$$

there exist $r_3$ and $r_4$ so that the maps

$$f, g : \text{Rips}_{r_1} \to \text{Rips}_{r_3}(Y)$$

are homotopic via a homotopy whose tracks have lengths $\leqslant r_4$.

We then say that a map $f : X \to Y$ determines a *coarse homotopy equivalence* (between the direct systems of Rips complexes of $X, Y$) if there exists a map $g$ :

$Y \to X$ so that the compositions $g \circ f, f \circ g$ are coarsely homotopic to the identity maps.

The next two corollaries, then, are immediate consequences of Proposition 6.15.

COROLLARY 6.17. *Let $f, g : X \to Y$ be $L$–Lipschitz maps within finite distance from each other. Then they are coarsely homotopic.*

COROLLARY 6.18. *If $f : X \to Y$ is a quasi-isometry, then $f$ induces a coarse homotopy-equivalence of the Rips complexes:* $\mathrm{Rips}_\bullet(X) \to \mathrm{Rips}_\bullet(Y)$.

The following corollary is a coarse analogue of the familiar fact that homotopy equivalence preserves connectivity properties of a space:

COROLLARY 6.19. *Coarse $k$-connectedness is a QI invariant.*

PROOF. Suppose that $X'$ is coarsely $k$-connected and $f : X \to X'$ is an $L$–Lipschitz quasi-isometry with $L$–Lipschitz quasi-inverse $\bar{f} : X' \to X$. Let $\gamma$ be a spherical $i$-cycle in $\mathrm{Rips}_r(X)$, $0 \leqslant i \leqslant k$. Then we have the spherical $i$-cycle $f(\gamma) \subset \mathrm{Rips}_{Lr}(X')$. Since $X'$ is coarsely $k$-connected, there exists $r' \geqslant Lr$ such that $f(\gamma)$ bounds a singular $(i+1)$–disk $\beta$ within $\mathrm{Rips}_{r'}(X')$. Consider now $\bar{f}(\beta) \subset \mathrm{Rips}_{L^2 r}(X)$. The boundary of this singular disk is a singular $i$-sphere $\bar{f}(\gamma)$. Since $\bar{f} \circ f$ is homotopic to $id$ within $\mathrm{Rips}_{r''}(X)$, $r'' \geqslant L^2 r$, there exists a singular cylinder $\sigma$ in $\mathrm{Rips}_{r''}(X)$ which cobounds $\gamma$ and $\bar{f}(\gamma)$. Note that $r''$ does not depend on $\gamma$. By combining $\sigma$ and $\bar{f}(\beta)$ we get a singular $(i+1)$–disk in $\mathrm{Rips}_{r''}(X)$ whose boundary is $\gamma$. Hence $X$ is coarsely $k$-connected. $\square$

### 6.3. Metric cell complexes

We now introduce a concept which generalizes simplicial complexes, where the notion of bounded geometry does not imply finite-dimensionality.

A *metric cell complex* is a cell complex $X$ together with a metric $d$ defined on its 0-skeleton $X^{(0)}$. Note that if $X$ is connected, its 1-skeleton $X^{(1)}$ us a graph, and, hence, can be equipped with the standard metric dist. Then the map $(X^{(0)}, d) \to (X^{(1)}, \mathrm{dist})$ in general need not be a quasi-isometry. However, in the most interesting cases, coming from finitely-generated groups, this map is actually an isometry. Therefore, we impose, from now on, the condition:

**Axiom 1.** The map $(X^{(0)}, d) \to (X^{(1)}, \mathrm{dist})$ is a quasi-isometry.

Even though this assumption could be avoided in what follows, restricting to complexes satisfying this axiom will make our discussion more intuitive.

Our first goal to define, using the metric $d$, certain metric concepts on the entire complex $X$. We define inductively a map $c$ which sends cells in $X$ to finite subsets of $X^{(0)}$ as follows. For $v \in X^{(0)}$ we set $c(v) = \{v\}$. Suppose that $c$ is defined on $X^{(i)}$. For each $i+1$-cell $e$, the *support* of $e$ is the smallest subcomplex $\mathrm{Supp}(e)$ of $X^{(i)}$ containing the image of the attaching map of $e$ to $X^{(i)}$. We then set

$$c(\sigma) = c(\mathrm{Supp}(e)).$$

For instance, for every 1-cell $\sigma$, $c(\sigma)$ consists of one or two vertices of $X$ to which $\sigma$ is attached.

Remark 6.20. The reader familiar with the concepts of *controlled topology*, see e.g. [**Ped95**], will realize that the coarsely defined map $c : X \to X^{(0)}$ is a *control map* for $X$ and $(X^{(0)}, d)$ is the *control space*. In particular, a metric cell complex is a special case of a *metric chain complex* defined in [**KK05**].

We now say that the *diameter* $\text{diam}(\sigma)$ of a cell $\sigma$ in $X$ is the diameter of $c(\sigma)$.

Example 6.21. Take a simplicial complex $X$ and restrict its standard metric to $X^{(0)}$. Then, the diameter of a cell in $X$ (as a simplicial complex) is the same as its diameter in the sense of metric cell complexes.

Definition 6.22. A metric cell complex $X$ is said to have *bounded geometry* if there exists a collections of increasing functions $\phi_k(r)$ and numbers $D_k < \infty$ so that the following axioms hold:

**Axiom 2.** For each ball $B(x, r) \subset X^{(0)}$, the set of $k$-cells $\sigma$ such that $c(\sigma) \subset B(x, r)$, contains at most $\phi_k(r)$ cells.

**Axiom 3.** The diameter of each $k$-cell is at most $D_k = D_{k,X}$, $k = 1, 2, 3, \ldots..$

**Axiom 4.** $D_0 := \inf\{d(x, x')|x \neq x' \in X^{(0)}\} > 0$.

Note that we allow $X$ to be infinite-dimensional. We will refer to the function $\phi_k(r)$ and the numbers $D_k$ as *geometric bounds* on $X$, and set

$$D_X = \sup_{k>0} D_{k,X}. \tag{6.1}$$

Exercise 6.23. 1. Suppose that $X$ is a simplicial complex. Then the two notions of bounded geometry coincide for $X$. We will use this special class of metric cell complexes in Section 6.6.

2. If $X$ is a metric cell complex of bounded geometry and $S \subset X$ is a connected subcomplex, then for every two vertices $u, v \in S$ there exists a chain $x_0 = u, x_1, \ldots, x_m = v$, so that $d(x_i, x_{i+1}) \leqslant D_1$ for every $i$. In particular, if $X$ is connected, the identity map $(X^{(0)}, d) \to (X^{(1)}, \text{dist})$ is $D_1$-Lipschitz.

3. Let $X^{(0)} := G$ be a finitely-generated group with its word metric, $X$ be the Cayley graph of $G$. Then $X$ is a metric cell complex of bounded geometry.

As a trivial example, consider spheres $S^n$ with the usual cell structure (single 0-cell and single $n$-cell). Thus, the cellular embeddings $S^n \hookrightarrow S^{n+1}$ give rise to an infinite-dimensional cell complex $S^\infty$. This complex has bounded geometry (since it has only one cell in every dimension). Therefore, the concept of metric cell complexes is more flexible than the one of simplicial complexes.

Exercise 6.24. Let $X, Y$ be metric cell complexes. Then the product cell–complex $X \times Y$ is also a metric cell complex, where we equip the zero-skeleton $X^{(0)} \times Y^{(0)}$ of $X \times Y$ with the product–metric. Furthermore, if $X, Y$ have bounded geometry, then so does $X \times Y$.

We now continue defining metric concepts for metric cell complexes. The (coarse) $R$-ball $\mathbf{B}(x, R)$ centered at a vertex $x \in X^{(0)}$ is the union of the cells $\sigma$ in $X$ so that $c(\sigma) \subset B(x, R)$.

We will say that the *diameter* $\text{diam}(S)$ of a subcomplex $S \subset X$ is the diameter of $c(S)$. Given a subcomplex $W \subset X$, we define the *closed $R$-neighborhood* $\bar{\mathcal{N}}_R(W)$

of $W$ in $X$ to be the largest subcomplex $S \subset X$ so that for every $\sigma \in S$, there exists a vertex $\tau \in W$ so that $\mathrm{dist}_{Haus}(c(v), c(w)) \leqslant R$. A cellular map $f : X \to Y$ between metric cell complexes is called *$L$–Lipschitz* if for every cell $\sigma$ in $X$, $\mathrm{diam}(f(\sigma)) \leqslant L$. In particular, $f : (X^{(0)}, d) \to (Y^{(0)}, d)$ is $L/D_0$-Lipschitz as a map of metric spaces.

Exercise 6.25. Suppose that $f_i : X_i \to X_{i+1}$ are $L_i$-Lipschitz for $i = 1, 2$. Show that $f_2 \circ f_1$ is $L_3$-Lipschitz with

$$L_3 = L_2 \max_k (\phi_{X_2,k}(L_1))$$

Exercise 6.26. Construct examples of a cellular map $f : X \to Y$ between metric graphs of bounded geometry, so that the restriction $f|X^{(0)}$ is $L$–Lipschitz but $f$ is not $L'$–Lipschitz, for any $L' < \infty$.

A map $f : X \to Y$ of metric cell complexes is called *uniformly proper* if $f$ is cellular, $L$-Lipschitz for some $L < \infty$ and $f|X^{(0)}$ is uniformly proper: There exists a proper monotonically increasing function $\eta(R)$ so that

$$\eta(d(x, x')) \leqslant d(f(x), f(x'))$$

for all $x, x' \in X$. The function $\eta(R)$ is called the *distortion function* of $f$.

We will now relate metric cell complexes of bounded geometry to simplicial complexes of bounded geometry:

Exercise 6.27. Let $X$ be a *finite-dimensional* metric cell complexes of bounded geometry. Then there exists a simplicial complex $Y$ of bounded geometry and a cellular homotopy-equivalence $X \to Y$ which is a quasi-isometry in the following sense: $f$ and has homotopy-inverse $\bar{f}$ so that:
1. Both $f, \bar{f}$ are $L$-Lipschitz for some $L < \infty$.
2. $f \circ \bar{f}, \bar{f} \circ f$ are homotopic to the identity.
3. $f : X^{(0)} \to Y^{(0)}, \bar{f} : Y^{(0)} \to X^{(0)}$ are quasi-inverse to each other:

$$d(f \circ \bar{f}, Id) \leqslant A, \quad d(\bar{f} \circ f, Id) \leqslant A.$$

Hint: Apply the usual construction which converts a finite-dimensional CW-complex to a simplicial complex.

Recall that quasi-isometries are not necessarily continuous. In order to use algebraic topology, we, thus, have to approximate quasi-isometries by cellular maps in the context of metric cell complexes. In general, this is of course impossible, since one complex in question can be, say, 0-dimensional and the other 1-dimensional. The *uniform contractibility* hypothesis allows one to resolve this issue.

Definition 6.28. A metric cell complex $X$ is said to be *uniformly contractible* if there exists a continuous function $\psi(R)$ so that for every $x \in X^{(0)}$ the map

$$\mathbf{B}(x, R) \to \mathbf{B}(x, \psi(R))$$

is null-homotopic.

Similarly, $X$ is *uniformly $k$-connected* if there exists a function $\psi_k(R)$ so that for every $x \in X^{(0)}$ the map

$$\mathbf{B}(x, R) \to \mathbf{B}(x, \psi_k(R))$$

induces trivial map on $\pi_i$, $0 \leqslant i \leqslant k$.

We will refer to $\psi, \psi_k$ as the *contractibility functions* of $X$.

Example 6.29. Suppose that $X$ is a connected metric graph with the standard metric. Then $X$ is uniformly 0-connected.

In general, even for simplicial complexes of bounded geometry, contractibility does not imply uniform contractibility. For instance, start with a triangulated 2-torus $T^2$, let $X$ be an infinite cyclic cover of $T^2$. Of course, $X$ is not contractible, but we attach a triangulated disk $D$ to $X$ along a simple homotopically nontrivial loop in $X^{(1)}$. The result is a contractible 2-dimensional simplicial complex $Y$ which clearly has bounded geometry.

Exercise 6.30. Show that $Y$ is not uniformly contractible.

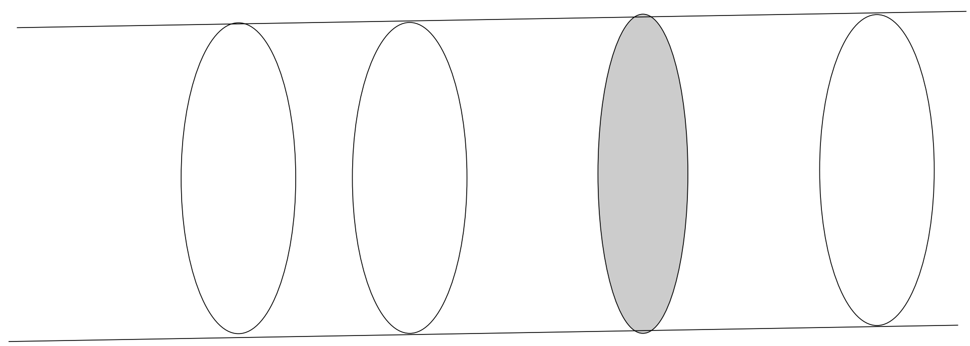



Figure 6.2. Contructible but not uniformly constructible space.

We will see, nevertheless, in Lemma 6.34, that under certain assumptions (presence of a cocompact group action) contractibility implies uniform contractibility.

The following proposition is a metric analogue of the cellular approximation theorem:

Proposition 6.31. *Suppose that $X, Y$ are metric cell complexes, where $X$ is finite-dimensional and has bounded geometry, $Y$ is uniformly contractible, and $f : X^{(0)} \to Y^{(0)}$ is an $L$–Lipschitz map. Then $f$ admits a (continuous) cellular extension $g : X \to Y$, which is an $L'$–Lipschitz map, where $L'$ depends on $L$ and geometric bounds on the complex $X$ and the uniform contractibility function of $Y$. Furthermore, $g(X) \subset \bar{\mathcal{N}}_{L'}(f(X^{(0)}))$.*

Proof. The proof of this proposition is a prototype of most of the proofs which appear in this chapter. We extend $f$ by induction on skeleta of $X$. We claim that (for certain constants $C_i, C'_{i+1}$, $i \geqslant 0$) we can construct a sequence of extensions $f_k : X^{(k)} \to Y^{(k)}$ so that

1. $\mathrm{diam}(f(\sigma)) \leqslant C_k$ for every $k$-cell $\sigma$.
2. $\mathrm{diam}(f(\partial\tau)) \leqslant C'_{k+1}$ for every $(k+1)$–cell $\tau$ in $X$.

Base of induction. We already have $f = f_0 : X^{(0)} \to Y^{(0)}$ satisfying (1) with $C_0 = 0$. If $x, x'$ belong to the boundary of a 1-cell $\tau$ in $X$ then $\mathrm{dist}(f(x), f(x')) \leqslant LD_1$, where $D_1 = D_{1,X}$ is the upper bound on the diameter of 1-cells in $X$. This establishes (2) in the base case.

Inductively, assume that $f = f_k$ was defined on $X^k$, so that (1) and (2) hold. Let $\sigma$ be a $(k+1)$–cell in $X$. Note that

$$\mathrm{diam}(f(\partial\sigma)) \leqslant C'_{k+1}$$

by the induction hypothesis. Then, using uniform contractibility of $Y$, we extend $f$ to $\sigma$ so that $\mathrm{diam}(f(\sigma)) \leqslant C_{k+1}$ where $C_{k+1} = \psi(C'_k)$. Let us verify that the extension $f : X^{k+1} \to Y^{k+1}$ satisfies (2).

Suppose that $\tau$ is a $(k+2)$–cell in $X$. Then, since $X$ has bounded geometry, $\mathrm{diam}(\tau) \leqslant D_{k+2} = D_{k+2,X}$. In particular, $\partial\tau$ is connected and is contained in the union of at most $\phi(D_{k+2}, k+1)$ cells of dimension $k+1$. Therefore,

$$\mathrm{diam}(f(\partial\tau)) \leqslant C_{k+1} \cdot \phi(D_{k+2}, k+1) =: C'_{k+2}.$$

This proves (2).

Since $X$ is, say, $n$-dimensional the induction terminates after $n$ steps. The resulting map $f : X \to Y$ satisfies

$$L' := \mathrm{diam}(f(\sigma)) \leqslant \max_{i=1,\ldots,n} C_i.$$

for every cell $\sigma$ in $X$. Therefore, $f : X \to Y$ is $L'$–Lipschitz. The second assertion of the proposition follows from the definition of $C_i$'s. □

We note that the above proposition can be *relativized*:

LEMMA 6.32. *Suppose that $X, Y$ are metric cell complexes, $X$ is finite-dimensional and has bounded geometry, $Y$ is uniformly contractible, and $Z \subset X$ is a subcomplex. Suppose that $f : Z \to Y$ is a continuous cellular map which extends to an $L$–Lipschitz map $f : X^{(0)} \to Y^{(0)}$. Then $f : Z \cup X^{(0)} \to Y$ admits a (continuous) cellular extension $g : X \to Y$, which is an $L'$–Lipschitz map, where $L'$ depends on $L$ and geometric bounds on $X$ and contractibility function of $Y$.*

PROOF. The proof is the same induction on skeleta argument as in Proposition 6.31. □

COROLLARY 6.33. *Suppose that $X, Y$ are as above and $f_0, f_1 : X \to Y$ are $L$–Lipschitz cellular maps so that $\mathrm{dist}(f_0, f_1) \leqslant C$ in the sense that $d(f_0(x), f_1(x)) \leqslant C$ for all $x \in X^{(0)}$. Then there exists an $L'$–Lipschitz homotopy $f : X \times I \to Y$ between the maps $f_0, f_1$.*

PROOF. Consider the map $f_0 \cup f_1 : X \times \{0,1\} \to Y$, where $X \times \{0,1\}$ is a subcomplex in the metric cell complex $W := X \times I$ (see Exercise 6.24). Then the required extension $f : W \to Y$ of this map exists by Lemma 6.32. □

## 6.4. Connectivity and coarse connectivity

Our next goal is to find a large supply of examples of metric spaces which are coarsely $k$-connected.

LEMMA 6.34. *If $X$ is a finite-dimensional $m$-connected complex which admits a geometric (properly discontinuous cocompact) cellular group action $G \curvearrowright X$, then $X$ is uniformly $m$-connected.*

PROOF. Existence of geometric action $G \curvearrowright X$ implies that $X$ is locally finite. Pick a base-vertex $x \in X$ and let $r < \infty$ be such that $G$-orbit of $B(x, r) \cap X^{(0)}$ is the entire $X^{(0)}$. Therefore, if $C \subset X$ has diameter $\leqslant R/2$, there exists $g \in G$ so that $C' = g(C) \subset \mathbf{B}(x, r+R)$.

Since $C$ is finite, $\pi_1(C')$ is finitely-generated. Thus, simple connectivity of $X$ implies that there exists a finite subcomplex $C'' \subset X$ so that each generator of $\pi_1(C')$ vanishes in $\pi_1(C'')$. Consider now $\pi_i(C'), 2 \leqslant i \leqslant m$. Then, by Hurewicz theorem, the image of $\pi_i(C')$ in $\pi_i(X) \cong H_i(X)$, is contained in the image of $H_i(C')$ in $H_i(X)$. Since $C'$ is a finite complex, we can choose $C''$ above so that the map $H_i(C') \to H_i(C'')$ is zero. To summarize, there exists a finite subcomplex $C''$ in $X$ containing $C'$, so that all maps $\pi_i(C') \to \pi_i(C'')$ are trivial, $1 \leqslant i \leqslant m$.

Since $C''$ is a finite complex, there exists $R' < \infty$ be such

$$C'' \subset \mathbf{B}(x, r + R + R').$$

Hence, the map

$$\pi_k(\mathbf{B}(x, r + R)) \to \pi_k(\mathbf{B}(x, r + R + R'))$$

is trivial for all $k \leqslant m$. Set $\psi(k, r) = \rho = r + R'$. Therefore, if $C \subset X$ is a subcomplex of diameter $\leqslant R/2$, then maps

$$\pi_k(C) \to \pi_k(\mathcal{N}_\rho(C))$$

are trivial for all $k \leqslant m$. □

THEOREM 6.35. *Suppose that $X$ is a uniformly $n$-connected metric cell complex of bounded geometry. Then $Z := X^{(0)}$ is coarsely $n$-connected.*

PROOF. Let $\gamma : S^k \to \mathrm{Rips}_R(Z)$ be a spherical $m$-cycle in $\mathrm{Rips}_R(Z)$, $0 \leqslant k \leqslant n$. Without loss of generality (using simplicial approximation) we can assume that $\gamma$ is a simplicial cycle, i.e. the sphere $S^k$ is given a triangulation $\tau$ so that $\gamma$ sends simplices of $S^k$ to simplices in $\mathrm{Rips}_R(Z)$ and the restriction of $\gamma$ to each simplex is a linear map.

LEMMA 6.36. *There exists a cellular map $\gamma' : (S^k, \tau) \to X$ which agrees with $\gamma$ on the vertex set of $\tau$ and so that $\mathrm{diam}(\gamma'(S^k)) \leqslant R'$, where $R' \geqslant R$ depends only on $R$ and contractibility functions $\psi_i(k, \cdot)$ of $X$, $i = 0, ..., k$.*

PROOF. We construct $\gamma'$ by induction on skeleta of $(S^k, \tau)$. The map is already defined on the 0-skeleton, namely, it is the map $\gamma$ and images of all vertices of $\tau$ are within distance $\leqslant R$ from each other. Suppose we constructed $\gamma'$ on $i$-skeleton $\tau^i$ of $\tau$ so that $\mathrm{diam}(\gamma'(\tau^i)) \leqslant R_i = R_i(R, \psi(k, \cdot))$. Let $\sigma$ be an $i+1$-simplex in $\tau$. We already have a map $\gamma'$ defined on the boundary of $\sigma$ and $\mathrm{diam}(\gamma'(\partial\sigma)) \leqslant R_i$. Then, using uniform contractibility of $X$ we extend $\geq'$ to $\sigma$, so that the resulting map satisfies

$$\mathrm{diam}(\gamma'(\sigma)) \leqslant \psi(i + 1, R_i),$$

which implies that the image is contained in $\mathbf{B}(\gamma(v), 2\psi(i + 1, R_i)$, where $v$ is a vertex of $\sigma$. Thus,

$$\mathrm{diam}(\gamma'(\tau^{i+1}) \leqslant R_{i+1} := R + \psi(i + 1, R_i).$$

Now, lemma follows by induction. Figure 6.3 illustrates the proof in the case $k = 1$. □

Since $X$ is $k$-connected, the map $\gamma'$ extends to a cellular map $\gamma' : D^{k+1} \to X^{(k+1)}$, where $D^{k+1}$ is a triangulated disk whose triangulation $\tau$ extends the triangulation $\tau$ of $S^k$. Our next goal is to "push" $\gamma'$ to a map $\gamma'' : D^{k+1} \to \mathrm{Rips}_{R'}(Z)$ relative to the boundary, where we want $\gamma''|S^k$. Let $\sigma$ be a simplex $D^{k+1}$. A simplicial map is determined by images of vertices. By definition of the number $R'$,

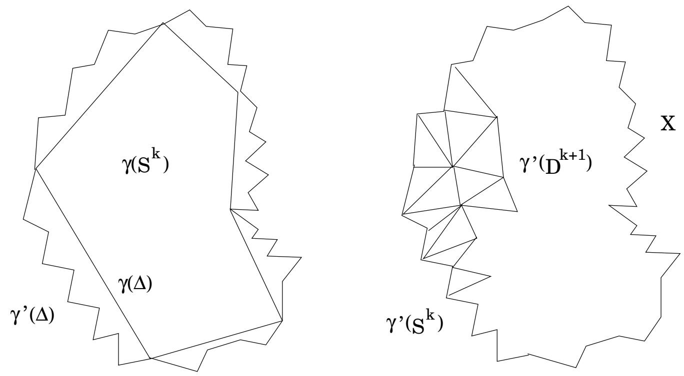



FIGURE 6.3.

images of vertices of $\sigma$ under $\gamma'$ are within distance $\leqslant R'$ from each other. Therefore, we have a canonical extension of $\gamma'|\sigma^{(0)}$ to a map $\sigma \to \mathrm{Rips}_{R'}(Z)$. If $\sigma_1 \subset \sigma_2$, then $\gamma'' : \sigma_1 \to \mathrm{Rips}_{R'}(Z)$ agrees with the restriction of $\gamma'' : \sigma_2 \to \mathrm{Rips}_{R'}(Z)$, since maps are determined by their vertex values. We thus obtain a simplicial map $D^{k+1} \to \mathrm{Rips}_{R'}(Z)$ which, by construction of $\gamma'$ and $\gamma''$, agrees with $\gamma$ on the boundary sphere.

Thus, the inclusion map $\mathrm{Rips}_R(Z) \to \mathrm{Rips}_{R'}(Z)$ induces trivial maps on $k$-th homotopy groups, $0 \leqslant k \leqslant n$. □

As a simple illustration of this theorem, consider the case $n = 0$.

COROLLARY 6.37. *If a bounded geometry metric cell complex $X$ is connected, then $X$ is quasi-isometric to a connected metric graph (with the standard metric).*

PROOF. By connectivity of $X$, for every pair of vertices $x, y \in Z$, there exists a path $\mathfrak{p}$ in $X$ connecting $x$ to $y$, so that $\mathfrak{p}$ is a concatenation of 1-cells in $X$. Since $X$ has bounded geometry diameter of each 1-cell is $\leqslant R = D_1$, where $D_1$ is a geometric bound on $X$ as in Definition 6.22. Therefore, consecutive vertices of $X$ which appear in $\mathfrak{p}$ are within distance $\leqslant R$ from each other. It follows that $\Gamma = \mathrm{Rips}_R(Z)$ is connected. Without loss of generality, we may assume that $R \geqslant 1$. Then the map $\iota : Z \to \mathrm{Rips}_R(Z)$ (sending $Z$ to the vertex set of the Rips complex) is 1-Lipschitz. It is also clear that this map is a $R^{-1}$-quasi-isometric embedding. Thus, $\iota$ is an $(R, 1)$-quasi-isometry. □

We saw, so far, how to go from uniform $k$-connectivity of a metric cell complex $X$ to coarse $k$-connectivity of its 0-skeleton. Our main goal now is to go in the opposite direction. This, of course, may require modifying the complex $X$. The simplest instance of the "inverse" relation is

EXERCISE 6.38. Suppose that $Z$ is a coarsely connected uniformly discrete metric space. Then $Z$ is the 0-skeleton of a connected metric graph $\Gamma$ of bounded geometry so that the inclusion map is a quasi-isometry. Hint: $\Gamma$ is the 1-skeleton of a connected Rips complex $\mathrm{Rips}_R(Z)$. Bounded valence property comes from the uniform discreteness assumption on $Z$.

Below we consider the situation $k \geqslant 1$ in the group-theoretic context, starting with $k = 1$.

Lemma 6.39. *Let $G$ be a finitely-generated group with word metric. Then $G$ is coarsely simply-connected if and only if* $\mathrm{Rips}_R(G)$ *is simply-connected for all sufficiently large $R$.*

Proof. One direction is clear, we only need to show that coarse simple connectivity of $G$ implies that $\mathrm{Rips}_R(G)$ is simply-connected for all sufficiently large $R$. Our argument is similar to the proof of Theorem 6.35. Note that 1-skeleton of $\mathrm{Rips}_1(G)$ is just the Cayley graph of $G$. Using coarse simple connectivity of $G$, we find $D \geqslant 1$ such that the map

$$\pi_1(\mathrm{Rips}_1(G)) \to \pi_1(\mathrm{Rips}_D(G))$$

is trivial. We claim that for all $R \geqslant D$ the Rips complex $\mathrm{Rips}_R(G)$ is simply-connected. Let $\gamma \subset \mathrm{Rips}_R(G)$ be a polygonal loop. For every edge $\gamma_i := [x_i, x_{i+1}]$ of $\gamma$ we let $\gamma_i' \subset \mathrm{Rips}_1(X)$ denote a geodesic path from $x_i$ to $x_{i+1}$. Then, by the triangle inequality, $\gamma_i'$ has length $\leqslant R$. Therefore, all the vertices of $\gamma_i'$ are contained in the ball $\mathbf{B}(x_i, R) \subset G$ and, hence, they span a simplex in $\mathrm{Rips}_R(G)$. Thus, the paths $\gamma_i, \gamma_i'$ are homotopic in $\mathrm{Rips}_R(G)$ rel. their end-points. Let $\gamma'$ denote the loop in $\mathrm{Rips}_1(G)$ which is the concatenation of the paths $\gamma_i'$. Then, by the above observation, $\gamma'$ is freely homotopic to $\gamma$ in $\mathrm{Rips}_R(G)$. On the other hand, $\gamma'$ is null-homotopic in $\mathrm{Rips}_R(G)$ since the map

$$\pi_1(\mathrm{Rips}_1(G)) \to \pi_1(\mathrm{Rips}_R(G))$$

is trivial. We conclude that $\gamma$ is null-homotopic in $\mathrm{Rips}_R(G)$ as well. □

Corollary 6.40. *Suppose that $G$ is a finitely generated group with the word metric. Then $G$ is finitely presented if and only if $G$ is coarsely simply-connected. In particular, finite-presentability is a QI invariant.*

Proof. Suppose that $G$ is finitely-presented and let $Y$ be its finite presentation complex (see Definition 4.80). Then the universal cover $X$ of $Y$ is simply-connected. Hence, by Lemma 6.34, $X$ is uniformly simply-connected and hence by Theorem 6.35, the group $G$ is coarsely simply-connected.

Conversely, suppose that $G$ is coarsely simply-connected. Then, by Lemma 6.39, the simplicial complex $\mathrm{Rips}_R(G)$ is simply-connected for some $R$. The group $G$ acts on $X := \mathrm{Rips}_R(G)$ simplicially, properly discontinuously and cocompactly. Therefore, by Corollary 3.28, $G$ admits a properly discontinuous, free cocompact action on another simply-connected cell complex $Z$. Therefore, $G$ is finitely-presented. □

We now proceed to $k \geqslant 2$. Recall (see Definition 3.26) that a group $G$ has type $\mathbf{F}_n$ ($n \leqslant \infty$) if its admits a free cellular action on a cell complex $X$ such that for each $k \leqslant n$: (1) $X^{(k+1)}/G$ is compact. (2) $X^{(k+1)}$ is $k$-connected.

Example 6.41 (See [**Bie76b**]). Let $\mathbb{F}_2$ be free group on 2 generators $a, b$. Consider the group $G = \mathbb{F}_2^n$ which is the direct product of $\mathbb{F}_2$ with itself $n$ times. Define a homomorphism $\phi : G \to \mathbb{Z}$ which sends each generator $a_i, b_i$ of $G$ to the same generator of $\mathbb{Z}$. Let $K := Ker(\phi)$. Then $K$ is of type $\mathbf{F}_{n-1}$ but not of type $\mathbf{F}_n$.

Analogously to Corollary 6.40 we obtain:

Theorem 6.42 (See 1.C2 in [**Gro93**]). *Type $\mathbf{F}_n$ is a QI invariant.*

Proof. Our argument is similar to the proof of Corollary 6.40, except we cannot rely on $n-1$-connectivity of Rips complexes $\mathrm{Rips}_R(G)$ for large $R$. If $G$ has type $\mathbf{F}_n$ then it admits a free cellular action $G \curvearrowright X$ on some (possibly infinite-dimensional) $n-1$-connected cell complex $X$ so that the quotient of each skeleton is a finite complex. By combining Lemma 6.34 and Theorem 6.35, we see that the group $G$ is coarsely $n-1$-connected. It remains, then to prove

Proposition 6.43. *If $G$ is a coarsely $n-1$-connected group, then $G$ has type $\mathbf{F}_n$.*

Proof. Note that we already proved this statement for $n = 2$: Coarsely simply-connected groups are finitely-presented (Corollary 6.40). The proof below follows [**KK05**].

Our goal is to build the complex $X$ on which $G$ would act as required by the definition of type $\mathbf{F}_n$. We construct this complex and the action by induction on skeleta $X^{(0)} \subset ... \subset X_{n-1} \subset X^n$. Furthermore, we will inductively construct cellular $G$-equivariant maps $f : X^{(i)} \to Y_{R_i} = \mathrm{Rips}_{R_i}(G)$ and equivariant "deformation retractions" $\rho_i : Y_{R_i}^{(i)} \to X^{(i)}, i = 0, ..., n$, which are $G$-equivariant cellular maps so that composition $h_i = \rho_i \circ f_i : X^{(i)} \to X^{(i)}$ is homotopic to the identity for $i = 0, ..., n-1$. We first explain the construction in the case when $G$ is torsion-free and then show how to modify the construction for groups with torsion.

**Torsion-free case.** In this case $G$-action on every Rips complex is free and cocompact. The construction is by induction on $i$.

$i = 0$. We let $X^{(0)} = G, R_0 = 0$ and let $f_0 = \rho_0 : G \to G$ be the identity map.

$i = 1$. We let $R_1 = 1$ and let $X_1 = Y_{R_1}^{(1)}$ be the Cayley graph of $G$. Again $f_1 = \rho_1 = Id$.

$i = 2$. According to Lemma 6.39, there exists $R_2$ so that $Y_R$ is simply-connected for all $R \geqslant R_2$. We then take $X_2 := Y_{R_2}^{(2)}$. Again, we let $f_2 = \rho_2 = Id$.

$i \Rightarrow i + 1$. Suppose now that $3 \leqslant i \leqslant n-1$, $X^{(i)}, f_i, \rho_i$ are constructed and $R_i$ chosen; we will construct $X^{(i+1)}, f_{i+1}, \rho_{i+1}$.

We first construct $X^{(i+1)}$.

Lemma 6.44. *There are finitely many spherical $i$-cycles $\sigma_\alpha, \alpha \in A'$, in $X^{(i)}$ such that their $G$-orbits generate $\pi_i(X^{(i)})$.*

Proof. Let $R' > R = R_i$ be such that the map

$$Y_R = \mathrm{Rips}_R(G) \to Y_{R'} = \mathrm{Rips}_{R'}(G)$$

induces zero map on $\pi_k$, $k = 0, ..., i$. Let $\tau_\alpha : S^i \to (Y_R)^{(i)}, \alpha \in A$, denote the attaching maps of the $i+1$-cells $\hat{\tau}_\alpha$ in $Y_{R'}^{(i+1)}$, these maps are just simplicial homeomorphic embeddings from the boundary $S^i$ of the standard $i+1$-simplex to the boundaries of the $i+1$-simplices in $Y_{R'}^{(i)}$. Since the map $H_i(Y_R) \to H_i(Y_{R+1})$ is zero, the spherical cycles $\tau_\alpha, \alpha \in A$, generate the image of the map

$$\eta_i : H_i(Y_R^{(i)}) \to H_i(Y_{R'}^{(i)}).$$

Since the action of $G$ on $Y_R$ is cocompact, there are finitely many of these spherical cycles $\{\tau_\alpha, \alpha \in A'\}$, whose $G$-images generate the entire image of $\eta_i$. We then let

$\sigma_\alpha := \rho_i(\tau_\alpha), \alpha \in A'$. We claim that this finite set of spherical cycles does the job. Note that for every $\sigma \in \pi_i(X^{(i)})$,

$$[f(\sigma)] = \sum_{\alpha \in A'} \sum_{g \in G} z_{g,\alpha} \cdot g([\tau_\alpha)]), \quad g \in G, z_{g,\alpha} \in \mathbb{Z},$$

in the group $H_i(Y_{R'})$. Applying the retraction $\rho_i$ and using the fact that $h_i = \rho_i \circ f_i$ is homotopic to the identity, we get

$$\sigma = \sum_{\alpha \in A'} \sum_{g \in G} z_{g,\alpha} \cdot g([\sigma_\alpha)]). \quad \square$$

We now equivariantly attach $i+1$-cells $\hat{\sigma}_{g,\alpha}$ along the spherical cycles $g(\sigma_\alpha), \alpha \in A'$. We let $X^{(i+1)}$ denote the resulting complex and we extend the $G$-action to $X^{(i+1)}$ in obvious fashion. It is clear that $G \curvearrowright X^{(i+1)}$ is properly discontinuous, free and cocompact. By the construction $X^{(i+1)}$ is $i$-connected.

We next construct maps $f_{i+1}$ and $\rho_{i+1}$. To construct the map $f_{i+1} : X^{(i+1)} \to Y_{R'}$ we extend $f_i|\sigma_{1,\alpha}$ to $\hat{\sigma}_{1,\alpha}$ using the fact that the map

$$\pi_i(Y_R) \to \pi_i(Y_{R'})$$

is trivial. We extend $f_{i+1}$ to the rest of the cells $\hat{\sigma}_{g,\alpha}, \alpha \in A'$, by $G$-equivariance. We extend $\rho_i$ to each $g\hat{\tau}_\alpha$ using the attaching map $g\hat{\sigma}_\alpha$. We extend the map to the rest of $Y_{R'}^{(i+1)}$ by induction on the skeleta, $G$-equivariance and using the fact that $X^{(i+1)}$ is $i$-connected. Lastly, we observe that $h_{i+1}$ is homotopic to the identity. Indeed, for each $i+1$-cell $g(\hat{\sigma}_\alpha)$, the map $f_i(g\sigma_\alpha)$ is homotopic to $g\tau_\alpha$ in $Y_{R'}$ (as $\pi_i(Y_R) \to \pi_i(Y_{R'})$ is zero) and $f_{i+1}(g\hat{\tau}_\alpha) = g(\hat{\sigma}_\alpha)$. (Note that we do not claim that $h_n$ is homotopic to the identity.)

If $n < \infty$, this construction terminates after finitely many steps, otherwise, it takes infinitely many steps. In either case, the result is $n-1$-connected complex $X$ and a free action $G \curvearrowright X$ which is cocompact on each skeleton. This concludes the proof in the case of torsion-free groups $G$.

**General Case.** We now explain what to do in the case when $G$ is not torsion-free. The main problem is that a group $G$ with torsion will not act freely on its Rips complexes. Thus, while equivariant maps $f_i$ would still exist, we would be unable to construct equivariant maps $\rho_i : \mathrm{Rips}_R(G) \to X^{(i)}$. Furthermore, it could happen that for large $R$ the complex $Y_R$ is contractible: This is clearly true if $G$ is finite, it also holds for all Gromov-hyperbolic groups. If were to have $f_i$ and $\rho_i$ as before, we would be able to conclude that $X^{(i)}$ is contractible for large $i$, while a group with torsion cannot act freely on a contractible cell complex.

We, therefore, have to modify the construction. For each $R$ we let $Z_R$ denote the barycentric subdivision of $Y_R^{(i)} = \mathrm{Rips}_R(G)^{(i)}$. Then $G$ acts on $Z_R$ without inversions (see Definition 3.22). Let $\widehat{Z_R}$ denote the regular cell complex obtained by applying the *Borel construction* to $Z_R$, see section 3.2. The complex $\widehat{Z_R}$ is infinite-dimensional if $G$ has torsion, but this does not cause problems since at each step of induction we work only with finite skeleta. The action $G \curvearrowright Z_R$ lifts to a free (properly discontinuous) action $G \curvearrowright \widehat{Z_R}$ which is cocompact on each skeleton. We then can apply the arguments from the torsion-free case to the complexes $\widehat{Z_R}$ instead of $\mathrm{Rips}_R(G)$. The key is that, since the action of $G$ on $\widehat{Z_R}$ is free, the construction of the equivariant retractions $\rho_i : Y_{R_i}^{(i)} \to X^{(i)}$ goes through. Note also that in the first steps of the induction we used the fact that $Y_R$ is simply-connected for sufficiently large $R$ in order to construct $X^{(2)}$. Since the projection $\widehat{Z_R} \to Z_R$ is homotopy-equivalence, 2-skeleton of $\widehat{Z_R}$ is simply-connected for the same values of $R$. □

This finishes the proof of Theorem 6.42 as well. □

There are other group-theoretic finiteness conditions, for instance, the condition $\mathbf{F}P_n$ which is a cohomological analogue of the finiteness condition $\mathbf{F}_n$. The arguments used in this section apply in the context of $\mathbf{FP}_n$-groups as well, see Proposition 11.4 in [**KK05**]. The main difference is that instead of metric cell complexes, one works with metric chain complexes and instead of $k$-connectedness of the system of Rips complexes, one uses acyclicity over $R$.

Theorem 6.45. *Let $R$ be a commutative ring with neutral element. Then the property of being $\mathbf{FP}_n$ over $R$ is QI invariant.*

Question 6.46. 1. Is the homological dimension of a group QI invariant?

2. Suppose that $G$ has geometric dimension $n < \infty$. Is there a bounded geometry uniformly contractible $n$-dimensional metric cell complex with free $G$-action $G \curvearrowright X$?

3. Is geometric dimension QI invariant for torsion-free groups?

Note that cohomological dimension is known to equal geometric dimension, except there could be groups satisfying

$$2 = cd(G) \leq gd(G) \leq 3,$$

see [**Bro82b**]. On the other hand,

$$cd(G) \leq hd(G) \leq cd(G) + 1,$$

see [**Bie76a**]. Here $cd$ stands for cohomological dimension, $gd$ is the geometric dimension and $hd$ is the homological dimension.

## 6.5. Retractions

The goal of this section is to give a non-equivariant version of the construction of the retractions $\rho_i$ from the proof of Proposition 6.43 in the previous section.

Suppose that $X, Y$ are uniformly contractible finite-dimensional metric cell complexes of bounded geometry. Consider a uniformly proper map $f : X \to Y$. Our goal is to define a *coarse left-inverse* to $f$, a *retraction* $\rho$ which maps an $r$-neighborhood of $V := f(X)$ back to $X$.

Lemma 6.47. *Under the above assumptions, there exist numbers $L, L', A$, function $R = R(r)$ which depend only on the distortion function of $f$ and on the geometry of $X$ and $Y$ so that:*

*1. For every $r \in \mathbb{N}$ there exists a cellular $L$–Lipschitz map $\rho = \rho_r : \mathcal{N}_r(V) \to X$ so that* $\operatorname{dist}(\rho \circ f, id_X) \leqslant A$. *Here and below we equip $W^{(0)}$ with the restriction of the path-metric on the metric graph $W^{(1)}$ in order to satisfy Axiom 1 of metric cell complexes.*

*2. $\rho \circ f$ is homotopic to the identity by an $L'$–Lipschitz cellular homotopy.*

*3. The composition $h = f \circ \rho : \mathcal{N}_r(V) \to V \subset \mathcal{N}_R(V)$ is homotopic to the identity embedding $id : V \to \mathcal{N}_R(V)$.*

*4. If $r_1 \leqslant r_2$ then $\rho_{r_2}|\mathcal{N}_{r_1}(V) = \rho_{r_1}$.*

Proof. Let $D_0 = 0, D_1, D_2, ...$ denote the geometric bounds on $Y$ and

$$\max_{k>0} D_k = D < \infty.$$

Since $f$ is uniformly proper, there exists a proper monotonic function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ so that

$$\eta(d(x, x')) \leqslant d(f(x), f(x')), \forall x, x' \in X^{(0)}.$$

Let $A_0, A_1$ denote numbers such that

$$\eta(t) > 0, \quad \forall t > A_0,$$

$$\eta(t) > 2r + D_1, \quad \forall t > A_1,$$

Recall that the neighborhood $W := \bar{\mathcal{N}}_r(V)$ is a subcomplex of $Y$. For each vertex $y \in W^{(0)}$ we pick a vertex $\rho(y) := x \in X^{(0)}$ such that the distance $\mathrm{dist}(y, f(x))$ is the smallest possible. If there are several such points $x$, we pick one of them arbitrarily. The fact that $f$ is uniformly proper, ensures that

$$\mathrm{dist}(\rho \circ f, id_{X^{(0)}}) \leqslant A := A_0.$$

Indeed, if $\rho(f(x)) = x'$, then $f(x) = f(x')$; if $d(x, x') > A_0$, then

$$0 < \eta(d(x, x')) \leqslant d(f(x), f(x')),$$

contradicting that $f(x) = f(x')$. Thus, by our choice of the metric on $W^{(0)}$ coming from $W^1$, we conclude that $\rho$ is $A_1$-Lipschitz.

Next, observe also that for each 1-cell $\sigma$ in $W$, $\mathrm{diam}(\rho(\partial\sigma)) \leqslant A_1$. Indeed, if $\partial\sigma = \{y_1, y_2\}$, $d(y_1, y_2) \leqslant D_1$ by the definition of a metric cell complex. For $y_i' := f(x_i)$, $d(y_i, y_i') \leqslant r$. Thus, $d(y_1', y_2') \leqslant 2r + D_1$ and $d(x_1, x_2) \leqslant A_1$ by the definition of $A_1$. Now, existence of $L$-Lipschitz extension $\rho : W \to X$ follows from Proposition 6.31. This proves (1).

Part (2) follows from Corollary 6.33. To prove Part (3), observe that $h = f \circ \rho : \bar{\mathcal{N}}_r(V) \to V$ is $L''$-Lipschitz (see Exercise 6.25), $\mathrm{dist}(h, Id) \leqslant r$. Now, (3) follows from Corollary 6.33 since $Y$ is also uniformly contractible.

Lastly, in order to guarantee (4), we can construct the retractions $\rho_r$ by induction on the values of $r$ and using the extension Lemma 6.32. ☐

Corollary 6.48. *There exists a function $\alpha(r) \geqslant r$ so that for every $r$ the map $h = f \circ \rho : \mathcal{N}_r(V) \to \mathcal{N}_{\alpha(r)}(V)$ is properly homotopic to the identity, where $V = f(X)$.*

We will think of this lemma and its corollary as a proper homotopy-equivalence between $X$ and the direct system of metric cell complexes $\mathcal{N}_R(V)$, $R \geqslant 1$. Recall that the usual proper homotopy-equivalence induces isomorphisms of compactly supported cohomology groups. In our case we get an "approximate isomorphism" of $H_c^*(X)$ to the inverse system of compactly supported cohomology groups $H_c^*(\mathcal{N}_R(V))$:

Corollary 6.49. *1. The induced maps $\rho_R^* : H_c^*(X) \to H_c^*(\mathcal{N}_R(W))$ are injective.*

*2. The induced maps $\rho_R^*$ are* approximately surjective *in the sense that the subgroup coker$(\rho_{\alpha(R)}^*)$ maps to zero under the map induced by restriction map*

$$rest_R : H_c^*(\mathcal{N}_{\alpha(R)}(V)) \to H_c^*(\mathcal{N}_R(V)).$$

Proof. 1. Follows from the fact that $\rho\circ f$ is properly homotopic to the identity and, hence, induces the identity map of $H_c^*(X)$, which means that $f^*$ is the right-inverse to $\rho_R^*$.

2. By Corollary 6.48 the restriction map $rest_R$ equals the map $\rho_R^* \circ f^*$. Therefore, the cohomology group $H_c^*(\mathcal{N}_{\alpha(R)}(W))$ maps *via* $rest_R$ to the image of $\rho_R^*$. The second claim follows. □

### 6.6. Poincaré duality and coarse separation

In this section we discuss coarse implications of Poincaré duality in the context of triangulated manifolds. For a more general version of Poincaré duality, we refer the reader to [**Roe03**]; this concept was coarsified in [**KK05**], where coarse Poincaré duality was introduced and used in the context of metric cell complexes. We will be working work with metric cell complexes which are simplicial complexes, the main reason being that Poincaré duality has cleaner statement in this case.

Let $X$ be a connected simplicial complex of bounded geometry which is a triangulation of a (possibly noncompact) $n$-dimensional manifold without boundary. Suppose that $W \subset X$ is a subcomplex, which is a triangulated manifold (possibly with boundary). We will use the notation $W'$ to denote its barycentric subdivision. We then have the Poincaré duality isomorphisms

$$P_k : H_c^k(W) \to H_{n-k}(W, \partial W) = H_{n-k}(X, X \setminus W).$$

Here $H_c^*$ are cohomology groups with compact support. The Poincaré duality isomorphisms are *natural* in the sense that they commute with proper embeddings of manifolds and manifold pairs. Furthermore, the isomorphisms $P_k$ move cocycles by *uniformly bounded amount*: Suppose that $\zeta \in Z_c^k(W)$ is a simplicial cocycle supported on a compact subcomplex $K \subset W$. Then the corresponding relative cycle $P_k(\zeta) \in Z_{n-k}(W, \partial W)$ is represented by a simplicial chain in $W'$ where each simplex has nonempty intersection with $K$.

Exercise 6.50. If $W \subsetneq X$ is a proper subcomplex, then $H_c^n(W) = 0$.

We will also have to use the Poincaré duality in the context of subcomplexes $V \subset X$ which are not submanifolds with boundary. Such $V$, nevertheless, admits a (closed) regular neighborhood $W = \mathcal{N}(V)$, which is a submanifold with boundary. The neighborhood $W$ is homotopy-equivalent to $V$.

We will present in this section two applications of Poincaré duality to the coarse topology of $X$.

**Coarse surjectivity**

Theorem 6.51. *Let $X, Y$ be uniformly contractible simplicial complexes of bounded geometry homeomorphic to $\mathbb{R}^n$. Then every uniformly cellular proper map $f : X \to Y$ is surjective.*

Proof. Assume to the contrary, i.e, $V = f(X) \neq Y$ is a proper subcomplex. Thus, $H_c^n(V) = 0$ by Exercise 6.50. Let $\rho : V \to X$ be a retraction constructed in Lemma 6.47. By Lemma 6.47, the composition $h = \rho \circ f : X \to X$ is properly homotopic to the identity. Thus, this map has to induce an isomorphism $H_c^*(X) \to H_c^*(X)$. However, $H_c^n(X) \cong \mathbb{Z}$ since $X$ is homeomorphic to $\mathbb{R}^n$, while $H_c^n(V) = 0$. Contradiction. □

Corollary 6.52. *Let $X, Y$ be as above an $f : X^{(0)} \to Y^{(0)}$ is a quasi-isometric embedding. Then $f$ is a quasi-isometry.*

Proof. Combine Proposition 6.31 with Theorem 6.51. $\square$

**Coarse separation.**

Suppose that $X$ is a simplicial complex and $W \subset X$ is a subcomplex. Consider, $\mathcal{N}_R(W)$, the open metric $R$-neighborhoods of $W$ in $X$ and their complements $C_R$ in $X$.

For a component $C \subset C_R$ define the *inradius*, $inrad(C)$, of $C$ to be the supremum of radii of metric balls $\mathbf{B}(x, R)$ in $X$ contained in $C$. A component $C$ is called *shallow* if $inrad(C)$ is $< \infty$ and *deep* if $inrad(C) = \infty$.

Example 6.53. Suppose that $W$ is compact. Then deep complementary components of $C_R$ are components of infinite diameter. These are the components which appears as neigborhoods of ends of $X$.

A subcomplex $W$ is said to *coarsely separate* $X$ if there is $R$ such that $\mathcal{N}_R(W)$ has at least two distinct deep complementary components.

Example 6.54. The simple properly embedded curve $\Gamma$ in $\mathbb{R}^2$ need not coarsely separate $\mathbb{R}^2$ (see Figure 6.4). A straight line in $\mathbb{R}^2$ coarsely separates $\mathbb{R}^2$.

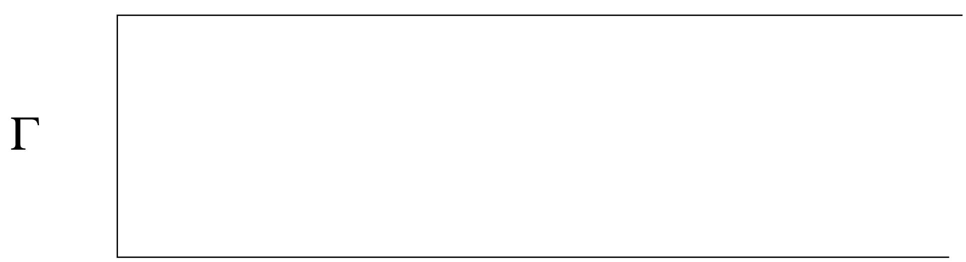



Figure 6.4. A separating curve which does not coarsely separate the plane.

Theorem 6.55. *Suppose that $X, Y$ are uniformly contractible simplicial complexes of bounded geometry which are homeomorphic to $\mathbb{R}^{n-1}$ and $\mathbb{R}^n$ respectively. Then for each uniformly proper cellular map $f : X \to Y$, the image $V = f(X)$ coarsely separates $Y$. Moreover, for all sufficiently large $R$, $Y \setminus \mathcal{N}_R(V)$ has exactly two deep components.*

Proof. Actually, our proof will use the assumption on the topology of $X$ only weakly: To get coarse separation it suffices to assume that $H_c^{n-1}(X) \neq 0$.

Recall that in Section 6.5 we constructed a system of retractions $\rho_R : \mathcal{N}_R(V) \to X$, $R \in \mathbb{N}$, and proper homotopy-equivalences $f \circ \rho \equiv Id$ and $\rho_R \circ f|_{\mathcal{N}_R(V)} \equiv Id : \mathcal{N}_R(V) \to \mathcal{N}_{\alpha(R)}(V)$. Furthermore, we have the *restriction maps*

$$rest_{R_1,R_2} : H_c^*(\bar{\mathcal{N}}_{R_2}(V)) \to H_c^*(\bar{\mathcal{N}}_{R_1}(V)), \quad R_1 \leqslant R_2.$$

These maps satisfy

$$rest_{R_1,R_2} \circ \rho_{R_2}^* = \rho_{R_1}^*$$

by Part 4 of Lemma 6.47. We also have the projection maps

$$proj_{R_1,R_2} : H_*(Y, Y - \bar{\mathcal{N}}_{R_2}(V)) \to H_*(Y, Y - \bar{\mathcal{N}}_{R_1}(V)) \quad R_1 \leqslant R_2.$$

induced by inclusion maps of pairs $(Y, Y - \bar{\mathcal{N}}_{R_2}(V)) \hookrightarrow (Y, Y - \bar{\mathcal{N}}_{R_1}(V))$. Poincaré duality in $\mathbb{R}^n$ also gives us a system of isomorphisms

$$P : H_c^{n-1}(\bar{\mathcal{N}}_R(V)) \cong H_1(X, X \setminus \mathcal{N}_R(V)).$$

By naturality of Poincaré duality we have a commutative diagram:

$$\begin{array}{ccc} H_c^*(\bar{\mathcal{N}}_{R_2}(V)) & \xrightarrow{P} & H_{n-*}(Y, C_{R_2}) \\ \Big\downarrow rest_{R_1,R_2} & & \Big\downarrow proj_{R_1,R_2} \\ H_c^*(\bar{\mathcal{N}}_{R_1}(V)) & \xrightarrow{P} & H_{n-*}(Y, C_{R_1}) \end{array}$$

Let $\omega$ be a generator of $H_c^{n-1}(X) \cong \mathbb{R}$. Given $R > 0$ consider the pull-back $\omega_R := \rho_R^*(\omega)$ and the relative cycle $\sigma_R = P(\omega_R)$. Then $\omega_r = rest_{r,R}(\omega_R)$ and

$$\sigma_r = proj_{r,R}(\sigma_R) \in H_1(Y, C_r)$$

for all $r < R$, see Figure 6.5. Observe that for every $r$, $\omega_r$ is non-zero, since $f^* \circ \rho^* = id$ on the compactly supported cohomology of $X$. Hence, every $\sigma_r$ is nonzero as well.

Contractibility of $Y$ and the long exact sequence of the homology groups of the pair $(Y, C_r)$ implies that

$$H_1(Y, C_r) \cong \tilde{H}_0(C_r).$$

We let $\tau_r$ denote the image of $\sigma_r$ under this isomorphism. Thus, each $\tau_r$ is represented by a 0-cycle, the boundary of the chain representing $\sigma_r$. Running the Poincaré duality in the reverse and using the fact that $\omega$ is a generator of $H_c^{n-1}(X)$, we see that $\tau_r$ is represented by the difference $y_r' - y_r''$, where $y_r', y_r'' \in C_r$. Nontriviality of $\tau_r$ means that $y_r', y_r''$ belong to distinct components $C_r', C_r''$ of $C_r$. Furthermore, since for $r < R$,

$$proj_{r,R}(\sigma_R) = \sigma_r,$$

it follows that

$$C_R' \subset C_r', \quad C_R'' \subset C_r''.$$

Since this could be done for arbitrarily large $r, R$, we conclude that components $C_r', C_r''$ are both deep.

The same argument run in the reverse implies that there are exactly two deep complementary components. □

We refer to [**FS96**], [**KK05**] for further discussion and generalization of coarse separation and coarse Poincaré/Alexander duality.

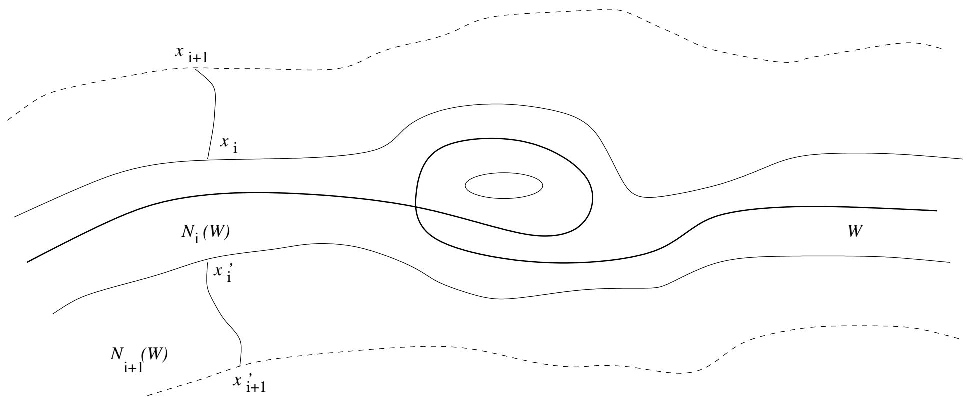



FIGURE 6.5. Coarse separation.

CHAPTER 7

# Ultralimits of Metric Spaces

Let $(X_i)_{i\in I}$ be an indexed family of metric spaces. One can describe the limiting behavior of the family $(X_i)$ by studying limits of indexed families of finite subsets $Y_i \subset X_i$. Ultrafilters are an efficient technical device for simultaneously taking limits of all such families of subspaces and putting them together to form one object, namely an ultralimit of $(X_i)$.

## 7.1. The axiom of choice and its weaker versions

We first recall that the *Zermelo-Fraenkel axioms (ZF)* form a list of axioms which are the basis of axiomatic set theory in its standard form. See for instance [**Kun80**], [**HJ99**], [**Jec03**].

The Axiom of Choice (AC) can be seen as a rule of building sets out of other sets. It was first formulated by E. Zermelo in [**Zer04**]. According to work of K. Gödel and P. Cohen, the axiom of choice is logically independent of the axioms of Zermelo-Fraenkel (i.e. neither it nor its negation can be proven in ZF).

Given a non-empty collection $\mathcal{S}$ of non-empty sets, a *choice function* defined on $\mathcal{S}$ is a function $f : \mathcal{S} \to \cup_{A\in\mathcal{S}} A$ such that for every set $A$ in $\mathcal{S}$, $f(A)$ is an element of $A$. A choice function on $\mathcal{S}$ can be viewed as an element of the Cartesian product $\prod_{A\in\mathcal{S}} A$.

**Axiom of choice**
*On any non-empty collection of non-empty sets one can define a choice function. Equivalently, an arbitrary Cartesian product of non-empty sets is non-empty.*

REMARK 7.1. If $\mathcal{S} = \{A\}$ then the existence of $f$ follows from the fact that $A$ is non-empty. If $\mathcal{S}$ is finite or countable the existence of a choice function can be proved by induction. Thus if the collection $\mathcal{S}$ is finite or countable then the existence of a choice function follows from ZF.

REMARK 7.2. Assuming ZF, the Axiom of choice is equivalent to each of the following statements (see [**HJ99**] and [**RR85**] for a much longer list):

(1) *Zorn's lemma.*
(2) Every vector space has a basis.
(3) Every ideal in a unitary ring is contained in a maximal ideal.
(4) If $A$ is a subset in a topological space $X$ and $B$ is a subset in a topological space $Y$ the closure of $A \times B$ in $X \times Y$ is equal to the product of the closure of $A$ in $X$ with the closure of $B$ in $Y$.
(5) (*Tychonoff's theorem:*) If $(X_i)_{i\in I}$ is a collection of non-empty compact topological spaces then $\prod_{i\in I} X_i$ is compact.

Remark 7.3. The following statements require the Axiom of Choice, see [**HJ99, RR85**]:

(1) every union of countably many countable sets is countable;
(2) The Nielsen–Schreier theorem: Every subgroup of a free group is free (see Corollary 4.74), to ensure the existence of a maximal subtree.

In ZF, we have the following irreversible sequence of implications:

Axiom of choice ⇒ Ultrafilter Lemma ⇒ the Hahn–Banach extension theorem.

The first implication is easy (see Lemma 7.16), it was proved to be irreversible in [**Hal64**]. The second implication is proved in ([**ŁRN51**], [**Lux62**], [**Lux67**], [**Lux69**]). Its irreversibility is proved in [**Pin72**] and [**Pin74**].

Thus, the Hahn–Banach extension theorem (see below) can be seen as the analyst's Axiom of Choice, in a weaker form.

Theorem 7.4 (Hahn–Banach Theorem, see e.g. [**Roy68**]). *Let $V$ be a real vector space, $U$ a subspace of $V$, and $\varphi : U \to \mathbb{R}$ a linear function. Let $p : V \to \mathbb{R}$ be a map with the following properties:*

$$p(\lambda x) = \lambda p(x) \text{ and } p(x+y) \leqslant p(x) + p(y)\,,\ \forall x, y \in V,\ \lambda \in [0, +\infty)\,,$$

*such that $\varphi(x) \leqslant p(x)$ for every $x \in U$. Then there exists a linear extension of $\varphi$, $\overline{\varphi} : V \to \mathbb{R}$ such that $\overline{\varphi}(x) \leqslant p(x)$ for every $x \in V$.*

Definition 7.5. A *filter* $\mathcal{F}$ on a set $I$ is a collection of subsets of $I$ satisfying the following conditions:

($F_1$) $\emptyset \notin \mathcal{F}$;
($F_2$) If $A, B \in \mathcal{F}$ then $A \cap B \in \mathcal{F}$;
($F_3$) If $A \in \mathcal{F}$, $A \subseteq B \subseteq I$, then $B \in \mathcal{F}$.

Exercise 7.6. Given an infinite set $I$, prove that the collection of all complementaries of finite sets is a filter on $I$. This filter is called *the Fréchet filter* (or the *Zariski filter*, which is used to define the Zariski topology on the affine line).

Definition 7.7. Subsets $A \subset I$ which belong to a filter $\mathcal{F}$ are called $\mathcal{F}$*-large.* We say that a property (P) holds for $\mathcal{F}$*-all* $i$ if (P) is satisfied for all $i$ in some $\mathcal{F}$-large set.

Definition 7.8. A *base of a filter* on a set $I$ is a set $\mathcal{B}$ of subsets of $I$ which satisfies the properties:

($B_1$) If $B_i \in \mathcal{B}$, $i = 1, 2$, then $B_1 \cap B_2$ contains a subset of $\mathcal{B}$;
($B_2$) $\emptyset \notin \mathcal{B}$ and $\mathcal{B}$ is not empty.

Exercise 7.9. If $\mathcal{B}$ is a base of a filter, then the set of subsets of $I$ containing some $B \in \mathcal{B}$ is a filter.

Definition 7.10. An *ultrafilter* on a set $I$ is a filter $\mathcal{U}$ on $I$ which is a maximal element in the ordered set of all filters on $I$ with respect to the inclusion. Equivalently, an ultrafilter can be defined [**Bou65**, §$I$.6.4] as a collection of subsets of $I$ satisfying the conditions ($F_1$), ($F_2$), ($F_3$) defining a filter and the additional condition:

$$\text{For every } A \subseteq I \text{ either } A \in \mathcal{U} \text{ or } I \setminus A \in \mathcal{U}. \qquad (F_4)$$

Exercise 7.11. Given a set $I$, take a point $x \in I$ and consider the collection $\mathcal{U}_x$ of subsets of $I$ containing $x$. Prove that $\mathcal{U}_x$ is an ultrafilter on $I$.

Exercise 7.12. Given the set $\mathbb{Z}$ of integers, prove, using Zorn's lemma, that there exists an ultrafilter containing all the non-trivial subgroups of $\mathbb{Z}$. Such an ultrafilter is called *profinite ultrafilter*.

Definition 7.13. An ultrafilter as in Exercise 7.11 is called *a principal (or atomic) ultrafilter*. A filter that cannot be defined in such a way is called a *a non-principal (or non-atomic) ultrafilter*.

Proposition 7.14. *An ultrafilter on an infinite set $I$ is non-principal if and only if it contains the Fréchet filter.*

Proof. We will prove the equivalence between the negations of the two statements.

A principal ultrafilter $\mathcal{U}$ on $I$ defined by a point $x$ contains $\{x\}$ hence by $(F_4)$ it does not contain $I \setminus \{x\}$ which is an element of the Fréchet filter.

Let now $\mathcal{U}$ be an ultrafilter that does not contain the Fréchet filter. This and property $(F_4)$ implies that it contains a finite subset $F$ of $I$.

If $F \cap \bigcap_{A\in\mathcal{U}} A = \emptyset$ then there exist $A_1, \dots, A_n \in \mathcal{U}$ such that $F \cap A_1 \cap \cdots \cap A_n = \emptyset$. This and property $(F_2)$ contradict property $(F_1)$.

It follows that $F \cap \bigcap_{A\in\mathcal{U}} A = F_1 \neq \emptyset$, in particular, given an element $x \in F_1$, $\mathcal{U}$ is contained in the principal ultrafilter $\mathcal{U}_x$. The maximality of $\mathcal{U}$ implies that $\mathcal{U} = \mathcal{U}_x$. □

Exercise 7.15.
(1) Let $S$ be an infinite subset of $I$. Prove (using Zorn's lemma) that there exists a non-principal ultrafilter $\mathcal{U}$ so that $S \in \mathcal{U}$.
(2) Let $\mathcal{S}_1 \supset \mathcal{S}_2 \supset \mathcal{S}_3 \supset \cdots \supset \mathcal{S}_m \supset \dots$ be an infinite sequence of infinite subsets of $I$. Prove that there exists a non-principal ultrafilter containing all $\mathcal{S}_m$, $\forall m \in \mathbb{N}$, as it elements.

Lemma 7.16 (The Ultrafilter Lemma). *Every filter on a set $I$ is a subset of some ultrafilter on $I$.*

Proof. Let $\mathcal{F}$ be the Fréchet filter of $I$. By Zorn's lemma, there exists a maximal filter $\mathcal{U}$ on $I$ containing $\mathcal{F}$. By maximality, $\mathcal{U}$ is an ultrafilter; $\mathcal{U}$ is nonprincipal by Proposition 7.14. □

In ZF, the Axiom of Choice is equivalent to Zorn's Lemma, and the latter clearly implies the Ultrafilter Lemma.

Definition 7.17. Equivalently, one can define *an ultrafilter on a set $I$* as a finitely additive measure $\omega$ defined on $\mathcal{P}(I)$ (the power set of $I$), taking only values 0 and 1 and such that $\omega(I) = 1$. Indeed, $\omega$ satisfies the previous properties if and only if it is the characteristic function $\mathbf{1}_{\mathcal{U}}$ of a collection $\mathcal{U}$ of subsets of $I$ which is an ultrafilter.

Note that for an atomic ultrafilter $\mathcal{U}_x$ defined as in Example 7.11, the corresponding measure is the Dirac measure $\delta_x$.

Definition 7.18. A *non-principal ultrafilter on a set $I$* is a finitely additive measure $\omega : \mathcal{P}(I) \to \{0, 1\}$ such that $\omega(I) = 1$ and $\omega(F) = 0$ for every finite subset $F$ of $I$.

Exercise 7.19. Prove the equivalence between Definitions 7.10 and 7.17, and between Definitions 7.13 and 7.18.

Remarks 7.20. (1) If $\omega(A_1 \sqcup \cdots \sqcup A_n) = 1$, then there exists $i_0 \in \{1, 2, \ldots, n\}$ such that $\omega(A_{i_0}) = 1$ and $\omega(A_j) = 0$ for every $j \neq i_0$.

(2) If $\omega(A) = 1$ and $\omega(B) = 1$ then $\omega(A \cap B) = 1$.

Notation 7.21. Let $(A_i)_{i \in I}$ and $(B_i)_{i \in I}$ be two sequences of sets indexed by $I$ and let $\mathcal{R}$ be a relation that exists between $A_i$ and $B_i$ for every $i \in I$. We write $A_i \, \mathcal{R}_\omega \, B_i$ if and only if $A_i \, \mathcal{R} \, B_i$ $\omega$-almost surely, that is

$$\omega\left(\{i \in I \mid A_i \, \mathcal{R} \, B_i\}\right) = 1 .$$

*Examples:* $=_\omega$ , $<_\omega$ , $\subset_\omega$.

For more details on filters and ultrafilters see [**Bou65**, §$I$.6.4].

We now explain how existence of non-principal ultrafilters implies Hahn–Banach in the following special case: $V$ is the real vector space of bounded sequences of real numbers $\mathbf{x} = (x_n)$, $U \subset V$ is the subspace of convergent sequences of real numbers, $p$ is the sup-norm

$$\|\mathbf{x}\|_\infty = \sup_{n \in \mathbb{N}} |x_n|$$

and $\varphi : U \to \mathbb{R}$ is the limit function, i.e.

$$\varphi(\mathbf{x}) = \lim_{n \to \infty} x_n .$$

In other words, we will show how, using a non-principal ultrafilter, one can extend the notion of limit from convergent sequences to bounded sequences.

Definition 7.22. [Ultralimit of a function] Given a function $f : I \to Y$ (where $Y$ is a topological space) define the $\omega$–limit

$$\omega\text{-}\lim_i f(i)$$

to be a point $y \in Y$ such that for every neighborhood $U$ of $y$, the pre-image $f^{-1}U$ belongs to $\omega$. The point $y$ is called the *ultralimit* of the function $f$.

Note that, in general, an ultralimit need not be unique. However, it is unique in the case when $Y$ is Hausdorff.

Lemma 7.23. *Suppose that $Y$ is compact and Hausdorff. Then for each function $f : I \to Y$ the ultralimit exists and is unique.*

Proof. To prove existence of a limit, assume that there is no point $y \in Y$ satisfying the definition of the ultralimit. Then each point $z \in Y$ possesses a neighborhood $U_z$ such that $f^{-1}U_z \notin \omega$. By compactness, we can cover $Y$ with finitely many of these neighborhoods $U_{z_i}$, $i = 1, \ldots, n$. Therefore,

$$I = \bigcup_{i=1}^n f^{-1}(U_{z_i})$$

and, thus

$$\emptyset = \bigcap_{i=1}^n (I \setminus f^{-1}(U_{z_i})) \in \omega.$$

This contradicts the definition of a filter. Uniqueness of the point $y$ follows, because $Y$ is Hausdorff. □

Note that the $\omega$–limit satisfies the usual "calculus" properties, e.g. linearity. In particular, the above lemma implies Hahn–Banach theorem for $U$, the space of convergent sequences, $V$ the space of all bounded sequences and $p := \lim$.

Exercise 7.24. Show that the $\omega$–limit of a function $f : I \to Y$ is an accumulation point of $f(I)$.

Conversely, if $y$ is an accumulation point of $\{f(i)\}_{i \in I}$ then there is a non-principal ultrafilter $\omega$ with $\omega\text{-}\lim f = y$, namely an ultrafilter containing the pull-back of the neighborhood basis of $y$.

Thus, an ultrafilter is a device to select an accumulation point for any set $A$ contained in a compact Hausdorff space $Y$, in a coherent manner.

Note that when the ultrafilter is principal, that is $\omega = \delta_{i_0}$ for some $i_0 \in I$, and $Y$ is Hausdorff the $\delta_{i_0}$–limit of a function $f : I \to Y$ is simply the element $f(i_0)$, so not very interesting. Thus when considering $\omega$–limits we shall always choose the ultrafilter $\omega$ to be non-principal.

Remark 7.25. Recall that when we have a countable collection of sequences

$$\mathbf{x}^{(k)} = \left(x_n^{(k)}\right)_{n \in \mathbb{N}}, k \in \mathbb{N}, x_n^{(k)} \in X,$$

where $X$ is a compact space, we can select a subset of indices $I \subset \mathbb{N}$, such that for every $k \in \mathbb{N}$ the subsequence $\left(x_i^{(k)}\right)_{i \in I}$ converges. This is achieved by the diagonal procedure. The $\omega$–limit allows, in some sense, to do the same for an uncountable collection of (uncountable) sets. Thus, it can be seen as an uncountable version of the diagonal procedure.

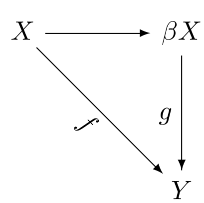
Note also that for applications in Geometric Group Theory, most of the time, one considers only countable index sets $I$. Thus, in principle, one can avoid using ultrafilters at the expense of getting complicated proofs involving passage to multiple subsequences.

Using ultralimits of maps we will later define ultralimits of sequences of metric spaces; in particular, given metric space $(X, \mathrm{dist})$, we will define an image of $(X, \mathrm{dist})$ seen from infinitely far away (the asymptotic cone of $(X, \mathrm{dist})$). Ultralimits and asymptotic cones will be among key technical tools used in this book. More details on this will appear in Chapter 7.

## 7.2. Ultrafilters and Stone–Čech compactification

Let $X$ be a Hausdorff topological space. The *Stone–Čech compactification* of $X$ is a pair consisting of a compact Hausdorff topological space $\beta X$ and a continuous map $X \to \beta X$ which satisfies the following universal property:

For every continuous map $f : X \to Y$, where $Y$ is a compact Hausdorff space, there exists a unique continuous map $g : X \to Y$, such that the following diagram commutes:



This universal property implies uniqueness of the Stone–Čech compactification.

EXERCISE 7.26. Show that $X \to \beta X$ is injective and its image is dense in $\beta X$.

In view of this exercise, we will regard $X$ as a subset of $\beta X$, so that $\beta X$ is a compactification of $X$.

We will now explain how to construct $\beta X$ using ultrafilters, provided that $X$ has discrete topology, e.g. $X = \mathbb{N}$. Declare $\beta X$ to be the set of all ultrafilters on $X$. Then, $\beta X$ is a subset of the power set

$$2^{2^X},$$

since every ultrafilter $\omega$ on $X$ is a subset of $2^X$. We equip $2^X$ and, hence, $2^{2^X}$, with the product topology and the subset $\beta X \subset 2^{2^X}$ with the subspace topology.

EXERCISE 7.27. Show that the subset $\beta X \subset 2^{2^X}$ is closed. Thus, by Tychonoff's theorem, $\beta X$ is compact. Since $X$ is Hausdorff, so is $2^X$ and, hence, $2^{2^X}$.

Every $x \in X$ determines the principal ultrafilter $\delta_x$; thus, we obtain an embedding $X \hookrightarrow \beta X$. This embedding is continuous since $X$ has discrete topology. Therefore, from now, on we will regard $X$ as a subset of $\beta X$.

EXERCISE 7.28. Let $\omega \in \beta X$ be a non-principal ultrafilter. Show that for every neighborhood $U$ of $\omega$ in $\beta X$, the intersection $X \cap U$ is an $\omega$-large set. Conversely, for every $\omega$-large set $A \subset X$, there exists a neighborhood $U$ of $\omega$ in $\beta X$ so that $A = U \cap X$. In particular, $X$ is dense in $\beta X$.

We will now verify the universal property of $\beta X$. Let $f : X \to Y$ be a continuous map to a compact Hausdorff space. For every $\omega \in \beta X \setminus X$ we set

$$g(\omega) := \omega\text{-}\lim f.$$

By definition of the ultralimit of a map, for every point $y \in Y$ and its neighborhood $V$ in $Y$, the preimage $A = f^{-1}(V)$ is $\omega$-large. Therefore, by Exercise 7.28, there exists a neighborhood $U$ of $\omega$ in $\beta X$ so that $A = U \cap X$. This proves that the map $g$ is continuous. Hence, $g$ is the required continuous extension of $f$. Uniqueness of $g$ follows from the fact that $X$ is dense in $\beta X$.

## 7.3. Elements of nonstandard algebra

Given an ultrafilter $\omega$ on $I$ and a collection of sets $X_i, i \in I$, define the *ultraproduct*

$$\prod_{i \in I} X_i/\omega$$

to be the collection of equivalence classes of maps $f : I \to \bigcup_{i \in I} X_i$ with $f(i) \in X_i$ for every $i \in I$, with respect to the equivalence relation $f \sim g$ defined by the property that $f(i) = g(i)$ for $\omega$–all $i$.

The equivalence class of a map $f$ is denoted by $f^\omega$. When the map is given by the indexed family of values $(x_i)_{i \in I}$, where $x_i = f(i)$, we also use the notation $(x_i)^\omega$ for the equivalence class.

When $X_i = X$ for all $i \in I$ the ultraproduct is called *ultrapower of* $X$ and denoted by $X^\omega$.

Our discussion here follows [**Gol98**], [**VdDW84**].

Note that any structure on $X$ (group, ring, order, total order) defines the same structure on $X^\omega$, e.g., if $G$ is a group then $G^\omega$ is a group, etc. When $X = \mathbb{K}$ is either $\mathbb{N}, \mathbb{Z}$ or $\mathbb{R}$, the ultrapower $\mathbb{K}^\omega$ is sometimes called the *nonstandard extension of* $\mathbb{K}$, and the elements in $\mathbb{K}^\omega \setminus \mathbb{K}$ are called *nonstandard elements.* If $X$ is totally ordered then $X^\omega$ is totally ordered as well: $f^\omega \leqslant g^\omega$ (for $f, g \in X^\omega$) if and only if $f(i) \leqslant_\omega g(i)$, with the Notation 7.21.

Every subset $A$ of $X$ can be embedded into $X^\omega$ by $a \mapsto (a)^\omega$. We denote its image by $\widehat{A}$. We denote the image of each element $a \in A$ by $\widehat{a}$.

Thus, we define the ordered semigroup $\mathbb{N}^\omega$ (the nonstandard natural numbers) and the ordered field $\mathbb{R}^\omega$ (the nonstandard real numbers).

Definition 7.29. An element $R \in \mathbb{R}^\omega$ is called *infinitely large* if given any $r \in \mathbb{R} \subset \mathbb{R}^\omega$, one has $R \geqslant \widehat{r}$. Note that given any $R \in \mathbb{R}^\omega$ there exists $n \in \mathbb{N}^\omega$ such that $n > R$.

Exercise 7.30. Prove that $R = (R_i)^\omega \in \mathbb{R}^\omega$ is infinitely large if and only if $\omega - \lim (R_i) = +\infty$ .

Definition 7.31 (Internal subsets). A subset $W^\omega \subset X^\omega$ is called *internal* if "membership in $W$ can be determined by coordinate-wise computation", i.e. if for each $i \in I$ there is a subset $W_i \subset X$ such that for $f \in X^I$

$$f^\omega \in W^\omega \iff f(i) \in_\omega W_i \, .$$

(Recall that the latter means that $f(i) \in W_i$ for $\omega$–all $i$.) The sets $W_i$ are called *coordinates* of $W$. We write $W^\omega = (W_i)^\omega$.

Lemma 7.32. (1) *If an internal subset $A^\omega$ is defined by a family of subsets of bounded cardinality $A_i = \{a_i^1, \dots, a_i^k\}$ then $A^\omega = \{a_\omega^1, \dots, a_\omega^k\}$, where $a_\omega^j = \left(a_i^j\right)^\omega$.*

(2) *In particular, if an internal subset $A^\omega$ is defined by a constant family of finite subsets $A_i = A \subseteq X$ then $A^\omega = \widehat{A}$.*

(3) *Every finite subset in $X^\omega$ is internal.*

Proof. (1) Let $x = (x_i)^\omega \in A^\omega$. The set of indices decomposes as $I = I_1 \sqcup \cdots \sqcup I_k$, where $I_j = \left\{ i \in I \, ; \, x_i = a_i^j \right\}$. Then there exists $j \in \{1, \dots, k\}$ such that $\omega(I_j) = 1$, that is $x_i =_\omega a_i^j$, and $x = a_\omega^j$.

(2) is an immediate consequence of (1).

(3) Let $U$ be a subset in $X^\omega$ of cardinality $k$, and let $x_1, \dots, x_k$ be its elements. Each element $x_r$ is of the form $(x_i^r)^\omega$ and $\omega$-almost surely $x_i^r \neq x_i^s$ when $r \neq s$. Therefore $\omega$-almost surely the set $A_i = \{x_i^1, \dots, x_i^k\}$ has cardinality $k$. It follows that $A^\omega = (A_i)^\omega$ has cardinality $k$, according to (1), and it contains $U$. Therefore $U = A^\omega$. □

Lemma 7.33. *If $A$ is an infinite subset in $X$ then $\widehat{A}$ is not internal.*

Proof. Assume $\widehat{A} = (B_i)^\omega$ for a family $(B_i)_{i \in I}$ of subsets. For every $a \in A$, $\hat{a} \in (B_i)^\omega$, i.e.

$$a \in B_i \quad \omega - \text{almost surely.} \tag{7.1}$$

Take an infinite sequence $a_1, a_2, \ldots, a_k, \ldots$ of distinct elements in $A$. Consider the nested sequence of sets $I_k = \{i \in I \mid \{a_1, a_2, \ldots, a_k\} \subseteq B_i\}$. From (7.1) and Remark 7.20, (2), it follows that $\omega(I_k) = 1$ for every $k$.

The intersection $J = \bigcap_{n \geqslant 1} I_k$ has $\omega$-measure either 0 or 1. Assume first that $\omega(J) = 0$. Since $I_1 = \bigsqcup_{k=1}^{\infty} (I_k \setminus I_{k+1}) \sqcup J$, it follows that $J' = \bigsqcup_{k=1}^{\infty} (I_k \setminus I_{k+1})$ has $\omega(J') = 1$.

Define the indexed family $(x_i)$ such that $x_i = a_k$ for every $i \in I_k \setminus I_{k+1}$. By definition $x_i \in B_i$ for every $i \in J'$. Thus $(x_i)^\omega \in (B_i)^\omega = \widehat{A}$, hence $x_i = a$ $\omega$-a.s. for some $a \in A$.

Let $E = \{i \in I \mid x_i = a\}$, $\omega(E) = 1$. Remark 7.20, (2), implies that $E \cap J' \neq \emptyset$, hence for some $k \in \mathbb{N}$, $E \cap (I_k \setminus I_{k+1}) \neq \emptyset$. For $i \in E \cap (I_k \setminus I_{k+1})$ we have $x_i = a = a_k$.

The fact that $\omega(I_{k+1}) = 1$ implies that $E \cap I_{k+1} \cap J' \neq \emptyset$. Hence $E \cap (I_j \setminus I_{j+1}) \neq \emptyset$ for some $j \geqslant k+1$. For an index $i$ in $E \cap (I_j \setminus I_{j+1})$ we have the equality $x_i = a = a_j$. But as $j > k$, $a_j \neq a_k$, so we obtain a contradiction.

Assume now that $\omega(J) = 1$. Assume that this occurs for every sequence $(a_k)$ of distinct elements in $A$. It follows that $\omega$-almost surely $A \subseteq B_i$.

DEFINITION 7.34 (internal maps). A map $f^\omega : X^\omega \to Y^\omega$ is *internal* if there exists an indexed family of maps $f_i : X_i \to Y_i$, $i \in I$, such that $f^\omega(x^\omega) = (f_i(x_i))^\omega$.

Note that the range of an internal map is an internal set.

For instance given a collection of metric spaces $(X_i, \mathrm{dist}_i)$ one can define a metric $\mathrm{dist}^\omega$ on $X^\omega$ as the internal function $\mathrm{dist}^\omega : X^\omega \times X^\omega \to \mathbb{R}^\omega$ defined by the collection of functions $(\mathrm{dist}_i)$, that is $\mathrm{dist}^\omega : X^\omega \times X^\omega \to \mathbb{R}^\omega$,

$$\mathrm{dist}^\omega\left((x_i)^\omega, (y_i)^\omega\right) = \left(\mathrm{dist}_i(x_i, y_i)\right)^\omega . \tag{7.2}$$

The main problem is that $\mathrm{dist}^\omega$ does not take values in $\mathbb{R}$ but in $\mathbb{R}^\omega$.

Let $(\Pi)$ be a property of a structure on the space $X$ that can be expressed using elements, subsets, $\in, \subset, \subseteq, =$ and the logical quantifiers $\exists, \forall, \wedge$ (and), $\vee$ (or), $\neg$ (not) and $\Rightarrow$ (implies).

The non-standard interpretation $(\Pi)^\omega$ of $(\Pi)$ is the statement obtained by replacing "$x \in X$" with "$x^\omega \in X^\omega$", and "$A$ subset of $X$" with "$A^\omega$ internal subset of $X^\omega$".

THEOREM 7.35 (Łoś' Theorem, see e.g. [**BS69**], [**Kei76**], Chapter 1, [**dDW84**], p.361). *A property $(\Pi)$ is true in $X$ if and only if its non-standard interpretation $(\Pi)^\omega$ is true in $X^\omega$.*

COROLLARY 7.36.
(1) *Every non-empty internal subset in $\mathbb{R}^\omega$ that is bounded from above (below) has a supremum (infimum).*
(2) *Every non-empty internal subset in $\mathbb{N}^\omega$ that is bounded from above (below) has a maximal (minimal) element.*

COROLLARY 7.37 (non-standard induction). *If a non-empty internal subset $A^\omega$ in $\mathbb{N}^\omega$ satisfies the properties:*

- $\widehat{1} \in A^\omega$;
- *for every $n^\omega \in A^\omega$, $n^\omega + 1 \in A^\omega$;*

*then $A^\omega = \mathbb{N}^\omega$.*

EXERCISE 7.38. (1) Give a direct proof of Corollary 7.36, (1), for $\mathbb{R}^\omega$.
(2) Deduce Corollary 7.36 from Theorem 7.35.
(3) Deduce Corollary 7.37 from Corollary 7.36.

Suppose we are given $a_n \in \mathbb{R}^\omega$, where $n \in \mathbb{N}^\omega$. Using the nonstandard induction principle on can define the nonstandard products:

$$a_1 \cdots a_n, n \in \mathbb{N}^\omega,$$

as an internal function $f : \mathbb{N}^\omega \to \mathbb{R}^\omega$ given by $f(1) = a_1,\ f(n+1) = f(n)a_{n+1}$.

Various properties of groups can be characterized in terms of ultrapowers, as explained below and in Chapter 16, Section 16.8.

**Ultrapowers and laws in groups**

It is easy to see that if $G$ satisfies a law then any ultrapower $G^\omega$ of $G$ satisfies the same law. Moreover, the following holds.

LEMMA 7.39 (Lemma 6.15 [**DS05**]). *A group $G$ satisfies a law if and only if there exists a non-principal ultrafilter $\omega$ on $\mathbb{N}$ such that the ultrapower $G^\omega$ does not contain free non-abelian subgroups (and in fact for every non-principal ultrafilter the statement is true).*

PROOF. For the direct implication note that if $G$ satisfies an identity then $G^\omega$ satisfies the same law. This is obviously true even for ultrafilters on an arbitrary infinite set $I$. Since a free nonabelian group cannot satisfy a law, claim follows.

For the converse implication, let $\omega$ be an arbitrary ultrafilter on $\mathbb{N}$, and assume that $G$ does not satisfy any law. Enumerate all words $u_1, u_2, \ldots,$ in two variables $x, y$ and the sequence of iterated left-commutators $v_1 = u_1, v_2 = [u_1, u_2], v_3 = [u_1, u_2, u_3], \ldots$, see Notation 10.26. Since $G$ does not satisfy any law, for every $n$ there exists a pair $(x_n, y_n)$ of elements in $G$ such that $v_n(x_n, y_n)$ is not 1 in $G$. Consider the corresponding elements $x = (x_n)^\omega$, $y = (y_n)^\omega$ in the ultrapower $G^\omega$. We claim that the subgroup $F \leqslant G^\omega$ generated by $x$ and $y$ is free. Suppose that the subgroup $F$ satisfies a reduced relation. That relation is a reduced word $u_i$ for some $i \in \mathbb{N}$. Hence, $u_i(x_n, y_n) = 1$ $\omega$–almost surely. In particular, since $\omega$ is a non-principal ultrafilter, for some $n > i$, $u_i(x_n, y_n) = 1$. But then $v_n(x_n, y_n) = 1$ since $u_i$ is a factor in the iterated commutator $v_n$, contradicting the choice of $x_n, y_n$. $\square$

## 7.4. Ultralimits of sequences of metric spaces

Let $(X_i)_{i\in I}$ be a family of metric spaces parameterized by an infinite set $I$.

CONVENTION 7.40. From now on, all ultrafilters are non-principal, and we will omit mentioning this property henceforth.

For an ultrafilter $\omega$ on $I$ we define the ultralimit

$$X_\omega = \omega\text{-}\lim_i X_i$$

as follows. Let $\prod_i X_i$ be the product of the spaces $X_i$, i.e. it is the space of indexed families of points $(x_i)_{i\in I}$ with $x_i \in X_i$. The distance between two points $(x_i), (y_i) \in \prod_i X_i$ is given by

$$\operatorname{dist}_\omega\big((x_i), (y_i)\big) := \omega\text{-}\lim\big(i \mapsto \operatorname{dist}_{X_i}(x_i, y_i)\big)$$

where we take the ultralimit of the function $i \mapsto \mathrm{dist}_{X_i}(x_i, y_i)$ with values in the compact set $[0, \infty]$. The function $\mathrm{dist}_\omega$ is a pseudo-distance on $\prod_i X_i$ with values in $[0, \infty]$. Set

$$(X_\omega, \mathrm{dist}_\omega) := (\prod_i X_i, \mathrm{dist}_i)/\sim$$

where we identify points with zero $\mathrm{dist}_\omega$–distance.

In the case when $X_i = Y$, for all $i$, the ultralimit $(X_\omega, \mathrm{dist}_\omega)$ is called a *constant ultralimit.*

Given an indexed family of points $(x_i)_{i\in I}$ with $x_i \in X_i$ we denote the equivalence class corresponding to it either by $x_\omega$ or by $\omega\text{-}\lim x_i$ .

Exercise 7.41. If $(X_\omega, \mathrm{dist}_\omega)$ is a constant ultralimit of a sequence of compact metric spaces $X_i = Y$, then $X_\omega \cong Y$ for all ultrafilters $\omega$.

Exercise 7.42. Let $(Y_i)_{i\in I}$ be a family of metric spaces parameterized by an infinite set $I$, and for every $i$ let $X_i$ be a dense subset in $Y_i$. For every ultrafilter $\omega$, the natural isometric embedding of the ultralimit $\omega\text{-}\lim_i X_i$ in the ultralimit $\omega\text{-}\lim_i Y_i$ is onto.

In particular, this is true when $Y_i = \widehat{X}_i$, the metric completion of $X_i$.

If the spaces $X_i$ do not have uniformly bounded diameter, then the ultralimit $X_\omega$ decomposes into (generically uncountably many) components consisting of points at mutually finite distance. We can pick out one of these components if the spaces $X_i$ have base-points $e_i$. The indexed family $(e_i)$ defines a base-point $e_\omega$ in $X_\omega$ and we set

$$X_{\omega,e} := \{x_\omega \in X_\omega \mid \mathrm{dist}_\omega(x_\omega, e_\omega) < \infty\}.$$

Define the *based ultralimit* as

$$\omega\text{-}\lim_i (X_i, e_i) := (X_{\omega,e}, e_\omega).$$

Example 7.43. For every proper metric space $Y$ with a base-point $y_0$, we have:

$$\omega\text{-}\lim_i (Y, y_0) \cong (Y, y_0).$$

Note that if $(X_i, x_i), (Y_i, y_i), i \in I$ are sequences of pointed metric spaces and $f_i : (X_i, x_i) \to (Y_i, y_i)$ is an isometry so that

$$\mathrm{dist}(f(x_i), y_i) \leqslant Const, \text{for } \omega - \text{all } i,$$

then $(f_i)$ yields an isometry $f_\omega : (X_\omega, x_\omega) \to (Y_\omega, y_\omega)$.

Proposition 7.44. *Every based ultralimit $\omega\text{-}\lim_i(X_i, e_i)$ of metric spaces is a complete metric space.*

Proof. According to Exercise 7.42, without loss of generality, we may assume that all $X_i$ are complete metric spaces. It suffices to prove that every Cauchy sequence $(x^{(k)})$ in $X_{\omega,e}$ contains a convergent subsequence. We select a subsequence (which we again denote $(x^{(k)})$) such that

$$\mathrm{dist}_\omega\left(x^{(k)}, x^{(k+1)}\right) < \frac{1}{2^k}.$$

Equivalently,

$$\omega - \lim\left(\mathrm{dist}_i\left(x_i^{(k)}, x_i^{(k+1)}\right)\right) < \frac{1}{2^k} \Rightarrow \mathrm{dist}_i\left(x_i^{(k)}, x_i^{(k+1)}\right) < \frac{1}{2^k}\, \omega - \text{a.s.}$$

It follows that we have $\omega(I_k) = 1$ for the set

$$I_k = \left\{ i \in I \ ; \ \mathrm{dist}_i \left( x_i^{(k)}, x_i^{(k+1)} \right) < \frac{1}{2^k} \right\} .$$

We can assume that $I_{k+1} \subseteq I_k$, otherwise we replace $I_{k+1}$ with $I_{k+1} \cap I_k$.

Thus, we obtain a nested sequence of subsets $I_k$ in $I$.

Assume that the set $J := \bigcap_{k \geqslant 1} I_k$ has the property that $\omega(J) = 1$. For every $i \in J$ the sequence $\left( x_i^{(k)} \right)$ is Cauchy, therefore, since the space $X_i$ is complete, there exists a limit $y_i \in X_i$ of the sequence $\left( x_i^{(k)} \right)$. The inequalities $\mathrm{dist}_i \left( x_i^{(k)}, x_i^{(k+1)} \right) < \frac{1}{2^k}$, $k \in \mathbb{N}$, imply that for every $m > k$, $\mathrm{dist}_i \left( x_i^{(k)}, x_i^{(m)} \right) < \frac{1}{2^{k-1}}$. The latter gives, when $m \to \infty$, that $\mathrm{dist}_i \left( x_i^{(k)}, y_i \right) < \frac{1}{2^{k-1}}$. Hence $\mathrm{dist}_\omega \left( x^{(k)}, y_\omega \right) \leqslant \frac{1}{2^{k-1}}$, where $y_\omega = \omega\text{-}\lim y_i$. We have thus obtained a limit $y_\omega$ for the sequence $\left( x^{(k)} \right)$

Assume now that $\omega(J) = 0$. Since for every $k \geqslant 1$ we have that $I_k = \bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1}) \sqcup J$ and $\omega(I_k) = 1$, it follows that $\omega \left( \bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1}) \right) = 1$. n what follows we denote $\bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1})$ by $J_k$.

We define what we claim is the limit point of $(x^{(k)})$ as $\omega\text{-}\lim y_i$, where $y_i = x_i^{(k)}$ when $i \in I_k \setminus I_{k+1}$. This defines $y_i$ for all $i \in J_1$. This suffices to completely define $\omega\text{-}\lim y_i$, because in all the ultralimit arguments, the values taken on sets of indices $i \in I$ of $\omega$-measure zero do not matter.

For every $i \in J_k = \bigsqcup_{j=k}^{\infty} (I_j \setminus I_{j+1})$ there exists $j \geqslant k$ such that $i \in I_j \setminus I_{j+1}$. By definition $y_i = x_i^{(j)}$.

Since $i \in I_j \subseteq I_{j-1} \subseteq \cdots \subseteq I_{k+1} \subseteq I_k$ we may write

$$\mathrm{dist}_i \left( x_i^{(k)}, y_i \right) \leqslant \mathrm{dist}_i \left( x_i^{(k)}, x_i^{(k+1)} \right) + \cdots + \mathrm{dist}_i \left( x_i^{(j-1)}, x_i^{(j)} \right) \leqslant$$

$$\frac{1}{2^k} + \frac{1}{2^{k+1}} + \cdots + \frac{1}{2^{j-1}} \leqslant \frac{1}{2^k} \frac{1}{1 - \frac{1}{2}} = \frac{1}{2^{k-1}} .$$

Thus we have $\mathrm{dist}_\omega \left( x^{(k)}, y_\omega \right) \leqslant \frac{1}{2^{k-1}}$, hence $x^{(k)} \to y_\omega$. □

A simple, but important, special case of ultralimits is the *constant* ultralimit, i.e., $\omega\text{-}\lim(X_i, e_i) = \omega\text{-}\lim(X, e)$, where $X_i = X$ and $e_i = e$.

EXERCISE 7.45. Let $X$ be a proper metric space, take a subset $Y \subset X$ equipped with the restriction metric. Then the constant ultralimit $\omega\text{-}\lim(Y, y)$ is naturally isometric to $(\bar{Y}, y)$, where $\bar{Y}$ is the closure of $Y$ in $X$. Furthermore, if $X$ is compact then for every $e_i \in X$, $\omega\text{-}\lim(X, e_i) = (X, \omega\text{-}\lim e_i)$.

EXERCISE 7.46. Let $X = \mathbb{R}^n$. Then for every sequence $e_i \in X$, $\omega\text{-}\lim(X, e_i) \cong (\mathbb{R}^n, 0)$.

LEMMA 7.47. *[Functoriality of ultralimits] Let $(X_i, p_i), (Y_i, q_i)$ be sequences of pointed metric spaces with ultralimits $X_\omega, Y_\omega$ respectively. Let $f_i : X_i \to Y_i$ be isometric embeddings so that*

$$\omega\text{-}lim\, d(f(p_i), q_i) < \infty.$$

*Then the maps $f_i$ yield an isometric embedding of the ultralimits $f_\omega : X_\omega \to Y_\omega$. If each $f_i$ is an isometry, then so is $f_\omega$.*

PROOF. Given $x_\omega = (x_i) \in X_\omega$ we define $f_\omega(x_\omega)$ to be the point $y_\omega \in Y_\omega$ represented by the sequence $(f_i(x_i))$. Triangle inequality immediately implies that $y_\omega$ indeed belongs to $Y_\omega$. By the definition of distances in $X_\omega$ and $Y_\omega$,

$$d(f_\omega(x_\omega), f_\omega(x'_\omega)) = \omega\text{-}\lim d(f_i(x_i), f_i(x'_i)) = \omega\text{-}\lim d(x_i, x'_i) = d(x_\omega, x'_\omega)$$

for any pair of points $x_\omega, x'_\omega \in X_\omega$. If each $f_i$ is surjective, then, clearly, $f_\omega$ is surjective as well. $\square$

The map $f_\omega$ is called the *ultralimit* of the sequence of maps $(f_i)$. An important example illustrating this lemma is the case when each $X_i$ is an interval in $\mathbb{R}$ and, hence, each $f_i$ is a geodesic in $Y_i$. Then the ultralimit $f_\omega : J_\omega \to X_\omega$ is a geodesic in $X_\omega$ (here $J_\omega$ is an interval in $\mathbb{R}$).

DEFINITION 7.48. Geodesics $f_\omega : J_\omega \to X_\omega$ are called *limit geodesics* in $X_\omega$.

In general, $X_\omega$ contains geodesics which are not limit geodesics. In the extreme case, $Y_i$ may contain only constant geodesics, while $Y_\omega$ is a geodesic metric space (containing more than one point). For instance, let $X = \mathbb{Q}$ with the metric induced from $\mathbb{R}$. Of course, $\mathbb{Q}$ contains no nonconstant geodesics, but

$$\omega\text{-}\lim(X, 0) \cong (\mathbb{R}, 0).$$

LEMMA 7.49. *Ultralimit $(X_\omega, e_\omega)$ of a sequence of pointed geodesic metric spaces $(X_i, e_i)$ is again a geodesic metric space.*

PROOF. Let $x_\omega = (x_i), y_\omega = (y_i)$ be points in $X_\omega$. Let $\gamma_i : [0, T_i] \to X_i$ be (unit speed) geodesics connecting $x_i$ to $y_i$. Thus,

$$\omega\text{-}\lim T_i = T = d((x_\omega), y_\omega)) = T < \infty.$$

We then define the map $\gamma_\omega : [0, T] \to X_\omega$ by

$$\gamma_\omega(t) = (\gamma_i(t_i)), where t = \omega\text{-}\lim t_i, t_i \in [0, T_i].$$

We leave it to the reader to verify that $\gamma_\omega$ is a geodesic connecting $x_\omega$ to $y_\omega$. $\square$

EXERCISE 7.50. Let $X$ be a path-metric space. Then every constant ultralimit of $X$ is a geodesic metric space.

LEMMA 7.51. *Let $(X_i, e_i)$ be pointed $CAT(\kappa_i)$ metric spaces, $\kappa_i \leqslant 0$, and $\kappa = \omega\text{-}\lim \kappa_i$. Then the ultralimit $(X_\omega, e_\omega)$ of the sequence $(X_i, e_i)$ is again a pointed $CAT(\kappa)$ space.*

PROOF. It is clear that comparison inequalities for triangles in $X_i$ yield comparison inequalities for limit triangles in $X_\omega$. What remains is to show that $X_\omega$ is a uniquely geodesic metric space, in which case every geodesic segment in $X_\omega$ is a limit geodesic. Suppose that $m_\omega \in X_\omega$ is a point so that

$$d(x_\omega, z_\omega) + d(z_\omega, y_\omega) = d(x_\omega, y_\omega).$$

Thus, if $z_i \in X_i$ is a sequence representing $z_\omega$, then

$$0 \leqslant d(x_i, z_i) + d(z_i, y_i) = d(x_i, y_i) \leqslant \eta_i, \quad \omega\text{-}\lim \eta_i = 0.$$

Let us assume that $s_i = d(x_i, z_i) \leqslant d(z_i, y_i)$ and consider the point $q_i \in [x_i, y_i]$ within distance $s_i$ from $x_i$. Compare the triangle $T_i = T(x_i, y_i, z_i)$ with the Euclidean triangle and using the comparison points $p_i = z_i$ and $q_i$.In the Euclidean comparison triangle $\tilde{T}_i$, we have

$$\omega\text{-lim}\, d(\tilde{z}_i, \tilde{q}_i) = 0$$

(since the constant ultralimit of the sequence of Euclidean planes is the Euclidean plane and, hence, is uniquely geodesic). Since, by the $CAT(0)$-comparison inequality,

$$d(z_i, q_i) \leqslant d(\tilde{z}_i, \tilde{q}_i)$$

we conclude that $(q_i) = z_\omega$ in the space $X_\omega$. Thus, $z_\omega$ lies on the limit geodesic connecting $x_\omega$ and $y_\omega$. □

### 7.5. The asymptotic cone of a metric space

A precursor to the notion of asymptotic cone appears in Gromov's paper [**Gro81**], the concept was formalized by van den Dries and Wilkie in [**dDW84**] (for groups) and by Gromov in [**Gro93**] for general metric spaces. The idea is to construct, for a metric space $(X, \text{dist})$, its "image" seen from "infinitely far." More precisely, one defines the notion of a *limit* of a sequence of metric spaces $(X, \varepsilon\text{dist})$, $\varepsilon > 0$, as $\varepsilon \to 0$. The main tool in this construction is a non-principal ultrafilter $\omega$ on an infinite set $I$.

Let $X$ be a metric space and $\omega$ be a non-principal ultrafilter on $I$. Suppose that we are given a family $\boldsymbol{\lambda} = (\lambda_i)_{i \in I}$ of positive real numbers indexed by $I$ so that $\omega\text{-lim}\, \lambda_i = 0$ and a family $\boldsymbol{e} = (e_i)_{i \in I}$ of base-points $e_i \in X$ indexed by $I$. Given this data, the *asymptotic cone* $\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ of $X$ is defined as the based ultralimit of rescaled copies of $X$:

$$\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}) := \omega\text{-}\lim_i (\lambda_i \cdot X, e_i).$$

Here $\lambda X$ is the metric space $(X, \lambda d_X)$, where $d_X$ is the metric on $X$.

Given a family of points $(x_i)_{i \in I}$ in $X$, the corresponding subset in the asymptotic cone $\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, which is either a one-point set, or the empty set if $\omega\text{-lim}\, \lambda_i \text{dist}(x_i, e_i) = \infty$, is denoted by $\omega\text{-lim}\, x_i$.

The family $\boldsymbol{\lambda} = (\lambda_i)_{i \in I}$ is called a *scaling family.* When either the scaling family or the family of base-points are irrelevant, they are omitted from the notation.

Thus, to every metric space $X$ we attach a collection of metric spaces $\text{Cones}(X)$ consisting of all asymptotic cones $\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ of $X$, that is of all "images of $X$ seen from infinitely far." The first questions to ask are: How large is the collection $\text{Cones}(X)$ for specific metric spaces $X$ or groups $G$, and what features of $X$ are inherited by the metric spaces in $\text{Cones}(X)$.

We begin by noting that the choice of base-points is irrelevant for spaces that are quasi-homogeneous:

EXERCISE 7.52. [See also Proposition 7.58.] When the space $X$ is quasi-homogeneous, all cones defined by the same fixed ultrafilter $\omega$ and sequence of scaling constants $\boldsymbol{\lambda}$, are isometric.

Another simple observation is

Remark 7.53. Let $\alpha$ be a positive real number. The map

$$I_\alpha : \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}) \to \mathrm{Cone}_\omega\left(X, \boldsymbol{e}, \frac{1}{\alpha}\boldsymbol{\lambda}\right), \; I_\alpha(\omega\text{-lim}\, x_i) = \omega\text{-lim}\, x_i$$

is a similarity with the factor $\alpha$. Thus, for a fixed metric space $X$, the collection of limit metric spaces $\mathrm{Cones}(X)$ is stable with respect to rescaling of the metrics.

Proofs of the following statements are straightforward and are left as an exercise to the reader:

Proposition 7.54. (1) $\mathrm{Cone}_\omega(X \times Y) = \mathrm{Cone}_\omega(X) \times \mathrm{Cone}_\omega(Y)$.
(2) *If* $\mathbb{R}^n$ *is endowed with a metric coming from a norm then* $\mathrm{Cone}_\omega \mathbb{R}^n \cong \mathbb{R}^n$.
(3) *The asymptotic cone of a geodesic space is a geodesic space.*

Definition 7.55. Given a family $(A_i)_{i\in I}$ of subsets of $(X, \mathrm{dist})$, we denote either by $\omega$-lim $A_i$ or by $A_\omega$ the subset of $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ that consists of all the elements $\omega$-lim $x_i$ such that $x_i \in A_i$ $\omega$–almost surely. We call $\omega$-lim $A_i$ the *limit set* of the family $(A_i)_{i\in I}$.

Note that if $\omega\text{-lim}\, \frac{\mathrm{dist}(e_i, A_i)}{\lambda_i} = \infty$ then the set $\omega$-lim $A_i$ is empty.

Proposition 7.56 (Van den Dries and Wilkie, [**dDW84**]). (1) *Any asymptotic cone of a metric space is complete.*
(2) *Every limit set* $\omega$*-lim* $A_i$ *is a closed subset of* $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$.

Proof. This is an immediate consequence of Proposition 7.44. □

In Definition 7.48 we introduced the notion of *limit geodesics* in the ultralimit of a sequence of metric spaces. Let $\gamma_i : [a_i, b_i] \to X$ be a family of geodesics with the limit geodesic $\gamma_\omega$ in $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$.

Exercise 7.57. Show that the image of $\gamma_\omega$ is the limit set of the sequence of images of the geodesics $\gamma_i$.

We saw earlier that geodesics in the ultralimit may fail to be limit geodesics. However, in our example, we took a sequence of metric spaces which were not geodesic. It turns out that, in general, there exist geodesics in $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ that are not limit geodesic, even when $X$ is the Cayley graph of a finitely generated group with a word metric. An example of this can be found in [**Dru09**].

Suppose that $X$ is a metric space and $G \subset \mathrm{Isom}(X)$ is a subgroup. Given a non-principal ultrafilter $\omega$ consider the ultraproduct $G^\omega = \prod_{i\in I} G/\omega$. For a family of positive real numbers $\boldsymbol{\lambda} = (\lambda_i)_{i\in I}$ so that $\omega$-lim $\lambda_i = 0$ and a family of base-points $\boldsymbol{e} = (e_i)$ in $X$, let $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ be the corresponding asymptotic cone. In view of Lemma 7.47, the group $G^\omega$ acts isometrically on the ultralimit

$$U := \omega\text{-}\lim_i (\lambda_i \cdot X).$$

Let $G^\omega_{\boldsymbol{e}} \subset G^\omega$ denote the stabilizer in $G^\omega$ of the component $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}) \subset U$. In other words,

$$G^\omega_{\boldsymbol{e}} = \{(g_i)^\omega \in G^\omega : \omega\text{-}\lim_i \lambda_i \mathrm{dist}(g_i(e_i), e_i) < \infty\}.$$

There is a natural homomorphism $G^\omega_{\boldsymbol{e}} \to \mathrm{Isom}(\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda}))$. Observe that if $(e_i)$ is a bounded sequence in $X$ then the group $G$ has a diagonal embedding in $G^\omega_{\boldsymbol{e}}$.

PROPOSITION 7.58. *Suppose that $G \subset \mathrm{Isom}(X)$ and the action $G \curvearrowright X$ is cobounded. Then for every asymptotic cone* $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ *the action* $G^\omega_{\boldsymbol{e}} \curvearrowright \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ *is transitive. In particular,* $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ *is a homogeneous metric space.*

PROOF. Let $D < \infty$ be such that $G \cdot x$ is a $D$-net in $X$. Given two indexed families $(x_i), (y_i)$ of points in $X$ there exists an indexed family $(g_i)$ of elements of $G$ such that

$$\mathrm{dist}(g_i(x_i), y_i) \leqslant 2D.$$

Therefore, if $g_\omega := (g_i)^\omega \in G^\omega$, then $g_\omega(\omega - \lim (x_i)) = \omega - \lim (y_i)$. Hence the action

$$G^\omega \curvearrowright X_\omega = \omega\text{-}\lim_i(\lambda_i \cdot X)$$

is transitive. It follows that the action $G^\omega_{\boldsymbol{e}} \curvearrowright \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is transitive as well. □

EXERCISE 7.59. 1. Construct an example of a metric space $X$ and an asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ so that for the isometry group $G = \mathrm{Isom}(X)$ the action $G^\omega_{\boldsymbol{e}} \curvearrowright \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is not effective, i.e. the homomorphism $G^\omega_{\boldsymbol{e}} \to \mathrm{Isom}\left(\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})\right)$ has nontrivial kernel. Construct an example when the kernel of the above homomorphism contains the entire group $G$ embedded diagonally in $G^\omega_{\boldsymbol{e}}$.

2. Show that $\mathrm{Ker}(G \to QI(X)) \leqslant \mathrm{Ker}(G \to \mathrm{Isom}(X_\omega))$.

Suppose that $X$ admits a cocompact discrete action by a group $G$ of isometries. The problem of how large the class of spaces $\mathrm{Cones}(X)$ can be, that is the problem of the dependence of the topological/metric type of $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ on the ultrafilter $\omega$ and the scaling sequence $\boldsymbol{\lambda}$, relates to the Continuum Hypothesis (the hypothesis stating that there is no cardinal number between $\aleph_0$ and $2^{\aleph_0}$). Kramer, Shelah, Tent and Thomas have shown in **[KSTT05]** that:

(1) if the Continuum Hypothesis (CH) is not true then $SL(n, \mathbb{R})$, $n \geqslant 3$ (as well as any uniform lattice, see Definition 3.15), has $2^{2^{\aleph_0}}$ non-isometric asymptotic cones;

(2) if the CH is true then all asymptotic cones of a uniform lattice in $SL(n, \mathbb{R})$, $n \geqslant 3$, are isometric. Moreover, a finitely generated group has at most a continuum of non-isometric asymptotic cones.

The case of $SL(2, \mathbb{R})$ was settled independently of the CH by A. Dyubina–Erschler and I. Polterovich (see Theorem 9.128).

Chronologically, the first non-trivial example of metric space $X$ such that the set $\mathrm{Cones}(X)$ contains very few elements is that of virtually nilpotent groups, and it is due to P. Pansu. In fact, this result comes as a strengthening of Gromov's Polynomial Growth Theorem that is proved in Chapter 14.

C. Druţu and M. Sapir constructed in **[DS05]** an example of two-generated and recursively presented (but not finitely presented) group with continuously many non-homeomorphic asymptotic cones. The construction is independent of the Continuum Hypothesis. The example can be adapted so that at least one asymptotic cone is a real tree.

Note that if a finitely presented group $G$ has *one asymptotic cone* which is a tree, then the group is hyperbolic and hence *every* asymptotic cone of $G$ is a tree, see **[KK07]**.

**Historical remarks.** The first instance (that we are aware of) where asymptotic cones of metric spaces were defined is a 1966 paper [**BDCK66**], where this is done in the context of normed vector spaces. Their definition, though, works for all metric spaces.

On the other hand, Gromov introduced the modified Hausdorff distance (see Section 5.1 for a definition) and corresponding limits of sequences of metric spaces in his work on groups of polynomial growth [**Gro81**]. This approach is no longer appropriate in the case of more general metric spaces, as we will explain below.

Firstly, the modified Hausdorff distance does not distinguish between a space and a dense subset in it, therefore in order to have a well defined limit one has to ask *a priori* that the limit be complete.

Secondly, if a pointed sequence of proper geodesic metric spaces $(X_n, \mathrm{dist}_n, x_n)$ converges to a complete geodesic metric space $(X, \mathrm{dist}, x)$ in the modified Hausdorff distance, then the limit space $X$ is proper. Indeed given a ball $B(x, R)$ in $X$, for every $\epsilon$ there exists an $n$ such that $B(x, R)$ is at Hausdorff distance at most $\epsilon$ from the ball $B(x_n, R)$ in $X_n$. From this and the fact that all spaces $X_n$ are proper it follows that for every sequence $(y_n)$ in $B(x, R)$ and every $\varepsilon$ there exists a subsequence of $(y_n)$ of diameter $\leqslant \varepsilon$. A diagonal argument and completeness of $X$ allow to conclude that $(y_n)$ has a convergent subsequence, and therefore that $B(x, R)$ is compact.

Thirdly, in [**KL95**] the following relation between Hausdorff limits and asymptotic cones is proved:

THEOREM 7.60 ([**KL95**]). *If $(X, \mathrm{dist}_X)$ and $(Y, \mathrm{dist}_Y)$ are proper metric spaces such that for a sequence of positive real numbers $(\epsilon_n)$ converging to zero and a sequence of points $(x_n)$, $(X, \epsilon_n \mathrm{dist}_X, x_n)$ converges to $(Y, \mathrm{dist}_Y, y_0)$ in the modified Hausdorff metric, then for all ultrafilters $\omega$ there exists an isometry between $\mathrm{Cone}_\omega(X, (x_n), (\epsilon_n))$ and $(Y, y_0)$ such that the image of $\omega$-lim $x_n$ is $y_0$.*

Thus, for a proper geodesic metric space $(X, \mathrm{dist})$, the existence of a sequence of pointed metric spaces of the form $(X, \epsilon_n \mathrm{dist}, x_n)$ convergent in the modified Hausdorff metric, implies the existence of proper asymptotic cones. On the other hand, if $X$ is, for instance, a non-elementary hyperbolic group, no asymptotic cone of $X$ is proper, see Theorem 9.128. Therefore, in such a case $(X, \varepsilon \mathrm{dist})$, $\varepsilon > 0$, has no subsequence convergent with respect to the modified Hausdorff metric.

### 7.6. Ultralimits of asymptotic cones are asymptotic cones

In this section we show that ultralimits of asymptotic cones are asymptotic cones, following [**DS05**]. To this end, we first describe a construction of ultrafilters on Cartesian products that generalizes the standard notion of product of ultrafilters, as defined in [**She78**, Definition 3.2 in Chapter VI]. In what follows, we view ultrafilters as in Definition 7.17. Throughout the section, $\omega$ will denote an ultrafilter on a set $I$ and $\mu = (\mu_i)_{i \in I}$ a family, indexed by $I$, of ultrafilters on a set $J$.

DEFINITION 7.61. We define a new ultrafilter $\omega\mu$ on $I \times J$ such that for every subset $A$ in $I \times J$, $\omega\mu(A)$ is equal to the $\omega$–measure of the set of all $i \in I$ such that $\mu_i(A \cap (\{i\} \times J)) = 1$.

LEMMA 7.62. *$\omega\mu$ is an ultrafilter over $I \times J$.*

Proof. It suffices to prove that $\omega\mu$ is finitely additive and that it takes the zero value on finite sets.

We first prove that $\omega\mu$ is finitely additive, using the fact that $\omega$ and $\mu_i$ are finitely additive. Let $A$ and $B$ be two disjoint subsets of $I \times J$. Fix $i \in I$ arbitrary. The sets $A \cap (\{i\} \times J)$ and $B \cap (\{i\} \times J)$ are disjoint, hence

$$\mu_i((A \cup B) \cap (\{i\} \times J)) = \mu_i(A \cap (\{i\} \times J)) + \mu_i(B \cap (\{i\} \times J)).$$

The finite additivity of $\omega$ implies that

$$\omega\mu(A \sqcup B) = \omega\mu(A) + \omega\mu(B).$$

Also, given a finite subset $A$ of $I \times J$, $\omega\mu(A) = 0$. Indeed, since the set of $i$'s for which $\mu_i(A \cap (\{i\} \times J)) = 1$ is empty, $\omega\mu(A) = 0$ by definition. □

Lemma 7.63 (double ultralimit of real numbers). *For every doubly indexed family of real numbers $\alpha_{ij}$, $i \in I, j \in J$ we have that*

$$\omega\mu\text{-}\lim \alpha_{ij} = \omega\text{-}\lim\left(\mu_i\text{-}\lim \alpha_{ij}\right), \tag{7.3}$$

*where the second limit on the right hand side is taken with respect to $j \in J$.*

Proof. Let $a$ be the limit $\omega\mu$-lim$\alpha_{ij}$. For every neighborhood $U$ of $a$

$$\omega\mu\left\{(i,j) \mid \alpha_{ij} \in U\right\} = 1 \Leftrightarrow$$

$$\omega\left\{i \in I \mid \mu_i\left\{j \mid \alpha_{ij} \in U\right\} = 1\right\} = 1\,.$$

This implies that

$$\omega\left\{i \in I \mid \mu_i\text{-}\lim \alpha_{ij} \in \overline{U}\right\} = 1\,,$$

which, in turn, implies that

$$\omega\text{-}\lim(\mu_i\text{-}\lim \alpha_{ij}) \in \overline{U}\,.$$

This holds for every neighborhood $U$ of $a \in \mathbb{R} \cup \{\pm\infty\}$. Therefore, we conclude that

$$\omega\text{-}\lim(\mu_i\text{-}\lim\ \alpha_{ij}) = a.$$

□

Lemma 7.63 implies a similar result for ultralimits of spaces.

Proposition 7.64 (double ultralimit of spaces). *Let $(X_{ij}, \mathrm{dist}_{ij})$ be a doubly indexed sequence of metric spaces, $(i,j) \in I \times J$, and let $e = (e_{ij})$ be a doubly indexed sequence of points $e_{ij} \in X_{ij}$. We denote by $e_i$ the sequence $(e_{ij})_{j \in J}$.*

*Then the map*

$$\omega\mu\text{-}\lim(x_{ij}) \mapsto \omega\text{-}\lim\left(\mu_i\text{-}\lim x_{ij}\right)\,, \tag{7.4}$$

*is an isometry from*

$$\omega\mu\text{-}\lim(X_{ij}, e_{ij})$$

*onto*

$$\omega\text{-}\lim\left(\mu_i\text{-}\lim\left(X_{ij}, e_{ij}\right)\,,\, e_i'\right)$$

*where, $e_i' = \mu_i\text{-}\lim e_{ij}$.*

Corollary 7.65 (ultralimits of asymptotic cones are as. cones). *Let $X$ be a metric space. Consider double indexed families of points $\boldsymbol{e} = (e_{ij})_{(i,j)\in I\times J}$ in $X$ and of positive real numbers $\boldsymbol{\lambda} = (\lambda_{ij})_{(i,j)\in I\times J}$ such that*

$$\mu_i\text{-lim}\lambda_{ij} = 0$$

*for every $i \in I$. Let $\text{Cone}_{\mu_i}(X,(e_{ij}),(\lambda_{ij}))$ be the corresponding asymptotic cone of $X$. The map*

$$\omega\mu\text{-lim}\,(x_{ij}) \mapsto \omega\text{-lim}\,(\mu_i\text{-lim}\,(x_{ij}))\,, \tag{7.5}$$

*is an isometry from $\text{Cone}_{\omega\mu}(X,\boldsymbol{e},\boldsymbol{\lambda})$ onto*

$$\omega\text{-lim}\left(\text{Cone}_{\mu_i}(X,(e_{ij}),(\lambda_{ij}))\,,\mu_i\text{-lim}e_{ij}\right)\,.$$

Proof. The statement follows from Proposition 7.64. The only thing to be proved here is that

$$\omega\mu\text{-lim}\lambda_{ij} = 0$$

Let $\varepsilon > 0$. For every $i \in I$ we have that

$$\mu_i\text{-lim}\lambda_{ij} = 0,$$

whence,

$$\mu_i\{j \in I \mid \lambda_{ij} < \varepsilon\} = 1\;.$$

It follows that

$$\{i \in I \mid \mu_i\{j \in I \mid \lambda_{ij} < \varepsilon\} = 1\} = I,$$

therefore, the $\omega$-measure of this set is 1. We conclude that

$$\omega\mu\{(i,j) \in I\times J \mid \lambda_{ij} < \varepsilon\} = 1. \quad \square$$

Corollary 7.66. *Let $X$ be a metric space. The collection of all asymptotic cones of $X$ is stable with respect to rescaling, ultralimits and taking asymptotic cones.*

Proof. It is an immediate consequence of Corollary 7.65 and Remark 7.53. $\square$

Corollary 7.67. *Let $X, Y$ be metric spaces such that all asymptotic cones of $X$ are isometric to $Y$. Then all asymptotic cones of $Y$ are isometric to $Y$.*

This, in particular, implies that the following are *examples of metric spaces isometric to all their asymptotic cones.*

Examples 7.68.
(1) The $2^{\aleph_0}$–universal real tree $T_C$, according to Theorem 9.128.
(2) A non-discrete Euclidean building that is the asymptotic cone of $SL(n,\mathbb{R})$, $n \geqslant 3$, under the Continuum Hypothesis, according to **[KSTT05]** and **[KL98b]**.
(3) A graded nilpotent Lie group with a Carnot-Caratheodory metric, according to Theorem 14.30 of P. Pansu.

## 7.7. Asymptotic cones and quasi-isometries

The following simple lemma shows why asymptotic cones are useful in studying quasi-isometries, since they become bi-Lipschitz maps of asymptotic cones, and the latter maps are much easier to handle.

LEMMA 7.69.
*Let $(X, x), (Y, y)$ be pointed metric spaces, let $f : X \to Y$ be an $(L, A)$–quasi-isometry and let $\lambda_i$ denote a scaling family. Then $f_\omega : X_\omega \to Y_\omega$, $f_\omega((x_i)) = (f(x_i))$, is an $L$–bi-Lipschitz map.*

PROOF. We have the inequalities:

$$L^{-1}\frac{1}{\lambda_i}dist(x,x') - \frac{A}{\lambda_i} \leqslant \frac{1}{\lambda_i}dist(f(x),f(x')) \leqslant L\frac{1}{\lambda_i}dist(x,x') + \frac{A}{\lambda_i}.$$

Passing to the $\omega$–limit, we obtain

$$L^{-1}dist_\omega(x_\omega, x'_\omega) \leqslant dist_\omega(f_\omega(x_\omega), f_\omega(x'_\omega)) \leqslant Ldist_\omega(x_\omega, x'_\omega)$$

where $f_\omega(z_\omega) = (f_i(z_i))$. Thus, $f_\omega$ is an $L$–bi-Lipschitz embedding. Since $f(X)$ is an $A$–net in $Y$, the same argument as above shows that $f_\omega$ is onto. □

EXERCISE 7.70. Extend this lemma to sequences of quasi-isometries and metric spaces.

One may ask if a converse to this lemma is true, for instance: Does the existence of a map between metric spaces that induces bi-Lipschitz maps between asymptotic cones imply quasi-isometry? We say that two spaces are *asymptotically bi-Lipschitz* if the latter holds. (This notion is introduced in **[dC09]**.) See Remark 14.31 for an example of asymptotically bi-Lipschitz spaces which are not quasi-isometric to each other.

Here is an example of application of asymptotic cones to the study of quasi-isometries.

LEMMA 7.71. *Suppose that $X = \mathbb{R}^n$ or $\mathbb{R}_+$ and $f : X \to X$ is an $(L, A)$–quasi-isometric embedding. Then $f$ is a quasi-isometry, furthermore, $\mathcal{N}_C(f(X)) = X$, for some $C = C(L, A)$.*

PROOF. We will give a proof in the case of $\mathbb{R}^n$ as the other case is analogous. Suppose that the assertion is false, i.e., there is a sequence of $(L, A)$–quasi-isometric embeddings $f_j : \mathbb{R}^n \to \mathbb{R}^n$, sequence of real numbers $r_j$ diverging to infinity and points $y_j \in \mathbb{R}^n$ such that $\text{dist}(y_j, Image(f)) = r_j$. Let $x_j \in \mathbb{R}^n$ be a point such that $\text{dist}(f(x_j), y_j) \leqslant r_j + 1$. Using $x_j, y_j$ as base-points on the domain and range for $f_j$, rescale the metrics on the domain and the range by $1/r_j$ and take the corresponding ultralimits. In the limit we get a bi-Lipschitz embedding

$$f_\omega : \mathbb{R}^n \to \mathbb{R}^n,$$

whose image misses the point $y_\omega \in \mathbb{R}^n$. However each bi-Lipschitz embedding is necessarily proper, therefore, by the invariance of domain theorem, the image of $f_\omega$ is both closed and open. Contradiction. □

REMARK 7.72. Alternatively, one can prove the above lemma as follows: Approximate $f$ by a continuous mapping $g$. Then, since $g$ is proper, it has to be onto.

CHAPTER 8

# Hyperbolic Space

The real hyperbolic space is the oldest and easiest example of hyperbolic space. A good reference for hyperbolic spaces in general is [**And05**]. The real-hyperbolic space has its origin in the following classical question that has challenged the geometers for nearly 2000 years:

QUESTION 8.1. Does Euclid's fifth postulate follow from the rest of the axioms of Euclidean geometry? (The fifth postulate is equivalent to the statement that given a line $L$ and a point $P$ in the plane, there exists exactly one line through $P$ parallel to $L$.)

After a long history of unsuccessful attempts to establish a positive answer to this question, N.I. Lobachevski, J. Bolyai and C.F. Gauss independently (in the early 19th century,) developed a theory of non-Euclidean geometry (which we now call "hyperbolic geometry"), where Euclid's fifth postulate is replaced by the axiom:

"For every point $P$ which does not belong to $L$, there are infinitely many lines through $P$ parallel to $L$."

Independence of the 5th postulate from the rest of the Euclidean axioms was proved by E. Beltrami in 1868, *via* a construction of a model of the hyperbolic geometry. In this chapter we will use the unit ball and the upper half-space models of hyperbolic geometry, the latter of which is due to H. Poincaré.

## 8.1. Moebius transformations

We will think of the sphere $S^n$ as the 1-point compactification of $\mathbb{R}^n$. Accordingly, we will regard the 1-point compactification of a hyperplane in $\mathbb{R}^n$ as a *round sphere* (of infinite radius) and the 1-points compactification of a line in $\mathbb{R}^n$ as a *round circle* (of infinite radius). Recall that the *inversion* in the $r$-sphere $\Sigma_r = \{x : \|x\| = r\}$ is the map

$$J_\Sigma : x \mapsto \frac{r^2 x}{\|x\|^2}, \quad J_\Sigma(0) = \infty, \quad J_\Sigma(\infty) = 0.$$

One defines the inversion $J_\Sigma$ in the sphere $\Sigma = \{x : \|x - a\| = r\}$ by the formula

$$T_a \circ J_{\Sigma_r} \circ T_{-a}$$

where $T_a$ is the translation by the vector $a$. Inversions map round spheres to round spheres and round circles to circles; inversions also preserve the Euclidean angles, and the *cross-ratio*

$$[x, y, z, w] := \frac{|x - y|}{|y - z|} \cdot \frac{|z - w|}{|w - x|},$$

see e.g. [**Rat94**, Theorem 4.3.1]. We will regard the reflection in a Euclidean hyperplane as an inversion (such inversion fixes $\infty$).

Definition 8.2. A *Moebius transformation* of $\mathbb{R}^n$ (or, rather, $S^n$) is a composition of finitely many inversions in $\mathbb{R}^n$. The group of all Moebius transformations of $\mathbb{R}^n$ is denoted $Mob(\mathbb{R}^n)$ or $Mob(S^n)$.

In particular, Moebius transformations preserve angles, cross-ratios and map circles to circles and spheres to spheres.

For instance, every translation is a Moebius transformation, since it is the composition of two reflections in parallel hyperplanes. Every rotation in $\mathbb{R}^n$ is the composition of at most $n$ inversions (reflections), since every rotation in $\mathbb{R}^2$ is the composition of two reflections. Every dilation $x \mapsto \lambda x, \lambda > 0$ is the composition of two inversions in spheres centered at 0.

Lemma 8.3. *The subgroup $Mob_{\infty,0}(\mathbb{R}^n)$ of $Mob(\mathbb{R}^n)$ fixing $\infty$ and $0$ equals the group $CO(n) = \mathbb{R}_+ \cdot O(n)$.*

Proof. We just observed that $CO(n)$ is contained in $Mob_{\infty,0}(\mathbb{R}^n)$. We, thus, need to prove the opposite inclusion. Consider the coordinate lines $L_1, ..., L_n$ in $\mathbb{R}^n$. Then every $g \in Mob_\infty(\mathbb{R}^n)$ sends these lines to pairwise orthogonal lines $L_1', ..., L_n'$ through the origin (since Moebius transformations map circles to circles and preserve angles). By postcomposing $g$ with an element of $O(n)$, we can assume that $g$ preserves each coordinate line $L_n$ and, furthermore, preserves the orientation on this line. By postcomposing $g$ with dilation we can also assume that $g$ maps the unit vector $e_1$ to itself. Thus, $g$ maps the unit sphere $\Sigma_1$ to the round sphere which is orthogonal to the coordinate lines and passes through the point $e_1$. Hence, $d(\Sigma_1) = \Sigma_1$. We claim that such $g$ is the identity. Indeed, if $L$ is a line through the origin, then the line $g(L)$ has the same angles with $L_i$ as $L$ for each $i = 1, ..., n$. Thus, $g(L) = L$ for every such $L$. By considering intersections of these lines with $\Sigma_1$, we conclude that $g$ restricts to the identity on $\Sigma_1$. It remains to show that $g$ is the identity on every sphere centered at the origin. Equivalently, we need to show that $g$ is the identity on the line $L_1$.

Let $x \in L_1$ be outside of $\Sigma_1$ and let $L$ be a line in the $x_1x_2$-plane through $x$ and tangent to $\Sigma_1$ at a point $y$. Then $g(L)$ is also a line through $g(x), y$, tangent to $\Sigma_1$ at $y$. Since $g$ preserves the orientation on $L_1$, $g(L) = L$ and, hence, $g(x) = x$. We leave the case of points $x \in L_1$ contained inside $\Sigma_1$ to the reader. □

Example 8.4. Let us construct a Moebius transformation $\sigma$ sending the unit ball $\mathbf{B}^n = B(0,1) \subset \mathbb{R}^n$ to the upper half-space $U^n = \mathbb{R}^n_+$,

$$\mathbb{R}^n_+ = \{(x_1, ...x_n) : x_n > 0\}.$$

We take $\sigma$ to be the composition of translation $x \mapsto x + e_n$, where $e_n = (0, ..., 0, 1)$, inversion $J_\Sigma$, where $\Sigma = \partial \mathbf{B}^n$, translation $x \mapsto x - \frac{1}{2}e_n$ and, lastly, the similarity $x \to 2x$. The reader will notice that the restriction of $\sigma$ to the boundary sphere $\Sigma$ of $\mathbf{B}^n$ is nothing but the stereographic projection with the pole at $-e_n$.

Note that the map $\sigma$ sends the origin $0 \in \mathbf{B}^n$ to the point $e_n \in U^n$.

**Low-dimensional Moebius transformations.** Suppose now that $n = 2$. The group $SL(2, \mathbb{C})$ acts on the extended complex plane $S^2 = \mathbb{C} \cup \infty$ by *linear-fractional transformations*:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}. \tag{8.1}$$

Note that the matrix $-I$ lies in the kernel of this action, thus, the above action factors through the group $PSL(2,\mathbb{C}) = SL(2,\mathbb{C})/\pm I$. If we identify the complex-projective line $\mathbb{CP}^1$ with the sphere $S^2 = \mathbb{C} \cup \infty$ *via* the map $[z : w] \mapsto z/w$, the above action of $SL(2,\mathbb{C})$ is nothing but the action of $SL(2,\mathbb{C})$ on $\mathbb{CP}^1$ obtained *via* projection of the linear action of $SL(2,\mathbb{C})$ on $\mathbb{C}^2 \setminus 0$.

EXERCISE 8.5. Show the group $PSL(2,\mathbb{C})$ acts faithfully on $S^2$.

EXERCISE 8.6. Prove that the subgroup $SL(2,\mathbb{R}) \subset SL(2,\mathbb{C})$ preserves the upper half-plane $U^2 = \{z : Im(z) > 0\}$. Moreover, $SL(2,\mathbb{R})$ is the stabilizer of $U^2$ in $SL(2,\mathbb{C})$.

EXERCISE 8.7. Prove that any matrix in $SL(2,\mathbb{C})$ is either of the form

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$$

or it can be written as a product

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

*Hint:* If a matrix is not of the first type then it is a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

such that $c \neq 0$. Use this information and multiplications on the left and on the right by matrices

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

to create zeroes on the diagonal in the matrix.

LEMMA 8.8. *$PSL(2,\mathbb{C})$ is the subgroup $Mob_+(S^2)$ of Moebius transformations of $S^2$ which preserve orientation.*

PROOF. 1. Every linear-fractional transformation is a composition of $j : z \mapsto z^{-1}$, translations, dilations and rotations (see Exercise 8.7). Note that $j(z)$ is the composition of the complex conjugation with the inversion in the unit circle. Thus, $PSL(2,\mathbb{C}) \subset Mob_+(S^2)$. Conversely, let $g \in Mob(S^2)$ and $z_0 := g(\infty)$. Then $h = j \circ \tau \circ g$ fixes the point $\infty$, where $\tau_0(z) = z - z_0$. Let $z_1 = h(0)$. Then composition $f$ of $h$ with the translation $\tau_1 : z \mapsto z - z_1$ has the property that $f(\infty) = \infty, f(0) = 0$. Thus, $f \in CO(2)$ and $h$ preserves orientation. It follows that $f$ has the form $f(z) = \lambda z$, for some $\lambda \in \mathbb{C} \setminus 0$. Since $f$, $\tau_0, \tau - 1, j$ are Moebius transformation, it follows that $g$ is also a Moebius transformation. $\square$

## 8.2. Real hyperbolic space

**Upper half-space model.** We equip $U^n = \mathbb{R}^n_+$ with the Riemannian metric

$$ds^2 = \frac{dx^2}{x_n^2} = \frac{dx_1^2 + ... + dx_n^2}{x_n^2} \tag{8.2}$$

The Riemannian manifold $(U^n, ds^2)$ is called the *n-dimensional hyperbolic space* and denoted $\mathbb{H}^n$. This space is also frequently called the *real-hyperbolic* space, in order to distinguish it from other spaces also called *hyperbolic* (e.g., complex-hyperbolic space, quaternionic-hyperbolic space, Gromov-hyperbolic space, etc.). We will use the terminology *hyperbolic space* for $\mathbb{H}^n$ and add adjective *real* in case when other notions of hyperbolicity are involved in the discussion. In case $n = 2$, we identify $\mathbb{R}^2$ with the complex plane, so that $U^2 = \{z | Im(z) > 0\}$, $z = x + iy$, and

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

Note that the hyperbolic Riemannian metric $ds^2$ on $U^n$ is conformally-Euclidean, hence, hyperbolic angles are equal to the Euclidean angles. One computes hyperbolic volumes of solids in $\mathbb{H}^n$ by the formula

$$Vol(\Omega) = \int_\Omega \frac{dx_1 ... dx_n}{x_n^n}$$

Consider the projection to the $x_n$-axis in $U^n$ given by the formula

$$\pi : (x_1, ..., x_n) \mapsto (0, ..., 0, x_n).$$

EXERCISE 8.9. 1. Verify that $d_x\pi$ does not increase the length of tangent vectors $v \in T_x\mathbb{H}^n$ for every $x \in \mathbb{H}^n$.

2. Verify that for a unit vector $v \in T_x\mathbb{H}^n$, $\|d_x\pi(v)\| = 1$ if and only if $v$ is "vertical", i.e., it has the form $(0, ..., 0, v_n)$.

EXERCISE 8.10. Suppose that $p = ae_n, q = be_n$, where $0 < a < b$. Let $\alpha$ be the vertical path $\alpha(t) = (1 - t)p + tq$, $t \in [0, 1]$ connecting $p$ to $q$. Show that $\alpha$ is the shortest path (with respect to the hyperbolic metric) connecting $p$ to $q$ in $\mathbb{H}^n$. In particular, $\alpha$ is a hyperbolic geodesic and

$$d(p, q) = \log(b/a).$$

Hint: Use the previous exercise.

We note that the metric $ds^2$ on $\mathbb{H}^n$ is clearly invariant under the "horizontal" Euclidean translations $x \mapsto x + v$, where $v = (v_1, ..., v_{n-1}, 0)$ (since they preserve the Euclidean metric and the $x_n$-coordinate). Similarly, $ds^2$ is invariant under the dilations

$$h : x \mapsto \lambda x, \lambda > 0$$

since $h$ scales both numerator and denominator in (8.2) by $\lambda^2$. Lastly, $ds^2$ is invariant under Euclidean rotations which fix the $x_n$-axis (since they preserve the $x_n$-coordinate). Clearly, compositions of such isometries of $\mathbb{H}^n$ act transitively on $\mathbb{H}^n$, which means that $\mathbb{H}^n$ is a *homogeneous Riemannian manifold.*

EXERCISE 8.11. Show that $\mathbb{H}^n$ is a complete Riemannian manifold. You can either use homogeneity of $\mathbb{H}^n$ or show directly that every Cauchy sequence in $\mathbb{H}^n$ lies in a compact subset of $\mathbb{H}^n$.

EXERCISE 8.12. Show that the inversion $J = J_\Sigma$ in the unit sphere $\Sigma$ centered at the origin, is an isometry of $\mathbb{H}^n$, i.e., $ds_B^2 = J^*(ds^2)$. The proof is easy but (somewhat) tedious calculation, which is best done using *calculus* interpretation of the pull-back Riemannian metric.

EXERCISE 8.13. Show that every inversion preserving $\mathbb{H}^n$ is an isometry of $\mathbb{H}^n$. To prove this, use compositions of the inversion $J_\Sigma$ in the unit sphere with translations and dilations.

In order to see clearly other isometries of $\mathbb{H}^n$, it is useful to consider the *unit ball model* of the hyperbolic space.

**Unit ball model.** Consider the open unit Euclidean $n$-ball $\mathbf{B}^n := \{x : |x| < 1\}$ in $\mathbb{R}^n$. We equip $\mathbf{B}^n$ with the Riemannian metric

$$ds_B^2 = 4\frac{dx_1^2 + ... + dx_n^2}{(1 - |x|^2)^2}.$$

The Riemannian manifold $(\mathbf{B}^n, ds^2)$ is called the *unit ball model* of the hyperbolic $n$-space. What is clear in this model is that the group $O(n)$ of orthogonal transformations of $\mathbb{R}^n$ preserves $ds_B^2$ (since its elements preserve $|x|$ and, hence, the denominator of $ds_B^2$). The two models of the hyperbolic space are related by the Moebius transformation $\sigma : \mathbf{B}^n \to U^n$ defined in the previous section.

EXERCISE 8.14. Show that $ds_B^2 = \sigma^*(ds^2)$. The proof is again a straightforward calculation similar to the Exercise 8.12. Namely, first, pull-back $ds^2$ *via* dilatation $x \to 2x$, then apply pull-back *via* the translation $x \mapsto x - \frac{1}{2}e_n$, etc. Thus, $\sigma$ is an isometry of the Riemannian manifolds $(\mathbf{B}^n, ds_B^2), (U^n, ds^2)$.

LEMMA 8.15. *The group $O(n)$ is the stabilizer of* 0 *in the group of isometries of $(\mathbf{B}^n, ds_B^2)$.*

PROOF. Note that if $g \in \mathrm{Isom}(\mathbf{B}^n)$ fixes 0, then its derivative at the origin $dg_0$ is an orthogonal transformation $u$. Thus, $h = u^{-1}g \in \mathrm{Isom}(\mathbf{B}^n)$ has the property $dh_0 = Id$. Therefore, for every geodesic $\gamma$ in $\mathbb{H}^n$ so that $\gamma(0) = 0$, $dh(\gamma'(0)) = \gamma'(0)$. Since geodesic in a Riemannian manifold is uniquely determined by its initial point and initial velocity, we conclude that $h(\gamma(t)) = \gamma(t)$ for every $t$. Since $\mathbf{B}^n$ is complete, for every $q \in \mathbf{B}^n$ there exists a geodesic hyperbolic $\gamma$ connecting $p$ to $q$. Thus, $h(q) = q$ and, therefore, $g = u \in O(n)$. ☐

COROLLARY 8.16. *The stabilizer of the point $p = e_n \in U^n$ in the group* $\mathrm{Isom}(\mathbb{H}^n)$ *is contained in the group of Moebius transformations.*

PROOF. Note that $\sigma$ sends $0 \in B^n$ to $p = e_n \in U^n$, and $\sigma$ is Moebius. Thus, $\sigma : \mathbf{B}^n \to U^n$ conjugates the stabilizer $O(n)$ of 0 in $\mathrm{Isom}(\mathbf{B}^n, ds_B^2)$ to the stabilizer $K = \sigma^{-1}O(n)\sigma$ of $p$ in $\mathrm{Isom}(U^n, ds^2)$. Since $O(n) \subset Mob(S^n), \sigma \in Mob(S^n)$, claim follows. ☐

COROLLARY 8.17. *a.* $\mathrm{Isom}(\mathbb{H}^n)$ *equals the group $Mob(\mathbb{H}^n)$ of Moebius transformations of $S^n$ preserving $\mathbb{H}^n$. b.* $\mathrm{Isom}(\mathbb{H}^n)$ *acts transitively on the unit tangent bundle $U\mathbb{H}^n$ of $\mathbb{H}^n$.*

PROOF. a. Since two models of $\mathbb{H}^n$ differ by a Moebius transformation, it suffices to work with $U^n$.

1. We already know that the $\mathrm{Isom}(\mathbb{H}^n) \cap Mob(\mathbb{H}^n)$ contains a subgroup acting transitively on $\mathbb{H}^n$. We also know, that the stabilizer $K$ of $p$ in $\mathrm{Isom}(\mathbb{H}^n)$ is contained in $Mob(\mathbb{H}^n)$. Thus, given $g \in \mathrm{Isom}(\mathbb{H}^n)$ we first find $h \in Mob(\mathbb{H}^n) \cap \mathrm{Isom}(\mathbb{H}^n)$ so that $k = h \circ g(p) = p$. Since $k \in Mob(\mathbb{H}^n)$, we conclude that $\mathrm{Isom}(\mathbb{H}^n) \subset Mob(\mathbb{H}^n)$.

2. We leave it to the reader to verify that the restriction homomorphism $Mob(\mathbb{H}^n) \to Mob(S^{n-1})$ is injective. Every $g \in Mob(S^{n-1})$ extends to a composition of inversions preserving $\mathbb{H}^n$. Thus, the above restriction map is a group

isomorphism. We already know that inversions $J \in Mob(\mathbb{H}^n)$ are hyperbolic isometries. Thus, $Mob(\mathbb{H}^n) \subset \text{Isom}(\mathbb{H}^n)$.

b. Transitivity of the action of $\text{Isom}(\mathbb{H}^n)$ on $U\mathbb{H}^n$ follows from the fact that this group acts transitively on $\mathbb{H}^n$ and that the stabilizer of $p$ acts transitively on the set of unit vectors in $T_p\mathbb{H}^n$. □

LEMMA 8.18. *Geodesics in $\mathbb{H}^n$ are arcs of circles orthogonal to the boundary sphere of $\mathbb{H}^n$. Furthermore, for every such arc $\alpha$ in $U^n$, there exists an isometry of $\mathbb{H}^n$ which carries $\alpha$ to a segment of the $x_n$-axis.*

PROOF. It suffices to consider complete hyperbolic geodesics $\alpha : \mathbb{R} \to \mathbb{H}^n$. Since $\sigma : \mathbf{B}^n \to U^n$ sends circles to circles and preserves angles, it again suffices to work with the upper half-space model. Let $\alpha$ be a hyperbolic geodesic in $U^n$. Since $\text{Isom}(\mathbb{H}^n)$ acts transitively on $U\mathbb{H}^n$, there exists a hyperbolic isometry $g$ so that the hyperbolic geodesic $\beta = g \circ \alpha$ satisfies: $\beta(0) = p = e_n$ and the vector $\beta'(0)$ has the form $e_n = (0, ..., 0, 1)$. We already know that the curve $\gamma(t) = e^t e_n$ is a hyperbolic geodesic, see Exercise 8.10. Furthermore, $\gamma'(0) = e_n$ and $\gamma(0) = p$. Thus, $\beta = \gamma$ is a (generalized) circle orthogonal to the boundary of $\mathbb{H}^n$. Since $\text{Isom}(\mathbb{H}^n) = Mob(\mathbb{H}^n)$ and Moebius transformations map circles to circles and preserve angles, lemma follows. □

COROLLARY 8.19. *The space $\mathbb{H}^n$ is* uniquely geodesic, *i.e., for every pair of points in $\mathbb{H}^n$ there exists a unique unit speed geodesic segment connecting these points.*

PROOF. By the above lemma, it suffices to consider points $p, q$ on the $x_n$-axis. But, according to Exercise 8.10, the vertical segment is the unique length-minimizing path between such $p$ and $q$. □

COROLLARY 8.20. *Let $H \subset \mathbb{H}^n$ be the intersection of $\mathbb{H}^n$ with a round $k$-sphere orthogonal to the boundary of $\mathbb{H}^n$. Then $H$ is a* totally-geodesic *subspace of $\mathbb{H}^n$, i.e., for every pair of points $p, q \in H$, the unique hyperbolic geodesic $\gamma$ connecting $p$ and $q$ in $\mathbb{H}^n$, is contained in $H$. Furthermore, if $\iota : H \to \mathbb{H}^n$ is the embedding, then the Riemannian manifold $(H, \iota^* ds^2)$ is isometric to $\mathbb{H}^k$.*

PROOF. The first assertion follows from the description of geodesics in $\mathbb{H}^n$. To prove the second assertion, by applying an appropriate isometry of $\mathbb{H}^n$, it suffices to consider the case when $H$ is contained in a coordinate $k$-dimensional subspace in $\mathbb{R}^n$:

$$H = \{(0, ..., 0, x_{n-k+1}, .., x_n) : x_n > 0\}.$$

Then

$$\iota^* ds^2 = \frac{dx_{n-k+1}^2 + ... + dx_n^2}{x_n^2}$$

is isometric to the hyperbolic metric on $\mathbb{H}^k$ (by relabeling the coordinates). □

We will refer to the submanifolds $H \subset \mathbb{H}^n$ as *hyperbolic subspaces.*

EXERCISE 8.21. Show that the hyperbolic plane violates the 5th Euclidean postulate: For every (geodesic) line $L \subset \mathbb{H}^2$ and every point $P \notin L$, there are infinitely many lines through $P$ which are parallel to $L$ (i., disjoint from $L$).

EXERCISE 8.22. Prove that

- the unit sphere $S^{n-1}$ is the ideal boundary (in the sense of Definition 2.44) of the hyperbolic space $\mathbb{H}^n$ in the unit ball model;

- the extended Euclidean space $\mathbb{R}^{n-1} \cup \{\infty\} = S^{n-1}$ is the ideal boundary of the hyperbolic space $\mathbb{H}^n$ in the upper half-space model.

Note that the Moebius transformation $\sigma : \mathbf{B}^n \to U^n$ carries the ideal boundary of $\mathbf{B}^n$ to the ideal boundary of $U^n$. Note also that all Moebius transformations which preserve $\mathbb{H}^n$ in either model, induce Moebius transformations of the ideal boundary of $\mathbb{H}^n$.

It follows from Corollaries 8.20 and 8.33 that $\mathbb{H}^n$ has sectional curvature $-1$, therefore all the considerations in Section 2.1.8, in particular those concerning the ideal boundary, apply to it. Later on, in Section 9.9 of Chapter 9, we will give another more intrinsic definition of ideal boundaries, for metric hyperbolic spaces in the sense of Gromov.

**Lorentzian model of $\mathbb{H}^n$.** We refer the reader to [**Rat94**] and [**Thu97**] for the material below.

Consider the *Lorentzian space* $\mathbb{R}^{n,1}$ which is $\mathbb{R}^{n+1}$ equipped with the quadratic form

$$q(x) = x_1^2 + \ldots + x_n^2 - x_{n+1}^2.$$

Let $H$ denote the upper sheet of the 2-sheeted hyperboloid in $\mathbb{R}^{n,1}$:

$$x_1^2 + \ldots + x_n^2 - x_{n+1}^2 = -1, x_{n+1} > 0.$$

Restriction of $q$ to the tangent bundle of $H$ is positive-definite and defines a Riemannian metric $ds^2$ on $H$. We identify the unit ball $\mathbf{B}^n$ in $\mathbb{R}^n$ with the ball

$$\{(x_1, \ldots, x_n, 0) : x_1^+ \ldots + x_n^2 < 1\} \subset \mathbb{R}^{n+1}.$$

Let $\pi : H \to \mathbf{B}^n$ denote the radial projection from the point $-e_{n+1}$:

$$\pi(x) = tx - (1-t)e_{n+1}, \quad t = \frac{1}{x_{n+1}+1}.$$

One then verifies that

$$\pi : (H, ds^2) \to \mathbb{H}^n = \left(\mathbf{B}^n, \frac{4dx^2}{(1-|x|^2)^2}\right)$$

is an isometry.

The stabilizer $PO(n,1)$ of $H$ in $O(n,1)$ acts isometrically on $H$. Furthermore, $PO(n,1)$ is the entire isometry group of $(H, ds^2)$. Thus, $\mathrm{Isom}(\mathbb{H}^n) \cong PO(n,1) \subset SO(n,1)$; in particular, the Lie group $\mathrm{Isom}(\mathbb{H}^n)$ is linear.

### 8.3. Hyperbolic trigonometry

In this section we consider geometry of triangles in the hyperbolic plane. We refer to [**Bea83, Rat94, Thu97**] for the proofs of the hyperbolic trigonometric formulae introduced in this section. Recall that a (geodesic) triangle $T = T(A, B, C)$ as a 1-*dimensional object*. From the Euclidean viewpoint, a hyperbolic triangle $T$ is a concatenations of circular arcs connecting points $A, B, C$ in $\mathbb{H}^2$, where the circles containing the arcs are orthogonal to the boundary of $\mathbb{H}^2$. Besides such "conventional" triangles, it is useful to consider *generalized* hyperbolic triangles where some vertices are *ideal*, i.e., they belong to the ideal boundary of $\mathbb{H}^2$. Such triangles are easiest to introduce by using Euclidean interpretation of hyperbolic triangles: One simply allows some (or, even all) vertices $A, B, C$ to be points on the boundary circle of $\mathbb{H}^2$, the rest of the definition is exactly the same. *However, we no longer allow two vertices which belong to the boundary circle $S^1$ to be the same.*

The vertices of $T$ which happen to be points of the boundary circle $S^1$ are called the *ideal vertices* of $T$. The *angle* of $T$ at its ideal vertex is just the Euclidean angle. In general, we will use the notation $\alpha = \angle_A(B, C)$ to denote the angle of $T$ at $a$. From now on, a *hyperbolic triangle* means either a usual triangle or a triangle where some vertices are ideal. We still refer to such triangles as *triangles in* $\mathbb{H}^2$, even though, some of the vertices could lie on the ideal boundary, so, strictly speaking, an ideal hyperbolic triangle in $\mathbb{H}^2$ is not a subset of $\mathbb{H}^2$. An *ideal* hyperbolic triangle, is a triangle where all the vertices are distinct ideal points in $\mathbb{H}^2$. The same conventions will be used for hyperbolic triangles in $\mathbb{H}^n$.

EXERCISE 8.23. If $A$ is an ideal vertex of a hyperbolic triangle $T$, then $T$ has zero angle at $A$. Hint: It suffices to consider the case when $A = 0$ and the side $[A, B]$ of $T$ is contained in the vertical line $L$. Show that the side $[A, C]$ of $T$ is a circular arc tangent to $L$ at $A$.

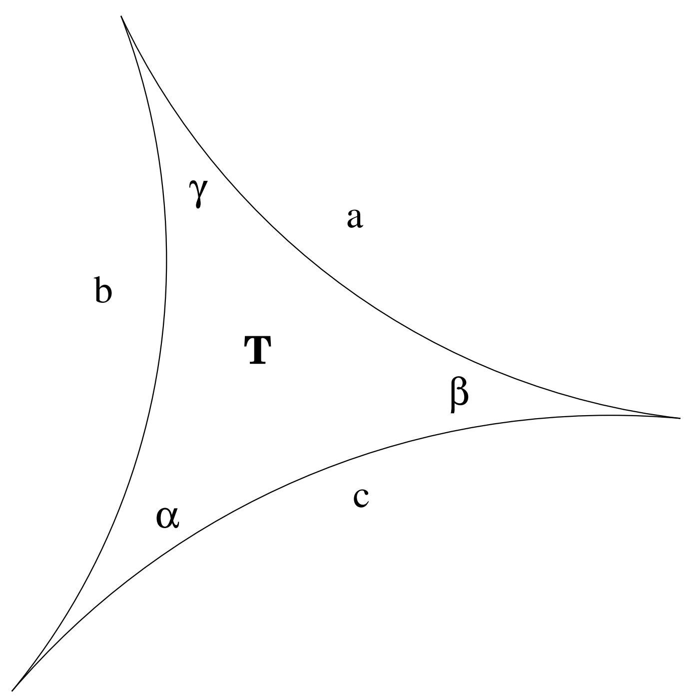



FIGURE 8.1. Geometry of a general hyperbolic triangle.

**1. General triangles.** Consider hyperbolic triangles $T$ in $\mathbb{H}^2$ with the side-lengths $a, b, c$ and the opposite angles $\alpha, \beta, \gamma$, see Figure 8.1.

**a. Hyperbolic Sine Law:**

$$\frac{\sinh(a)}{\sin(\alpha)} = \frac{\sinh(b)}{\sin(\beta)} = \frac{\sinh(c)}{\sin(\gamma)}. \tag{8.3}$$

**b. Hyperbolic Cosine Law:**

$$\cosh(c) = \cosh(a)\cosh(b) - \sinh(a)\sinh(b)\cos(\gamma) \tag{8.4}$$

**c. Dual Hyperbolic Cosine Law:**

$$\cos(\gamma) = -\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)\cosh(c) \tag{8.5}$$

**2. Right triangles.** Consider a right-angled hyperbolic triangle with the hypotenuse $c$, the other side-lengths $a, b$ and the opposite angles $\alpha, \beta$. Then, hyperbolic cosine laws become:

$$\cosh(c) = \cosh(a)\cosh(b), \tag{8.6}$$

$$\cos(\alpha) = \sin(\beta)\cosh(a), \tag{8.7}$$

$$\cos(\alpha) = \frac{\tanh b}{\tanh c} \tag{8.8}$$

In particular,

$$\cos(\alpha) = \frac{\cosh(a)\sinh(b)}{\sinh(c)}. \tag{8.9}$$

**3. First variation formula for right triangles.** We now hold the side $a$ fixed and vary the hypotenuse in the above right-angled triangle. By combining (8.6) and (8.4) we obtain the *First Variation Formula*:

$$c'(0) = \frac{\cosh(a)\sinh(b)}{\sinh(c)} b'(0) = \cos(\alpha) b'(0). \tag{8.10}$$

The equation $c'(0) = \cos(\alpha)b'(0)$ is a special case of the *First Variation Formula* in Riemannian geometry, which applies to general Riemannian manifolds.

As an application of the first variation formula, consider a hyperbolic triangle with vertices $A, B, C$, side-lengths $a, b, c$ and the angles $\beta, \gamma$ opposite to the sides $b, c$. Then

LEMMA 8.24. *$a + b - c \geqslant ma$, where*

$$m = \min\{|1 - \cos(\beta)|, |1 - \cos(\gamma)|\}.$$

PROOF. We let $g(t)$ denote the unit speed parameterizations of the segment $[BC]$, so that $g(0) = C, g(a) = B$. Let $c(t)$ denote the distance $\mathrm{dist}(A, g(t))$ (so that $b = c(0), c = c(a)$) and let $\beta(t)$ denote the angle $\angle Ag(t)B$. We leave it to the reader to verify that

$$|1 - \cos(\beta(t))| \geqslant m.$$

Consider the function

$$f(t) = t + b - c(t), \quad f(0) = 0, \quad f(a) = a + b - c.$$

By the 1st variation formula,

$$c'(t) = \cos(\beta(t))$$

and, hence,

$$f'(t) = 1 - \cos(\beta(t)) \geqslant m$$

Thus,

$$a + b - c = f(a) \geqslant ma \quad \square$$

EXERCISE 8.25. [Monotonicity of the hyperbolic distance] Let $T_i, i = 1, 2$ be right hyperbolic triangles with vertices $A_i, B_i, C_i$ (where $A_i$ or $B_i$ could be ideal vertices) so that $A = A_1 = A_2$, $[A_1, B_1] \subset [A_2, B_2]$, $\alpha_1 = \alpha_2$ and $\gamma_1 = \gamma_2 = \pi/2$. See Figure 8.2. Then $a_1 \leqslant a_2$. Hint: Use either (8.8).

In other words, if $\sigma(t), \tau(t)$ are hyperbolic geodesic with unit speed parameterizations, so that $\sigma(0) = \tau(0) = A \in \mathbb{H}^2$, then the distance $d(\sigma(t), \tau)$ from the point $\sigma(t)$ to the geodesic $\tau$, is a monotonically increasing function of $t$.

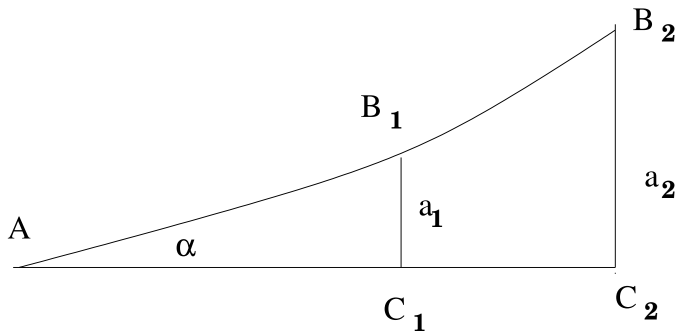



FIGURE 8.2. *M*onotonicity of distance.

## 8.4. Triangles and curvature of $\mathbb{H}^n$

Given points $A, B, C \in \mathbb{H}^n$ we define the *hyperbolic triangle* $T = [A, B, C] = \Delta ABC$ with vertices $A, B, C$. We topologize the set $Tri(\mathbb{H}^n)$ of hyperbolic triangles $T$ in $\mathbb{H}^n$ by using topology on triples of vertices of $T$, i.e., a subset topology in $(\bar{\mathbf{B}}^n)^3$.

EXERCISE 8.26. Angles of hyperbolic triangles are continuous functions on $Tri(\mathbb{H}^n)$.

EXERCISE 8.27. Every hyperbolic triangle $T$ in $\mathbb{H}^n$ is contained in (the compactification of) a 2-dimensional hyperbolic subspace $H \subset \mathbb{H}^n$. Hint: Consider a triangle $T = [A, B, C]$, where $A, B$ belong to a common vertical line.

So far, we considered only geodesic hyperbolic triangles, we now introduce their 2-dimensional counterparts. First, let $T = T(A, B, C)$ be a generalized hyperbolic triangle in $\mathbb{H}^2$. We will assume that $T$ is *nondegenerate*, i.e., is not contained in a hyperbolic geodesic. Such triangle $T$ cuts $\mathbb{H}^2$ in several (2, 3 or 4) convex regions, one of which has the property that its boundary is the triangle $T$. The closure of this region is called *solid* (generalized) hyperbolic triangle and denoted $\blacktriangle = \blacktriangle(A, B, C)$. It $T$ is degenerate, we set $\blacktriangle = T$. More generally, if $T \subset \mathbb{H}^n$ is a hyperbolic triangle, then the *solid triangle* bounded by $T$ is the solid triangle bounded by $T$ in the hyperbolic plane $H \subset \mathbb{H}^n$ containing $T$. We will retain the notation $\blacktriangle$ for solid triangles in $\mathbb{H}^n$.

Exercise 8.28. Let $S$ be a hyperbolic triangle with the sides $\sigma_i, i = 1, 2, 3$. Then there exists an ideal hyperbolic triangle $T$ in $\mathbb{H}^2$ with the sides $\tau_i, i = 1, 2, 3$, bounding solid triangle $\blacktriangle$, so that $S \subset \blacktriangle$ and $\sigma_1$ is contained in the side $\tau_1$ of $T$. See Figure 8.3.

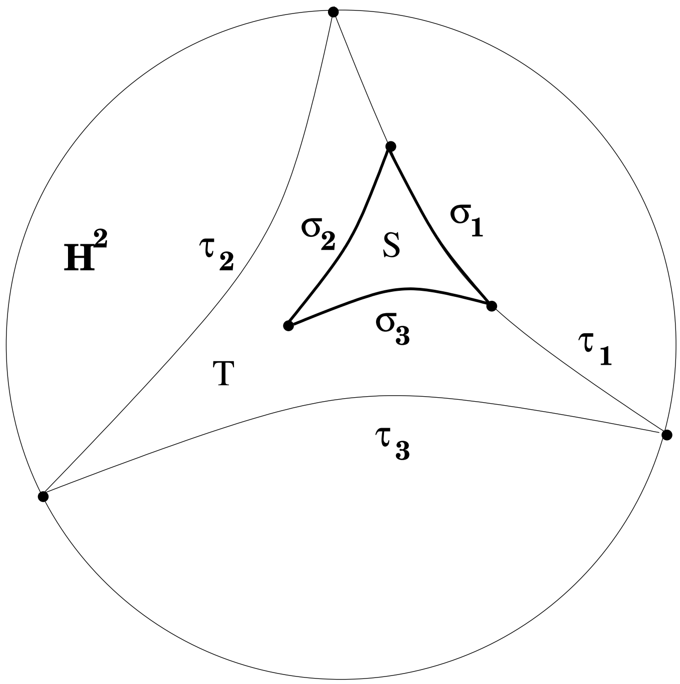



Figure 8.3. *T*riangles in the hyperbolic plane.

Lemma 8.29. Isom($\mathbb{H}^2$) *acts transitively on the set of ordered triples of pairwise distinct points in* $\mathbb{H}^2$.

Proof. Let $a, b, c \in \mathbb{R} \cup \infty$ be distinct points. By applying inversion we send $a$ to $\infty$, so we can assume $a = \infty$. By applying a translation in $\mathbb{R}$ we get $b = 0$. Lastly, composing a map of the type $x \to \lambda x$, $\lambda \in \mathbb{R} \setminus 0$, we send $c$ to 1. The composition of the above maps is a Moebius transformation of $S^1$ and, hence, equals to the restriction of an isometry of $\mathbb{H}^2$. □

Corollary 8.30. *All ideal hyperbolic triangles are congruent to each other.*

Exercise 8.31. Generalize the above corollary to: Every hyperbolic triangle is uniquely determined by its angles. Hint: Use hyperbolic trigonometry.

We will use the notation $T_{\alpha,\beta,\gamma}$ to denote unique (up to congruence) triangle with the angles $\alpha, \beta, \gamma$.

Given a hyperbolic triangle $T$ bounding a solid triangle $\blacktriangle$, the *area of* $T$ is the area of $\blacktriangle$

$$Area(T) = \iint_{\blacktriangle} \frac{dxdy}{y^2}.$$

Area of a degenerate hyperbolic triangle is, of course, zero. Here is an example of the area calculation. Consider the triangle $T = T_{0,\alpha,\pi/2}$ (which has angles $\pi/2, 0, \alpha$). We can realize $T$ as the triangle with the vertices $i, \infty, e^{i\alpha}$. Computing hyperbolic area of this triangle (and using the substitution $x = \cos(t), \alpha \leqslant t \leqslant \pi/2$), we obtain

$$Area(T) = \iint_{\blacktriangle} \frac{dxdy}{y^2} = \frac{\pi}{2} - \alpha.$$

For $T = T_{0,0,\alpha}$, we subdivide $T$ in two right triangles congruent to $T_{0,\alpha/2,\pi/2}$ and, thus, obtain

$$Area(T_{0,0,\alpha}) = \pi - \alpha. \tag{8.11}$$

In particular, area of the ideal triangle equals $\pi$.

LEMMA 8.32. *$Area(T_{\alpha,\beta,\gamma}) = \pi - (\alpha + \beta + \gamma)$.*

PROOF. The proof given here is due to Gauss, it appears in the letter from Gauss to Bolyai, see [**Gau73**]. We realize $T = T_{\alpha,\beta,\gamma}$ as a part of the subdivision of an ideal triangle $T_{0,0,0}$ in four triangles, the rest of which are $T_{0,0,\alpha'}, T_{0,0,\beta'}, T_{0,0,\gamma'}$, where $\theta' = \pi - \theta$ is the complementary angle. See Figure 8.4. Using additivity of area and equation (8.11), we obtain the area formula for $T$. ☐

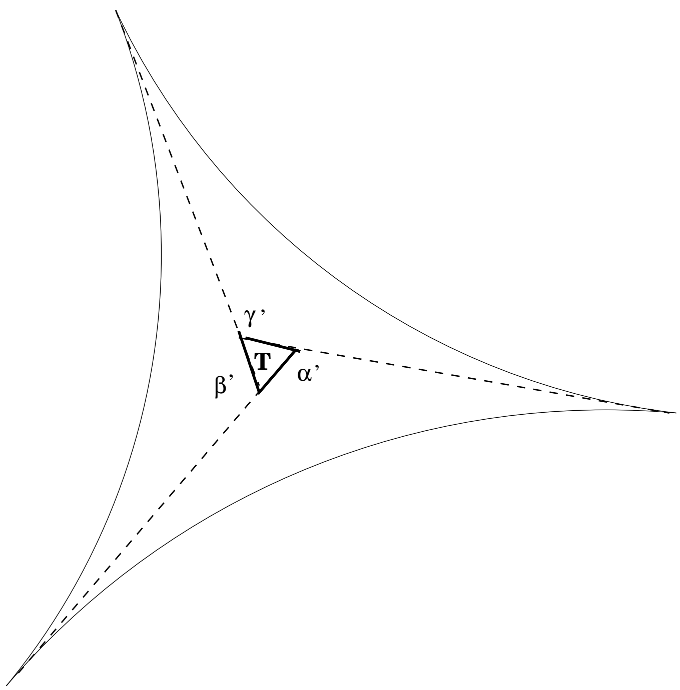



FIGURE 8.4. *Computation of area of the triangle $T$.*

**Curvature computation.** Our next goal is to compute sectional curvature of $\mathbb{H}^n$. Since $\mathrm{Isom}(\mathbb{H}^n)$ acts transitively on pairs $(p, P)$, where $P \subset T_pM$ is a 2-dimensional subspace, it follows that $\mathbb{H}^n$ has *constant* sectional curvature $\kappa$ (see Section 2.1.6). Since $\mathbb{H}^2 \subset \mathbb{H}^n$ is a totally-geodesic isometric embedding (in the sense of Riemannian geometry), $\kappa$ is the same for $\mathbb{H}^n$ and $\mathbb{H}^2$.

COROLLARY 8.33. *The Gaussian curvature* $\kappa$ *of* $\mathbb{H}^2$ *equals* $-1$.

PROOF. Instead of computing curvature tensor (see e.g. **[dC92]** for the computation), we will use Gauss-Bonnet formula. Comparing the area computation given in Lemma 8.32 with Gauss-Bonnet formula (Theorem 2.21) we conclude that $\kappa = -1$. □

Note that scaling properties of the sectional curvature (see Section 2.1.6) imply that sectional curvature of

$$\left(U^n, \frac{adx^2}{x_n^2}\right)$$

equals $-a^2$ for every $a > 0$.

## 8.5. Distance function on $\mathbb{H}^n$

We begin by defining the following quantities:

$$\text{dist}\,(z,w) = \operatorname{arccosh}\left(1 + \frac{|z-w|^2}{2\operatorname{Im} z \operatorname{Im} w}\right)\; z, w \in U^2 \tag{8.12}$$

and, more generally,

$$\text{dist}\,(p,q) = \operatorname{arccosh}\left(1 + \frac{|p-q|^2}{2p_n q_n}\right)\; p, q \in U^n \tag{8.13}$$

It is immediate that $dist(p,q) = dist(q,p)$ and that $dist(p,q) = 0$ if and only if $p = q$. However, it is, *a priori*, far from clear that *dist* satisfies the triangle inequality.

LEMMA 8.34. *dist is invariant under* $\operatorname{Isom}(\mathbb{H}^n) = Mob(U^n)$.

PROOF. First, it is clear that *dist* is invariant under the group $Euc(U^n)$ of Euclidean isometries which preserve $U^n$. Next, any two points in $U^n$ belong to a vertical half-plane in $U^n$. Applying elements of $Euc(U^n)$ to this half-plane, we can transform it to the coordinate half-plane $U^2 \subset U^n$. Thus, the problem reduces to the case $n = 2$ and orientation-preserving Moebius transformations of $\mathbb{H}^2$. We leave it to the reader as an exercise to show that the map $z \mapsto -\frac{1}{z}$ (which is an element of $PSL(2,\mathbb{R})$) preserves the quantity

$$\frac{|z-w|^2}{\operatorname{Im} z \operatorname{Im} w}$$

and, hence, *dist*. Now, the assertion follows from Exercise 8.7 and Lemma 8.8. □

Recall that $d(p,q)$ denotes the hyperbolic distance between points $p, q \in U^n$.

PROPOSITION 8.35. $dist(p,q) = d(p,q)$ *for all points* $p, q \in \mathbb{H}^n$. *In particular, the function dist is indeed a metric on* $\mathbb{H}^n$.

PROOF. As in the above lemma, it suffices to consider the case $n = 2$. We can also assume that $p \neq q$. First, suppose that $p = i$ and $q = ib$, $b > 1$. Then, by Exercise 8.10,

$$dist(p,q) = \int_1^b \frac{dt}{t} = \log(b), \quad \exp(d(p,q)) = b.$$

On the other hand, the formula (8.12) yields:

$$dist(p,q) = \operatorname{arccosh}\left(1 + \frac{(b-1)^2}{2b}\right).$$

Hence,

$$\cosh(dist(p,q)) = \frac{e^{dist(p,q)} + e^{-dist(p,q)}}{2} = 1 + \frac{(b-1)^2}{2b}.$$

Now, the equality $dist(p,q) = d(p,q)$ follows from the identity

$$1 + \frac{(b-1)^2}{2b} = \frac{b + b^{-1}}{2}.$$

For general points $p, q$ in $\mathbb{H}^2$, by Lemma 8.18, there exists a hyperbolic isometry which sends $p$ to $i$ and $q$ to a point of the form $ib, b \geqslant 1$. We already know that both hyperbolic distance $d$ and the quantity $dist$ are invariant under the action of $\mathrm{Isom}(\mathbb{H}^2)$. Thus, the equality $d(p,q) = dist(p,q)$ follows from the special case of points on the $y$-axis. □

EXERCISE 8.36. Deduce from (8.12) that

$$\ln\left(1 + \frac{|z-w|^2}{2\,\mathrm{Im}\, z\,\mathrm{Im}\, w}\right) \leq d(z,w) \leq \ln\left(1 + \frac{|z-w|^2}{2\,\mathrm{Im}\, z\,\mathrm{Im}\, w}\right) + \ln 2$$

for all points $z, w \in U^2$.

## 8.6. Hyperbolic balls and spheres

Pick a point $p \in \mathbb{H}^n$ and a positive real number $R$. Then the *hyperbolic sphere* of radius $R$ centered at $p$ is the set

$$S_h(p,R) = \{x \in \mathbb{H}^n : d(x,p) = R\}.$$

EXERCISE 8.37. 1. Prove that $S_h(e_n, R) \subset \mathbb{H}^n = U^n$ equals the Euclidean sphere of center $\cosh(R)e_n$ and radius $\sinh(R)$. *Hint.* It follows immediately from the distance formula (8.12).

2. Suppose that $S = S(x,R) \subset U^n$ is a Euclidean sphere with Euclidean radius $R$ and the center $x$ so that $x_n = a$. Then $S = S_h(p,r)$, where the hyperbolic radius $r$ equals

$$\frac{1}{2}\left(\log(a+R) - \log(a-R)\right).$$

Since group generated by dilations and horizontal translations acts transitively on $U^n$, it follows that every hyperbolic sphere is also a Euclidean sphere. A non-computational proof of this fact is as follows: Since the hyperbolic metric $ds_B^2$ on $\mathbf{B}^n$ is invariant under $O(n)$, it follows that hyperbolic spheres centered at 0 in $\mathbf{B}^n$ are also Euclidean spheres. The general case follows from transitivity of $\mathrm{Isom}(\mathbb{H}^n)$ and the fact that isometries of $\mathbb{H}^n$ are Moebius transformations, which, therefore, send Euclidean spheres to Euclidean spheres.

LEMMA 8.38. *Suppose that $B(x_1,R_1) \subset B(x_2,R_2)$ are hyperbolic balls. Then $R_1 \leqslant R_2$.*

PROOF. It follows from the triangle inequality that the diameter of a metric ball $B(x,R)$ is the longest geodesic segment contained in $B(x,R)$. Therefore, let $\gamma \subset B(x_1,R_1)$ be a diameter. Then $\gamma$ is contained in $B(x_2,R_2)$ and, hence, its length is $\leqslant 2R_2$. However, length of $\gamma$ is $2R_1$, therefore, $R_1 \leqslant R_2$. □

## 8.7. Horoballs and horospheres in $\mathbb{H}^n$

Consider the unit ball model $\mathbf{B}^n$ of $\mathbb{H}^n$, $\alpha$ a point in the ideal boundary (here identified with the unit sphere $S^{n-1}$) and $r$ a geodesic ray with $r(\infty) = \alpha$, i.e. according to Lemma 8.18, an arc of circle orthogonal to $S^{n-1}$ in $\alpha$ with the other endpoint $x$ in the interior of $\mathbf{B}^n$. By Lemma 2.52, the open horoball $B(\alpha)$ defined by the inequality $f_r < 0$, where $f_r$ is the Busemann function for the ray $r$, equals the union of open balls $\bigcup_{t\geqslant 0} B(r(t), t)$ . The discussion in Section 8.6, in particular Exercise 8.37, implies that each ball $B(r(t), t)$ is a Euclidean ball with center in a point $r(T_t)$ with $T_t > t$ . Therefore, the above union is the open Euclidean ball with boundary tangent to $S^{n-1}$ at $\alpha$, and containing the point $x$. According to Lemma 2.54, the closed horoball and the horosphere defined by $f_r \leqslant 0$ and $f_r = 0$, respectively, are the closed Euclidean ball and the boundary sphere, both with the point $\alpha$ removed.

We conclude that the set of horoballs (closed or open) with center $\alpha$ is the same as the set of Euclidean balls (closed or open) tangent to $S^{n-1}$ at $\alpha$, with the point $\alpha$ removed.

Applying the map $\sigma : \mathbf{B}^n \to U^n$ to horoballs and horospheres in $\mathbf{B}^n$, we obtain horoballs and horospheres in the upper-half space model $U^n$ of $\mathbb{H}^n$. Being a Moebius transformation, $\sigma$ carries Euclidean spheres to Euclidean spheres (recall that a compactified Euclidean hyperplane is also regarded as a Euclidean sphere). It is then clear that hyperbolic isometries carry horoballs/horospheres to horoballs/horospheres.

Recall that $\sigma(-e_n) = \infty$. Therefore, every horosphere in $\mathbf{B}^n$ centered at $-e_n$ is sent by $\sigma$ to an $n-1$-dimensional Euclidean subspace $E$ of $U^n$ whose compactification contains the point $\infty$. Hence, $E$ has to be a horizontal Euclidean subspace, i.e., a subspace of the form

$$\{x \in U^n : x_n = t\}$$

for some fixed $t > 0$. Restricting the metric $ds^2$ to such $E$ we obtain the Euclidean metric rescaled by $t^{-2}$. Thus, the restriction of $ds^2$ to every horosphere is isometric to the flat metric on $\mathbb{R}^{n-1}$.

EXERCISE 8.39. Consider the upper half-space model for the hyperbolic space $\mathbb{H}^n$ and the vertical geodesic ray $r$ in $\mathbb{H}^n$:

$$r = \{(0, \dots, 0, x_n) : x_n \geqslant 1\}.$$

Show that the Busemann function $f_r$ for the ray $r$ is given by

$$f_r(x_1, \dots, x_n) = -\log(x_n).$$

## 8.8. $\mathbb{H}^n$ is a symmetric space

A *symmetric space* is a complete simply connected Riemannian manifold $X$ such that for every point $p$ there exists a global isometry of $X$ which is a geodesic symmetry $\sigma_p$ with respect to $p$, that is for every geodesic $\mathfrak{g}$ through $p$, $\sigma_p(\mathfrak{g}(t)) = \mathfrak{g}(-t)$. Let us verify that such $X$ is a homogeneous Riemannian manifold. Indeed, given points $p, q \in X$, let $m$ denote the midpoint of a geodesic connecting $p$ to $q$. Then $\sigma_m(p) = q$. Besides being homogeneous, symmetric spaces also admit large discrete isometry groups: For every symmetric space $X$, there exists a subgroup $\Gamma \subset \mathrm{Isom}(X)$ which acts geometrically on $X$.

Details on symmetric spaces can be found for instance in [**Hel01**] and [**Ebe72**]. The *rank* of a symmetric space $X$ is the largest number $r$ so that $X$ contains a totally-geodesic submanifold $F \subset X$ which is isometric to an open disk in $\mathbb{R}^r$.

We note that in the unit ball model of $\mathbb{H}^n$ we clearly have the symmetry $\sigma_p$ with respect to $p = 0$, namely, $\sigma_0 : x \mapsto -x$. Since $\mathbb{H}^n$ is homogeneous, it follows that it has a symmetry at every point. Thus, $\mathbb{H}^n$ is a symmetric space.

EXERCISE 8.40. Prove that the linear-fractional transformation $\sigma_i \in PSL(2, \mathbb{R})$ defined by $\pm S_i$, where $S_i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ fixes $i$ and is a symmetry with respect to $i$.

We proved in Section 8.4 that $\mathbb{H}^n$ has negative curvature $-1$. In particular, it contains no totally-geodesic Euclidean subspaces of dimension $\geqslant 2$ and, thus, $\mathbb{H}^n$ has rank 1.

It turns out that besides real-hyperbolic space $\mathbb{H}^n$, there are three other families of rank 1 negatively curved symmetric spaces: $\mathbf{C}\mathbb{H}^n, n \geqslant 2$ (complex-hyperbolic spaces) $\mathbf{H}\mathbb{H}^n, n \geqslant 2$ (quaternionic hyperbolic spaces) and $\mathbf{O}\mathbb{H}^2$ (octonionic hyperbolic plane). The rank 1 symmetric spaces $X$ are also characterized among symmetric spaces by the property that any two segments of the same length are congruent in $X$. Below is a brief discussion of these spaces, we refer to Mostow's book [**Mos73**] and Parker's survey [**Par08**] for a more detailed discussion.

In all four cases, the symmetric $X$ will appear as a projectivization of a certain cone equipped with a hermitian form $\langle \cdot, \cdot \rangle$ and the distance function in $X$ will be given by the formula:

$$\cosh^2(\text{dist}(p, q)) = \frac{\langle p, q \rangle \langle q, p \rangle}{\langle p, p \rangle \langle q, q \rangle}, \tag{8.14}$$

where $p, q \in C$ represent points in $X$.

**Complex-hyperbolic space.** Consider $\mathbb{C}^{n+1}$ equipped with the Hermitian bilinear form

$$\langle v, w \rangle = \sum_{k=1}^{n} v_k \bar{w}_k - v_{n+1} \bar{w}_{n+1}.$$

The group $U(n, 1)$ is the group of complex-linear automorphisms of $\mathbb{C}^{n+1}$ preserving this bilinear form. Consider the negative light cone

$$C = \{v : \langle v, v \rangle < 0\} \subset \mathbb{C}^{n+1}.$$

Then the *complex-hyperbolic space* $\mathbb{CH}^n$ is the projectivization of $C$. The group $PU(n, 1)$ acts naturally on $X = \mathbb{CH}^n$. One can describe the Riemannian metric on $\mathbb{CH}^n$ as follows. Let $p \in C$ be such that $\langle p, p \rangle = 1$; tangent space at the projection of $p$ to $X$ is the projection of the orthogonal complement $p^{\perp}$ in $\mathbb{C}^{n+1}$. Let $v, w \in \mathbb{C}^{n+1}$ be such that $\langle p, v \rangle = 0, \langle p, w \rangle$. Then set

$$(v, w)_p := -Im \langle v, w \rangle .$$

This determines a $PU(n, 1)$-invariant Riemannian metric on $X$. The corresponding distance function (8.14) will be $G$-invariant.

**Quaternionic-hyperbolic space.** Consider the ring $\mathbf{H}$ of *quaternions*; the elements of the quaternion ring have the form

$$q = x + iy + jz + kw, \quad x, y, z, w \in \mathbb{R}.$$

The quaternionic conjugation is given by

$$\bar{q} = x - iy - jz - kw$$

and

$$|q| = (q\bar{q})^{1/2} \in \mathbb{R}_+$$

is the quaternionic norm. A *unit quaternions* is a quaternion of the unit norm. Let $V$ be a left $n+1$-dimensional free module over $\mathbf{H}$:

$$V = \{\mathbf{q} = (q_1, \ldots, q_{n+1}) : q_m \in \mathbf{H}\}.$$

Consider the quaternionic-hermitian inner product of signature $(n, 1)$:

$$\langle \mathbf{p}, \mathbf{q} \rangle = \sum_{m=1}^{n} p_m \bar{q}_m - p_{n+1}\bar{q}_{n+1}.$$

Then the group $G = Sp(n, 1)$ is the group of automorphisms of the module $V$ preserving this inner product. The quotient of $V$ by the group of nonzero quaternions $\mathbf{H}^\times$ (with respect to the multiplication action) is the $n$-dimensional *quaternionic-projective space* $PV$. Analogously to the case of real and complex hyperbolic spaces, we consider the negative light cone

$$C = \{\mathbf{q} \in V : \langle q, q \rangle < 0\}.$$

The group $G$ acts naturally on $PC \subset PV$ through the group $PSp(n, 1)$ (the quotient of $G$ by the subgroup of unit quaternions embedded in the subgroup of diagonal matrices in $G$). The space $PC$ is called the $n$-dimensional *quaternionic-hyperbolic space* $\mathbf{H}\mathbb{H}^n$

**Octonionic-hyperbolic plane.** One defines *octonionic-hyperbolic plane* $\mathbf{O}\mathbb{H}^2$ analogously to $\mathbf{H}\mathbb{H}^n$, only using the algebra $\mathbf{O}$ of Cayley octonions instead of quaternions. An extra complication comes from the fact that the algebra $\mathbf{O}$ is not associative, so one cannot talk about free $\mathbf{O}$-modules; we refer the reader to [**Mos73, Par08**] for the details.

## 8.9. Inscribed radius and thinness of hyperbolic triangles

Suppose that $T$ is a hyperbolic triangle in the hyperbolic plane $\mathbb{H}^2$ with the sides $\tau_i, i = 1, 2, 3$, so that $T$ bounds the solid triangle $\blacktriangle$. For a point $x \in \blacktriangle$ define the quantities

$$\Delta_x(T) := \max_{i=1,2,3} d(x, \tau_i).$$

and

$$\Delta(T) := \inf_{x \in \blacktriangle} \Delta_x(T).$$

The goal of this section is to estimate $\Delta(T)$ from above. It is immediate that the infimum in the definition of $\Delta(T)$ is realized by a point $x_o \in \blacktriangle$ which is equidistant from all the three sides of $T$, i.e., by the intersection point of the angle bisectors.

Define the *inscribed radius* $inrad(T)$ of $T$ is the supremum of radii of hyperbolic disks contained in $\blacktriangle$.

LEMMA 8.41. $\Delta(T) = Inrad(T)$.

PROOF. Suppose that $D = B(X,R) \subset \blacktriangle$ is a hyperbolic disk. Unless $D$ touches two sides of $T$, there exists a disk $D' = B(X',R') \subset \blacktriangle$ which contains $D$ and, hence, has larger radius, see Lemma 8.38. Suppose, therefore, that $D \subset \blacktriangle$ touches two boundary edges of $T$, hence, center $X$ of $D$ belongs to the bisector $\sigma$ of the corner $ABC$ of $T$. Unless $D$ touches all three sides of $T$, we can move the center $X$ of $D$ along the bisector $\sigma$ away from the vertex $B$ so that the resulting disk $D' = B(X',R')$ still touches only the sides $[A,B],[B,C]$ of $T$. We claim that the (radius $R'$ of $D'$ is larger than the radius $R$ of $D$. In order to prove this, consider hyperbolic triangles $[X,Y,B]$ and $[X',Y',B']$, where $Y,Y'$ are the points of tangency between $D,D'$ and the side $[BA]$. These right-angled triangles have the common angle $\angle_b xy$ and satisfy

$$d(B,X) \leqslant d(B,X').$$

Thus, the inequality $R \leqslant R'$ follows from the Exercise 8.25. □

Thus, we need to estimate inradius of hyperbolic triangles from above. Recall that by Exercise 8.28, for every hyperbolic triangle $S$ in $\mathbb{H}^2$ there exists an ideal hyperbolic triangle $T$, so that $S \subset \blacktriangle$. Clearly, $inrad(S) \leqslant inrad(T)$. Since all ideal hyperbolic triangles are congruent, it suffices to consider the ideal hyperbolic triangle $T$ in $U^2$ with the vertices $-1, 1, \infty$. The inscribed circle $C$ in $T$ has Euclidean center $(0,2)$ and Euclidean radius 1. Therefore, by Exercise 8.37, its hyperbolic radius equals $\log(3)/2$. By combining these observations with Exercise 8.27, we obtain

PROPOSITION 8.42. *For every hyperbolic triangle $T$, $\Delta(T) = inrad(T) \leqslant \frac{\log(3)}{2}$. In particular, for every hyperbolic triangle in $\mathbb{H}^n$, there exists a point $p \in H^n$ so that distance from $p$ to all three sides of $T$ is $\leqslant \frac{\log(3)}{2}$.*

Another way to measure thinness of a hyperbolic triangle $T$ is to compute distance from points of one side of $T$ to the union of the two other sides. Let $T$ be a hyperbolic triangle with sides $\tau_j, j = 1,2,3$. Define

$$\delta(T) := \max_j \sup_{p\in\tau_j} d(p, \tau_{j+1} \cup \tau_{j+2}),$$

where indices of the sides of $T$ are taken modulo 3. In other words, if $\delta = \delta(T)$ then each side of $T$ is contained in the $\delta$-neighborhood of the union of the other two sides.

PROPOSITION 8.43. *For every geodesic triangle $S$ in $\mathbb{H}^n$, $\delta(S) \leqslant \operatorname{arccosh}(\sqrt{2})$.*

PROOF. First of all, as above, it suffices to consider the case $n = 2$. Let $\sigma_j, j = 1,2,3$ denote the edges of $S$. We will estimate $d(p, \sigma_2 \cup \sigma_3)$ (from above) for $p \in \sigma_1$. We enlarge the hyperbolic triangle $S$ to an ideal hyperbolic triangle $T$ as in Figure 8.5. For every $p \in \sigma_1$, every geodesic segment $g$ connecting $p$ to a point of $\tau_2 \cup \tau_3$ has to cross $\sigma_2 \cup \sigma_3$. In particular,

$$d(p, \sigma_2 \cup \sigma_3) \leqslant d(p, \tau_2 \cup \tau_3).$$

Thus, it suffices to show that $\delta(T) \leqslant \operatorname{arccosh}(\sqrt{2})$ for the ideal triangle $T$ as above. We realize $T$ as the triangle with the (ideal) vertices $A_1 = \infty, A_2 = -1, A_3 = 1$ in $\partial_\infty \mathbb{H}^2$. We parameterize sides $\tau_i = [A_{j-1}, A_{j+1}], j = 1,2,3$ modulo 3, according to their orientation. Then, by the Exercise 8.25, for every $i$,

$$d(\tau_j(t), \tau_{j-1})$$

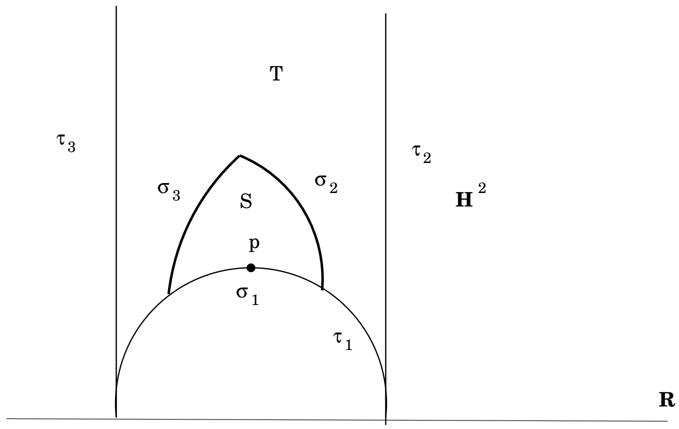



FIGURE 8.5. *E*nlarging hyperbolic triangle $S$.

is monotonically increasing. Thus,

$$\sup_t d(\tau_1(t), \tau_2 \cup \tau_3)$$

is achieved at the point $p = \tau_1(t) = i = \sqrt{-1}$ and equals $d(p,q)$, where $q = -1+\sqrt{2}i$. Then, using formula 8.13, we get $d(p,q) = \operatorname{arccosh}(\sqrt{2})$. Note that alternatively, one can get the formula for $d(p,q)$ from (8.7) by considering the right triangle $[p,q,-1]$ where the angle at $p$ equals $\pi/4$. $\square$

As we will see in Section 9.1, the above propositions mean that *all hyperbolic triangles are uniformly thin.*

## 8.10. Existence-uniqueness theorem for triangles

*Proof of Lemma 2.31.* We will prove this result for the hyperbolic plane $\mathbb{H}^2$, this will imply lemma for all $\kappa < 0$ by rescaling the metric on $\mathbb{H}^2$. We leave the cases $\kappa \ge 0$ to the reader as the proof is similar. The proof below is goes back to Euclid (in the case of $\mathbb{R}^2$). Let $c$ denote the largest of the numbers $a, b, c$. Draw a geodesic $\gamma \subset \mathbb{H}^2$ through points $x, y$ so that $d(x,y) = c$. Then

$$\gamma = \gamma_x \cup [x,y] \cap \gamma_y,$$

where $\gamma_x, \gamma_y$ are geodesic rays emanating from $x$ and $y$ respectively. Now, consider circles $S(x,b)$ and $S(y,a)$ centered at $x, y$ and having radii $b, a$ respectively. Since $c \ge \max(a,b)$,

$$\gamma_x \cap S(y,a) \subset \{x\}, \quad \gamma_y \cap S(x,b) \subset \{y\},$$

while

$$S(x,b) \cap [x,y] = p, \quad S(y,a) \cap [x,y] = y.$$

By the triangle inequality on $c \le a+b$, $p$ separates $q$ from $y$ (and $q$ separates $x$ from $p$). Therefore, both the ball $B(x,b)$ and its complement contain points of the circle $S(y,a)$, which (by connectivity) implies that $S(x,b) \cap S(y,a) \ne \emptyset$. Therefore, the triangle with the side-lengths $a, b, c$ exists. Uniqueness (up to congruence) of this triangle follows, for instance, from the hyperbolic cosine law. $\square$

CHAPTER 9

# Gromov-hyperbolic spaces and groups

The goal of this chapter is to define and review basic properties of $\delta$-hyperbolic spaces and word-hyperbolic groups, which are far-reaching generalizations of the real-hyperbolic space $\mathbb{H}^n$ and groups acting geometrically on $\mathbb{H}^n$. The advantage of $\delta$-hyperbolicity is that it can be defined in the context of arbitrary metric spaces which need not even be geodesic. These spaces were introduced in the seminal essay by Mikhail Gromov on hyperbolic groups, although ideas of *combinatorial* curvature and (in retrospect) hyperbolic properties of finitely-generated groups are much older. They go back to work of Max Dehn (on word problem in groups), Martin Grindlinger (small cancelation theory), Alexandr Ol'shanskii (who used what we now would call *relative hyperbolicity* in order to construct finitely-generated groups with exotic properties) and many others.

## 9.1. Hyperbolicity according to Rips

We begin our discussion of $\delta$-hyperbolic spaces with the notion of hyperbolicity in the context of geodesic metric spaces, which (according to Gromov) is due to Ilya (Eliyahu) Rips. This definitions will be then applied to Cayley graphs of groups, leading to the concept of a *hyperbolic group* discussed later in this chapter. Rips notion of hyperbolicity is based on the thinness properties of hyperbolic triangles which are established in section 8.9.

Let $(X, d)$ be a geodesic metric space. As in section 8.4, a geodesic triangle $T$ in $X$ is a concatenation of three geodesic segments $\tau_1, \tau_2, \tau_3$ connecting the points $A_1, A_2, A_3$ (vertices of $T$) in the natural cyclic order. Unlike the real-hyperbolic space, we no longer have uniqueness of geodesics, thus $T$ is not (in general) determined by its vertices. We define a measure of the thinness of $T$ similar to the one in Section 8.9 of Chapter 8.

Definition 9.1. The *thinness radius* of the geodesic triangle $T$ is the number

$$\delta(T) := \max_{j=1,2,3} \left( \sup_{p \in \tau_j} d(p, \tau_{j+1} \cup \tau_{j+2}) \right),$$

A triangle $T$ is called $\delta$*-thin* if $\delta(T) \leqslant \delta$.

Definition 9.2 (Rips' definition of hyperbolicity). A geodesic hyperbolic space $X$ is called $\delta$-hyperbolic (in the sense of Rips) if every geodesic triangle $T$ in $X$ is $\delta$-thin. A space $X$ which is $\delta$-hyperbolic for some $\delta < \infty$ is called *Rips–hyperbolic.* In what follows, we will refer to $\delta$–hyperbolic spaces in the sense of Rips simply as being $\delta$–hyperbolic.

Below are few simple but important geometric features of $\delta$-hyperbolic spaces.

First, not that general Rips–hyperbolic metric spaces $X$ are by no means uniquely geodesics. Nevertheless, next lemma shows that geodesics in $X$ between given pair of points are "almost unique":

LEMMA 9.3. *If $X$ is $\delta$–hyperbolic, then every pair of geodesics $[x,y],[x,z]$ with $d(y,z) \leqslant D$ are at Hausdorff distance at most $D+\delta$. In particular, if $\alpha,\beta$ are geodesic segments connecting points $x,y \in X$, then $\mathrm{dist}_{Haus}(\alpha,\beta) \leqslant \delta$.*

PROOF. Every point $p$ on $[x,y]$ is, either at distance at distance at most $\delta$ from $[x,z]$, or at distance at most $\delta$ from $[y,z]$; in the latter case $p$ is at distance at most $D+\delta$ from $[x,z]$. $\square$

The next lemma, the *fellow-traveling property of hyperbolic geodesics* sharpens the conclusion of Lemma 9.3.

LEMMA 9.4. *Let $\alpha(t),\beta(t)$ be geodesics in a $\delta$-hyperbolic space $X$, so that $\alpha(0)=\beta(0)=o$ and $d(\alpha(t_0),\beta(t_0)) \leqslant D$ for some $t_0 \geqslant 0$. Then for all $t\in[0,t_0]$,*

$$d(\alpha(t),\beta(t)) \leqslant 2(D+\delta).$$

PROOF. By previous lemma, for every $t \in [0,t_0]$ there exists $s\in[0,t_0]$ so that

$$d(\beta(t),\alpha(s)) \leqslant c=\delta+D.$$

By applying the triangle inequality, we see that

$$|t-s| \leqslant c,$$

hence, $d(\alpha(t),\beta(t)) \leqslant 2c = 2(\delta+D)$. $\square$

The notion of thin triangles generalizes naturally to the concept of thin polygons. A *geodesic $n$-gon* in a metric space $X$ is a concatenation of geodesic segments $\sigma_i, i=1,\ldots,n$, connecting points $P_i, i=1,\ldots,n$, in the natural cyclic order. A polygon $P$ is called $\eta$-thin if every side of $P$ is contained in the $\eta$-neighborhood of the union of the other sides.

EXERCISE 9.5. Suppose that $X$ is a $\delta$-hyperbolic metric space. Show that every $n$-gon in $X$ is $\delta(n-2)$-thin. Hint: Triangulate an $n$-gon $P$ by $n-3$ diagonals emanating from a single vertex . Now, use $\delta$-thinness of triangles in $X$ inductively.

We next improve the estimate provided by this exercise.

LEMMA 9.6 (thin polygons). *If $X$ is $\delta$–hyperbolic then every geodesic $n$-gon in $X$ is $\eta_n$-thin for*

$$\eta_n = 2\delta \log_2 n.$$

PROOF. We prove the estimate on thinness of $n$-gons by induction on $m$. For $n \leqslant 3$ the statement follows from $\delta$-thinness of bigons and triangles. Suppose $n \geqslant 4$ and the inequality holds for all $m \leqslant n-1$. Consider a geodesic $n$-gon $P$ which has edges $\tau_i = [A_i, A_{i+1}]$ and consider its edge $\tau=\tau_n$ of $P$. We will consider the case when $n$ is odd, $n=2k+1$, since the other case is similar. We subdivide $P$ in two $k+1$-gons $P', P''$ and one triangle $T$ by introducing the diagonals $[A_1, A_{k+1}]$ and $[A_{k+1}, A_n]$. By the induction hypothesis, $P', P''$ are $\eta_{k+1}$-thin, while the triangle $T$ is $\delta$-thin. Therefore, $\tau$ is within distance $\leqslant \eta_{k+1}+\delta$ from the union of the other sides of $P$. We leave it to the reader to check that

$$2\log_2(k+1)+1 \leqslant 2\log(n) = 2\log_2(2k+1). \quad \square$$

We now give some examples of Rips–hyperbolic metric spaces.

EXAMPLE 9.7. (1) Proposition 8.42 implies that $\mathbb{H}^n$ is $\delta$-hyperbolic for $\delta = \operatorname{arccos}(\sqrt{2})$.

(2) Suppose that $(X,d)$ is $\delta$-hyperbolic and $a > 0$. Then the metric space $(X, a\cdot d)$ is $a\delta$-hyperbolic. Indeed, distances in $(X, a\cdot d)$ are obtained from distances in $(X,d)$ by multiplication by $a$. Therefore, the same is true for distances between the edges of geodesic triangles.

(3) Let $X_\kappa$ is the model surface of curvature $\kappa < 0$ as in section 2.1.8. Then $X_\kappa$ is $\delta$-hyperbolic for

$$\delta_\kappa = |\kappa|^{-1/4}\operatorname{arccos}(\sqrt{2}).$$

Indeed, the Riemannian metric on $X_\kappa$ is obtained by multiplying the Riemannian metric on $\mathbb{H}^2$ by $|\kappa|^{-1/2}$. This has effect of multiplying all distances in $\mathbb{H}^2$ by $|\kappa|^{-1/4}$. Hence, if $d$ is the distance function on $\mathbb{H}^2$ then $|\kappa|^{-1/4}d$ is the distance function on $X_\kappa$.

(4) Suppose that $X$ is a $CAT(\kappa)$-space where $\kappa < 0$, see section 2.1.8. Then $X$ is $\delta_\kappa$-hyperbolic. Indeed, all triangles in $X$ are thinner then triangles in $X_\kappa$. Therefore, given a geodesic triangle $T$ with edges $\tau_i, i = 1,2,3$ and a points $P_1 \in \tau_1$ we take the comparison triangle $\tilde{T} \subset X_\kappa$ and the comparison point $\tilde{P}_1 \in \tilde{\tau}_1 \subset \tilde{T}$. Since $\tilde{T}$ is $\delta_\kappa$-thin, there exists a point $\tilde{P}_i \in \tilde{\tau}_i, i = 2$ or $i = 3$, so that $d(\tilde{P}_1, \tilde{P}_i) \leqslant \delta_\kappa$. Let $P_i \in \tau_i$ be the comparison point of $\tilde{P}_i$. Then, by the comparison inequality

$$d(P_1, P_i) \leqslant d(\tilde{P}_1, \tilde{P}_i) \leqslant \delta_\kappa.$$

Hence, $T$ is $\delta_\kappa$-thin. In particular, if $X$ is a simply-connected complete Riemannian manifold of sectional curvature $\leqslant \kappa < 0$, then $X$ is $\delta_\kappa$-hyperbolic.

(5) Let $X$ be a simplicial tree, and $d$ be a path-metric on $X$. Then, by the Exercise 2.36, $X$ is $CAT(-\infty)$. Thus, by (4), $X$ is $\delta_\kappa$-hyperbolic for every $\delta_\kappa = |\kappa|^{-1/4}\operatorname{arccos}(\sqrt{2})$. Since

$$\inf_\kappa \delta_\kappa = 0,$$

it follows that $X$ is 0-hyperbolic. Of course, this fact one can easily see directly by observing that every triangle in $X$ is a tripod.

(6) Every geodesic metric space of diameter $\leqslant \delta < \infty$ is $\delta$-hyperbolic.

EXERCISE 9.8. Let $X$ be the circle of radius $R$ in $\mathbb{R}^2$ with the induced path-metric $d$. Thus, $(X,d)$ has diameter $\pi R$. Show that $X$ is $\pi R/2$-hyperbolic and is not $\delta$-hyperbolic for any $\delta < \pi R/2$.

Not every geodesic metric space is hyperbolic:

EXAMPLE 9.9. For instance, let us verify that $\mathbb{R}^2$ is not $\delta$-hyperbolic for any $\delta$. Pick a nondegenerate triangle $T \subset \mathbb{R}^2$. Then $\delta(T) = k > 0$ for some $k$. Therefore, if we scale $T$ by a positive constant $c$, then $\delta(cT) = ck$. Sending $c \to \infty$, show that $\mathbb{R}^2$ is not $\delta$-hyperbolic for any $\delta > 0$. More generally, if a metric space $X$ contains an isometrically embedded copy of $\mathbb{R}^2$, then $X$ is not hyperbolic.

Here is an example of a metric space which is not hyperbolic, but does not contain a quasi-isometrically embedded copy of $\mathbb{R}^2$ either. Consider the wedge $X$ of countably many circles $C_i$ each given with path-metric of overall length $2\pi i$, $i \in \mathbb{N}$. We equip $X$ with the path-metric so that each $C_i$ is isometrically embedded. Exercise 9.8 shows that $X$ is not hyperbolic.

Exercise 9.10. Show that $X$ contains no quasi-isometrically embedded copy of $\mathbb{R}^2$. Hint: Use coarse topology.

More interesting examples of non-hyperbolic spaces containing no quasi–isometrically embedded copies of $\mathbb{R}^2$ are given by various solvable groups, e.g. the $Sol_3$ group and Cayley graph of the Baumslag–Solitar group $BS(n,1)$, see [**Bur99**].

Below we describe briefly another measure of thinness of triangles which can be used as an alternative definition of Rips–hyperbolicity. It is also related to the minimal size of the triangle, described in Definition 5.49, consequently it is related to the filling area of the triangle *via* a Besikovitch type inequality as described in Proposition 5.50.

Definition 9.11. For a geodesic triangle $T \subset X$ with the sides $\tau_1, \tau_2, \tau_3$, define the *inradius* of $T$ to be

$$\Delta(T) := \inf_{x \in X} \max_{i=1,2,3} d(x, \tau_i).$$

In the case of the real-hyperbolic plane, as we saw in Lemma 8.41, this definition coincides with the radius of the largest circle inscribed in $T$. Clearly, $\Delta(T) \leqslant \delta(T)$ and

$$\Delta(T) \leqslant \text{minsize}(T) \leqslant 2\Delta(T) + 1\,.$$

It turns out that

$$\text{minsize}(T) \leqslant 2\delta. \tag{9.1}$$

Indeed, let $\tau_1, \tau_2, \tau_3$ be the sides of $T$, we will assume that $\tau_1$ is parameterized so that

$$\tau_1(0) \in Im(\tau_3), \tau_1(a_1) = Im(\tau_2),$$

where $a_1$ is the length of $\tau_1$. Then by the intermediate value theorem, applied to the difference

$$d(\tau_1(t) - Im(\tau_2)) - d(\tau_1(t) - Im(\tau_3))$$

we conclude that there exists $t_1$ so that $d(\tau_1(t_1), Im(\tau_2)) = d(\tau_1(t_1), Im(\tau_3)) \leqslant \delta$. Taking $p_1 = \tau_1(t_1)$ and $p_i \in Im(\tau_i), i = 2, 3$, the points nearest to $p_1$, we get

$$d(p_1, p_2) \leqslant \delta, d(p_1, p_3) \leqslant \delta,$$

hence,

$$minsize(T) \leqslant 2\delta.$$

## 9.2. Geometry and topology of real trees

In this section we consider a special type of hyperbolic spaces, the *real trees*.

Definition 9.12. A 0–hyperbolic (geodesic) metric space is called a *real tree*.

Exercise 9.13. 1. Show that every real tree is a $CAT(0)$ space.
2. Show that every real tree is a $CAT(\kappa)$ space for every $\kappa$.

It follows from Exercise 9.5 that every polygon in a real tree is 0-thin.

Lemma 9.14. *If $X$ is a real tree then any two points in $X$ are connected by a unique topological arc in $X$.*

PROOF. Let $D = d(x, y)$. Consider a continuous injective map (i.e., a topological arc) $x = \alpha(0), y = \alpha(1)$. Let $\alpha^* = [x, y], \alpha^* : [0, D] \to X$ be the geodesic connecting $x$ to $y$. We claim that the image of $\alpha$ contains the image of $\alpha^*$. Indeed, we can approximate $\alpha$ by piecewise-geodesic (nonembedded!) arcs

$$\alpha_n = [x_0, x_1] \cup ... \cup [x_{n-1}, x_n], \quad x_0 = x, x_n = y.$$

Since the $n + 1$-gon $P$ in $X$,which is the concatenation of $\alpha_n$ with $[y, x]$ is 0-thin, $\alpha^* \subset \alpha_n$. Therefore, the image of $\alpha$ also contains the image of $\alpha^*$. Consider the continuous map $(\alpha^*)^{-1} \circ \alpha : [0, D] \to [0, D]$. Applying the intermediate value theorem to this function, we see that the images of $\alpha$ and $\alpha^*$ are equal. □

EXERCISE 9.15. Prove the converse to the above lemma.

DEFINITION 9.16. Let $T$ be a real tree and $p$ be a point in $T$. The *space of directions* at $p$, denoted $\Sigma_p$, is defined as the space of germs of geodesics in $T$ emanating from $p$, i.e., the quotient $\Sigma_p := \Re_p / \sim$, where

$$\Re_p = \{r : [0, a) \to T \mid a > 0, \ r \text{ isometry}, \ r(0) = p\}$$

and

$$r_1 \sim r_2 \iff \exists \, \varepsilon > 0 \text{ such that } r_1|_{[0,\varepsilon)} \equiv r_2|_{[0,\varepsilon)}.$$

Simplest examples of real trees are given by simplicial trees equipped with path-metrics. We will see, however, that other real trees also arise naturally in geometric group theory.

By Lemma 9.14, for every homeomorphism $c : [a, b] \to T$ the image $c([a, b])$ coincides with the geodesic segment $[c(a), c(b)]$. It follows that we may also define $\Sigma_p$ as the space of germs of topological arcs $\Im_p / \sim$, where

$$\Im_p = \{c : [0, a) \to T \mid a > 0, \ c \text{ homeomorphism}, \ c(0) = p\}$$

and

$$c_1 \sim c_2 \iff \exists \, \varepsilon_1 > 0, \, \varepsilon_2 > 0 \text{ such that } c_1([0, \varepsilon_1)) = c_2([0, \varepsilon_2)).$$

DEFINITION 9.17. Define *valence* $val(p)$ of a point $p$ in a real tree $T$ to be the cardinality of the set $\Sigma_p$. A *branch–point* of $T$ is a point $p$ of valence $\geqslant 3$. The *valence* of $T$ is the supremum of valences of points in $T$.

EXERCISE 9.18. Show that $val(p)$ equals the number of connected components of $T \setminus \{p\}$.

DEFINITION 9.19. A real tree $T$ is called *$\alpha$–universal* if every real tree with valence at most $\alpha$ can be isometrically embedded into $T$.

See **[MNLGO92]** for a study of universal trees. In particular, the following holds:

THEOREM 9.20 ([**MNLGO92**]). *For every cardinal number $\alpha > 2$ there exists an $\alpha$–universal tree, and it is unique up to isometry.*

**Fixed-point properties.**

Part 1 of Exercise 9.13 together with Corollary 2.43 implies:

COROLLARY 9.21. *If $G$ is a finite group acting isometrically on a complete real tree $T$, then $G$ fixes a point in $T$.*

DEFINITION 9.22. A group $G$ is said to have *Property FA* if for every isometric action $G \curvearrowright T$ on a complete real tree $T$, $G$ fixes a point in $T$.

Thus, all finite groups have property FA.

### 9.3. Gromov hyperbolicity

One drawback of the Rips definition of hyperbolicity is that it uses geodesics. Below is an alternative definition of hyperbolicity, due to Gromov, where one needs to verify certain inequalities only for quadruples of points in a metric space (which need not be geodesic). Gromov's definition is less intuitive than the one of Rips, but, as we will see, it is more suitable in certain situations.

Let $(X, \mathrm{dist})$ be a metric space (*which is no longer required to be geodesic*). Pick a base-point $p \in X$. For each $x \in X$ set $|x|_p := \mathrm{dist}(x, p)$ and define the *Gromov product*

$$(x, y)_p := \frac{1}{2}\left(|x|_p + |y|_p - \mathrm{dist}(x, y)\right).$$

Note that the triangle inequality immediately implies that $(x, y)_p \geqslant 0$ for all $x, y, p$; the Gromov product measures how far the triangle inequality for the points $x, y, p$ is from being an equality.

REMARK 9.23. The Gromov product is a generalization of the inner product in vector spaces with $p$ serving as the origin. For instance, suppose that $X = \mathbb{R}^n$ with the usual inner product, $p = 0$ and $|v|_p := \|v\|$ for $v \in \mathbb{R}^n$. Then

$$\frac{1}{2}\left(|x|_p^2 + |y|_p^2 - \|x - y\|^2\right) = x \cdot y.$$

EXERCISE 9.24. Suppose that $X$ is a metric tree. Then $(x, y)_p$ is the distance $\mathrm{dist}(p, \gamma)$ from $p$ to the geodesic segment $\gamma = [xy]$.

In general a direct calculation shows that for each point $z \in X$

$$(p, x)_z + (p, y)_z \leqslant |z|_p - (x, y)_p$$

with equality

$$(p, x)_z + (p, y)_z = |z|_p - (x, y)_p. \tag{9.2}$$

if and only $d(x, z) + d(z, y) = d(x, y)$. Thus, for every $z \in \gamma = [x, y]$,

$$(x, y)_p = d(z, p) - (p, x)_z - (p, y)_z \leqslant d(z, p).$$

In particular, $(x, y)_p \leqslant \mathrm{dist}(p, \gamma)$.

LEMMA 9.25. *Suppose that $X$ is $\delta$–hyperbolic in the sense of Rips. Then the Gromov product in $X$ is "comparable" to $\mathrm{dist}(p, \gamma)$: For every $x, y, p \in X$ and geodesic $\gamma = [x, y]$,*

$$(x, y)_p \leqslant \mathrm{dist}(p, \gamma) \leqslant (x, y)_p + 2\delta.$$

PROOF. The inequality $(x, y)_p \leqslant \mathrm{dist}(p, \gamma)$ was proved above; so we have to establish the other inequality. Note that since the triangle $\Delta(pxy)$ is $\delta$–thin, for each point $z \in \gamma = [x, y]$ we have

$$\min\{(x, p)_z, (y, p)_z\} \leqslant \min\{\mathrm{dist}(z, [p, x]), \mathrm{dist}(z, [p, y])\} \leqslant \delta.$$

By continuity of the distance function, there exists a point $z \in \gamma$ such that $(x, p)_z, (y, p)_z \leqslant \delta$. By applying the equality (9.2) we get:

$$|z|_p - (x, y)_p = (p, x)_z + (p, y)_z \leqslant 2\delta.$$

Since $|z|_p \leqslant \text{dist}(p, \gamma)$, we conclude that $\text{dist}(p, \gamma) \leqslant (x, y)_p + 2\delta$. □

Now, for a metric space $X$ define a number $\delta_p = \delta_p(X) \in [0, \infty]$ as follows:

$$\delta_p := \sup\{\min((x, z)_p, (y, z)_p) - (x, y)_p\}$$

where the supremum is taken over all triples of points $x, y, z \in X$.

EXERCISE 9.26. If $\delta_p \leqslant \delta$ then $\delta_q \leqslant 2\delta$ for all $q \in X$.

DEFINITION 9.27. A metric space $X$ is said to be *$\delta$–hyperbolic in the sense of Gromov*, if $\delta_p \leqslant \delta < \infty$ for all $p \in X$. In other words, for every quadruple $x, y, z, p \in X$, we have

$$(x, y)_p \geqslant \min((x, z)_p, (y, z)_p) - \delta.$$

EXERCISE 9.28. The real line with the usual metric is 0-hyperbolic in the sense of Gromov.

EXERCISE 9.29. Gromov–hyperbolicity is invariant under $(1, A)$-quasi-isometries.

EXERCISE 9.30. Let $X$ be a metric space and $N \subset X$ be an $R$-net. Show that the embedding $N \hookrightarrow X$ is an $(1, R)$-quasi-isometry. In particular, $X$ is Gromov–hyperbolic if and only if $N$ is Gromov–hyperbolic. In particular, a group $(G, d_S)$ with word metric $d_S$ is Gromov–hyperbolic if and only if the Cayley graph $\Gamma_{G,S}$ of $G$ is Rips–hyperbolic.

LEMMA 9.31. *Suppose that $X$ is $\delta$–hyperbolic in the sense of Rips. Then it is $3\delta$–hyperbolic in the sense of Gromov. In particular, a geodesic metric space is a real tree if and only if it is 0-hyperbolic in the sense of Gromov.*

PROOF. Consider points $x, y, z, p \in X$ and the geodesic triangle $T(xyz) \subset X$ with vertices $x, y, z$. Let $m \in [x, y]$ be the point nearest to $p$. Then, since the triangle $T(x, y, z)$ is $\delta$–thin, there exists a point $n \in [x, z] \cup [y, z]$ so that $\text{dist}(n, m) \leq \delta$. Assume that $n \in [y, z]$. Then, by Lemma 9.25,

$$(y, z)_p \leqslant \text{dist}(p, [y, z]) \leqslant \text{dist}(p, [x, y]) + \delta.$$

On the other hand, by Lemma 9.25,

$$\text{dist}(p, [x, y]) \leqslant (x, y)_p - 2\delta.$$

By combining these two inequalities, we obtain

$$(y, z)_p \leqslant (x, y)_p - 3\delta.$$

Therefore, $(x, y)_p \geqslant \min\left((x, z)_p, (y, z)_p\right) - 3\delta$. □

We now prove the "converse" to the above lemma:

LEMMA 9.32. *Suppose that $X$ is a geodesic metric space which is $\delta$–hyperbolic in the sense Gromov, then $X$ is $2\delta$–hyperbolic in the sense of Rips.*

PROOF. 1. We first show that in such space geodesics connecting any pair of points are "almost" unique, i.e., if $\alpha$ is a geodesic connecting $x$ to $y$ and $p$ is a point in $X$ such that

$$\text{dist}(x, p) + \text{dist}(p, y) \leqslant \text{dist}(x, y) + 2\delta$$

then $\text{dist}(p, \alpha) \leqslant 2\delta$. We suppose that $\text{dist}(p, x) \leqslant \text{dist}(p, y)$. If $\text{dist}(p, x) \geqslant \text{dist}(x, y)$ then $\text{dist}(x, y) \leqslant 2\delta$ and thus $\min(\text{dist}(p, x), p(y)) \leqslant 2\delta$ and we are done.

Therefore, assume that $\text{dist}(p,x) < \text{dist}(x,y)$ and let $z \in \alpha$ be such that $\text{dist}(z,y) = \text{dist}(p,y)$. Since $X$ is $\delta$–hyperbolic in the sense Gromov,

$$(x,y)_p \geqslant \min((x,z)_p,(y,z)_p) - \delta.$$

Thus we can assume that $(x,y)_p \geqslant (x,z)_p$. Then

$$\text{dist}(y,p) - \text{dist}(x,y) \geqslant \text{dist}(z,p) - \text{dist}(x,z) - 2\delta \iff$$
$$\text{dist}(z,p) \leqslant 2\delta.$$

Thus $\text{dist}(p,\alpha) \leqslant 2\delta$.

2. Consider now a geodesic triangle $[x,y,p] \subset X$ and let $z \in [x,y]$. Our goal is to show that $z$ belongs to $\mathcal{N}_{4\delta}([p,x] \cup [p,y])$. We have:

$$(x,y)_p \geqslant \min((x,z)_p,(y,z)_p) - \delta.$$

Assume that $(x,y)_p \geqslant (x,z)_p - \delta$. Set $\alpha := [p,y]$. We will show that $z \in \mathcal{N}_{2\delta}(\alpha)$.

By combining $\text{dist}(x,z) + \text{dist}(y,z) = \text{dist}(x,y)$ and $(x,y)_p \geqslant (x,z)_p - \delta$, we obtain

$$\text{dist}(y,p) \geqslant \text{dist}(y,z) + \text{dist}(z,p) - 2\delta.$$

Therefore, by Part 1, $z \in \mathcal{N}_{2\delta}(\alpha)$ and hence the triangle $T(x,y,z)$ is $2\delta$–thin. □

Corollary 9.33 (M. Gromov, **[Gro87]**, section 6.3C.). *For geodesic metric spaces, Gromov–hyperbolicity is equivalent to Rips–hyperbolicity.*

The drawback is that in this generality, Gromov–hyperbolicity fails to be QI invariant:

Example 9.34 (Gromov–hyperbolicity is not QI invariant ). This example is taken from **[Väi05]**. Consider the graph $X$ of the function $y = |x|$, where the metric on $X$ is the restriction of the metric on $\mathbb{R}^2$. (This is not a path-metric!) Then the map $f : \mathbb{R} \to X, f(x) = (x, |x|)$ is a quasi-isometry:

$$|x - x'| \leqslant d(f(x), f(x')) \leqslant \sqrt{2}|x - x'|.$$

Let $p = (0,0)$ be the base-point in $X$ and for $t > 0$ we let $x := (2t, 2t)$, $y := (-2t, 2t)$ and $z := (t,t)$. The reader will verify that

$$\min((x,z)_p,(y,z)_p) - (x,y)_p) = t\left(\frac{7\sqrt{2}}{2} - 3\right) > t.$$

Therefore, the quantity $\min((x,z)_p,(y,z)_p) - (x,y)_p)$ is not bounded from above as $t \to \infty$ and hence $X$ is not $\delta$-hyperbolic for any $\delta < \infty$. Thus $X$ is QI to a Gromov–hyperbolic space $\mathbb{R}$, but is not Gromov–hyperbolic itself. We will see, as a corollary of Morse Lemma (Corollary 9.39), that in the context of geodesic spaces, Gromov–hyperbolicity is a QI invariant.

## 9.4. Ultralimits and stability of geodesics in Rips–hyperbolic spaces

In this section we will see that every hyperbolic geodesic metric spaces $X$ *globally* resembles a tree. This property will be used to prove *Morse Lemma*, which establishes that quasi-geodesics in $\delta$-hyperbolic spaces are uniformly close to geodesics.

Lemma 9.35. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of geodesic $\delta_i$–hyperbolic spaces with $\delta_i$ tending to 0. Then for every non-principal ultrafilter $\omega$ each component of the ultralimit $X_\omega$ is a metric tree.*

PROOF. First, according to Lemma 7.49, ultralimit of geodesic metric spaces is again a geodesic metric space. Thus, in view of Lemma 9.32, it suffices to verify that $X_\omega$ is 0-hyperbolic in the sense of Gromov (since it will be 0-hyperbolic in the sense of Rips and, hence, a metric tree). This is one of the few cases where Gromov–hyperbolicity is superior to Rips–hyperbolicity: It suffices to check hyperbolicity condition only for quadruples of points.

We know that for every quadruple $x_i, y_i, z_i, p_i$ in $X_i$,

$$(x_i, y_i)_{p_i} \geqslant \min((x_i, z_i)_{p_i}, (y_i, z_i)_{p_i}) - \delta_i.$$

By taking $\omega$-lim of this inequality, we obtain (for every quadruple of points $x_\omega, y_\omega, z_\omega, p_\omega$ in $X_\omega$):

$$(x_\omega, y_\omega)_{p_\omega} \geqslant \min((x_\omega, z_\omega)_{p_\omega}, (y_\omega, z_\omega)_{p_\omega}),$$

since $\omega\text{-}\lim \delta_i = 0$. Thus, $X_\omega$ is 0-hyperbolic. □

EXERCISE 9.36. Find a flaw in the following "proof" of this lemma: Since $X_i$ is $\delta_i$-hyperbolic, it follows that every geodesic triangle $T_i$ in $X_i$ is $\delta_i$-thin. Suppose that $\omega\text{-}\lim d(x_i, e_i) < \infty$, $\omega\text{-}\lim d(p_i, e_i) < \infty$. Taking limit in the definition of thinness of triangles, we conclude that the ultralimit of triangles $T_\omega = \omega\text{-}\lim T_i \subset X_\pm$ is 0-thin. Therefore, every geodesic triangle in $X_\omega$ is 0-thin.

COROLLARY 9.37. *Every geodesic in the tree $X_\omega$ is a limit geodesic.*

The following fundamental theorem in the theory of hyperbolic spaces is called *Morse Lemma* or *stability of hyperbolic geodesics*.

THEOREM 9.38 (Morse Lemma). *There exists a function $\theta = \theta(L, A, \delta)$, so that the following holds. If $X$ be a $\delta$–hyperbolic geodesic space, then for every $(L, A)$–quasigeodesic $f : [a, b] \to X$ the Hausdorff distance between the image of $f$ and the geodesic segment $[f(a), f(b)] \subset X$ is at most $\theta$.*

PROOF. Set $c = d(f(a), f(b))$. Given quasi-geodesic $f$ and geodesic $f^* : [0, c] \to X$ parameterizing $[f(a), f(b)]$, we define two numbers:

$$D_f = \sup_{t \in [a,b]} d(f(t), Im(f^*))$$

and

$$D_f^* = \sup_{t \in [0,c]} d(f^*(t), Im(f)).$$

Then $dist_{Haus}(Im(f), Im(f^*))$ is $\max(D_f, D_f^*)$. We will prove that $D_f$ is uniformly bounded in terms of $L, A, \delta$, since the proof for $D_f^*$ is completely analogous.

Suppose that the quantities $D_f$ are not uniformly bounded, that is, exists a sequence of $(L, A)$–quasigeodesics $f_n : [-n, n] \to X_n$ in $\delta$-hyperbolic geodesic metric spaces $X_n$, such that

$$\lim_{n \to \infty} D_n = \infty.$$

where $D_n = D_{f_n}$. Pick points $t_n \in [-n, n]$ such that

$$|\text{dist}(f_n(t_n), [f(-n), f(n)]) - D_n| \leqslant 1.$$

As in the definition of asymptotic cones, consider two sequences of pointed metric spaces

$$\left(\frac{1}{D_n} X_n, f_n(t_n)\right), \quad \left(\frac{1}{D_n}[-n, n], t_n\right).$$

Note that $\omega$-lim $\frac{n}{D_n}$ could be infinite. Let

$$(X_\omega, x_\omega) = \omega\text{-lim}\left(\frac{1}{D_n} X_n, f_n(t_n)\right)$$

and

$$(Y, y) := \omega\text{-lim}\left(\frac{1}{D_n}[-n, n], t_n\right).$$

The metric space $Y$ is either a nondegenerate segment in $\mathbb{R}$ or a closed geodesic ray in $\mathbb{R}$ or the whole real line. Note that the distance from points $Im(f_n)$ to $Im(f_n^*)$ in the rescaled metric space $\frac{1}{D_n} X_n$ is at most $1 + 1/d_n$. Each map

$$f_n : Y_n \to \frac{1}{d_n} X_n$$

is an $(L, A/D_n)$–quasi-geodesic. Therefore the ultralimit

$$f_\omega = \omega\text{-lim}\, f_n : (Y, y) \to (X_\omega, x_\omega)$$

is an $(L, 0)$–quasi-isometric embedding, i.e. it is a $L$-bi-Lipschitz map. In particular this map is a continuous embedding. Therefore, the image of $f_\omega$ is a geodesic $\gamma$ in $X_\omega$, see Lemma 9.14.

On the other hand, the sequence of geodesic segments $[f_n(-n), f_n(n)] \subset \frac{1}{d_n} X_n$ also $\omega$–converges to a geodesic $\gamma^* \subset X_\omega$, this geodesic is either a finite geodesic segment or a geodesic ray or a complete geodesic. In any case, by our choice of the points $x_n$, $\gamma$ is contained in 1-neighborhood of the geodesic $\gamma^*$ and, at the same time, $\gamma \neq \gamma^*$ since $x_\omega \in \gamma \setminus \gamma^*$. This contradicts the fact that $X_\omega$ is a real tree. □

**Historical Remark.** Morse [**Mor24**] proved a special case of this theorem in the case of $\mathbb{H}^2$ where the quasi-geodesics in question where geodesics in another Riemannian metric on $\mathbb{H}^2$, which admits a cocompact group of isometries. Busemann, [**Bus65**], proved a version of this lemma in the case of $\mathbb{H}^n$, where metrics in question were not necessarily Riemannian. A version in terms of quasi-geodesics is due to Mostow [**Mos73**], in the context of negatively curved symmetric spaces, although his proof is general.

Corollary 9.39 (QI invariance of hyperbolicity). *Suppose that $X, X'$ are quasi–isometric geodesic metric spaces and $X'$ is hyperbolic. Then $X$ is also hyperbolic.*

Proof. Suppose that $X'$ is $\delta'$-hyperbolic and $f : X \to X'$ is an $(L, A)$–quasi-isometry and $f' : X' \to X$ is its quasi-inverse. Pick a geodesic triangle $T \subset X$. Its image under $f$ is a quasi-geodesic triangle $S$ in $X'$ whose sides are $(L, A)$–quasi-geodesic. Therefore each of the quasi-geodesic sides $\sigma_i$ of $S$ is within distance $\leqslant \theta = \theta(L, A, \delta')$ from a geodesic $\sigma_i^*$ connecting the end-points of this side. See Figure 9.1. The geodesic triangle $S^*$ formed by the segments $\sigma_1^*, \sigma_2^*, \sigma_3^*$ is $\delta'$-thin. Therefore, the quasi-geodesic triangle $f'(S^*) \subset X$ is $\epsilon := L\delta' + A$–thin, i.e. each quasi-geodesic $\tau_i' := f'(\sigma_i^*)$ is within distance $\leqslant \epsilon$ from the union $\tau_{i-1}', \tau_{i+1}'$. However,

$$dist_{Haus}(\tau_i, \tau_i') \leqslant L\theta + 2A.$$

Putting this all together, we conclude that the triangle $T$ is $\delta$-thin with

$$\delta = 2(L\theta + 2A) + \epsilon = 2(L\theta + 2A) + L\delta' + A. \quad \square$$

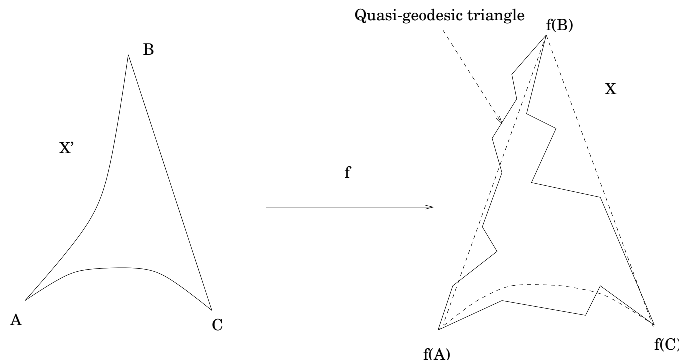



FIGURE 9.1. *I*mage of a geodesic triangle.

Note that in Morse Lemma, we are not claiming, of course, that the distance $d(f(t), f^*(t))$ is uniformly bounded, only that for every $t$ there exist $s$ and $s^*$ so that

$$d(f(t), f^*(s)) \leqslant \theta,$$

and

$$d(f^*(t), f(s^*)) \leqslant \theta.$$

Here $s = s(t), s^* = s^*(t)$. However, applying triangle inequalities one gets for $B = A + \theta$ the following estimates:

$$L^{-1}t - B \leqslant s \leqslant Lt + B \tag{9.3}$$

and

$$L^{-1}(t - B) \leqslant s^* \leqslant L(t + B) \tag{9.4}$$

### 9.5. Quasi-convexity in hyperbolic spaces

The usual notion of convexity does not make much sense in the context of hyperbolic geodesic metric spaces. For instance, there is an example of a geodesic Gromov–hyperbolic metric space $X$ where the convex hull of a finite subset is the entire $X$. The notion of convex hull is then replaces with

DEFINITION 9.40. Let $X$ be a geodesic metric space and $Y \subset X$. Then the *quasiconvex hull* $H(Y)$ of $Y$ in $X$ is the union of all geodesics $[y_1, y_2] \subset X$, where $y_1, y_2 \in Y$.

Accordingly, a subset $Y \subset X$ is *R-quasiconvex* if $H(Y) \subset \mathcal{N}_R(Y)$. A subset $Y$ is called *quasiconvex* if it is quasiconvex for some $R < \infty$.

EXAMPLE 9.41. Let $X$ be a $\delta$-hyperbolic geodesic metric space. Then thin triangle property immediately implies:

1. Every metric ball $B(x, R)$ in is $\delta$-quasiconvex.
2. let $Y_i \subset X$ be $R_i$-quasiconvex, $i = 1, 2$, and $Y_1 \cap Y_2 \neq \emptyset$. Then $Y_1 \cup Y_2$ is $R_1 + R_2 + \delta$-quasiconvex.
3. Intersection of any family of $R$-quasiconvex sets is again $R$-quasiconvex.

An example of a non-quasiconvex subset is a horosphere in $\mathbb{H}^n$: Its quasiconvex hull is the horoball bounded by this horosphere.

The construction of quasiconvex hull could be iterated and, by applying the fact that quadrilaterals in $X$ are $2\delta$-thin, we obtain:

LEMMA 9.42. *Let $Y \subset X$ be a subset. Then $H(Y)$ is $2\delta$-quasiconvex in $X$.*

The following results connect quasiconvexity and quasi-isometry for subsets of Gromov–hyperbolic geodesic metric spaces.

THEOREM 9.43. *Let $X, Y$ be geodesic metric spaces, so that $X$ is $\delta$-hyperbolic geodesic metric space. Then for every quasi-isometric embedding $f : Y \to X$, the image $f(Y)$ is quasiconvex in $X$.*

PROOF. Let $y_1, y_2 \in Y$ and $\alpha = [y_1, y_2] \subset Y$ be a geodesic connecting $y_1$ to $y_2$. Since $f$ is an $(L, A)$ quasi-isometric embedding, $\beta = f(\alpha)$ is an $(L, A)$ quasi-geodesic in $X$. By Morse Lemma,

$$dist_{Haus}(\beta, \beta^*) \leqslant R = \theta(L, A, \delta),$$

where $\beta^*$ is any geodesic in $X$ connecting $x_1 = f(y_1)$ to $x_2 = f(y_2)$. Therefore, $\beta^* \subset \mathcal{N}_R(f(Y)$, and $f(Y)$ is $R$-quasi-convex. □

The map $f : Y \to f(Y)$ is a quasi-isometry, where we use the restriction of the metric from $X$ to define a metric on $f(Y)$. Of course, $f(Y)$ is not a geodesic metric space, but it is quasi-convex, so applying the same arguments as in the proof of Theorem 9.39, we conclude that $Y$ is also hyperbolic.

Conversely, let $Y \subset X$ be a coarsely connected subset, i.e., there exists a constant $c < \infty$ so that the complex $Rips_C(Y)$ is connected for all $C \geqslant c$, where we again use the restriction of the metric $d$ from $X$ to $Y$ to define the Rips complex. Then we define a path-metric $d_{Y,C}$ on $Y$ by looking at infima of lengths of paths in $Rips_C(Y)$ connecting points of $Y$. The following is a converse to Theorem 9.43:

THEOREM 9.44. *Suppose that $Y \subset X$ is coarsely connected and $Y$ is quasi-convex in $X$. Then the identity map $f : (Y, d_{Y,C}) \to (X, \mathrm{dist}_X)$ is a quasi-isometric embedding for all $C \geqslant 2c + 1$.*

PROOF. Let $C$ be such that $H(Y) \subset \mathcal{N}_C(Y)$. First, if $d_Y(y, y') \leqslant C$ then $dist_X(y, y') \leqslant C$ as well. Hence, $f$ is coarsely Lipschitz. Let $y, y' \in Y$ and $\gamma$ is a geodesic in $X$ of length $L$ connecting $y, y'$. Subdivide $\gamma$ in $n = [L]$ subintervals of unit intervals and an interval of the length $L - n$:

$$[z_0, z_1], ..., [z_{n-1}, z_n], [z_n, z_{n+1}],$$

where $z_0 = y, z_{n+1} = y'$. Since each $z_i$ belongs to $\mathcal{N}_c(Y)$, there exist points $y_i \in Y$ so that $dist_X(y_i, z_i) \leqslant c$, where we take $y_0 = z_0, y_{n+1} = z_{n+1}$. Then

$$dist_X(z_i, z_{i+1}) \leqslant 2c + 1 \leqslant C$$

and, hence, $z_i, z_{i+1}$ are connected by an edge (of length $C$) in $Rips_C(Y)$. Now it is clear that

$$d_{Y,C}(y, y') \leqslant C(n + 1) \leqslant C dist_X(y, y') + C. \quad \square$$

REMARK 9.45. It is proven in [**Bow94**] that in the context of subsets of negatively pinched complete simply-connected Riemannian manifolds $X$, quasi-convex hulls $Hull(Y)$ are essentially the same as convex hulls:

There exists a function $L = L(C)$ so that for every $C$-quasiconvex subset $Y \subset X$,

$$H(Y) \subset Hull(Y) \subset \mathcal{N}_{L(C)}(Y).$$

## 9.6. Nearest-point projections

In general, nearest-point projections to geodesics in $\delta$-hyperbolic geodesic spaces are not well defined. The following lemma shows, nevertheless, that they are *coarsely-well defined*:

Let $\gamma$ be a geodesic in $\delta$-hyperbolic geodesic space $X$. For a point $x \in X$ let $p = \pi_\gamma(x)$ be a point nearest to $x$.

LEMMA 9.46. *Let $p' \in \gamma$ be such that $d(x, p') < d(x, p) + R$. Then*

$$d(p, p') \leqslant 2(R + 2\delta).$$

*In particular, if $p, p' \in \gamma$ are both nearest to $x$ then*

$$d(p, p') \leqslant 4\delta.$$

PROOF. Consider the geodesics $\alpha, \alpha'$ connecting $x$ to $p$ and $p'$ respectively. Let $q' \in \alpha'$ be the point within distance $\delta + R$ from $p'$ (this point exists unless $d(x, p) < \delta + R$ in which case $d(p, p') \leqslant 2(\delta + R)$ by the triangle inequality). Since the triangle $\Delta(x, p, p')$ is $\delta$-thin, there exists a point $q \in [xp] \cup [pp'] \subset [xp] \cup \gamma$ within distance $\delta$ from $q$. If $q \in \gamma$, we obtain a contradiction with the fact that the point $p$ is nearest to $x$ on $\gamma$ (the point $q$ will be closer). Thus, $q \in [xp]$. By the triangle inequality

$$d(x, p') - (R + \delta) = d(x, q') \leqslant d(x, q) + \delta \leqslant d(x, p) - d(q, p) + \delta.$$

Thus,

$$d(q, p) \leqslant d(x, p) - d(x, p') + R + 2\delta \leqslant R + 2\delta.$$

Since $d(p', q) \leqslant R + 2\delta$, we obtain $d(p', p) \leqslant 2(R + 2\delta)$. □

This lemma can be strengthened, we now show that the nearest-point projection to a quasi-geodesic subspace in a hyperbolic space is coarse Lipschitz:

LEMMA 9.47. *Let $X' \subset X$ be an $R$-quasiconvex subset. Then the nearest-point projection $\pi = \pi_{X'} : X \to X'$ is $(2, 2R + 9\delta)$-coarse Lipschitz.*

PROOF. Suppose that $x, y \in X$ so that $d(x, y) = D$. Let $x' = \pi(x), y' = \pi(y)$. Consider the quadrilateral formed by geodesic segments $[x, y] \cup [y, y'], [y', x'] \cup [x', x]$. Since this quadrilateral is $2\delta$-thin, there exists a point $q \in [x', y']$ which is within distance $\leqslant 2\delta$ from $[x', x] \cup [xy]$ and $[x, y] \cup [y, y]$.

**Case 1.** We first assume that there are points $x'' \in [x, x'], y'' \in [y, y]$ so that

$$d(q, x'') \leqslant 2\delta, d(q, y'') \leqslant 2\delta.$$

Let $q' \in X'$ be a point within distance $\leqslant R$ from $q$. By considering the paths

$$[x, x''] \cup [x'', q] \cup [q, q'], \quad [y, y''] \cup [y'', q] \cup [q, q']$$

and using the fact that $x' = \pi(x), y' = \pi(y)$, we conclude that

$$d(x', x'') \leqslant R + 2\delta, \quad d(y', y'') \leqslant R + 2\delta.$$

Therefore,

$$d(x', y') \leqslant 2R + 9\delta.$$

**Case 2.** Suppose that there exists a point $q'' \in [x, y]$ so that $d(q, q'') \leqslant 2\delta$. Setting $D_1 = d(x, q''), D_2 = d(y, q'')$, we obtain

$$d(x, x') \leqslant d(x, q') \leqslant D_1 + R + 2\delta, d(y, y') \leqslant d(y, q') \leqslant D_2 + R + 2\delta$$

which implies that

$$d(x', y') \leqslant 2D + 2R + 4\delta.$$

In either case, $d(x'.y') \leqslant 2d(x, y) + 2R + 9\delta$. $\square$

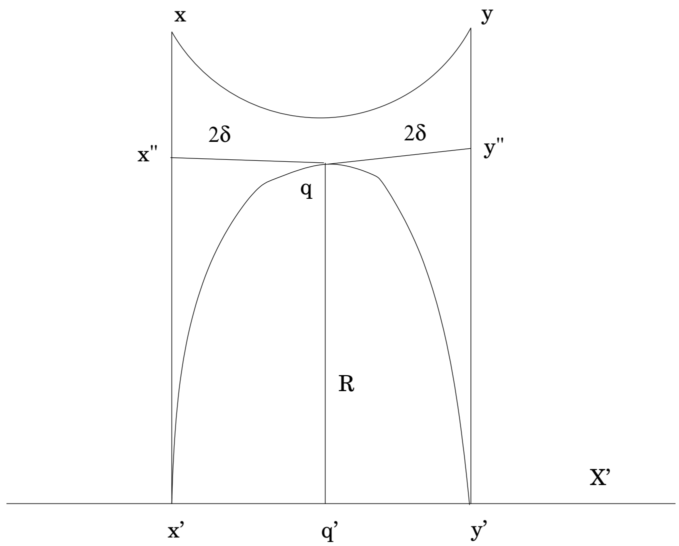



FIGURE 9.2. *P*rojection to a quasiconvex subset.

### 9.7. Geometry of triangles in Rips–hyperbolic spaces

In the case of real-hyperbolic space we relied upon hyperbolic trigonometry in order to study geodesic triangles. Trigonometry no longer makes sense in the context of Rips–hyperbolic spaces $X$, so instead one compares geodesic triangles in $X$ to geodesic triangles in real trees, i.e., to tripods, in the manner similar to the comparison theorems for $CAT(\kappa)$-spaces. In this section we describe comparison maps to tripods, called *collapsing maps*. We will see that such maps are $(1, 14\delta)$-quasi-isometries. We will use the collapsing maps in order to get a detailed information about geometry of triangles in $X$.

A *tripod* $\tilde{T}$ is a metric graph which is the union of three Euclidean line segments (called *legs* of the tripod) joined at a common vertex $o$, called the *centroid* of $\tilde{T}$. By abusing the notation, we will regard a tripod $\tilde{T}$ as a geodesic triangle whose vertices are the extreme points (leaves) $\tilde{x}_i$ of $\tilde{T}$; hence, we will use the notation $\mathcal{T} = \tilde{T} = T(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$.

REMARK 9.48. Using the symbol $\sim$ in the notation for a tripod is motivated by the comparison geometry, as we will compare geodesic triangles in $\delta$-hyperbolic spaces with the tripods $\tilde{T}$: This is analogous to comparing geodesic triangles in metric spaces to geodesic triangles in constant curvature spaces, see Definition 2.33.

EXERCISE 9.49. Given three numbers $a_i \in \mathbb{R}_+, i = 1, 2, 3$ satisfying the triangle inequalities $a_i \leqslant a_j + a_k$ ($\{1, 2, 3\} = \{i, j, k\}$), there exists a unique (up to isometry) tripod $\tilde{T} = \mathcal{T}_{a_1,a_2,a_3}$ with the side-lengths $a_1, a_2, a_3$.

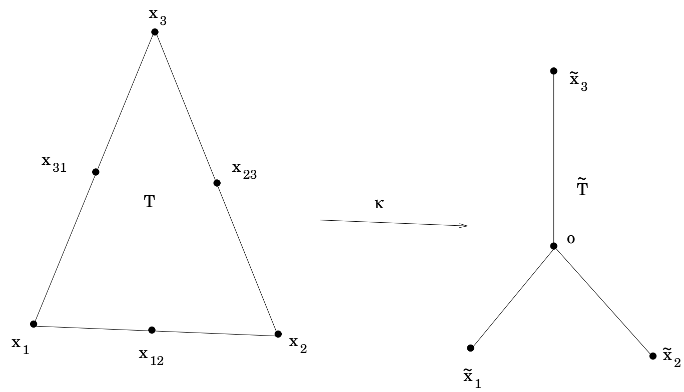



FIGURE 9.3. *C*ollapsing map of triangle to a tripod.

Now, given a geodesic triangle $T = T(x_1, x_2, x_3)$ with side-lengths $a_i, i = 1, 2, 3$ in a metric space $X$, there exists a unique (possibly up to postcomposition with an isometry $\tilde{T} \to \tilde{T}$) map $\kappa$ to the "comparison" tripod $\tilde{T}$,

$$\kappa : T \to \tilde{T} = \mathcal{T}_{a_1,a_2,a_3}$$

which is isometric on every edge of $T$: The map $\kappa$ sends the vertices $x_i$ of $T$ to the leaves $\tilde{x}_i$ of the tripod $\tilde{T}$. The map $\kappa$ is called the *collapsing map* for $T$. We say that points $x, y \in T$ are *dual* to each other if $\kappa(x) = \kappa(y)$.

EXERCISE 9.50. 1. The collapsing map $\kappa$ preserves the Gromov-products $(x_i, x_j)_{x_k}$.

2. $\kappa$ is 1-Lipschitz.

Then,

$$(x_i, x_j)_{x_k} = d(\tilde{x}_k, [\tilde{x}_i, \tilde{x}_j]) = d(\tilde{x}_k, o).$$

By taking the preimage of $o \in \tilde{T}$ under the maps $\kappa|[x_i, x_j]$ we obtain points

$$x_{ij} \in [x_i, x_j]$$

called the *central points* of the triangle $T$:

$$d(x_i, x_{ij}) = (x_j, x_k)_{x_i}.$$

LEMMA 9.51 (Approximation of triangles by tripods). *Assume that a geodesic metric space $X$ is $\delta$–hyperbolic in the sense of Rips, and consider an arbitrary geodesic triangle $T = \Delta(x_1, x_2, x_3)$ with the central points $x_{ij} \in [x_i, x_j]$. Then for every $\{i, j, k\} = \{1, 2, 3\}$ we have:*

*1. $d(x_{ij}, x_{jk}) \leqslant 6\delta$.*

*2. $d_{Haus}([x_j, x_{ji}], [x_j, x_{kj}]) \leqslant 7\delta$.*

*3. Distances between dual points in $T$ are $\leqslant 14\delta$. In detail: Suppose that $\alpha_{ji}, \alpha_{jk} : [0, t_j] \to X$ ($t_j = d(x_j, x_{ij}) = d(x_j, x_{jk})$) are unit speed parameterizations of geodesic segments $[x_j, x_{ji}], [x_j, x_{jk}]$. Then*

$$d(\alpha_{ji}(t), \alpha_{jk}(t)) \leqslant 14\delta$$

*for all $t \in [0, t_j]$.*

Proof. The geodesic $[x_i, x_j]$ is covered by the closed subsets $\overline{\mathcal{N}}_\delta([x_i, x_k])$ and $\overline{\mathcal{N}}_\delta([x_j, x_k])$, hence by connectedness there exists a point $p$ on $[x_i, x_j]$ at distance at most $\delta$ from both $[x_i, x_k]$ and $[x_j, x_k]$. Let $p' \in [x_i, x_k]$ and $p'' \in [x_j, x_k]$ be points at distance at most $\delta$ from $p$. The inequality

$$(x_j, x_k)_{x_i} = \frac{1}{2}\left[d(x_i, p) + d(p, x_j) + d(x_i, p') + d(p', x_k) - d(x_j, p'') - d(p'', x_k)\right]$$

combined with the triangle inequality implies that

$$|(x_j, x_k)_{x_i} - d(x_i, p)| \leqslant 2\delta,$$

and, hence $d(x_{ij}, p) \leqslant 2\delta$. Then $d(x_{ik}, p') \leqslant 3\delta$, whence $d(x_{ij}, x_{ik}) \leqslant 6\delta$. It remains to apply Lemma 9.3 to obtain 2 and Lemma 9.4 to obtain 3. □

We thus obtain

Proposition 9.52. *$\kappa$ is a $(1, 14\delta)$-quasi-isometry.*

Proof. The map $\kappa$ is a surjective 1-Lipschitz map. On the other hand, Part 3 of the above lemma implies that

$$d(x, y) - 14\delta \leqslant d(\kappa(x), \kappa(y))$$

for all $x, y \in T$. □

Proposition 9.52 allows one to reduce (up to a uniformly bounded error) study of geodesic triangles in $\delta$-hyperbolic spaces to study of tripods. For instance suppose that $m_{ij} \in [x_i, x_j]$ be points so that

$$d(m_{ij}, m_{jk}) \leqslant r$$

for all $i, j, k$. We already know that this property holds for the central points $x_{ij}$ of $T$ (with $r = 6\delta$). Next result shows that points $m_{ij}$ have to be uniformly close to the central points:

Corollary 9.53. *Under the above assumptions, $d(m_{ij}, x_{ij}) \leqslant r + 14\delta$.*

Proof. Since $\kappa$ is 1-Lipschitz,

$$d(\kappa(m_{ik}), \kappa(m_{jk})) \leqslant r$$

for all $i, j, k$. By definition of the map $\kappa$, all three points $\kappa(m_{ij})$ cannot lie in the same leg of the tripod $\tilde{T}$, except when one of them is the center $o$ of the tripod. Therefore, $d(\kappa(m_{ij}), o) \leqslant r$ for all $i, j$. Since $\kappa$ is $(1, 14\delta)$-quasi-isometry,

$$d(m_{ij}, x_{ij}) \leqslant d(\kappa(m_{ik}), \kappa(m_{jk})) + 14\delta \leqslant r + 14\delta.$$

Definition 9.54. We say that a point $p \in X$ is an $R$-centroid of a triangle $T \subset X$ if distances from $p$ to all three sides of $T$ are $\leqslant R$.

Corollary 9.55. *Every two $R$-centroids of $T$ are within distance at most $\phi(R) = 4R + 28\delta$ from each other.*

Proof. Given an $R$-centroid $p$, let $m_{ij} \in [x_i, x_j]$ be the nearest points to $p$. Then

$$d(m_{ij}, m_{jk}) \leqslant 2R$$

for all $i, j, k$. By previous corollary,

$$d(m_{ij}, x_{ij}) \leqslant 2R + 14\delta.$$

Thus, triangle inequalities imply that every two centroids are within distance at most $2(2R + 14\delta)$ from each other. $\square$

Let $p_3 \in \gamma_{12} = [x_1, x_2]$ be a point closest to $x_3$. Taking $R = 2\delta$ and combining Lemma 9.25 with Lemma 9.46, we obtain:

COROLLARY 9.56. $d(p_3, x_{12}) \leqslant 2(2\delta + 2\delta) = 6\delta$.

We now can define a continuous quasi-inverse $\bar{\kappa}$ to $\kappa$ as follows: We map $[\tilde{x}_1, \tilde{x}_2] \subset \tilde{T}]$ isometrically to a geodesic $[x_1, x_2]$. We send $[o, \tilde{x}_3]$ onto a geodesic $[x_{12}, x_3]$ by an affine map. Since

$$d(x_{12}, x_{32}) \leqslant 6\delta$$

and

$$d(x_3, x_{32}) = d(\tilde{x}_3, 0),$$

we conclude that the map $\bar{\kappa}$ is $(1, 6\delta)$-Lipschitz.

EXERCISE 9.57.

$$d(\bar{\kappa} \circ \kappa, Id) \leqslant 32\delta.$$

## 9.8. Divergence of geodesics in hyperbolic metric spaces

Another important feature of hyperbolic spaces is the *exponential divergence* of its geodesic rays. This can be deduced from the thinness of polygons described in Lemma 9.6, as shown below. Our arguments are inspired by those in [**Pap**].

LEMMA 9.58. *Let $X$ be a geodesic metric space, $\delta$–hyperbolic in the sense of Rips' definition. If $[x, y]$ is a geodesic of length $2r$ and $m$ is its midpoint then every path joining $x, y$ outside the open ball $B(m, r)$ has length at least $2^{\frac{r-1}{2\delta}}$.*

PROOF. Consider such a path $\mathfrak{p}$, of length $\ell$. Divide it first into two arcs of length $\frac{\ell}{2}$, then into four arcs of length $\frac{\ell}{4}$ etc, until we obtain $k$ arcs of length $\frac{\ell}{2^k} \leqslant 1$. Consider the minimal $k$ satisfying this, i.e. $k$ is the integer part $\lfloor \log_2 \ell \rfloor$. Let $x_0 = x, x_1, ..., x_k = y$ be the consecutive points on $\mathfrak{p}$ obtained after this procedure. Lemma 9.6 applied to a geodesic polygon with vertices $x_0 = x, x_1, ..., x_k = y$ with $[x, y]$ as an edge, implies that $m$ is contained in the $(2\delta k)$–tubular neighborhood of $\bigcup_{i=0}^{k-1} [x_i, x_{i+1}]$, hence in the $(2\delta k + 1)$–tubular neighborhood of $\mathfrak{p}$. However, we assumed that $\text{dist}(m, \mathfrak{p}) \geqslant r$. Thus,

$$r \leqslant 2\delta k + 1 \leqslant 2 \log_2 \ell + 1 \Rightarrow \ell \geqslant 2^{\frac{r-1}{2\delta}}.$$

$\square$

LEMMA 9.59. *Let $X$ be a geodesic metric space, $\delta$–hyperbolic in the sense of Rips' definition, and let $x$ and $y$ be two points on the sphere $S(o, R)$ such that* $\text{dist}(x, y) = 2r$. *Every path joining $x$ and $y$ outside $\overline{B}(o, R)$ has length at least* $\psi(r) = 2^{\frac{r-1}{2\delta} - 3} - 12\delta$.

PROOF. Let $m \in [x, y]$ be the midpoint. Since $d(o, x) = d(o, y)$, it follows that $m$ is also one of the center-points of the triangle $\Delta(x, y, o)$ in the sense of Section 9.7. Then, by using Lemma 9.51 (Part 1), we see that $d(m, o) \leqslant (R - r) + 6\delta$. Therefore, the closed ball $\overline{B}(m, r - 6\delta)$ is contained in $\overline{B}(o, R)$. Let $\mathfrak{p}$ be a path joining $x$ and $y$ outside $\overline{B}(o,R)$, and let $[x,x']$ and $[y',y]$ be subsegments of $[x,y]$ of length $6\delta$. Lemma 9.58 implies that the path $[x',x]\cup \mathfrak{p}\cup [y,y']$ has length at least

$$2^{\frac{r-6\delta-1}{2\delta}}$$

whence $\mathfrak{p}$ has length at least

$$2^{\frac{r-1}{\delta}-3}-12\delta.$$

□

LEMMA 9.60. *Let $X$ be a $\delta$–hyperbolic in the sense of Rips, and let $x$ and $y$ be two points on the sphere $S(o,r_1+r_2)$ such that there exist two geodesics $[x,o]$ and $[y,o]$ intersecting the sphere $S(o,r_1)$ in two points $x',y'$ at distance larger than $14\delta$. Then every path joining $x$ and $y$ outside $B(o,r_1+r_2)$ has length at least $\psi(r_2-15\delta)=2^{\frac{r_2-1}{\delta}-18}-12\delta$.*

PROOF. Let $m$ be the midpoint $m$ of $[x,y]$, since $\Delta(x,y,o)$ is isosceles, $m$ is one of the centroids of this triangle. Since $d(x',y')>14\delta$, they cannot be dual point on $\Delta(x,y,o)$ in the sense of Section 9.7. Let $x'',y''\in[x,y]$ be dual to $x',y'$. Thus (by Lemma 9.51 (Part 3)),

$$d(o,x'')\leqslant r_1+14\delta, d(o,x'')\leqslant r_1+14\delta.$$

Furthermore, by the definition of dual points, since $m$ is a centroid of $\Delta(x,y,o)$, $m$ belongs to the segment $[x'',y'']\subset[x,y]$. Thus, by quasiconvexity of metric balls, see Section 9.5,

$$d(m,o)\leqslant r_1+14\delta+\delta=r_1+15\delta.$$

By the triangle inequality,

$$r_1+r_2=d(x,o)\leqslant r+d(m,o)\leqslant r+r_1+15\delta,\quad r_2-15\delta\leqslant r.$$

Since the function $\psi$ in Lemma 9.59 is increasing,

$$\psi(r_2-15\delta)\leqslant\psi(r).$$

Combining this with Lemma 9.59 (where we take $R=r_1+r_2$), we get the required inequality. □

For a more detailed treatment of divergence in metric spaces, see [**Ger94, KL98a, Mac, DR09, BC12, AK11**].

## 9.9. Ideal boundaries

We consider the general notion of ideal boundary defined in Section 2.1.10 of Chapter 1 in the special case when $X$ is geodesic, $\delta$–hyperbolic and locally compact (equivalently, proper).

LEMMA 9.61. *For each $p\in X$ and each element $\alpha\in\partial_\infty X$ there exists a geodesic ray $\rho$ with initial point $p$ and such that $\rho(\infty)=\alpha$.*

PROOF. Let $\rho'$ be a geodesic ray from the equivalence class $\alpha$, with initial point $x_0$. Consider a sequence of geodesic segments $\gamma_n:[0,D_n]\to X$, connecting $p$ to $x_n=\rho'(n)$, where $D_n=d(p,\rho'(n))$. The $\delta$-hyperbolicity of $X$ implies that $Im(\gamma_n)$ is at Hausdorff distance at most $\delta+\mathrm{dist}(p,x_0)$ from $[x_0,x_n]$, where $[x_0,x_n]$ is the initial subsegment of $\rho'$.

Combining the properness of $X$ with the Arzela-Ascoli theorem, we see that the geodesic maps $\gamma_n$ subconverge to a geodesic ray $\rho$, $\rho(0) = p$. Clearly, $Im(\rho)$ is at Hausdorff distance at most $\delta + \mathrm{dist}(p, x_0)$ from $Im(\rho)$). In particular, $\rho \sim \rho$. $\square$

Lemma 9.61 is very similar to the result in the case of $X$ $CAT(0)$–space. The important difference with respect to that case is that the ray $\rho$ may not be unique. Nevertheless we shall still use the notation $[p, \alpha)$ to designate a geodesic (one of the geodesics) with initial point $x$ in the equivalence class $\alpha$.

In view of this lemma, in order to understand $\partial_\infty X$ it suffices to restrict to the set $Ray_p(X)$ of geodesic rays in $X$ emanating from $p \in X$.

It is convenient to extend the topology $\tau$ defined on $\partial_\infty X$ (i.e. the quotient topology of the compact-open topology on the set of rays) to a topology on $\bar{X} = X \cup \partial_\infty X$. Namely, we say that a sequence $x_n \in X$ converges to a point $\xi \in \partial_\infty X$ if a sequence of geodesics $[p, x_n]$ converges (uniformly on compacts) to a ray $[p, \xi)$. Then $\partial_\infty X \subset \bar{X}$ is a closed subset. Consider the set $Geo_p(X)$ consisting of geodesics in $X$ (finite or half-infinite) emanating from $p$. We again quip $Geo_p(X)$ with the compact-open topology. There is a natural quotient map $Geo_p(X) \to \bar{X}$ which sends a finite geodesic or a geodesic ray emanating from $p$ to its terminal point in $\bar{X}$.

COROLLARY 9.62. *If $X$ is geodesic, hyperbolic and proper, then $\bar{X}$ is compact.*

PROOF. The space $Geo_p(X)$ is compact by Arzela-Ascoli theorem. Since a quotient of a compact is compact, the claim follows. $\square$

LEMMA 9.63 (Asymptotic rays are uniformly close). *Let $\rho_1, \rho_2$ be asymptotic geodesic rays in $X$ such that $\rho_1(0) = \rho_2(0) = p$. Then for each $t$,*

$$d(\rho_1(t), \rho_2(t)) \leqslant 2\delta.$$

PROOF. Suppose that the rays $\rho_1, \rho_2$ are within distance $\leqslant C$ from each other. Take $T \succ t$. Then (since the rays are asymptotic) there exists $S \in \mathbb{R}_+$ such that

$$d(\rho_1(T), \rho_2(S)) \leqslant C.$$

By $\delta$–thinness of the triangle $\Delta(p\rho_1(T)\rho_2(S))$, the point $\rho_1(t)$ is within distance $\leqslant \delta$ from a point either on $[p, \rho_2(S)]$ or on $[\rho_1(T), \rho_2(S)]$. Since the length of $[\rho_1(T), \rho_2(S)]$ is $\leqslant C$ and $T \succ t$, it follows that there exists $t'$ such that

$$\mathrm{dist}(\rho_1(t), \rho_2(t')) \leqslant \delta.$$

By the triangle inequality, $|t - t'| \leqslant \delta$. It follows that $\mathrm{dist}(\rho_1(t), \rho_2(t)) \leqslant 2\delta$. $\square$

COROLLARY 9.64. *$\partial_\infty X$ is Hausdorff.*

PROOF. Let $\rho_n, \rho_n'$ be sequences of rays emanating from $p \in X$, so that $\rho_n \sim \rho_n'$ and

$$\lim_{n\to\infty} \rho_n = \rho, \quad \lim_{n\to\infty} \rho_n' = \rho'.$$

We claim that $\rho \sim \rho'$. Suppose not. Then there exists $a > 0$ so that $d(\rho(a), \rho'(a)) \geqslant 2\delta + 1$. For all sufficiently large $n$

$$d(\rho_n(a), \rho(a)) < 1/2, \quad d(\rho_n'(a), \rho'(a)) < 1/2,$$

while

$$d(\rho_n(a), \rho_n'(a)) \leqslant 2\delta.$$

Thus, $d(\rho(a), \rho'(a)) < 2\delta + 1$, contradicting our choice of $a$. $\square$

EXERCISE 9.65. Show that $\bar{X}$ is also Hausdorff.

Given a number $k > 2\delta$, define the topology $\tau_k$ on $Ray_p(X)/\sim$, where the basis of neighborhoods of a point $\rho(\infty)$ given by

$$U_{k,n}(\rho) := \{\rho' : \mathrm{dist}(\rho'(t), \rho(t)) < k, t \in [0, n]\}, n \in \mathbb{R}_+. \tag{9.5}$$

LEMMA 9.66. *Topologies $\tau$ and $\tau_k$ coincide.*

PROOF. 1. Suppose that $\rho_j$ is a sequence of rays emanating from $p$ such that $\rho_j \notin U_{k,n}(\rho)$ for some $n$. If $\lim_j \rho_j = \rho'$ then $\rho' \notin U_{k,n}$ and by Lemma 9.63, $\rho'(\infty) \neq \rho(\infty)$.

2. Conversely, if for each $n$, $\rho_j \in U_{k,n}(\rho)$ (provided that $j$ is large enough), then the sequence $\rho_j$ subconverges to a ray $\rho'$ which belongs to each $U_{k,n}(\rho)$. Hence $\rho'(\infty) = \rho(\infty)$. □

LEMMA 9.67. *Suppose that $\rho, \rho' \in Ray_p(X)$ are inequivalent rays. Then for every sequence $t_n$ diverging to $\infty$,*

$$\lim_{i\to\infty} d(\rho(t_i), \rho'(t_i)) = \infty.$$

PROOF. Suppose to the contrary, there exists a divergent sequence $t_i$ so that $d(\rho(t_i), \rho'(t_i)) \leqslant D$. Then, by Lemma 9.4, for every $t \leqslant t_i$,

$$d(\rho(t), \rho'(t)) \leqslant 2(D + \delta).$$

Since $\lim t_i = \infty$, it follows that $\rho \sim \rho'$. Contradiction. □

LEMMA 9.68. *Let $X$ be a proper geodesic Gromov–hyperbolic space. Then for each pair of distinct points $\xi, \eta \in \partial_\infty X$ there exists a geodesic $\gamma$ in $X$ which is asymptotic to both $\xi$ and $\eta$.*

PROOF. Consider geodesic rays $\rho, \rho'$ emanating from the same point $p \in X$ and asymptotic to $\xi, \eta$ respectively. Since $\xi \neq \eta$, by previous lemma, for each $R < \infty$ the set

$$K(R) := \{x \in X : \mathrm{dist}(x, \rho) \leqslant R, \mathrm{dist}(x, \rho') \leqslant R\}$$

is compact. Consider the sequences $x_n := \rho(n), x_n' := \rho'(n)$ on $\rho, \rho'$ respectively. Since the triangles $[p, x_n, x_n']$ are $\delta$–thin, each segment $\gamma_n := [x_n, x_n']$ contains a point within distance $\leqslant \delta$ from both $[p, x_n], [p, x_n']$, i.e. $\gamma_n \cap K(\delta) \neq \emptyset$. Therefore, by Arzela-Ascoli theorem, the sequence of geodesic segments $\gamma_n$ subconverges to a complete geodesic $\gamma$ in $X$. Since $\gamma \subset \mathcal{N}_\delta(\rho \cup \rho')$ it follows that $\gamma$ is asymptotic to $\xi$ and $\eta$. □

EXERCISE 9.69. Suppose that $X$ is $\delta$-hyperbolic. Show that there are no complete geodesics $\gamma$ in $X$ so that

$$\lim_{n\to\infty} \gamma(-n) = \lim_{n\to\infty} \gamma(n).$$

Hint: Use the fact that geodesic bigons in $X$ are $\delta$-thin.

EXERCISE 9.70 (Ideal bigons are $2\delta$-thin). Suppose that $\alpha, \beta$ are geodesics in $X$ which are both asymptotic to points $\xi, \eta \in \partial_\infty X$. Then $dist_{Haus}(\alpha, \beta) \leqslant 2\delta$. Hint: For $n \in \mathbb{N}$ define $z_n, w_n \in Im(\beta)$ to be the nearest points to $x_n = \alpha(n), y_n = \alpha(-n)$. Let $[x_n, y_n], [z_n, w_n]$ be the subsegments of $\alpha, \beta$ between $x_n, y_n$ and $y_n, z_n$ respectively. Now use the fact that the quadrilateral in $X$ with the edges $[x_n, y_n], [y_n, w_n], [w_n, z_n], [z_n, x_n]$ is $2\delta$-thin.

We now compute two examples of ideal boundaries of hyperbolic spaces.

1. Suppose that $X = \mathbb{H}^n$ is the real-hyperbolic space. We claim that $\partial_\infty X$ is naturally homeomorphic to the sphere $S^{n-1}$, the boundary sphere of $\mathbb{H}^n$ in the unit ball model. Every ray $\rho \in Ray_o(X)$ (which is a Euclidean line segment $[o, \xi)$, $\xi \in S^{n-1}$) determines a unique point on the boundary sphere $S^{n-1}$, namely the point $\xi$. Furthermore, we claim that distinct rays $\rho_1, \rho_2 \in Ray_o(X)$ are never asymptotic. Indeed, consider the equilateral triangle $[o, \rho_1(t), \rho_2(t)]$ with the angle $\gamma > 0$ at $o$. Then the hyperbolic cosine law (8.4), implies that

$$\cosh(d(\rho_1(t), \rho_2(t))) = 1 + \sinh^2(t)(1 - \cos(\gamma)).$$

It is clear that this quantity diverges to $\infty$ as $t \to \infty$. We, thus, obtain a bijection

$$Ray_o(X) \to \partial_\infty(X).$$

We equip $Ray_p(X)$ with the topology given by the initial velocities $\rho'(0)$ of the geodesic rays $\rho \in Ray_o(X)$. Clearly, the map $Ray_o(X) \to S^{n-1}$, sending each ray $\rho = [o, \xi)$ to $\xi \in S^{n-1}$ is a homeomorphism. It is also clear that the above topology on $Ray_o(X)$ coincides with the compact-open topology on geodesic rays since the latter depend continuously on their initial velocities. Thus, the composition

$$S^{n-1} \to Ray_o(X) \to \partial_\infty X$$

is a homeomorphism.

2. Suppose that $X$ is a simplicial tree of finite constant valence $val(X) \geqslant 3$, metrized so that every edge has unit length. As before, it suffices to restrict to rays in $Ray_p(X)$, where $p \in X$ is a fixed vertex. Note that $\rho, \rho' \in Ray_p(X)$ are equivalent if and only they are equal. We know that $X$ is 0-hyperbolic. Our claim then is that $\partial_\infty X$ is homeomorphic to the Cantor set. Since we know that $\partial_\infty X$ is compact and Hausdorff, it suffices to verify that $\partial_\infty X$ is totally disconnected and contains no isolated points. Let $\rho \in Ray_p(X)$ be a ray. For each $n$ pick a ray $\rho_n \in Ray_p(X)$ which coincides with $\rho$ on $[0, n]$ , but $\rho_n(t) \neq \rho(t)$ for all $t > n$ (this is where we use the fact that $val(X) \geqslant 3$. It is then clear that

$$\lim_{n \to \infty} \rho_n = \rho$$

uniformly on compacts. Hence, $\partial_\infty X$ has no isolated points. Recall that for $k = \frac{1}{2}$, we have open sets $U_{n,k}(\rho)$ forming a basis of neighborhoods of $\rho$. We also note that each $U_{n,k}(\rho)$ is also closed, since (for a tree $X$ as in our example) it is also given by

$$\{\rho' : \rho(t) = \rho'(t), t \in [0, n]\}.$$

Therefore, $\partial_\infty X$ is totally-disconnected as for any pair of distinct points $\rho, \rho' \in Ray_p(X)$, there exist open, closed and disjoint neighborhoods $U_{n,k}(\rho), U_{n,k}(\rho')$ of the points $\rho, \rho'$. Thus, $\partial_\infty X$ is compact, Hausdorff, perfect, consists of at least 2 points and is totally-disconnected. Therefore, $\partial_\infty X$ is homeomorphic to the Cantor set.

**Gromov topology on $\bar{X} = X \cup \partial_\infty X$.** The above definition of $\bar{X}$ was worked fine for geodesic hyperbolic metric spaces. Gromov extended this definition to the case when $X$ is an arbitrary hyperbolic metric space. Pick a base-point $p \in X$. Gromov boundary $\partial_{Gromov} X$ of $X$ consists of equivalence classes of sequences $(x_n)$ in $X$ so that $\lim d(p, x_n) = \infty$, where $(x_n) \sim (y_n)$ if

$$\lim_{n \to \infty} (x_n, y_n)_p = \infty.$$

One then defines the Gromov–product $(\xi,\eta)_p \in [0,\infty]$ for points $\xi,\eta$ in Gromov-boundary of $X$ by

$$(\xi,\eta)_p = \lim sup_{n\to\infty}(x_n,y_n)_p$$

where $(x_n)$ and $(y_n)$ are sequences representing $\xi,\eta$ respectively. Then, Gromov topologizes $\bar{X} = X \cup \partial_{Gromov}X$ by:

$$\lim x_n = \xi, \xi \in \partial_{Gromov}X$$

if and only if

$$\lim_{n\to\infty}(x_n,\xi)_p = \infty.$$

It turns out that this topology is independent of the choice of $p$. In case when $X$ is also a geodesic metric space, there is a natural map

$$X \cup \partial_\infty X \to X \cup \partial_{Gromov}X$$

which is the identity on $X$ and which sends $\xi = [\rho]$ in $\partial_\infty X$ to the equivalence class of the sequence $(\rho(n))$. This map is a homeomorphism provided that $X$ is proper.

**Hyperbolic triangles with ideal vertices.** We return to the case when $X$ is a $\delta$-hyperbolic proper geodesic metric space. We now generalize (geodesic) triangles in $X$ to triangles where some vertices are in $\partial_\infty X$, similarly to the definitions made in section 8.3. Namely a (generalized) geodesic triangle in $\bar{X}$ is a concatenation of geodesics connecting (consecutively) three points $A, B, C$ in $\bar{X}$; geodesics are now allowed to be finite, half-infinite and infinite. The points $A, B, C$ are called vertices of the triangle. As in the case of $\mathbb{H}^n$, *we do not allow two ideal vertices of a triangle $T$ to be the same.* By abusing terminology, we will again refer to such generalized triangles as *hyperbolic triangles*.

An ideal triangle is a triangle where all three vertices are in $\partial_\infty X$. We topologize the set $Tri(X)$ of hyperbolic triangles in $X$ by compact-open topology on the set of their geodesic edges. Given a hyperbolic triangle $T = T(A,B,C)$ in $X$, we find a sequence of finite triangles $T_i \subset X$ whose vertices converge to the respective vertices of $T$. Passing to a subsequence if necessary and taking a limit of the sides of the triangles $T_i$, we obtain limit geodesics connecting vertices $A, B, C$ of $T$. The resulting triangle $T'$, of course, need not be equal to $T$ (since geodesics connecting points in $\bar{X}$ need not be unique), however, in view of Exercise 9.70, sides of $T'$ are thin distance $\leqslant 2\delta$ from the respective sides of $T$. We will say that the sequence of triangles $T_i$ *coarsely converges* to the triangle $T$ (cf. Definition 5.25).

EXERCISE 9.71. Every (generalized) hyperbolic triangle $T$ in $X$ is $5\delta$-thin. In particular,

$$minsize(T) \leqslant 4\delta.$$

Hint: Use a sequence of finite triangles which coarsely converges to $T$ and the fact that finite triangles are $\delta$-thin.

This exercise allows one to define a *centroid* of a triangle $T$ in $X$ (with sides $\tau_i, i = 1,2,3$) to be a point $p \in X$ so that

$$d(p,\tau_i) \leqslant 5\delta, i = 1,2,3.$$

More generally, as in Definition 9.54, we say that a point $p \in X$ is an $R$-centroid $T$ it $p$ is within distance $\leqslant R$ from all three sides of $T$.

LEMMA 9.72. *Distance between any two $R$-centroids of a hyperbolic triangle $T$ is at most*

$$r(R,\delta) = 4R + 32\delta.$$

PROOF. Let $p, q$ be $R$-centroids of $T$. We coarsely approximate $T$ by a sequence of finite triangles $T_i \subset X$. Then for every $\epsilon > 0$, for all sufficiently large $i$, the points $p, q$ are $R + 2\delta + \epsilon$-centroids of $T_i$. Therefore, by Corollary 9.55 applied to triangles $T_i$,

$$d(p,q) \leqslant \phi(R + 2\delta + \epsilon) = 4(R + 2\delta + \epsilon) + 28\delta = 4R + 32\delta + 2\epsilon$$

Since this holds for every $\epsilon > 0$, we conclude that $d(p,q) \leqslant 4R + 32\delta$. □

We thus, define the correspondence

$$center : Trip(\partial_\infty X) \to X$$

which sends every triple of distinct points in $\partial_\infty X$ first to the set of ideal triangle $T$ that they span and then to the set of centroid of these ideal triangles. Then Lemma 9.72 implies

COROLLARY 9.73. *For every $\xi \in Trip(\partial_\infty X)$,*

$$\mathrm{diam}(center(\xi)) \leqslant r(7\delta, \delta) = 60\delta.$$

EXERCISE 9.74. Suppose that $\gamma_n$ are geodesics in $X$ which limit to points $\zeta_n, eta_n \in \partial_\infty X$ and

$$\lim_n \zeta_n = \zeta, \lim_n \eta_n = \eta, \eta \neq \zeta.$$

Show that geodesics $\gamma_n$ subconverge to a geodesic which is asymptotic to both $\xi$ and $\eta$.

Use this exercise to conclude:

EXERCISE 9.75. If $K \subset Trip(\partial_\infty X)$ is a compact subset, then $center(K)$ is a bounded subset of $X$.

Conversely,

EXERCISE 9.76. Let $B \subset X$ be a bounded subset and $K \subset Trip(\partial_\infty X)$ is a subset such that $center(K) \subset B$. Show that $K$ is relatively compact in $Trip(\partial_\infty X)$. Hint: For every $\xi \in K$, every ideal edge of a triangle spanned by $\xi$ intersects $5\delta$-neighborhood of $B$. Now, use Arzela-Ascoli theorem.

Loosely speaking, the two exercises show that the correspondence *center* is coarsely continuous (image of a compact is bounded) and coarsely proper (preimage of a bounded subset is relatively compact).

**Cone topology.** Suppose that $X$ is a proper geodesic hyperbolic metric space. Later on, it will be convenient to use another topology on $\bar{X}$, called *cone topology*. This topology is not equivalent to the topology $\tau$: With few exceptions, $\bar{X}$ is noncompact with respect to this topology (even if $X = \mathbb{H}^n, n \geqslant 2$).

DEFINITION 9.77. We say that a sequence $x_n \in X$ converges to a point $\xi = \rho(\infty) \in \partial_\infty X$ in the **cone topology** if there is a constant $C$ such that $x_n \in \mathcal{N}_C(\rho)$ and the geodesic segments $[x_1 x_n]$ converge to a geodesic ray asymptotic to $\xi$.

EXERCISE 9.78. If a sequence $x_n$ converges to $\xi \in \partial_\infty X$ in the cone topology, then it also converges to $\xi$ in the topology $\tau$ on $\bar{X}$.

As an example, consider $X = \mathbb{H}^m$ in the upper half-space model, $\xi = 0 \in \mathbb{R}^{m-1}$, $L$ is the vertical geodesic from the origin. Then a sequence $x_n \in X$ converges $\xi$ in the cone topology if and only if all the points $x_n$ belong to the Euclidean cone with the axis $L$ and the Euclidean distance from $x_n$ to 0 tends to zero. See Figure 9.4. This explains the name *cone topology.*

EXERCISE 9.79. Suppose that a sequence $(x_i)$ converges to a point $\xi \in \partial_\infty \mathbb{H}^n$ along a horosphere centered at $\xi$. Show that the sequence $(x_i)$ contains no convergent subsequence in the cone topology on $\bar{X}$.

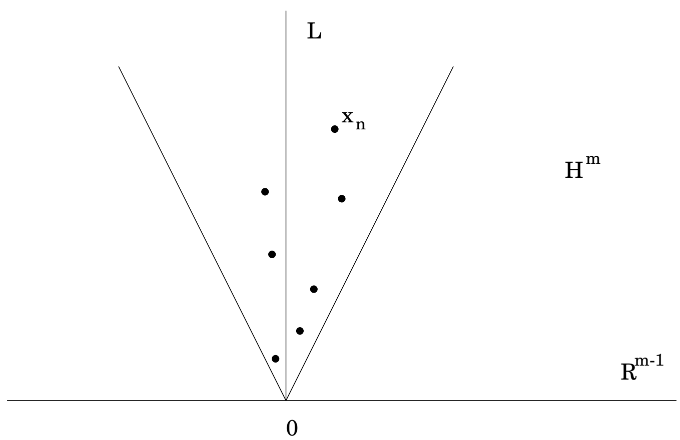



FIGURE 9.4. Convergence in the cone topology.

## 9.10. Extension of quasi-isometries of hyperbolic spaces to the ideal boundary

The goal of this section is to explain how quasi-isometries of Rips–hyperbolic spaces extend to their ideal boundaries.

We first extend Morse lemma to the case of quasi-geodesic rays and complete geodesics.

LEMMA 9.80 (Extended Morse Lemma). *Suppose that $X$ is a proper $\delta$–hyperbolic geodesic space. Let $\rho$ be an $(L, A)$–quasigeodesic ray or a complete $(L, A)$–quasi-geodesic. Then there is $\rho^*$ which is either a geodesic ray or a complete geodesic in $X$ so that the Hausdorff distance between $Im(\rho)$ and $Im(\rho^*)$ is $\leqslant \theta(L, A, \delta)$. Here $\theta$ is the function which appears in Morse lemma.*

*Moreover, there are two functions $s = s(t), s^* = s^*(t)$ so that*

$$L^{-1}t - B \leqslant s \leqslant Lt + B \tag{9.6}$$

*and*

$$L^{-1}(t - B) \leqslant s^* \leqslant L(t + B) \tag{9.7}$$

*and for every $t$, $d(\rho(t), \rho^*(s)) \leqslant \theta$, $d(\rho^*(t), \rho(s^*)) \leqslant \theta$. Here $B = A + \theta$.*

Proof. We will consider only the case of quasigeodesic rays $\rho : [0, \infty) \to X$ as the other case is similar. Let $\rho_i := \rho|[0, i]$, $i \in \mathbb{N}$. Consider the sequence of geodesic segments $\rho_i^* = [\rho(0)\rho(i)]$ as in Morse lemma. By Morse lemma,

$$dist_{Haus}(\rho_i, \rho_i^*) \leqslant \theta(L, A, \delta).$$

By properness, the geodesic segments $\rho_i^*$ subconverge to a complete geodesic ray $\rho^*$. It is now clear that

$$dist_{Haus}(\rho, \rho^*) \leqslant \theta(L, A, \delta).$$

Estimates (9.6) and (9.7) follow from the estimates (9.3) and (9.4) in the case of finite geodesic segments. □

Corollary 9.81. *If $\rho$ is a quasi-geodesic ray as in the above lemma, there exists a point $\xi \in \partial_\infty X$ so that $\lim_{t\to\infty} \rho(t) = \xi$.*

Proof. Take $\xi = \rho^*(\infty)$. Since $d(\rho(t), Im(\rho^*)) \leqslant \theta$, it follows that

$$\lim_{t\to\infty} \rho(t) = \xi. \quad \square$$

We will refer to the point $\eta$ as $\rho(\infty)$. Note that if $\rho'$ is another quasi-geodesic ray which is Hausdorff-close to $\rho$ then $\rho(\infty) = \rho'(\infty)$.

Below is another useful application of the Extended Morse Lemma. Given a geodesic $\gamma$ in $X$ we let $\pi_\gamma : X \to \gamma$ denote the nearest-point projection.

Proposition 9.82 (Quasi-isometries commute with projections). *There exists $C = C(L, A, \delta)$ so that the following holds. Let $X$ be a $\delta$-hyperbolic geodesic metric space and let $f : X \to X$ be an $(L, A)$-quasi-isometry. Let $\alpha$ be a (finite or infinite) geodesic in $X$, and $\beta \subset X$ be a geodesic which is $\theta(L, A, \delta)$-close to $f(\alpha)$. Then the map $f$* almost commutes *with the nearest-point projections $\pi_\alpha, \pi_\beta$:*

$$d(f(\pi_\alpha(x)), \pi_\beta f(x)) \leqslant C, \quad \forall x \in X.$$

Proof. For a (finite or infinite) geodesic $\gamma \subset X$ consider the triangle $\Delta = \Delta_{x,\gamma}$ where one side is $\gamma$ and $x$ is a vertex: The other two sides are geodesics connecting $x$ to the (finite or ideal) end-points of $\gamma$. Let $c = center(\Delta) \in \overline{\gamma}$ denote a centroid of $\Delta$: The distance from $c$ to each side of $\Delta$ is $\leqslant 6\delta$. By Corollary 9.56,

$$d(c, \pi_\gamma(x)) \leqslant 21\delta.$$

Applying $f$ to the centroid $c(\Delta_{x,\alpha})$ we obtain a point $a \in X$ whose distance to each side of the quasi-geodesic triangle $f(\Delta_{x,\alpha})$ is $\leqslant 2\delta L + A$. Hence, the distance from $a$ to each side of the geodesic triangle $\Delta_{y,\beta}, y = f(x)$ is at most $R := 2\delta L + A + D(L, A, \delta)$. Hence, $a$ is an $R$-centroid of $\Delta_{y,\beta}$. By Lemma 9.72, it follows that

$$d(a, c(\Delta_{y,\beta})) \leqslant 8R + 32\delta.$$

Since $d(\pi_\beta(y), c(\Delta_{y,\beta})) \leqslant 21\delta$, we obtain:

$$d(f(\pi_\alpha(x)), \pi_\beta f(x)) \leqslant C := 21\delta + 8R + 27\delta + 21\delta L + A. \quad \square$$

Below is the main theorem of this section, which is a fundamental fact of the theory of hyperbolic spaces:

Theorem 9.83 (Extension Theorem). *Suppose that $X$ and $X'$ are Rips–hyperbolic proper metric spaces. Let $f : X \to X'$ be a quasi-isometry. Then $f$ admits a homeomorphic extension $f_\infty : \partial_\infty X \to \partial_\infty X'$. This extension is such that the map*

*$f \cup f_\infty$ is continuous at each point $\eta \in \partial_\infty X$ with respect to the topology $\tau$ on $\bar{X}$. The extension satisfies the following functoriality properties:*

*1. For every pair of quasi-isometries $f_i : X_i \to X_{i+1}, i = 1, 2$, we have*

$$(f_2 \circ f_1)_\infty = (f_2)_\infty \circ (f_1)_\infty.$$

*2. For every pair of quasi-isometries $f_1, f_2 : X \to X'$ satisfying $\operatorname{dist}(f_1, f_2) < \infty$, we have $(f_2)_\infty = (f_1)_\infty$.*

PROOF. First, we construct the extension $f_\infty$. Let $\eta \in \partial_\infty X$, $\eta = \rho(\infty)$ where $\rho$ is a geodesic ray in $X$. The image of this ray $f \circ \rho : \mathbb{R}_+ \to X'$ is a quasi-geodesic ray, hence we set $f_\infty(\eta) := f\rho(\infty)$. Observe that $f_\infty(\eta)$ does not depend on the choice of a geodesic ray asymptotic to $\eta$.

We will verify continuity for the map $f_\infty : \partial_\infty X \to \partial_\infty X$ and leave the case of $\bar{X}$ as an exercise to the reader. Let $\eta_n \in \partial_\infty X$ be a sequence which converges to $\eta$. Let $\rho_n$ be a sequence of geodesic rays asymptotic to $\eta_n$ with $\rho_n(0) = \rho(0) = x_0$. Then, by Lemma 9.66, for each $a \in \mathbb{R}_+$ there exists $n_0$ such that for all $n \geqslant n_0$ and $t \in [0, a]$ we have

$$d(\rho(t), \rho_n(t)) \leqslant 3\delta,$$

where $\delta$ is the hyperbolicity constant of $X$. Let $\rho'_n := (f \circ \rho_n)^*, \rho' := (f\rho)^*$ denote a geodesic rays given by Lemma 9.80. Thus, for all $t \in [0, a]$ there exist $s$ and $s_n$,

$$L^{-1}t - A - \theta \leqslant \min(s, s_n),$$

so that

$$d(f\rho_n(t), \rho'_n(s_n)) \leqslant \theta,$$
$$d(f\rho(t), \rho'(s)) \leqslant \theta,$$

and for all $t \in [0, a]$,

$$d(f\rho_n(t), f\rho(t)) \leqslant 3\delta L + A.$$

Thus, by the triangle inequalities, for the above $s, s_n$ we get

$$d(\rho'_n(s_n), \rho'(s)) \leqslant C = 3\delta L + A + 2\theta.$$

Since $\rho'_n, \rho'$ are geodesic, $|s - s_n| \leqslant C$. In particular, for $t = a$, and $b$ the corresponding value of $s$, we obtain

$$d(\rho'(b), \rho'_n(b)) \leqslant 2C.$$

By the fellow-traveling property of hyperbolic geodesics, for all $u \in [0, b]$,

$$d(\rho'(u), \rho'_n(u)) \leqslant k := 2(2C + \delta).$$

Since $b \geqslant L^{-1}a - A - \theta$ and

$$\lim_{a \to \infty} (L^{-1}a - A - \theta) = \infty,$$

it follows that $\lim \rho'_n(\infty) = \rho'(\infty)$ in the topology $\tau_k$. Since topologies $\tau$ and $\tau_k$ agree, it follows that $\lim_n f_\infty(\xi_n) = f_\infty(\xi)$. Hence, $f_\infty$ is continuous.

Functoriality properties (1) and (2) of the extension are clear from the construction (in view of Morse Lemma). They also follow from continuity of the extension.

Let $\bar{f}$ be a quasi-inverse of $f : X \to X'$. Then, by the functoriality properties, $(\bar{f})_\infty$ is inverse of $f_\infty$. Thus, extension of a quasi-isometry $X \to X'$ is a homeomorphism $\partial_\infty X \to \partial_\infty X'$. □

EXERCISE 9.84. Suppose that $f$ is merely a QI embedding $X \to X'$. Show that the continuous extension $f_\infty$ given by this theorem is 1-1.

Remark 9.85. The above extension theorem was first proven by Efremovich and Tikhomirova in [**ET64**] for the real-hyperbolic space and, soon afterwards, reproved by Mostow [**Mos73**]. We will see later on that the homeomorphisms $f_\infty$ are *quasi-symmetric*, in particular, they enjoy certain regularity properties which are critical for proving QI rigidity theorems in the context of hyperbolic groups and spaces.

We thus obtained a functor from quasi-isometries between Rips–hyperbolic spaces to homeomorphisms between their boundaries.

The following lemma is a "converse" to the 2nd functoriality property in Theorem 9.83:

Lemma 9.86. *Let $X$ and $Y$ be proper geodesic $\delta$–hyperbolic spaces. In addition we assume that centroids of ideal triangles in $X$ form an $R$-net in $X$. Suppose that $f, f' : X \to Y$ are $(L, A)$–quasi-isometries such that $f_\infty = f'_\infty$ Then* $\operatorname{dist}(f, f') \leqslant$ *$D(L, A, R, \delta)$,*

Proof. Let $x \in X$ and $p \in X$ be a centroid of an ideal triangle $T$ in $X$, so that $d(x, p) \leqslant R$. (Recall that $p$ is a centroid of $T$ if $p$ is within distance $\leqslant 4\delta$ from all three sides of $T$). Then, by Lemma 9.80, $q = f(p), q' = f'(q')$ are $C$-centroids of the ideal geodesic triangle $S \subset Y$ whose ideal vertices are the images of the ideal vertices of $T$ under $f_\infty$. Here $C = 4\delta L + A + \theta(L, A, \delta)$. By Lemma 9.72, $d(q, q') \leqslant r(C, \delta)$. Therefore,

$$d(f(x), f'(x)) \leqslant D(L, A, R, \delta) = 2(LR + A) + r(C, \delta). \quad \square$$

Suppose that $X$ is Gromov–hyperbolic and $\partial_\infty X$ contains at least 3 points. Then $X$ has at least one ideal triangle and, hence, at least one centroid of an ideal triangle. If, in addition, $X$ is quasi-homogeneous, then centroids of ideal triangles in $X$ form a net. Thus, the above lemma applies to the real-hyperbolic space and, as we will sees soon, every non-elementary hyperbolic group.

Example 9.87. The line $X = \mathbb{R}$ is 0-hyperbolic, its ideal boundary consists of 2 points. Take a translation $f : X \to X$, $f(x) = x + a$. Then $f_\infty$ is the identity map of $\{-\infty, \infty\}$ but there is no bound on the distance from $f$ to the identity.

Corollary 9.88. *Let $X$ be a Rips–hyperbolic space. Then the map $f \mapsto f_\infty$ (where $f : X \to X$ are quasi-isometries) descends to a homomorphism $QI(X) \to Homeo(X)$. Furthermore, under the hypothesis of Lemma 9.86, this homomorphism is injective.*

In Section 20.5 we will identify the image of this homomorphism in the case of real-hyperbolic space $\mathbb{H}^n$, it will be a subgroup of $Homeo(S^{n-1})$ consisting of *quasi-Moebius* homeomorphisms.

**Boundary extension and quasi-actions.** In view of Corollary 9.88, we have

Corollary 9.89. *Every quasi-action $\phi$ of a group $G$ on $X$ extends (by $g \mapsto \phi(g)_\infty$) to an action $\phi_\infty$ of $G$ on $\partial_\infty X$ by homeomorphisms.*

Lemma 9.90. *Suppose that $X$ satisfies the hypothesis of Lemma 9.86 and the quasi-action $G \curvearrowright X$ is properly discontinuous. Then the kernel for the action $\phi_\infty$ is finite.*

Proof. The kernel $K$ of $\phi_\infty$ consists of the elements $g \in G$ such that the distance from $\phi(g)$ to the identity is finite. Since $\phi(g)$ is an $(L, A)$-quasi-isometry of $X$, it follows from Lemma 9.86, that

$$\mathrm{dist}(\phi(g), id) \leqslant D(L, A, R, \delta).$$

Since $\phi$ was properly discontinuous, $K$ is finite. $\square$

**Conical limit points of quasi-actions.**

Suppose that $\phi$ is a quasi-action of a group $G$ on a Rips–hyperbolic space $X$. A point $\xi \in \partial_\infty X$ is called a *conical limit point* for the quasi-action $\phi$ if there exists a sequence $g_i \in G$ so that $\phi(g_i)(x)$ converges to $\xi$ in the conical topology. In other words, for some (equivalently every) geodesic ray $\gamma \subset X$ asymptotic to $\xi$, and some (equivalently every) point $x \in X$, there exists a constant $R < \infty$ so that:

- $\lim_{i\to\infty} \phi(g)(x) = \xi$.
- $d(\phi(g_i)(x), \gamma) \leqslant R$ for all $i$.

Lemma 9.91. *Suppose that $\psi : G \curvearrowright X$ is a cobounded quasi-action. Then every point of the ideal boundary $\partial_\infty X$ is a conical limit point for $\psi$.*

Proof. Let $\xi \in \partial_\infty X$ and let $x_i \in X$ be a sequence converging to $\xi$ in conical topology (e.g., we can take $x_i = \gamma(i)$, where $\gamma$ is a geodesic ray in $X$ asymptotic to $\xi$). Fix a point $x \in X$ and a ball $B = B_R(x)$ so that for every $x' \in X$ there exists $g \in G$ so that $d(x', \phi(g)(x)) \leqslant R$. Then, by coboundedness of the quasi-action $\psi$, there exists a sequence $g_i \in G$ so that

$$d(x_i, \phi(g_i)(x)) \leqslant R.$$

Thus, $\xi$ is a conical limit point of the quasi-action. $\square$

Corollary 9.92. *Suppose that $G$ is a group and $f : X \to G$ is a quasi-isometry, $G \curvearrowright G$ is isometric action by left multiplication. Let $\psi : G \curvearrowright X$ be the quasi-action, obtained by conjugating $G \curvearrowright G$ via $f$. Then every point of $\partial_\infty X$ is a conical limit point for the quasi-action $\psi$.*

Proof. The action $G \curvearrowright G$ by left multiplication is cobounded, hence, the conjugate quasi-action $\psi : G \curvearrowright X$ is also cobounded. $\square$

If $\phi_\infty$ is a topological action of a group $G$ on $\partial_\infty X$ which is obtained by extension of a quasi-action $\phi$ of $G$ on $X$, then we will say that *conical limit points* of the action $G \curvearrowright \partial_\infty X$ are the conical limit points for the quasi-action $G \curvearrowright X$.

## 9.11. Hyperbolic groups

We now come to the *raison d'être* for $\delta$-hyperbolic spaces, namely, *hyperbolic groups*.

Definition 9.93. A finitely-generated group $G$ is called *Gromov–hyperbolic* or *word-hyperbolic*, or simply *hyperbolic* if one of its Cayley graphs is hyperbolic.

Example 9.94. 1. Every finitely-generated free groups is hyperbolic: Taking Cayley graphs corresponding to a free generating set, we obtain a simplicial tree, which is 0-hyperbolic.

2. Finite groups are hyperbolic.

Many examples of hyperbolic groups can be constructed *via small cancelation* theory, see e.g. [**GS90, IS98**]. For instance, let $G$ be a 1-relator group with the presentation

$$\langle x_1, \dots, x_n | w^m \rangle ,$$

where $m \geqslant 2$ and $w$ is a cyclically reduced word in the generators $x_i$. Then $G$ is hyperbolic. (This was proven by B. B. Newman in [**New68**, Theorem 3] before the notion of hyperbolic groups was introduced; Newman proved that for such groups $G$ *Dehn's algorithm* applies, which is equivalent to hyperbolicity, see §9.13.)

Below is a combinatorial characterization of hyperbolic groups among Coxeter groups. Let $\Gamma$ be a finite Coxeter graph and $G = C_\Gamma$ be the corresponding Coxeter group. A *parabolic subgroup* of $\Gamma$ is the Coxeter subgroup defined by a subgraph $\Lambda$ of $\Gamma$. It is clear that every parabolic subgroup of $G$ admits a natural homomorphism to $G$; it turns out that such homomorphisms are always injective.

THEOREM 9.95 (G. Moussong [**Mou88**]). *A Coxeter group $G$ is Gromov–hyperbolic if and only if the following condition holds:*

*No parabolic subgroup of $G$ is virtually isomorphic to the direct product of two infinite groups.*

*In particular, a Coxeter group is hyperbolic if and only if it contains no free abelian subgroup of rank* 2.

PROBLEM 9.96. Is there a similar characterization of Gromov–hyperbolic groups among Shephard groups and generalized von Dyck groups?

Since changing generating set does not alter the quasi-isometry type of the Cayley graph and Rips–hyperbolicity is invariant under quasi-isometries (Corollary 9.39), we conclude that a group $G$ is hyperbolic if and only if all its Cayley graphs are hyperbolic. Furthermore, if groups $G, G'$ are quasi-isometric then $G$ is hyperbolic if and only if $G'$ is hyperbolic. In particular, if $G, G'$ are virtually isomorphic, then $G$ is hyperbolic if and only if $G'$ is hyperbolic. For instance, all virtually free groups are hyperbolic.

In view of Milnor-Schwarz lemma,

OBSERVATION 9.97. If $G$ is a group acting geometrically on a Rips–hyperbolic metric space, then $G$ is also hyperbolic.

DEFINITION 9.98. A group $G$ is called $CAT(\kappa)$ if it admits a geometric action on a $CAT(\kappa)$ space.

Thus, every $CAT(-1)$ group is hyperbolic. In particular, fundamental groups of compact Riemannian manifolds of negative curvature are hyperbolic.

The following is an outstanding open problem in geometric group theory:

OPEN PROBLEM 9.99. Construct a hyperbolic group $G$ which is not a $CAT(-1)$ or even a $CAT(0)$ group.

DEFINITION 9.100. A hyperbolic group is called *elementary* if it is virtually cyclic. A hyperbolic group is called non-elementary otherwise.

Here are some examples of non-hyperbolic groups:

1. $\mathbb{Z}^n$ is not hyperbolic for every $n \geqslant 2$. Indeed, $\mathbb{Z}^n$ is QI to $\mathbb{R}^n$ and $\mathbb{R}^n$ is not hyperbolic (see Example 9.9).

2. A deeper fact is that if a group $G$ contains a subgroup isomorphic to $\mathbb{Z}^2$ then $G$ is not hyperbolic, see e.g. [**BH99**].

3. More generally, if $G$ contains a solvable subgroup $S$ then $G$ is not hyperbolic unless $S$ is virtually cyclic.

4. Even more generally, for every subgroup $S$ of a hyperbolic group $G$, the group $S$ is either elementary hyperbolic or contains a nonabelian free subgroup. In particular, every amenable subgroup of a hyperbolic group is virtually cyclic. See e.g. [**BH99**].

5. Furthermore, if $Z \leqslant G$ is a central subgroup of a hyperbolic group, then either $Z$ is finite, or $G/Z$ is finite.

REMARK 9.101. There are hyperbolic groups which contain non-hyperbolic finitely-generated subgroups, see Theorem 9.142. A subgroup $H \leqslant G$ of a hyperbolic group $G$ is called *quasiconvex* if it is a quasiconvex subset of a Cayley graph of $G$. If $H \leqslant G$ is a quasiconvex subgroup , then, according to Theorem 9.44, $H$ is quasi-isometrically embedded in $G$ and, hence, is hyperbolic itself.

Examples of quasiconvex subgroups are given by finite subgroups (which is clear) and (less obviously) infinite cyclic subgroups. Let $G$ be a hyperbolic group with a word metric $d$. Define the *translation length* of $g \in G$ to be

$$\|g\| := \lim_{n\to\infty} \frac{d(g^n, e)}{n}.$$

It is clear that $\|g\| = 0$ if $g$ has finite order. On the other hand, every cyclic subgroup $\langle g\rangle \subset G$ is quasiconvex and $\|g\| > 0$ for every $g$ of infinite order, see Chapter III.Γ, Propositions 3.10, 3.15 of [**BH99**].

## 9.12. Ideal boundaries of hyperbolic groups

We define the *ideal boundary* $\partial_\infty G$ of a hyperbolic group $G$ as the ideal boundary of some (every) Cayley graph of $G$: It follows from Theorem 9.83, that boundaries of different Cayley graphs are equivariantly homeomorphic. Here are two simple examples of computation of the ideal boundary.

Since $\partial_\infty \mathbb{H}^n = S^{n-1}$, we conclude that for the fundamental group $G$ of a closed hyperbolic $n$-manifold, $\partial_\infty G \cong S^{n-1}$. Similarly, if $G = F_n$ is the free group of rank $n$, then free generating set $S$ of $G$ yields Cayley graph $X = \Gamma_{G,S}$ which is a simplicial tree of constant valence. Therefore, as we saw in Section 9.9, $\partial_\infty X$ is homeomorphic to the Cantor set. Thus, $\partial_\infty F_n$ is the Cantor set.

LEMMA 9.102. *Let $G$ be a hyperbolic group and $Z = \partial_\infty G$. Then $Z$ consists of 0, 2 or continuum of points, in which case it is perfect. In the first two cases $G$ is elementary, otherwise $G$ is non-elementary.*

PROOF. Let $X$ be a Cayley graph of $G$. If $G$ is finite, then $X$ is bounded and, hence $Z =$. Thus, we assume that $G$ is infinite. By Exercise 4.74, $X$ contains a complete geodesic $\gamma$, thus, $Z$ has at least two distinct points, the limit points of $\gamma$. If $\mathrm{dist}_{Haus}(\gamma, X) < \infty$, $X$ is quasi-isometric to $\mathbb{R}$ and, hence, $G$ is 2-ended. Therefore, $G$ is virtually cyclic by Part 3 of Theorem 6.8.

We assume, therefore, that $\mathrm{dist}_{Haus}(\gamma, X) = \infty$. Then there exists a sequence of vertices $x_n \in X$ so that $\lim \mathrm{dist}(x_n, \gamma) = \infty$. Let $y_n \in \gamma$ be a nearest vertex to $x_n$. Let $g_n \in G$ be such that $g_n(y_n) = e \in G$. Then applying $g_n$ to the union of geodesics $[x_n, y_n] \cup \gamma$ and taking limit as $n \to \infty$, we obtain a complete geodesic $\beta \subset X$ (the limit of a subsequence $g_n(\gamma)$) and a geodesic ray $\rho$ meeting $\beta$ at $e$, so that for every $x \in \rho$, $e$ is a nearest point on $\gamma$ to $x$. Therefore, $\rho(\infty)$ is a point different from $\gamma(\pm\infty)$, so $Z$ contains at least 3 distinct points. Let $p$ be a centroid of a corresponding ideal triangle. Then $G \cdot o$ is a 1-net in $X$ and, we are, therefore, in the situation described in Lemma 9.86. Let $K$ denote the kernel of the action $G \curvearrowright Z$. Then every $k \in K$ moves every point in $X$ by $\leqslant D(1, 0, 1, \delta)$, where $D$ is the function defined in Lemma 9.86. It follows that $K$ is a finite group. Since $G$ is infinite, $Z$ is also infinite.

Let $\xi \in Z$ and let $\rho$ be a ray asymptotic to $\xi$. Then, there exists a sequence $g_n \in G$ so that $g_n(e) = x_n \in \rho$. Let $\gamma \subset X$ be a complete geodesic asymptotic to points $\eta, \zeta$ different from $\xi$. We leave it to the reader to verify that either

$$\lim_n g_n(\eta) = \xi,$$

or

$$\lim_n g_n(\zeta) = \xi,$$

Since $Z$ is infinite, we can choose $\xi, \eta$ so that their images under the given sequence $g_n$ are not all equal to $\xi$. Thus, $\xi$ is an accumulation point of $Z$ and $Z$ is perfect. Since $Z$ is infinite, it follows that it has cardinality continuum. □

DEFINITION 9.103. Let $Z$ be a compact and $G \subset Homeo(Z)$ be a subgroup. The group $G$ is said to be a *convergence group* if $G$ acts properly discontinuously on $Trip(Z)$, where $Trip(Z)$ is the set of triples of distinct elements of $Z$. A convergence group $G$ is said to be a *uniform* if $Trip(Z)/G$ is compact.

THEOREM 9.104 (P. Tukia, [**Tuk94**]). *Suppose that $X$ is a proper $\delta$-hyperbolic geodesic metric space with the ideal boundary $Z = \partial_\infty X$ which consists of at least 3 points. Let $G \curvearrowright X$ be an isometric action and $G \curvearrowright Z$ be the corresponding topological action. Then the action $G \curvearrowright X$ is geometric if and only if $G \curvearrowright Z$ is a uniform convergence action.*

PROOF. Recall that we have a correspondence *center* : $Trip(Z) \to X$ sending each triple of distinct points in $Z$ to the set of centroids of the corresponding ideal triangles. Furthermore, by Corollary 9.73, for every $\xi \in Trip(Z)$,

$$\operatorname{diam}(center(\xi)) \leqslant 60\delta.$$

Clearly, the correspondence *center* is $G$-equivariant. Moreover, the image of every compact $K$ in $Trip(Z)$ under *center* is bounded (see Exercise 9.75).

Assume now that the action $G \curvearrowright X$ is geometric. Given a compact subset $K \subset Trip(Z)$, suppose that the set

$$G_K := \{g \in G | gK \cap K \neq \emptyset\}$$

is infinite. Then there exists a sequence $\xi_n \in K$ and an infinite sequence $g_n \in G$ so that $g_n(\xi_n) \in K$. Then the diameter of the set

$$E = \left( \bigcup_n center(\xi_n) \cup center(g_n(\xi_n)) \right) \subset X$$

is bounded and each $g_n$ sends some $p_n \in E$ to an element of $E$. This, however, contradicts proper discontinuity of the action of $G$ on $X$. Thus, the action $G \curvearrowright Trip(Z)$ is properly discontinuous.

Similarly, since $G \curvearrowright X$ is cobounded, the $G$-orbit of some metric ball $B(p, R)$ covers the entire $X$. Thus, using equivariance of *center*, for every $\xi \in Trip(Z)$, there exists $g \in G$ so that

$$center(g\xi) \subset B = B(x, R + 60\delta).$$

Since $center^{-1}(B)$ is relatively compact in $Trip(Z)$ (see Exercise 9.76), we conclude that $G$ acts cocompactly on $Trip(Z)$. Thus, $G \subset Homeo(Z)$ is a uniform convergence group.

The proof of the converse is essentially the same argument run in the reverse. Let $K \subset Trip(Z)$ be a compact, so that $G$-orbit of $K$ is the entire $Trip(Z)$. Then the set $center(K)$, which is the union of sets of centroids of points $\xi' \in K$, is a bounded subset $B \subset X$. Now, by equivariance of the correspondence *center*, it follows that $G$-orbit of $B$ is the entire $X$. Hence, $G \curvearrowright X$ is cobounded. The argument for proper discontinuity of the action $G \curvearrowright Trip(Z)$ is similar, we just use the fact that the preimage of a sufficiently large metric ball $B \subset X$ under the correspondence *center* is nonempty and relatively compact in $Trip(Z)$. Then proper discontinuity of the action $G \curvearrowright X$ follows from proper discontinuity of $G \curvearrowright Trip(Z)$. □

COROLLARY 9.105. *Every hyperbolic group $G$ acts by homeomorphisms on $\partial_\infty G$ as a uniform convergence group.*

The converse to Theorem 9.104 is a deep theorem of B. Bowditch [**Bow98**]:

THEOREM 9.106. *Let $Z$ be a perfect compact Hausdorff space consisting of more than one point. Suppose that $G \subset Homeo(Z)$ is a uniform convergence group. Then $G$ is hyperbolic and, moreover, there exists an equivariant homeomorphism $Z \to \partial_\infty G$.*

Note that in the proof of Part 1 of Theorem 9.104 we did not really need the property that the action of $G$ on itself was isometric, a geometric quasi-action (see Definition 5.59) suffices:

THEOREM 9.107. *Suppose that $X$ is a $\delta$-hyperbolic proper geodesic metric space. Assume that there exists $R$ so that every point in $X$ is within distance $\leqslant R$ from a centroid of an ideal triangle in $X$. Let $\phi : G \curvearrowright X$ be a geometric quasi-action. Then the extension $\phi_\infty : G \to Homeo(Z)$, $Z = \partial_\infty X$, of the quasi-action $\phi$ to a topological action of $G$ on $Z$ is a uniform convergence action.*

PROOF. The proof of this result closely follows the proof of Theorem 9.104; the only difference is that ideal triangles $T \subset X$ are not mapped to ideal triangles by quasi-isometries $\phi(g), g \in G$. However, ideal quasi-geodesic triangles $\phi(g)(T)$ are uniformly close to ideal triangles which suffices for the proof. □

## 9.13. Linear isoperimetric inequality and Dehn algorithm for hyperbolic groups

Let $G$ be a hyperbolic group, we suppose that $\Gamma$ is a $\delta$-hyperbolic Cayley graph of $G$. We will assume that $\delta \geqslant 2$ is a natural number. Recall that a *loop* in $\Gamma$ is required to be a closed edge-path. Since the group $G$ acts transitively on the vertices of $X$, the number of $G$-orbits of loops of length $\leqslant 10\delta$ in $\Gamma$ is bounded. We attach a 2-cell along every such loop. Let $X$ denote the resulting cell complex. Recall that for a loop $\gamma$ in $X$, $\ell(\gamma)$ is the length of $\gamma$ and $A(\gamma)$ is the least combinatorial area of a disk in $X$ bounding $\gamma$, see Section 4.9.

Our goal is to show that $X$ is simply-connected and satisfies a linear isoperimetric inequality. We will prove a somewhat stronger statement. Namely, suppose that $X$ is a connected 2-dimensional cell complex whose 1-skeleton $X^1$ (metrized to have unit edges) is $\delta$-hyperbolic (with $\delta \geqslant 2$ an integer) and so that for every loop $\gamma$ of length $\leqslant 10\delta$ in $X$, $A(\gamma) \leqslant K < \infty$. Then:

The following theorem was first proven by Gromov in Section 2.3 of [**Gro87**]:

THEOREM 9.108 (Hyperbolicity implies linear isoperimetric inequality). *Under the above assumptions, for every loop $\gamma \subset X$,*

$$A(\gamma) \leqslant K\ell(\gamma). \tag{9.8}$$

Since the argument in the proof of the theorem is by induction on the length of $\gamma$, the following result is the main tool.

PROPOSITION 9.109. *Every loop $\gamma$ in $X^{(1)}$ of length larger than $10\delta$ is a product of two loops, one of length $\leqslant 10\delta$ and another one of length $< \ell(\gamma)$.*

PROOF. We assume that $\gamma$ is parameterized by its arc-length, and that it has length $n$.

Without loss of generality we may also assume that $\delta > 2$.

*Case 1.* Assume that there exists a vertex $u = \gamma(t)$ such that the vertex $v = \gamma(t + 5\delta)$ satisfies $d(u, v) < 5\delta$. By a circular change of the parameterizations of $\gamma$ we may assume that $t = 0$. Let $p$ denote the geodesic $[v, u]$ in $X^{(1)}$. We then obtain two new loops

$$\gamma_1 = \gamma([0, 5\delta]) \cup p$$

and

$$\gamma_2 = (-p) \cup \gamma([5\delta, n]).$$

Here $-p$ is the geodesic $p$ with the reversed orientation. Since $\ell(p) < \ell(\gamma([0, 5\delta]))$, we have $\ell(\gamma_1) \leqslant 10\delta$ and $\ell(\gamma_2) < \ell(\gamma_1)$ .

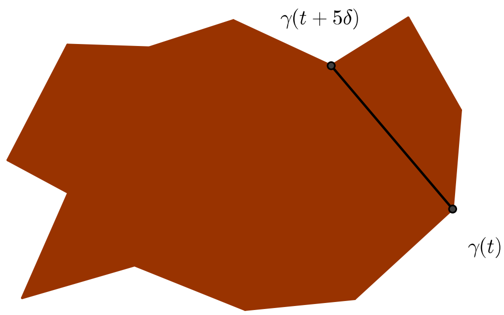



FIGURE 9.5. Case 1.

*Case 2.* Assume now that for every $t$, $d(\gamma(t), \gamma(t + 5\delta)) = 5\delta$, where $t + 5\delta$ is considered modulo $n$. In other words, every sub-arc of $\gamma$ of length $5\delta$ is a geodesic segment.

Let $v_0 = \gamma(0)$. Assume that $v = \gamma(t)$ is a vertex on $\gamma$ whose distance to $v_0$ is the largest possible, in particular it is at least $5\delta$.

Consider the triangles $\Delta_\pm$ with the vertices $v_0, v = \gamma(t), v_\pm = \gamma(t \pm 5\delta)$. Each triangle in $X^{(1)}$ is $\delta$-thin, therefore, $u_\pm = \gamma(t \pm (\delta+1))$ is within distance $\leqslant \delta$ of a vertex on one of the sides $[v_0, v]$, $[v_0, v_\pm]$. If, say, $u_+$ is within $\leqslant \delta$ of some $w \in [v_0, v_+]$, then

$$d(v_0, v) \leqslant r + \delta + (\delta + 1) = r + 2\delta + 1,$$

$$d(v_0, v_+) = r + s \geqslant r + 3\delta - 1 > r + 2\delta + 1$$

where $r = d(v_0, w), s = d(w, v_+)$. Hence, $d(v_0, v_+) > d(v_0, v)$ which contradicts our choice of $v$ as being farthest away from $v_0$. Therefore both $u_\pm$ are within distance $\leq \delta$ from *the same* point on the geodesic $[v_0, v]$ and, hence, $d(u_+, u_-) \leqslant 2\delta$. On the other hand, the distance between these vertices along the path $\gamma$ is $2\delta + 2$. This contradicts our working hypothesis that every sub-arc of $\gamma$ of length at most $5\delta$ is a geodesic segment.

We have thus obtained that Case 2 is impossible. □

*Proof of Theorem 9.108.*

The proof of the inequality (9.8) is by induction on the length of $\gamma$.

1. If $\ell(\gamma) \leqslant 10\delta$ then $A(\gamma) \leqslant K \leqslant K\ell(\gamma)$.

2. Suppose the inequality holds for $\ell(\gamma) \leqslant n$, $n \geqslant 10\delta$. If $\ell(\gamma) = n+1$, then $\gamma$ is the product of loops $\gamma', \gamma''$ as in Proposition 9.109: $\ell(\gamma') < \ell(\gamma)$, $\ell(\gamma'') \leqslant 10\delta$. Then, inductively,

$$A(\gamma') \leqslant K\ell(\gamma'), \quad A(\gamma'') \leqslant K,$$

and, thus,

$$A(\gamma) \leqslant A(\gamma') + A(\gamma'') \leqslant K\ell(\gamma') + K \leqslant K\ell(\gamma). \quad \square$$

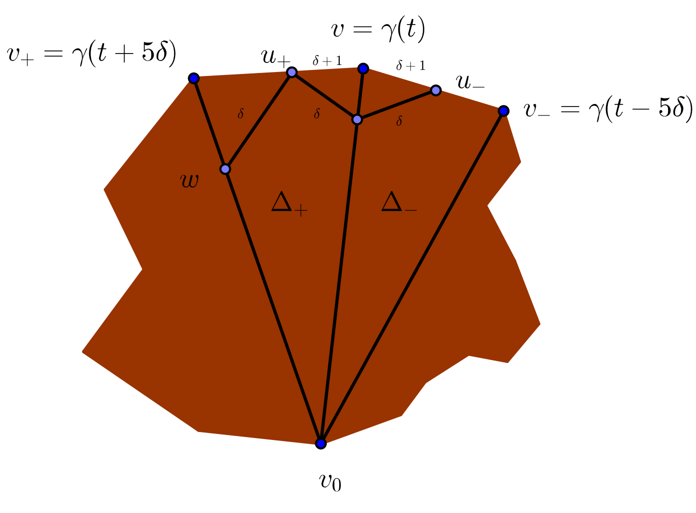



FIGURE 9.6. Case 2.

Below are two corollaries of Proposition 9.109, which was the key to the proof of the linear isoperimetric inequality.

COROLLARY 9.110 (M. Gromov, **[Gro87]**). *Every hyperbolic group is finitely-presented.*

PROOF. Proposition 9.109 means that every loop in the Cayley graph of $\Gamma$ is a product of loops of length $\leqslant 10\delta$. Attaching 2-cells to $\Gamma$ along the $G$-images of these loops we obtain a simply-connected complex $Y$ on which $G$ acts geometrically. Thus, $G$ is finitely-presented. $\square$

COROLLARY 9.111 (M. Gromov, **[Gro87]**, section 6.8N). *Let $Y$ be a coarsely connected Rips–hyperbolic metric space. Then $X$ satisfies linear isoperimetric inequality:*

$$Ar_\mu(\mathfrak{c}) \leqslant K\ell(\mathfrak{c})$$

*for all sufficiently large $\mu$ and for appropriate $K = K(\mu)$.*

PROOF. Quasi-isometry invariance of isoperimetric functions implies that it suffices to prove the assertion for $\Gamma$, the 1-skeleton of a connected $R$-Rips complex $Rips_R(X)$ of $X$. By Proposition 9.109, every loop $\gamma$ in $\Gamma$ is a product of $\leqslant \ell(\gamma)$ loops of length $\leqslant 10\delta$, where $\Gamma$ is $\delta$-hyperbolic in the sense of Rips. Therefore, for any $\mu \geqslant 10\delta$, we get

$$Ar_\mu(\gamma) \leqslant \ell(\gamma). \quad \square$$

**Dehn algorithm.** A (finite) presentation $\langle X|R\rangle$ is called *Dehn* if for every nontrivial word $w$ representing $1 \in G$, the word $w$ contains more than half of a relator. A word $w$ is called *Dehn-reduced* if it does not contain more than half of any relator. Given a word $w$, we can inductively reduce the length of $w$ by replacing subwords $u$ in $w$ with $u'$ so that $u'u^{-1}$ is a relator so that $|u'| < |u|$. This, of course, does not change the element $g$ of $G$ represented by $w$. Since the length of $w$ is decreasing on each step, eventually, we get a Dehn-reduced word $v$ representing $g \in G$. Since $\langle X|R\rangle$ is Dehn, either $v = 1$ (in which case $g = 1$) or $v \neq 1$ in which case $g \neq 1$. This algorithm is, probably, the simplest way to solve word problems in groups. It is also, historically, the oldest: Max Dehn introduced it in order to solve the word problem for hyperbolic surface groups.

Geometrically, Dehn reduction represents a based homotopy of the path in $X$ represented by the word $w$ (the base-point is $1 \in G$). Similarly, one defines *cyclic* Dehn reduction, where the reduction is applied to the (unbased) loop represented by $w$ and the *cyclically Dehn* presentation: If $w$ is a null-homotopic loop in $X$ then this loop contains a subarc which is more than half of a relator. Again, if $G$ admits a cyclically Dehn presentation then the word problem in $G$ is solvable.

LEMMA 9.112. *If $G$ is $\delta$-hyperbolic finitely-presented group then it admits a finite (cyclically) Dehn presentation.*

PROOF. Start with an arbitrary finite presentation of $G$. Then add to the list of relators all the words of length $\leqslant 10\delta$ representing the identity in $G$. Since the set of such words is finite, we obtain a new finite presentation of the group $G$. The fact that the new presentation is (cyclically) Dehn is just the induction step of the proof of Proposition 9.108. $\square$

Note, however, that the construction of a (cyclically) Dehn presentation requires solvability of the word problem for $G$ (or, rather, for the words of the length

$\leqslant 10\delta$) and, hence, is not *a priori* algorithmic. Nevertheless, the word problem in $\delta$-hyperbolic groups (with known $\delta$) is solvable as we will see below, and, hence, a Dehn presentation is algorithmically computable.

The converse of Proposition 9.108 is true as well, i.e. if a finitely-presented group satisfies a linear isoperimetric inequality then it is hyperbolic. We shall discuss this in Section 9.17.

## 9.14. Central co-extensions of hyperbolic groups and quasi-isometries

We now consider a central co-extension

$$1 \to A \to \tilde{G} \xrightarrow{r} G \to 1 \tag{9.9}$$

with $A$ a finitely-generated abelian group and $G$ hyperbolic.

THEOREM 9.113. *$\tilde{G}$ is QI to $A \times G$.*

PROOF. In the case when $A \cong \mathbb{Z}$, the first published proof belongs to S. Gersten **[Ger92]**, although, it appears that D.B.A. Epstein and G. Mess also knew this result. Our proof follows the one in **[NR97]**. First of all, since an epimorphism with finite kernel is a quasi-isometry, it suffices to consider the case when $A$ is free abelian of finite rank.

Our main goal is to construct a Lipschitz section (which is not a homomorphism!) $s : G \to \tilde{G}$ of the sequence (9.9). We first consider the case when $A \cong \mathbb{Z}$. Each fiber $r^{-1}(g), g \in G$, is a copy of $\mathbb{Z}$ and, therefore, has a natural order denoted $\leq$. We let $\iota$ denote the embedding $\mathbb{Z} \cong A \to \tilde{G}$. We let $\mathcal{X}$ denote a symmetric generating set of $\tilde{G}$ and use the same name for its image under $s$. We let $\langle \mathcal{X} | \mathcal{R} \rangle$ be a finite presentation of $G$. Let $|w|$ denote the word length with respect to this generating set, for $w \in \mathcal{X}^*$, where $\mathcal{X}^*$ is the set of all words in $\mathcal{X}$, as in Section 4.2. Lastly, let $\tilde{w}$ and $\bar{w}$ denote the elements of $\tilde{G}$ and $G$ represented by $w \in \mathcal{X}^*$.

LEMMA 9.114. *There exists $C \in \mathbb{N}$ so that for every $g \in G$ there exists*

$$r(g) := \max\{\tilde{w}\iota(-C|w|) : w \in X^*, \bar{w} = g\}.$$

*Here the maximum is taken with respect to the natural order on $s^{-1}(g)$.*

PROOF. We will use the fact that $G$ satisfies the linear isoperimetric inequality

$$Area(\alpha) \leqslant K|\alpha|$$

for every $\alpha \in \mathcal{X}^*$ representing the identity in $G$. We will assume that $K \in \mathbb{N}$. For each $R \in \mathcal{X}^*$ so that $R^{\pm 1}$ is a defining relator for $G$, the word $R$ represents some $\tilde{R} \in A$. Therefore, since $G$ is finitely-presented, we define a natural number $T$ so that

$$\iota(T) = \max\{\tilde{R} : R^{\pm 1} \text{ is a defining relator of } G\}.$$

We then claim that for each $u \in \mathcal{X}^*$ representing the identity in $G$,

$$\iota(T Area(u)) \geqslant \tilde{u} \in A. \tag{9.10}$$

Since general relators $u$ of $G$ are products of words of the form $hRh^{-1}$, $R \in \mathcal{R}$, (where $Area(u)$ is at most the number of these terms in the product) it suffices to verify that for $w = h^{-1}Rh$,

$$\tilde{w} \leqslant \iota(T)$$

where $R$ is a defining relator of $G$ and $h \in \mathcal{X}^*$. The latter inequality follows from the fact that the multiplications by $\bar{h}$ and $\bar{h}^{-1}$ determine an order isomorphism and its inverse between $r^{-1}(1)$ and $r^{-1}(\bar{h})$.

Set $C := TK$. We are now ready to prove lemma. Let $w, v$ be in $\mathcal{X}^*$ representing the same element $g \in G$. Set $u := v^{-1}$. Then $q = wu$ represents the identity and, hence, by (9.10),

$$\tilde{q} = \tilde{w}\tilde{u} \leqslant \iota(C|q|) = \iota(C|w|) + \iota(C|u|).$$

We now switch to the addition notation for $A \cong \mathbb{Z}$. Then,

$$w - v \leqslant \iota(C|w|) + \iota(C|v|),$$

and

$$w - \iota(C|w|) \leqslant v + \iota(C|v|).$$

Therefore, taking $v$ to be a fixed word representing $g$, we conclude that all the differences $w - \iota(C|w|)$ are bounded from above. Hence their maximum exists. $\square$

Consider the section $s$ (given by Lemma 9.114) of the exact sequence (9.9). A word $w = w_g$ realizing the maximum in the definition of $s$ is called *maximizing*. The section $s$, of course, need not be a group homomorphism. We will see nevertheless that it is not far from being one. Define the cocycle

$$\sigma(g_1, g_2) := s(g_1)s(g_2) - s(g_1 g_2)$$

where the difference is taking place in $r^{-1}(g_1 g_2)$. The next lemma does not use hyperbolicity of $G$, only the definition of $s$.

LEMMA 9.115. *The set $\sigma(G, X)$ is finite.*

PROOF. Let $x \in \mathcal{X}$, $g \in G$. We have to estimate the difference

$$s(g)x - s(gx).$$

Let $w_1$ and $w_2$ denote maximizing words for $g$ and $gx$ respectively. Note that the word $w_1 x$ also represents $gx$. Therefore, by the definition of $s$,

$$\widetilde{w_1 x}\iota(-C(|w_1| + 1)) \leqslant \tilde{w}_2\iota(-C|w_2|).$$

Hence, there exists $a \in A, a \geqslant 0$, so that

$$\widetilde{w_1}\iota(-C(|w_1|)\widetilde{x}\iota(-C)a = \tilde{w}_2\iota(-C|w_2|)$$

and, thus

$$s(g)\widetilde{x}\iota(-C)a = s(gx). \tag{9.11}$$

Since $w_2 x^{-1}$ represents $g$, we similarly obtain

$$s(gx)\widetilde{x}^{-1}\iota(-C)b = s(g), b \geqslant 0, b \in A. \tag{9.12}$$

By combining equations (9.11) and (9.12) and switching to the additive notation for the group operation in $A$ we get

$$a + b = \iota(2C).$$

Since $a \geqslant 0, b \geqslant 0$, we conclude that $-\iota(C) \leqslant a - \iota(C) \leq \iota(C)$. Therefore, (9.11) implies that

$$|s(g)x - s(gx)| \leqslant C.$$

Since the finite interval $[-\iota(C), \iota(C)]$ in $A$ is a finite set, lemma follows. $\square$

Remark 9.116. Actually, more is true: There exists a section $s' : G \to \tilde{G}$ so that $\sigma'(G, G)$ is a finite set. This follows from the fact that all (degree $\geqslant 2$) cohomology classes of hyperbolic groups are *bounded* (see [**Min01**]). However, the proof is more difficult and we will not need this fact.

Letting $L$ denote the maximum of the word lengths (with respect to the generating set $\mathcal{X}$) of the elements in the sets $\sigma(G, \mathcal{X}), \sigma(\mathcal{X}, G)$, we conclude that the map $s : G \to \tilde{G}$ is $(L+1)$–Lipschitz. Given the section $s : G \to \tilde{G}$, we define the projection $\phi = \phi_s : \tilde{G} \to A$ by

$$\phi(\tilde{g}) = \tilde{g} - s \circ r(\tilde{g}). \tag{9.13}$$

It is immediate that $\phi$ is Lipschitz since $s$ is Lipschitz.

We now extend the above construction to the case of central co-extensions with free abelian kernel of finite rank. Let $A = \prod_{i=1}^{n} A_i, A_i \cong \mathbb{Z}$. Consider a central co-extension (9.9). The homomorphisms $A \to A_i$ induce quotient maps $\eta_i : \tilde{G} \to \tilde{G}_i$ with the kernels $\prod_{j \neq i} A_j$. Each $\tilde{G}_i$, in turn, is a central co-extension

$$1 \to A_i \to \tilde{G}_i \xrightarrow{r_i} G \to 1. \tag{9.14}$$

Assuming that each central co-extension (9.14) has a Lipschitz section $s_i$, we obtain the corresponding Lipschitz projection $\phi_i : \tilde{G}_i \to A_i$ given by (9.13). This yields a Lipschitz projection

$$\Phi : \tilde{G} \to A, \Phi = (\phi_1 \circ \eta_1, ...., \phi_n \circ \eta_n).$$

We now set

$$s(r(\tilde{g})) := \tilde{g} - \Phi(\tilde{g}).$$

It is straightforward to verify that $s$ is well-defined and that it is Lipschitz provided that each $s_i$ is. We thus obtain

Corollary 9.117. *Given a finitely-generated free abelian group $A$ and a hyperbolic group $G$, each central co-extension (9.9) admits a Lipschitz section $s : G \to \tilde{G}$ and a Lipschitz projection $\Phi : \tilde{G} \to A$ given by*

$$\Phi(\tilde{g}) = \tilde{g} - s(r(\tilde{g})).$$

We then define the map

$$h : G \times A \to \tilde{G}, \quad h(g, a) = s(g) + \iota(a)$$

and its inverse

$$h^{-1} : \tilde{G} \to G \times A, \quad \hat{h}(\tilde{g}) = (r(\tilde{g}), \Phi(\tilde{g})).$$

Since homomorphisms are 1-Lipschitz while the maps $r$ and $\Phi$ are Lipschitz, we conclude that $h$ is a bi-Lipschitz quasi-isometry. □

Remark 9.118. The above proof easily generalizes to the case of an arbitrary finitely-generated group $G$ and a central co-extension (9.9) given by a *bounded 2-nd cohomology class* (see e.g. [**Bro81b, Gro82, EF97a**] for the definition): One has to observe only that each cyclic central co-extension

$$1 \to A_i \to \tilde{G}_i \to G \to 1$$

is still given by a bounded cohomology class. We refer the reader to [**Ger92**] for the details.

Example 9.119. Let $G = \mathbb{Z}^2$, $A = \mathbb{Z}$. Since $H^2(G, \mathbb{Z}) = H^2(T^2, \mathbb{Z}) \cong \mathbb{Z}$, the group $G$ admits nontrivial central co-extensions with the kernel $A$, for instance, the integer Heisenberg group $H_3$. The group $\tilde{G}$ for such an co-extension is nilpotent but not virtually abelian. Hence, by Pansu's theorem (Theorem 14.33), it is not quasi-isometric to $G \times A = \mathbb{Z}^3$.

One can ask if Theorem 9.113 generalizes to other normal co-extensions of hyperbolic groups $G$. We note that Theorem 9.113 does not extend, say, to the case where $A$ is a non-elementary hyperbolic group and the action $G \curvearrowright A$ is trivial. The reason is the *quasi-isometric rigidity* for products of certain types of groups proven in [**KKL98**]. A special case of this theorem says that if $G_1, ..., G_n$ are non-elementary hyperbolic groups, then quasi-isometries of the product $G = G_1 \times ... \times G_n$ quasi-preserve the product structure:

Theorem 9.120. *Let $\pi_j : G \to G_j, j = 1, ..., n$ be natural projections. Then for each $(L, A)$–quasi-isometry $f : G \to G$, there is $C = C(G, L, A) < \infty$, so that, up to a composition with a permutation of quasi-isometric factors $G_k$, the map $f$ is within distance $\leqslant C$ from a product map $f_1 \times ... \times f_n$, where each $f_i : G_i \to G_i$ is a quasi-isometry.*

### 9.15. Characterization of hyperbolicity using asymptotic cones

The goal of this section is to strengthen the relation between hyperbolicity of geodesic metric spaces and 0-hyperbolicity of their asymptotic cones.

Proposition 9.121 (§2.$A$, [**Gro93**]). *Let $(X, \mathrm{dist})$ be a geodesic metric space. Assume that either of the following two conditions holds:*

(a) *There exists a non-principal ultrafilter $\omega$ such that for all sequences $\boldsymbol{e} = (e_n)_{n\in\mathbb{N}}$ of base-points $e_n \in X$ and $\boldsymbol{\lambda} = (\lambda_n)_{n\in\mathbb{N}}$ of scaling constants with $\omega$-$\lim \lambda_n = 0$, the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is a real tree.*

(b) *For every non-principal ultrafilter $\omega$ and every sequence $\boldsymbol{e} = (e_n)_{n\in\mathbb{N}}$ of base-points, the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, (n))$ is a real tree.*

*Then $(X, \mathrm{dist})$ is hyperbolic.*

The proof of Proposition 9.121 relies on the following lemma.

Lemma 9.122. *Assume that a geodesic metric space $(X, \mathrm{dist})$ satisfies either property (a) or property (b) in Proposition 9.121. Then there exists $M > 0$ such that for every geodesic triangle $\Delta(x, y, z)$ with $\mathrm{dist}(y, z) \geqslant 1$ ,the two edges with endpoint $x$ are at Hausdorff distance at most $M\mathrm{dist}(y, z)$.*

Proof. Suppose to the contrary that there exist sequences of triples of points $x_n$, $y_n$, $z_n$, such that $\mathrm{dist}(y_n, z_n) \geqslant 1$ and

$$\mathrm{dist}_{Haus}([x_n, y_n], [x_n, z_n]) = M_n \mathrm{dist}(y_n, z_n),$$

such that $M_n \to \infty$. Let $a_n$ be a point on $[x_n, y_n]$ such that

$$\delta_n := \mathrm{dist}(a_n, [x_n, z_n]) = \mathrm{dist}_{Haus}([x_n, y_n], [x_n, z_n]).$$

Since $\delta_n \geqslant M_n$, it follows that $\delta_n \to \infty$.

Suppose condition **(a)** holds. Consider the sequence of base-points $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}}$ and the sequence of scaling constants $\boldsymbol{\delta}' = (1/\delta_n)_{n\in\mathbb{N}}$. In the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{a}, \boldsymbol{\delta}')$, the limits of $[x_n, y_n]$ and $[x_n, z_n]$ are at Hausdorff distance 1.

The triangle inequalities imply that the limits

$$\omega\text{-lim}\,\frac{\mathrm{dist}(y_n,a_n)}{\delta_n}\ \text{and}\ \omega\text{-lim}\,\frac{\mathrm{dist}(z_n,a_n)}{\delta_n}$$

are either both finite or both infinite. It follows that the limits of $[x_n, y_n]$ and $[x_n, z_n]$ are either two distinct geodesics joining the points $x_\omega = (x_n)$ and the point $y_\omega = (y_n) = z_\omega(z_n)$, or two distinct asymptotic rays with common origin, or two distinct geodesics asymptotic on both sides. All these cases are impossible in a real tree.

Suppose condition **(b)** holds. Let $\mathcal{S} = \{\lfloor \delta_n \rfloor \;;\; n \in \mathbb{N}\}$, where $\lfloor \delta_n \rfloor$ is the integer part of $\delta_n$. By Exercise 7.15, there exists $\omega$ such that $\omega(\mathcal{S}) = 1$. Consider $(x'_m)$, $(y'_m)$, $(z'_m)$ and $(a'_m)$ defined as follows. For every $m$ in the set $\mathcal{S}$ choose an $n \in \mathbb{N}$ with $\lfloor \delta_n \rfloor = m$ and set $(x'_m, y'_m, z'_m, a'_m) = (x_n, y_n, z_n, a_n)$. For $m$ not in $\mathcal{S}$ make an arbitrary choice for the entries of all four sequences.

In $\mathrm{Cone}_\omega(X, \boldsymbol{a}', (m))$ the limits $\omega\text{-lim}[x'_m, y'_m]$ and $\omega\text{-lim}[x'_m, z'_m]$ are as in one of the three cases discussed in the previous case, all cases being forbidden in a real tree. □

Proof of Proposition 9.121. Suppose that the geodesic space $X$ is not hyperbolic. For every triangle $\Delta$ in $X$ and a point $a \in \Delta$ we define the quantity $d_\Delta(a)$, which is the distance from $a$ to union of the two sides of $\Delta$ which do not contain $a$ (if $a$ lies on all three sides then we set $\epsilon(a) = 0$). Then for every $n \in \mathbb{N}$ there exists a geodesic triangle $\Delta_n = \Delta(x_n,\, y_n,\, z_n)$, and a point $a_n$ on the edge $[x_n, y_n]$ such that

$$d_n = d_{\Delta_n}(a_n) \geqslant n.$$

For each $\Delta_n$ we then will choose the point $a_n$ in $\Delta_n$ which maximizes the function $d_{\Delta_n}$. After relabelling the vertices, we may assume that $a_n \in [x_n, z_n]$ and that $d_n = \mathrm{dist}(a_n, [y_n, z_n]) = \mathrm{dist}(a_n, b_n)$, where $b_n \in [y_n, z_n]$. Let $\delta_n$ be equal to $\mathrm{dist}(a_n, [x_n, z_n]) = \mathrm{dist}(a_n, c_n)$, for some $c_n \in [x_n, z_n]$. By hypothesis $\delta_n \geqslant d_n$ .

Suppose condition **(a)** is satisfied. In the asymptotic cone $\mathbf{K} = \mathrm{Cone}_\omega(X, \boldsymbol{a}, \boldsymbol{\lambda})$ , where $\boldsymbol{a} = (a_n)$ and $\boldsymbol{\lambda} = (1/d_n)$ we look at the limit of $\Delta_n$. There are two cases:

A) $\omega\text{-lim}\,\frac{\delta_n}{d_n} < +\infty$.

By Lemma 9.122, we have that $\mathrm{dist}_{Haus}([a_n, x_n], [c_n, x_n]) \leqslant M \cdot \delta_n$. Therefore the limits of $[a_n, x_n]$ and $[c_n, x_n]$ are either two geodesic segments with a common endpoint or two asymptotic rays. The same is true of the pairs of segments $[a_n, y_n]$, $[b_n, y_n]$ and $[b_n, z_n], [c_n, z_n]$, respectively. It follows that the limit $\omega\text{-lim}\,\Delta_n$ is a geodesic triangle $\Delta$ with vertices $x, y, z \in \mathbf{K} \cup \partial_\infty \mathbf{K}$. The point $a = \omega\text{-lim}\, a_n \in [x, y]$ is such that $\mathrm{dist}(a, [x, z] \cup [y, z]) \geqslant 1$, which implies that $\Delta$ is not a tripod. This contradicts the fact that $\mathbf{K}$ is a real tree.

B) $\omega\text{-lim}\,\frac{\delta_n}{d_n} = +\infty$.

This also implies that

$$\omega\text{-lim}\,\frac{\mathrm{dist}(a_n,x_n)}{d_n} = +\infty \ \text{and}\ \omega\text{-lim}\,\frac{\mathrm{dist}(a_n,z_n)}{d_n} = +\infty.$$

By Lemma 9.122, we have $\mathrm{dist}_{Haus}([a_n, y_n], [b_n, y_n]) \leqslant M \cdot d_n$. Thus, the respective limits of the sequences of segments $[x_n, y_n]$ and $[y_n, z_n]$ are either two rays of origin $y = \omega\text{-lim}\, y_n$ or two complete geodesics asymptotic on one side. We denote them $\overline{xy}$ and $\overline{yz}$, respectively, with $y \in \mathbf{K} \cup \partial_\infty \mathbf{K}, x, z \in \partial_\infty \mathbf{K}$. The limit of $[x_n, z_n]$ is empty (it is "out of sight").

The choice of $a_n$ implies that any point of $[b_n, z_n]$ must be at a distance at most $d_n$ from $[x_n, y_n] \cup [x_n, z_n]$. This implies that all points on the ray $\overline{bz}$ are at distance at most 1 from $\overline{xy}$. It follows that $\overline{xy}$ and $\overline{yz}$ are either asymptotic rays emanating from $y$ or complete geodesics asymptotic on both sides, and they are at Hausdorff distance 1. We again obtain a contradiction with the fact that $\mathbf{K}$ is a real tree.

We conclude that the condition in **(a)** implies that $X$ is $\delta$-hyperbolic, for some $\delta > 0$.

Suppose the condition **(b)** holds. Let $\mathcal{S} = \{\lfloor d_n \rfloor \; ; \; n \in \mathbb{N}\}$, and let $\omega$ be a non-principal ultrafilter such that $\omega(\mathcal{S}) = 1$ (see Exercise 7.15). We consider a sequence $(\Delta'_m)$ of geodesic triangles and a sequence $(a'_m)$ of points on these triangles with the property that whenever $m \in \mathcal{S}$, $\Delta'_m = \Delta_n$ and $a'_m = a_n$, for some $n$ such that $\lfloor d_n \rfloor = m$.

In the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{a}', (m))$, with $\boldsymbol{a}' = (a'_m)$ we may consider the limit of the triangles $(\Delta'_m)$, argue as previously, and obtain a contradiction to the fact that the cone is a real tree. It follows that the condition **(b)** also implies the hyperbolicity of $X$. □

REMARK 9.123. An immediate consequence of Proposition 9.121 is an alternative proof of the quasi-isometric invariance of Rips-hyperbolicity among geodesic metric spaces: A quasi-isometry between two spaces induces a bi-Lipschitz map between asymptotic cones, and a topological space bi-Lipschitz equivalent to a real tree is a real tree.

As a special case, consider Proposition 9.121 in the context of hyperbolic groups: A finitely-generated group is hyperbolic if and only if every asymptotic cone of $G$ is a real tree. A finitely-generated group $G$ is called *lacunary-hyperbolic* if at least one asymptotic cone of $G$ is a tree. Theory of such groups is developed in [**OOS09**], where many examples of non-hyperbolic lacunary hyperbolic groups are constructed. Thus, having one tree as an asymptotic cone is not enough to guarantee hyperbolicity of a finitely-generated group. On the other hand:

THEOREM 9.124 (M.Kapovich, B.Kleiner [**OOS09**]). *Let $G$ be a finitely–presented group. Then $G$ is hyperbolic if and only if one asymptotic cone of $G$ is a tree.*

PROOF. Below we present a of this theorem which we owe to Thomas Delzant. We will need the following

THEOREM 9.125 (B. Bowditch, [**Bow91**], Theorem 8.1.2). *For every $\delta$ there exists $\delta'$ so that for every $m$ there exists $R$ for which the following holds. If $Y$ be an $m$-locally simply-connected $R$-locally $\delta$-hyperbolic geodesic metric space, then $Y$ is $\delta'$-hyperbolic.*

Here, a space $Y$ is *$R$-locally $\delta$-hyperbolic* if every $R$-ball with the path-metric induced from $Y$ is $\delta$-hyperbolic. Instead of defining $m$-locally simply-connected spaces, we note that every simply-connected simplicial complex where each cell is isometric to a Euclidean simplex, satisfies this condition for every $m > 0$. We refer to [**Bow91**, Section 8.1] for the precise definition. We will be applying this theorem in the case when $\delta = 1$, $m = 1$ and let $\delta'$ and $R$ denote the resulting constants.

We now proceed with the proof suggested to us by Thomas Delzant. Suppose that $G$ is a finitely-presented group, so that one of its asymptotic cones is a tree. Let $X$ be a simply-connected simplicial complex on which $G$ acts freely, simplicially and cocompactly. We equip $X$ with the standard path-metric dist. Then $(X, \mathrm{dist})$ is quasi-isometric to $G$. Suppose that $\omega$ is an ultrafilter, $(\lambda_n)$ is a scaling sequence converging to zero, and $X_\omega$ is the asymptotic cone of $X$ with respect to this sequence, so that $X_\omega$ is isometric to a tree. Consider the sequence of metric spaces $X_n = (X, \lambda_n \mathrm{dist})$. Then, since $X_\omega$ is a tree, by taking a diagonal sequence, there exists a pair of sequences $r_n, \delta_n$ with

$$\omega\text{-}\lim r_n = \infty, \quad \omega\text{-}\lim \delta_n = 0$$

so that for $\omega$-all $n$, the every $r_n$-ball in $X_n$ is $\delta_n$-hyperbolic. In particular, for for $\omega$-all $n$, every $R$-ball in $X_n$ is 1-hyperbolic. Therefore, by Theorem 9.125, the space $X_n$ is $\delta'$-hyperbolic for $\omega$-all $n$. Since $X_n$ is a rescaled copy of $X$, it follows that $X$ (and, hence, $G$) is Gromov-hyperbolic as well. □

We now continue discussion of properties of trees which appear as asymptotic cones of hyperbolic spaces.

PROPOSITION 9.126. *Let $X$ be a geodesic hyperbolic space which admits a geometric action of a group $G$. Then all the asymptotic cones of $X$ are real trees where every point is a branch-points with valence continuum.*

PROOF. STEP 1. By Theorem 5.29, the group $G$ is finitely generated and hyperbolic and every Cayley graph $\Gamma$ of $G$ is quasi-isometric to $X$. It follows that there exists a bi-Lipschitz bijection between asymptotic cones

$$\Phi : \mathrm{Cone}_\omega(G, \mathbf{1}, \boldsymbol{\lambda}) \to \mathrm{Cone}_\omega(X, \boldsymbol{x}, \boldsymbol{\lambda}),$$

where $x$ is an arbitrary base-point in $X$, and $\mathbf{1}, \boldsymbol{x}$ denote the constant sequences equal to $1 \in G$, respectively to $x \in X$. Moreover, $\Phi(\mathbf{1}_\omega) = \boldsymbol{x}_\omega$. The map $\Phi$ thus determines a bijection between the space of directions $\Sigma_{\mathbf{1}_\omega}$ in the cone of $\Gamma$ and the space of directions $\Sigma_{\boldsymbol{x}_\omega}$ in the cone of $X$. It suffices therefore to prove that the set $\Sigma_{\mathbf{1}_\omega}$ has the cardinality of continuum. For simplicity, in what follows we denote the asymptotic cone $\mathrm{Cone}_\omega(G, \mathbf{1}, \boldsymbol{\lambda})$ by $G_\omega$.

STEP 2. We show that the geodesic rays joining 1 to distinct points of $\partial_\infty \Gamma$ give distinct directions in $\mathbf{1}_\omega$ in the asymptotic cone.

Let $\rho_i : [0, \infty) \to \Gamma, i = 1, 2$ be geodesic rays, $\rho_i(0) = 1, i \in \{1, 2\}, \rho_1(\infty) = \alpha, \rho_2(\infty) = \beta$, where $\alpha \neq \beta$. For every $t$ and $s$ in $[0, \infty)$, we consider

$$a_t = \omega\text{-}\lim \rho_1(t/\lambda_n) \text{ and } b_s = \omega\text{-}\lim \rho_2(s/\lambda_n),\ a_t,\ b_s \in \Gamma_\omega\,.$$

We have

$$\mathrm{dist}(a_t, b_s) = \omega\text{-}\lim \lambda_n \mathrm{dist}(\rho_1(t/\lambda_n), \rho_2(s/\lambda_n)) =$$

$$\omega\text{-}\lim\, [t + s - 2\lambda_n(\rho_1(t/\lambda_n), \rho_2(s/\lambda_n))_1] = t + s,$$

because the sequence of Gromov products

$$(\rho_1(t/\lambda_n), \rho_2(s/\lambda_n))_1$$

$\omega$-converges to a constant. The two limit rays, $\rho_1^\omega$ and $\rho_2^\omega$, of the rays $\rho_1$ and $\rho_2$, defined by $\rho_1^\omega(t) = a_t$, $\rho_2^\omega(s) = b_s$, have only the origin in common and give therefore distinct directions in $\mathbf{1}_\omega$.

We thus have found an injective map from $\partial_\infty \Gamma$ to $\mathbf{1}_\omega$

STEP 3. We argue that every direction of $\Gamma_\omega$ in $\mathbf{1}_\omega$ is determined by a sequence of geodesic rays emanating from 1 in $\Gamma$. This argument was suggested to us by P. Papasoglu.

An arbitrary direction of $\Gamma_\omega$ in $\mathbf{1}_\omega$ is the germ of a geodesic segment with one endpoint in $\mathbf{1}_\omega$, and this segment is the limit set of a sequence of geodesic segments of $\Gamma$ with one endpoint in 1, with lengths growing linearly in $\frac{1}{\lambda_n}$.

LEMMA 9.127. *Every sufficiently long geodesic segment in a Cayley graph of a hyperbolic group is contained in the $M$-neighborhood of a geodesic ray, where $M$ depends only on the Cayley graph.*

PROOF. According to [**Can84**] and to [**ECH$^+$92**, Chapter 3, §2], given a Cayley graph $\Gamma$ of a hyperbolic group $G$, there exists a finite directed graph $\mathcal{G}$ with edges labeled by the generators of $G$ such that every geodesic segment in $\Gamma$ corresponds to a path in $\mathcal{G}$. If a geodesic segment is long enough, the corresponding path contains at least one loop in $\mathcal{G}$. The distance from the endpoint of the path to the last loop is bounded by a constant $M$ which depends only on the graph $\mathcal{G}$. Let $\rho$ be the geodesic ray obtained by going around this loop infinitely many times. The initial segment is contained in $\mathcal{N}_M(\rho)$. □

We conclude that every direction of $\Gamma_\omega$ in $\mathbf{1}_\omega$ is the germ of a limit ray. We then have a surjective map from the set of sequences in $\partial_\infty G$ to $\Sigma_{[\mathbf{1}_\omega]}$:

$$\{(\alpha_n)_{n\in\mathbb{N}} \,;\, \alpha_n \in \partial_\infty \Gamma\} = (\partial_\infty \Gamma)^{\mathbb{N}} \to \Sigma_{[\mathbf{1}_\omega]}.$$

Steps 2 and 3 imply that for a non-elementary hyperbolic group, the cardinality of $\Sigma_{[\mathbf{1}_\omega]}$ is continuum, . □

A. Dyubina–Erschler and I. Polterovich ([**DP01**], [**DP98**]) have shown a stronger result than Proposition 9.126:

THEOREM 9.128 ([**DP01**], [**DP98**]). *Let $\mathcal{A}$ be the $2^{\aleph_0}$–universal tree, as defined in Theorem 9.20.*

(a) *Every asymptotic cone of a non-elementary hyperbolic group is isometric to $\mathcal{A}$.*
(b) *Every asymptotic cone of a complete, simply connected Riemannian manifold with sectional curvature at most $-k$, $k > 0$ a fixed constant, is isometric to $\mathcal{A}$.*

A consequence of Theorem 9.128 is that asymptotic cones of non-elementary hyperbolic groups and of complete, simply connected Riemannian manifold with strictly negative sectional curvature cannot be distinguished from one another.

## 9.16. Size of loops

The characterization of hyperbolicity with asymptotic cones allows one to define hyperbolicity in terms of size of its closed loops, in particular of the size of its geodesic triangles. Throughout this section $X$ denotes a geodesic metric space.

One parameter that measures the size of geodesic triangles is the minimal size introduced in Definition 5.49 for topological triangles. Only now, the three arcs that we consider are the three geodesic edges of the triangles. With this we can define the *minsize* function of a geodesic metric space $X$:

DEFINITION 9.129. The *m*inimal size function,

$$\text{minsize} = \text{minsize}_X : \mathbb{R}_+^* \to \mathbb{R}_+^*,$$

$$\text{minsize}(\ell) = \sup\{\text{minsize}(\Delta) \; ; \; \Delta \text{ a geodesic triangle of perimeter } \leqslant \ell\}\, .$$

Note that according to (9.1), if $X$ is $\delta$–hyperbolic in the sense of Rips, the function minsize is bounded by $2\delta$. We will see below that the "converse" is also true, i.e. when the function minsize is bounded, the space $X$ is hyperbolic. Moreover, M. Gromov proved [**Gro87**, §6] that a sublinear growth of minsize is enough to conclude that a space is hyperbolic. With the characterization of hyperbolicity using asymptotic cones, the proof of this statement is straightforward:

PROPOSITION 9.130. *A geodesic metric space $X$ is hyperbolic if and only if* $\text{minsize}(\ell) = o(\ell)$.

PROOF. As noted above, the direct part follows from Lemma 9.51. Conversely, assume that $\text{minsize}(\ell) = o(\ell)$. We begin by proving that in an arbitrary asymptotic cone of $X$ every finite geodesic is a limit geodesic, in the sense of Definition 7.48. More precisely:

LEMMA 9.131. *Let $\mathfrak{g} = [a_\pm, b_\omega]$ be a finite geodesic in* $\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ *and assume that $a_\omega = (a_i), b_\omega = (b_i)$. Then for every geodesic $[a_i, b_i] \subset X$ connecting $a_i$ to $b_i$, $\omega$-lim$[a_i, b_i] = \mathfrak{g}$.*

PROOF. Let $c_\omega = (c_i)$ be an arbitrary point on $\mathfrak{g}$. Consider an arbitrary triangle $\Delta_i \subset X$ with vertices $a_i, b_c, c_i$. Let $\ell_i$ be the perimeter of $\Delta_i$. Since $\omega\text{-lim}\, \lambda_i \ell_i < \infty$ and $minsize(\Delta_i) = o(\ell_i)$, we get

$$\omega\text{-lim}\, \lambda_i minsize(\Delta_i) = 0.$$

Taking the points $x_i, y_i, z_i$ on the sides of $\Delta_i$ realizing the minsize of $\Delta_i$, we conclude:

$$\omega\text{-lim}\, \lambda_i \,\text{diam}(x_i, y_i, z_i) = 0.$$

Let $\{x_\omega\} = \omega\text{-lim}\{x_i, y_i, z_i\}$ . Then

$$\text{dist}(a_\omega, b_\omega) \leqslant \text{dist}(a_\omega, x_\omega) + \text{dist}(x_\omega, b_\omega) \leqslant$$

$$\text{dist}(a_\omega, x_\omega) + \text{dist}(x_\omega, b_\omega) + 2\text{dist}(x_\omega, c_\omega) = \text{dist}(a_\omega, c_\omega) + \text{dist}(c_\omega, b_\omega)\, .$$

The first and the last term in the above sequence of inequalities are equal, hence all inequalities become equalities, in particular $c_\omega = x_\omega$ . Thus $c_\omega \in \omega\text{-lim}[a_i, b_i]$ and lemma follows. □

If one asymptotic cone $\text{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is not a real tree then it contains a geodesic triangle $\Delta$ which is not a tripod. Without loss of generality we may assume that the geodesic triangle is a simple loop. By the above lemma, the geodesic triangle is an ultralimit of a family of geodesic triangles $(\Delta_i)_{i \in I}$ with perimeters of the order $O\left(\frac{1}{\lambda_i}\right)$. The fact that $\text{minsize}(\Delta_i) = o\left(\frac{1}{\lambda_i}\right)$ implies that the three edges of $\Delta$ have a common point, a contradiction. □

M. Gromov in [**Gro87**, Proposition 6.6.F] proved the following version of Proposition 9.130:

THEOREM 9.132. *There exists a universal constant* $\varepsilon_0 > 0$ *such that if in a geodesic metric space* $X$ *all geodesic triangles with length* $\geqslant L_0$*, for some* $L_0$*, have*

$$\mathrm{minsize}(\Delta) \leqslant \varepsilon_0 \cdot \mathrm{perimeter}(\Delta)\,,$$

*then* $X$ *is hyperbolic.*

Another way of measuring the size of loops in a space $X$ is through their *constriction* function. We define the constriction function only for simple loops in $X$ primarily for the notational convenience, the definition and the results generalize without difficulty if one considers non-simple loops.

Let $\lambda \in \left(0, \frac{1}{2}\right)$. For a simple Lipschitz loop $c : \mathbb{S}^1 \to X$ of length $\ell$, we define the $\lambda$–*constriction of the loop* $c$ as $\mathrm{constr}_\lambda(c)$, which is the infimum of $d(x, y)$, where the infimum is taken over all all points $x, y$ separating $c(\mathbb{S}^1)$ into two arcs of length at least $\lambda\ell$.

The $\lambda$–*constriction function*, $\mathrm{constr}_\lambda : \mathbb{R}_+ \to \mathbb{R}_+$, of a metric space $X$ is defined as

$$\mathrm{constr}_\lambda(\ell) = \sup\{\mathrm{constr}_\lambda(c)\ ;\ c \text{ is a Lipschitz simple loop in } X \text{ of length } \leqslant \ell\}\,.$$

Note that when $\lambda \leqslant \mu$, $\mathrm{constr}_\lambda \leqslant \mathrm{constr}_\mu$ and $\mathrm{constr}_\lambda(\ell) \leqslant \ell$.

PROPOSITION 9.133 ([**Dru01**], Proposition 3.5). *For geodesic metric spaces* $X$ *the following are equivalent:*

(1) $X$ *is* $\delta$*–hyperbolic in the sense of Rips, for some* $\delta > 0$*;*
(2) *there exists* $\lambda \in \left(0, \frac{1}{4}\right]$ *such that* $\mathrm{constr}_\lambda(\ell) = o(\ell)$*;*
(3) *for all* $\lambda \in \left(0, \frac{1}{4}\right]$ *and* $\ell > 1$*,*

$$\mathrm{constr}_\lambda(\ell) \leqslant 2\delta\left[\log_2(\ell + 28\delta) + 6\right] + 2\,.$$

REMARK 9.134. One cannot obtain a better order than $O(\log \ell)$ for the general constriction function. This can be seen by considering, in the half-space model of $\mathbb{H}^3$, the horizontal circle of length $\ell$.

PROOF. We begin by arguing that (2) implies (1). In what follows we define *limit triangles* in an asymptotic cone $Cone(X) = \mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, to be the triangles in $Cone(X)$ whose edges are limit geodesics. Note that such triangles *a priori* need not be themselves limits of sequences of geodesic triangles in $X$.

First note that (2) implies that every limit triangle in every asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ is a tripod. Indeed, if one assumes that one limit triangle is not a tripod, without loss of generality one can assume that it is a simple triangle. This triangle is the limit of a family of geodesic hexagons $(H_i)_{i\in I}$, with three edges of lengths of order $O\left(\frac{1}{\lambda_i}\right)$ alternating with three edges of lengths of order $o\left(\frac{1}{\lambda_i}\right)$. (We leave it to the reader to verify that such hexagons may be chosen to be simple.) Since $\mathrm{constr}_\lambda(H_i) = o\left(\frac{1}{\lambda_i}\right)$ we obtain that $\omega$-$\lim H_i$ is not simple, a contradiction.

It remains to prove that every finite geodesic in every asymptotic cone is a limit geodesic. Let $\mathfrak{g}([a_\omega, b_\omega])$ be a geodesic in a cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$, where $a_\omega = (a_i)$ and $b_\omega = (b_i)$; let $c_\omega = (c_i)$ be an arbitrary point on $\mathfrak{g}$. By the previous argument every limit geodesic triangle with vertices $a_\omega, b_\omega, c_\omega$ is a tripod. If $c_\omega$ does not coincide with the center of this tripod then this implies that

$$\operatorname{dist}(a_\omega, c_\omega) + \operatorname{dist}(c_\omega, b_\omega) \geqslant \operatorname{dist}(a_\omega, b_\omega),$$

a contradiction. Thus, $c_\omega \in \omega\text{-}\lim[a_i, b_i]$ and, hence, $\mathfrak{g} = \omega\text{-}\lim[a_i, b_i]$.

We thus proved that every geodesic triangle in every asymptotic cone of $X$ is a tripod, hence every asymptotic cone is a real tree. Hence, $X$ is hyperbolic.

Clearly, (3) implies (2). We will prove that (1) implies (3). By monotonicity of the constriction function (as a function of $\lambda$), it suffices to prove (3) for $\lambda = \frac{1}{4}$. Consider an arbitrary simple closed Lipschitz curve $\mathfrak{c} : \mathbb{S}^1 \to X$ of length $\ell$. We orient the circle and will use the notation $\alpha_{pq}$ to denote the oriented arc of the image of $\mathfrak{c}$ connecting $p$ to $q$. We denote $\operatorname{constr}_{\frac{1}{4}}(\mathfrak{c})$ simply by constr. Let $x, y, z$ be three points on $\mathfrak{c}(\mathbb{S}^1)$ which are endpoints of arcs $\alpha_{xy}, \alpha_{yz}, \alpha_{zx}$ in $\mathfrak{c}(\mathbb{S}^1)$ so that the first two arcs have lenth $\frac{\ell}{4}$. Let $t \in \alpha_{zx}$ be the point minimizing the distance to $y$ in $X$. Clearly,

$$R := \operatorname{dist}(y, t) \geqslant \operatorname{constr}, \quad R \leqslant d(x, y), \quad R \leqslant d(z, y)\,.$$

The point $t$ splits the arc $\alpha_{z,x}$ into two sub-arcs $\alpha_{z,t}, \alpha_{t,x}$. Without loss of generality, we can assume that length of $\alpha_{t,x}$ is $\geqslant \frac{\ell}{4}$. In partcicular, $d(x', t) = 2r \geqslant \operatorname{constr}$. Let $\alpha_{xx'}$ be the maximal subarc of $\alpha_{xy}$ disjoint from the interior of $B(y, r)$ (we allow $x = x'$). As $d(x', t) \geqslant \operatorname{constr}$, lemma 9.59 implies that

$$\ell \geqslant \ell(\alpha_{tx'}) \geqslant 2^{\frac{r-1}{2\delta} - 3} - 12\delta,$$

and, thus,

$$\operatorname{constr} \leqslant 4\delta\left(\log_2(\ell + 12\delta) + 3\right) + 2$$

The inequality in (3) follows. □

## 9.17. Filling invariants

Recall that for every $\mu$-simply connected geodesic metric space $X$ we defined (in Section 5.4) the *filling area function* (or, *isoperimetric function*) $A(\ell) = A_X(\ell)$ (this function, technically speaking, depends on the choice of $\mu$), which computes upper bound on the areas of disks bounding loops of lengths $\leqslant \ell$ in $X$. We also defined the *filling radius function* $r(\ell)$ which computes upper bounds on radii of such disks. The goal of this section is to relate both invariants to hyperbolicity of the sapce $X$. Recall also that hyperbolicity implies linearity of $A_X(\ell)$, see Corollary 9.111.

There is a stronger version of this (converse) statement. This version states that there is a gap between the quadratic filling order and the linear isoperimetric order: As soon as the isoperimetric inequality is less than quadratic, it has to be linear and the space has to be hyperbolic:

THEOREM 9.135 (Subquadratic filling, §2.3, §6.8, [**Gro87**]). *If a coarsely simply-connected geodesic metric space $X$ the isoperimetric function $A_X(\ell) = o(\ell^2)$, then the space is hyperbolic.*

Note that there is a second gap for the possible filling orders of groups.

REMARK 9.136 ([**Ol'91b**], [**Bat99**]). If a finitely presented group $G$ has Dehn function $D(\ell) = o(\ell)$, then $G$ is either free or finite.

Proofs of Theorem 9.135 can be found in [**Ol'91b**], [**Pap95b**], [**Bow95**] and [**Dru01**]. B. Bowditch makes use of only two properties of the area function in his proof: The *quadrangle* (or *Besikovitch*) inequality (see Proposition 5.48) and a certain *theta–property*. In fact, as we will see below, only the quadrangle inequality or its triangle counterpart, the minsize inequality (see Proposition5.50) are needed. Also, we will see it suffices to have subquadratic isoperimetric function for geodesic triangles.

*Proof of Theorem 9.135.* Let $X$ be a $\mu$-simply-connected geodesic metric space and $A_X$ be its isoperimetric function and $\text{minsize}_X : \mathbb{R}_+ \to \mathbb{R}_+$ be the minsize function, see Definition 9.129. According to Proposition 5.50, for every $\delta \geqslant \mu$,

$$[\text{minsize}_X(\ell)]^2 \leqslant \frac{\delta^2}{2\pi} A_X(\ell)\,,$$

whence $A_X(\ell) = o(\ell)$ implies $\text{minsize}_X(\ell) = o(\ell)$. Proposition 9.130 then implies that $X$ is hyperbolic. □

The strongest known version of the converse to Corollary 9.111 is:

THEOREM 9.137 (Strong subquadratic filling theorem,see §2.3, §6.8 of [**Gro87**], and also [**Ol'91b**], [**Pap96**]). *Let $X$ be a $\delta$-simply connected geodesic metric space. If there exist sufficiently large $N$ and $L$ $\epsilon > 0$ sufficiently small, such that every loop $\mathfrak{c}$ in $X$ with $N \leqslant \text{Ar}_\delta(\mathfrak{c}) \leqslant LN$ satisfies*

$$\text{Ar}_\delta(\mathfrak{c}) \leqslant \epsilon[\mathit{length}(\mathfrak{c})]^2\,,$$

*then the space $X$ is hyperbolic.*

It seems impossible to prove this theorem using asymptotic cones.

In Theorem 9.137 it suffices to consider only geodesic triangles $\Delta$ instead of all closed curves, and to replace the condition $N \leqslant Ar_\delta(\Delta) \leqslant LN$ by *length* $(\Delta) \geqslant N$. This follows immediately from Theorem 9.132 and the minsize inequality in Proposition 5.50.

M. Coornaert, T. Delzant and A. Papadopoulos have shown that if $X$ is a complete simply connected Riemannian manifold which is *reasonable* (see [**CDP90**, Chapter 6, §1] for a definition of this notion; for instance if $X$ admits a geometric group action, then $X$ is reasonable) then the constant $\epsilon$ in the previous theorem only has to be smaller than $\frac{1}{16\pi}$, see [**CDP90**, Chapter 6, Theorem 2.1].

In terms of the multiplicative constant, a sharp inequality was proved by S. Wenger.

THEOREM 9.138 (S. Wenger [**Wen08**]). *Let $X$ be a geodesic metric space. Assume that there exists $\varepsilon > 0$ and $\ell_0 > 0$ such that every Lipschitz loop $\mathfrak{c}$ of length* $\text{length}(\mathfrak{c})$ *at least $\ell_0$ in $X$ bounds a Lipschitz disk $\mathfrak{d} : D^2 \to X$ with*

$$\mathit{Area}(\mathfrak{d}) \leqslant \frac{1-\varepsilon}{4\pi}\,\text{length}(\mathfrak{c})^2\,.$$

*Then $X$ is Gromov hyperbolic.*

In the Euclidean space one has the classical isoperimetric inequality

$$\mathit{Area}(\mathfrak{d}) \leqslant \frac{1}{4\pi}\,\text{length}(\mathfrak{c})^2\,,$$

with equality if and only if $\mathfrak{c}$ is a circle and $\mathfrak{d}$ a planar disk.

Note that the quantity $Area(\mathfrak{d})$ appearing in Theorem 9.138 is a generalization of the notion of the geometric area used in this book. If the Lipschitz map $\phi : D^2 \to X$ is injective almost everywhere then $Area(\phi)$ is the 2-dimensional Hausdorff measure of its image. In the case of a Lipschitz map to a Riemannian manifold, $Area(\phi)$ is the *area of a map* defined in Section 2.1.4. When the target is a general geodesic metric space, $Area(\phi)$ is obtained by suitably interpreting the *Jacobian* $J_x(\phi)$ in the integral formula

$$Area(\phi) = \int_{D^2} |J_x\phi(x)|.$$

Another application of the results of Section 9.16 is a description of asymptotic behavior of the filling radius in hyperbolic spaces.

PROPOSITION 9.139 ([**Gro87**], §6, [**Dru01**], §3). *In a geodesic $\mu$-simply connected metric space $X$ the following statements are equivalent:*

(1) *$X$ is hyperbolic;*

(2) *the filling radius $r(\ell) = o(\ell)$;*

(3) *the filling radius $r(\ell) = O(\log \ell)$.*

Furthermore, in (3) one can say that given a loop $\mathfrak{c} : \mathbb{S}^1 \to X$ of length $\ell$, a filling disk $\mathfrak{d}$ minimizing the area has the filling radius $r(\mathfrak{d}) = O(\log \ell)$.

REMARK 9.140. The logarithmic order in (3) cannot be improved, as shown by the example of the horizontal circle in the half-space model of $\mathbb{H}^3$. We note that the previous result shows that, as in the case of the filling area, there is a gap between the linear order of the filling radius and the logarithmical one.

PROOF. In what follows, we let $\mathrm{Ar} = \mathrm{Ar}_\mu$ denote the $\mu$-filling area function in the sense of Section 5.4, defined for loops in the space $X$.

We first prove that (1) $\Rightarrow$ (3). According to the linear isoperimetric inequality for hyperbolic spaces (see Corollary 9.111), there exists a constant $K$ depending only on $X$ such that

$$\mathrm{Ar}(\mathfrak{c}) \leqslant K\ell_X(\mathfrak{c}) \tag{9.15}$$

Here $\mathrm{Ar}(\mathfrak{c})$ is the $\mu$-area of a least-area $\mu$-disk $\mathfrak{d} : \mathcal{D}^{(0)} \to X$ bounding $\mathfrak{c}$. Recall also that the *combinatorial length* and *area* of a simplicial complex is the number of 1-simplices and 2-simpleces respectively in this complex. Thus, for a loop $\mathfrak{c}$ as above, we have

$$\ell_X(\mathfrak{c}) \leqslant \mu \; length(\mathcal{C}),$$

where $\mathcal{C}$ is the triangulation of the circle $S^1$ so that vertices of any edge are mapped by $\mathfrak{c}$ to points within distance $\leqslant \mu$ in $X$.

Consider now a loop $\mathfrak{c} : \mathbb{S}^1 \to X$ of metric length $\ell$ and a least area $\mu$-disk $\mathfrak{d} : \mathcal{D}^{(0)} \to X$ filling $\mathfrak{c}$; thus, $\mathrm{Ar}(\mathfrak{c}) \leqslant K\ell$.

Let $v \in \mathcal{D}^{(0)}$ be a vertex such that its image $a = \mathfrak{d}(v)$ is at maximal distance $r$ from $\mathfrak{c}(\mathbb{S}^1)$. For every $1 \leqslant j \leqslant k$, with

$$k = \lfloor \frac{r}{\mu} \rfloor$$

we denfine a subcomplex $\mathcal{D}_j$ of $\mathcal{D}$: $\mathcal{D}_j$ is the maximal connected subcomplex in $\mathcal{D}$ containing $v$, so that every vertex in $\mathcal{D}_j$ could be connected to $v$ by a *gallery* (in the sense of Section 3.2.1) of 2-dimensional simplices $\sigma$ in $\mathcal{D}$ so that

$$\mathfrak{d}\left(\sigma^{(0)}\right) \subset \overline{B}(a, j\mu).$$

For instance $\mathcal{D}_1$ contains the star of $v$ in $\mathcal{D}$. Let $\mathrm{Ar}_j$ be the number of 2-simplices in $\mathcal{D}_j$.

For each $j \leqslant k-1$ the geometric realization $\mathcal{D}_j$ of the subcomplex $\mathcal{D}_j$ is homeomorphic to a 2-dimensional disk with several disks removed from the interior. (As usual, we will conflate a simplicial complex and its geometric realization.) Therefore the boundary $\partial\mathcal{D}_j$ of $\mathcal{D}_j$ in $D^2$ is a union of several disjont topological circles, while all the edges of $\mathcal{D}_j$ are interior edges for $\mathcal{D}$. We denote by $s_j$ the outermost circle in $\partial\mathcal{D}_j$, i.e., $s_j$ bounds a triangulated disk $\mathcal{D}'_j \subset \mathcal{D}$, so that $\mathcal{D}_j \subset \mathcal{D}'_j$. Let $\mathrm{length}(\partial\mathcal{D}_j)$ and $\mathrm{length}(s_j)$ denote the number of edges of $\partial\mathcal{D}_j$ and of $s_j$ respectively.

By definition, every edge of $\mathcal{D}_j$ is an interior edge of $\mathcal{D}_{j+1}$ and belongs to a 2-simplex of $\mathcal{D}_{j+1}$. Note also that if $\sigma$ is a 2-simplex in $\mathcal{D}$ and two edges of $\sigma$ belong to $\mathcal{D}_j$, then $\sigma$ belongs to $\mathcal{D}_j$ as well. Therefore,

$$\mathrm{Ar}_{j+1} \geqslant \mathrm{Ar}_j + \frac{1}{3}\mathrm{length}(\partial\mathcal{D}_j) \geqslant \mathrm{Ar}_j + \frac{1}{3}\mathrm{length}(s_j).$$

Since $\mathfrak{d}$ is a least area filling disk for $\mathfrak{c}$ it follows that each disk $\mathfrak{d}|_{\mathcal{D}'_j}$ is a least area disk bounding the loop $\mathfrak{d}|_{\int_j}$. In particular, by the isoperimetric inequality in $X$,

$$\mathrm{Ar}_j = Area(\mathcal{D}_j) \leqslant Area(\mathcal{D}'_j) \leqslant K\ell_X(\mathfrak{d}(s_j)) \leqslant K\mu length(s_j)$$

We have thus obtained that

$$\mathrm{Ar}_{j+1} \geqslant \left(1 + \frac{1}{3\mu K}\right)\mathrm{Ar}_j\,.$$

It follows that

$$K\ell \geqslant \mathrm{Ar}(\mathfrak{d}) \geqslant \left(1 + \frac{1}{3\mu K}\right)^k$$

whence,

$$r \leqslant \mu(k+1) \leqslant \mu\left(\frac{\ln\ell + \ln K}{\ln\left(1 + \frac{1}{3\mu K}\right)} + 1\right).$$

Clearly (3) $\Rightarrow$ (2). It remains to prove that (2) $\Rightarrow$ (1).

We first show that (2) implies that in an every asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ all geodesic triangles that are limits of geodesic triangles in $X$ (i.e. $\boldsymbol{\Delta} = \omega\text{-lim}\,\Delta_i$) are tripods. We assume that $\boldsymbol{\Delta}$ is not a point. Every geodesic triangle $\Delta_i$ can be seen as a loop $\mathfrak{c}_i : \mathbb{S}^1 \to \Delta_i$, and can be filled with a $\mu$-disk $\mathfrak{d}_i : \mathcal{D}^{(1)} \to X$ of filling radius $r_i = r(\mathfrak{d}_i) = o\left(\mathrm{length}\left(\Delta_i\right)\right)$. In particular, $\omega\text{-lim}_i\, \lambda_i r_i = 0$.

Let $[x_i, y_i], [y_i, z_i]$ and $[z_i, x_i]$ be the three geodesic edges of $\Delta_i$, and let $\overline{x}_i, \overline{y}_i, \overline{z}_i$ be the three points on $\mathbb{S}^1$ corresponding to the three vertices $x_i, y_i, z_i$. Consider a path $\overline{\mathfrak{p}}_i$ in the 1-skeleton of $\mathcal{D}$ with endpoints $\overline{y}_i$ and $\overline{z}_i$ such that $\overline{\mathfrak{p}}_i$ together with the arc of $\mathbb{S}^1$ with endpoints $\overline{y}_i, \overline{z}_i$ encloses a maximal number of triangles with $\mathfrak{d}_i$–images in the $r_i$–neighborhood of $[y_i, z_i]$. Every edge of $\overline{\mathfrak{p}}_i$ that is not in $\mathbb{S}^1$ is contained in a 2-simplex whose third vertex has $\mathfrak{d}_i$–image in the $r_i$–neighborhood of

$[y_i, x_i] \cup [x_i, z_i]$. The edges in $\overline{\mathfrak{p}}_i$ that are in $\mathbb{S}^1$ are either between $\overline{x}_i, \overline{y}_i$ or between $\overline{x}_i, \overline{z}_i$.

Thus $\overline{\mathfrak{p}}_i$ has $\mathfrak{d}_i$–image $\mathfrak{p}_i$ in the $(r_i + \mu)$–neighborhood of $[y_i, x_i] \cup [x_i, z_i]$. See Figure 9.7.

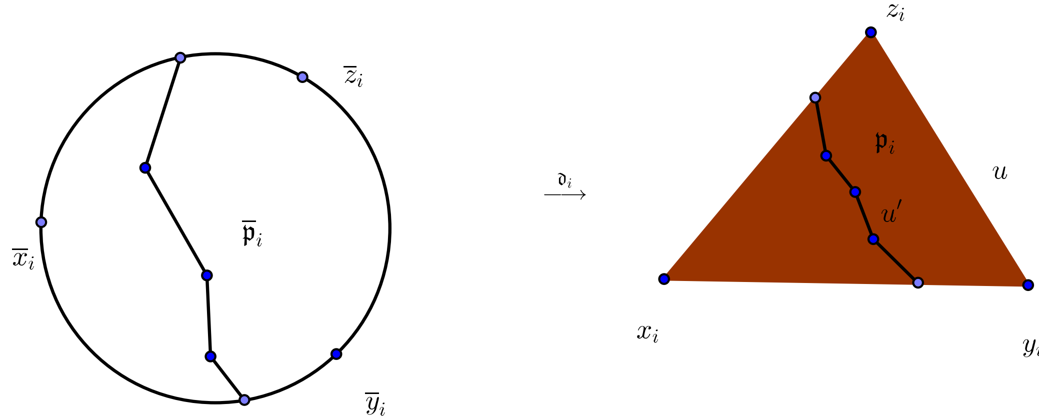



FIGURE 9.7. The path $\overline{\mathfrak{p}}_i$ and its image $\mathfrak{p}_i$.

Consider an arbitrary vertex $\overline{u}$ on $\mathbb{S}^1$ between $\overline{y}_i, \overline{z}_i$ and its image $u \in [y_i, z_i]$. We have that $\mathfrak{p}_i \subset \overline{\mathcal{N}}_{r_i+\mu}([y_i, u]) \cup \overline{\mathcal{N}}_{r_i+\mu}([u, z_i])$, where $[y_i, u]$ and $[u, z_i]$ are subgeodesics of $[y_i, z_i]$.

By connectedness, there exists a point $u' \in \mathfrak{p}_i$ at distance at most $r_i + \mu$ from a point $u_1 \in [y_i, u]$, and from a point $u_2 \in [u, z_i]$. As the three points $u_1, u, u_2$ are aligned on a geodesic and $\operatorname{dist}(u_1, u_2) \leqslant 2(r_i + \mu)$ it follow that, say, $\operatorname{dist}(u_1, u) \leqslant r_i + \mu$, whence $\operatorname{dist}(u, u') \leqslant 3(r_i + \mu)$. Since the point $\overline{u}$ was arbitrary, we have thus proved that $[y_i, z_i]$ is in $\overline{\mathcal{N}}_{3r_i+3\mu}(\mathfrak{p}_i)$, therefore it is in $\overline{\mathcal{N}}_{4r_i+4\mu}([y_i, x_i] \cup [x_i, z_i])$. This implies that in $\boldsymbol{\Delta}$ one edge is contained in the union of the other two. The same argument done for each edge implies that $\boldsymbol{\Delta}$ is a tripod.

From this, one can deduce that every triangle in the cone is a tripod. In order to do this it suffices to show that every geodesic in the cone is a limit geodesic. Consider a geodesic in $\operatorname{Cone}_\omega(X, \boldsymbol{e}, \boldsymbol{\lambda})$ with the endpoints $x_\omega = (x_i)$ and $y_\omega = (y_i)$ and an arbitrary point $z_\omega = (z_i)$ on this geodesic. Geodesic triangles $\Delta_i$ with vertices $x_i, y_i, z_i$ yield a tripod $\Delta_\omega = \Delta(x_\omega, y_\omega, z_\omega)$ in the asymptotic cone, but since,

$$\operatorname{dist}(x_\omega, z_\omega) + \operatorname{dist}(z_\omega, y_\omega) = \operatorname{dist}(x_\omega, y_\omega),$$

it follows that the tripod must be degenerate. Thus $z_\omega \in \omega\text{-}\lim[x_i, y_i]$. □

Like for the area, for the radius too there is a stronger version of the implication *sublinear radius* $\Longrightarrow$ *hyperbolicity*, similar to Theorem 9.137.

Proposition 9.141 (M. Gromov; P. Papasoglou [**Pap98**]). *Let $\Gamma$ be a finitely presented group. If there exists $\ell_0 > 0$ such that*

$$r(\ell) \leqslant \frac{\ell}{73}, \ \forall \ell \geqslant \ell_0 \,,$$

*then the group $\Gamma$ is hyperbolic.*

According to [**Pap98**], the best possible constant expected is not $\frac{1}{73}$, but $\frac{1}{8}$. Note that the proof of Proposition 9.141 cannot be extended from groups to metric spaces, because it relies on the bigon criterion for hyperbolicity [**Pap95c**], which only works for groups. There is probably a similar statement for general metric spaces, with a constant that can be made effective for complete simply connected Riemannian manifolds.

## 9.18. Rips construction

The goal of this section is to describe Rips construction which associates a hyperbolic group with to an arbitrary finite presentation.

Theorem 9.142 (Rips Construction, I. Rips [**Rip82**]). *Let $Q$ be a group with a finite presentation $\langle A|R\rangle$. Then, with such presentation of $Q$ one can associate a short exact sequence*

$$1 \to K \to G \to Q \to 1$$

*where $G$ is hyperbolic and $K$ is finitely generated. Furthermore, the group $K$ in this construction is finitely-presentable if and only if $Q$ is finite.*

Proof. We will give here only a sketch of the argument. Let $A = \{a_1, ..., a_m\}$, $R = \{R_1, ..., R_n\}$. For $i = 1, ..., m, j = 1, 2$, pick even natural numbers $r_i < s_i$, $p_{ij} < q_{ij}$, $u_{ij} < v_{ij}$, so that all the intervals

$$[r_i, s_i], \quad [p_{ij}, q_{ij}]. \quad [u_{ij}, v_{ij}], i = 1, ..., m, j = 1, 2$$

are pairwise disjoint and all the numbers $r_i, s_i, p_{ij}, q_{ij}, u_{ij}, v_{ij}$ are at least 10 times larger than the lengths of the words $R_k$. Define the group $G$ by the presentation $P$ where generators are $a_1, ..., a_m, b_1, b_2$, and relators are:

$$R_i b_1 b_2^{r_2} b_1 b_2^{r_i+1} \cdots b_1 b_2^{s_i}, i = 1, ..., n \tag{9.16}$$

$$a_i^{-1} b_j a_i b_1 b_2^{u_{ij}} b_1 b_2^{u_{ij}+1} \cdots b_1 b_2^{v_{ij}}, i = 1, ..., m, j = 1, 2, \tag{9.17}$$

$$a_i b_j a_i^{-1} b_1 b_2^{p_{ij}} b_1 b_2^{p_{ij}+1} \cdots b_1 b_2^{q_{ij}}, i = 1, ..., m, j = 1, 2. \tag{9.18}$$

Now, define the map $\phi : G \to Q$, $\phi(a_i) = a_i, \phi(b_j) = 1, j = 1, 2$. Clearly, $\phi$ respects all the relators and, hence, it determines an epimomorphism $\phi : G \to Q$. We claim that the kernel $K$ of $\phi$ is generated by $b_1, b_2$. First, the kernel, of course, contains $b_1, b_2$. The subgroup generated by $b_1, b_2$ is clearly normal in $G$ because of the relators (9.17) and (9.18). Thus, indeed, $b_1, b_2$ generate $K$.

The reason that the group $G$ is hyperbolic is that the presentation written above is Dehn: because of the choices of the numbers $r_i$ etc., when we multiply conjugates of the relators of $G$, we cannot cancel more than half of one of the relators (9.16) — (9.18), namely, the product of generators $b_1, b_2$ appearing in the end of each relator. This argument is a typical example of application of the small cancelation theory, see [**LS77**]. Rips in his paper [**Rip82**], did not use the language of hyperbolic groups, but the language of the small cancelation theory.

One then verifies that $G$ has cohomological dimension 2 by showing that the presentation complex $Z$ of the presentation $P$ of the group $G$ is aspherical, for this one can use, for instance, [**Ger87**].

Now, R. Bieri proved in [**Bie76b**, Theorem B] that if $G$ is a group of cohomological dimension 2 and $H \lhd G$ is a normal subgroup of infinite index, then $H$ is free.

Suppose that the subgroup $K$ is free. Then rank of $K$ is at most 2 since $K$ is 2-generated. The elements $a_1, a_2 \in G$ act on $K$ as automorphisms (by conjugation). However, considering action of $a_1, a_2$ on the abelianization, we see that because $p_{ij}, q_{ij}$ are even, the images of the generators $b_1, b_2$ cannot generate the abelianization of $K$. Similar argument shows that $K$ cannot be cyclic, so $K$ is trivial and, hence, $b_1 = b_2 = 1$ in $G$. However, this clearly contradicts the fact that the presentation (9.16) — (9.18) is a Dehn presentation (since the words $b_1, b_2$ obviously do not contain more than half of the length of any relator). ☐

In particular, there are hyperbolic groups which contain non-hyperbolic finitely-generated subgroups. Furthermore,

COROLLARY 9.143. *Hyperbolic groups could have unsolvable membership problem.*

PROOF. Indeed, start with a finitely-presented group $Q$ with unsolvable word problem and apply the Rips construction to $Q$. Then $g \in G$ belongs to $N$ if and only if $g$ maps trivially to $Q$. Since $Q$ has unsolvable word problem, the problem of membership of $g$ in $N$ is unsolvable as well. ☐

On the other hand, the membership problem is solvable for quasiconvex subgroups, see Theorem 9.163.

## 9.19. Asymptotic cones, actions on trees and isometric actions on hyperbolic spaces

Let $G$ be a finitely-generated group with the generating set $g_1, .., g_m$; let $X$ be a metric space. Given a homomorphism $\rho : G \to \mathrm{Isom}(X)$, we define the following function:

$$d_\rho(x) := \max_k d(\rho(g_k)(x), x) \tag{9.19}$$

and set

$$d_\rho := \inf_{x \in X} d_\rho(x).$$

This function does not necessarily have minimum, so we choose $x_\rho \in X$ to be a point so that

$$d_\rho(x) - d_\rho \leqslant 1.$$

Such points $x_\rho$ are called *min-max* points of $\rho$ for obvious reason. The set of min-max points could be unbounded, but, as we will see, this does not matter. Thus, high value of $d_\rho$ means that all points of $X$ move a lot by at least one of the generators of $\rho(G)$.

Example 9.144. 1. Let $X = \mathbb{H}^n$, $G = \langle g \rangle$ be infinite cyclic group, $\rho(g) \in \mathrm{Isom}(X)$ is a hyperbolic translation along a geodesic $L \subset X$ by some amount $t > 1$, e.g. $\rho(g)(x) = e^t x$ in the upper half-space model. Then $d_\rho = t$ and we can take $x_\rho \in L$, since the set of points of minima of $d_\rho(x)$ is $L$.

2. Suppose that $X = \mathbb{H}^n = U^n$ and $G$ are the same but $\rho(g)$ is a parabolic translation, e.g. $\rho(g)(x) = x+u$, where $u \in \mathbb{R}^{n-1}$ is a unit vector. Then $d_\rho$ does not attain minimum, $d_\rho = 0$ and we can take as $x_\rho$ any point $x \in U^n$ so that $x_n \geqslant 1$.

3. Suppose that $X$ is the same, but $G$ is no longer required to be cyclic. Assume that $\rho(G)$ fixes a unique point $x_o \in X$. Then $d_\rho = 0$ and the set of min-max points is contained in a metric ball centered at $x_o$. The radius of this ball could be estimated from above independently of $G$ and $\rho$. (The latter is nontrivial.)

Suppose $\sigma \in \mathrm{Isom}(X)$ and we replace the original representation $\rho$ with the conjugate representation $\rho' = \rho^\sigma : g \mapsto \sigma\rho(g)\sigma^{-1}, g \in G$.

Exercise 9.145. Verify that $d_\rho = d_{\rho'}$ and that as $x_{\rho'}$ one can take $\sigma(x_\rho)$.

Thus, conjugating $\rho$ by an isometry, does not change the geometry of the action, but moves min-max points in a predictable manner.

The set $Hom(G, \mathrm{Isom}(X))$ embeds in $(\mathrm{Isom}(X))^m$ since every $\rho$ is determined by the $m$-tuple

$$(\rho(g_1), ..., \rho(g_m)).$$

As usual, we equip the group $\mathrm{Isom}(X)$ with the topology of uniform convergence on compacts and the set $Hom(G, \mathrm{Isom}(X))$ with the subset topology.

Exercise 9.146. Show that topology on $Hom(G, \mathrm{Isom}(X))$ is independent of the finite generating set. Hint: Embed $Hom(G, \mathrm{Isom}(X))$ in the product of countably many copies of $\mathrm{Isom}(X)$ (indexed by the elements of $G$) and relate topology on $Hom(G, \mathrm{Isom}(X))$ to the Tychonoff topology on the infinite product.

Suppose now that the metric space $X$ is proper. Pick a base-point $o \in X$. Then Arzela-Ascoli theorem implies that for every $D$ the subset

$$Hom(G, \mathrm{Isom}(X))_{o,D} = \{\rho : G \to \mathrm{Isom}(X) | d_\rho(o) \leqslant D\}$$

is compact. We next consider the quotient

$$Rep(G, \mathrm{Isom}(X)) = Hom(G, \mathrm{Isom}(X)) / \mathrm{Isom}(X)$$

where $\mathrm{Isom}(X)$ acts on $Hom(G, \mathrm{Isom}(X))$ by conjugation $\rho \mapsto \rho^\sigma$. We equip $Rep(G, \mathrm{Isom}(X))$ with the quotient topology. In general, this topology is not Hausdorff.

Example 9.147. Let $G = \langle g \rangle$ is infinite cyclic, $X = \mathbb{H}^n$. Show that trivial representation $\rho_0 : G \to 1 \in \mathrm{Isom}(X)$ and representation $\rho_1$ where $\rho_1(g)$ acts as a parabolic translation, project to points $[\rho_i]$ in $Rep(G, \mathrm{Isom}(X))$, so that every neighborhood of $[\rho_0]$ contains $[\rho_1]$.

Exercise 9.148. Let $X$ be a graph (not necessarily locally-finite) with the standard metric and consider the subset $Hom_f(G, \mathrm{Isom}(X))$ consisting of representations $\rho$ which give rise to the free actions $G/Ker(\rho) \curvearrowright X$. Then

$$Rep_f(G, \mathrm{Isom}(X)) = Hom_f(G, \mathrm{Isom}(X)) / \mathrm{Isom}(X)$$

is Hausdorff.

We will be primarily interested in compactness rather than Hausdorff properties of $Rep(G, \mathrm{Isom}(X))$. Define

$$Hom_D(G, \mathrm{Isom}(X)) = \{\rho : G \to \mathrm{Isom}(X) | d_\rho \leqslant D\}.$$

Similarly, for a subgroup $H \subset \mathrm{Isom}(X)$, one defines

$$Hom_D(G, H) = Hom_D(G, \mathrm{Isom}(X)) \cap Hom(G, H).$$

Lemma 9.149. *Suppose that $H \subset \mathrm{Isom}(X)$ is a closed subgroup whose action on $X$ is cobounded. Then for every $D \in \mathbb{R}_+$, $Rep_D(G, H) = Hom_D(G, H)/H$ is compact.*

Proof. Let $o \in X, R < \infty$ be such that the orbit of $\bar{B}(o, R)$ under the $H$-action is the entire space $X$. For every $\rho \in Hom(G, H)$ we pick $\sigma \in H$ so that *some* min-max point $x_\rho$ of $\rho$ satisfies:

$$\sigma(x_\rho) \in \bar{B}(o, R).$$

Then, using conjugation by such $\sigma$'s, for each equivalence class $[\rho] \in Rep_D(G, H)$ we choose a representative $\rho$ so that $x_\rho \in \bar{B}(o, R)$. It follows that for every such $\rho$

$$\rho \in Hom(G, H) \cap Hom(G, \mathrm{Isom}(X))_{o,D'}, \quad D' = D + 2R.$$

This set is compact and, hence, its projection $Rep_D(G, H)$ is also compact. □

In view of this lemma, even if $X$ is not proper, we say that a sequence $\rho_i : G \to \mathrm{Isom}(X)$ *diverges* if

$$\lim_{i\to\infty} d_{\rho_i} = \infty.$$

Definition 9.150. We say that an isometric action of a group on a real tree $T$ is *nontrivial* if the group does not fix a point in $T$.

Proposition 9.151 (M.Bestvina; F. Paulin). *Suppose that $(\rho_i)$ is a diverging sequence of representations $\rho_i : G \to H \subset \mathrm{Isom}(X)$, where $X$ is a Rips–hyperbolic metric space. Then $G$ admits a nontrivial isometric action on a real tree.*

Proof. Let $p_i = x_{\rho_i}$ be min-max points of $\rho_i$'s. Take $\lambda_i := (d_{\rho_i})^{-1}$ and consider the corresponding asymptotic cone $Cone_\omega(X, \mathbf{P}, \lambda)$ of the space $X$; here $\mathbf{p} = (p_i)$. According to Lemma 9.35, the metric space $\mathbf{X}$ in this asymptotic cone is a real tree $T$. Furthermore, the sequence of group actions $\rho_i$ converges to an isometric action $\rho_\omega : G \curvearrowright T$:

$$\rho_\omega(g)(x_\omega) = (\rho_i(x_i)),$$

the key here is that *all generators $\rho_i(g_k)$ of $\rho_i(G)$ move the base-point $p_i \in \lambda_i X$ by* $\leqslant \lambda_i(d_{\rho_i} + 1)$. The ultralimit of the latter quantity is equal to 1. Furthermore, for $\omega$-all $i$ one of the generators, say $g = g_k$, satisfies

$$|d_{\rho_i} - d(\rho_i(g)(p_i)| \leqslant 1$$

in $X$. Thus, the element $\rho_\omega(g)$ will move the point $\mathbf{p} \in T$ exactly by 1. Because $p_i$ was a min-max point of $\rho_i$, it follows that

$$d_{\rho_\omega} = 1.$$

In particular, the action $\rho_\omega : G \curvearrowright T$ has no fixed point, i.e., is nontrivial. □

One of the important applications of this proposition is

Theorem 9.152 (F. Paulin, [**Pau91**]). *Suppose that $G$ is a finitely-generated group with property FA and $H$ is a hyperbolic group. Then, up to conjugation in $H$, there are only finitely many homomorphisms $G \to H$.*

Proof. Let $X$ be a Cayley graph of $H$, then $H \subset \mathrm{Isom}(X)$, $X$ is proper and Rips–hyperbolic. Then, by the above proposition, if $Hom(G, H)/H$ is noncompact, then $G$ has a nontrivial action on a real tree. This contradicts the assumption that $G$ has the property FA. Suppose, therefore, that $Hom(G, H)/H$ is compact. If this quotient is infinite, pick a sequence $\rho_i \in Hom(G, H)$ of pairwise non-conjugate representations. Without loss of generality, by replacing $\rho_i$'s by their conjugates, we can assume that min-max points $p_i$ of $\rho_i$'s are in $\overline{B}(e, 1)$. Therefore, after passing to a subsequence if necessary, the sequence of representations $\rho_i$ converges. However, the action of $H$ on itself is free, so for every generator $g$ of $G$, the sequence $\rho_i(g)$ is eventually constant. Therefore, the entire sequence $(\rho_i)$ consists of only finitely many representations. Contradiction. Thus, $Hom(G, H)/H$ is finite. □

This theorem is one of many results of this type: Bounding number of homomorphisms from a group to a hyperbolic group. Having Property FA is a very strong restriction on the group, so, typically one improves Proposition 9.151 by making stronger assumptions on representations $G \to H$ and, accordingly, stronger conclusions about the action of $G$ on the tree, for instance:

Theorem 9.153. *Suppose that $H$ is a hyperbolic group, $X$ is its Cayley graph and all the representations $\rho_i : G \to H$ are faithful. Then the resulting nontrivial action of $G$ on a real tree is* small*, i.e., stabilizer of every nontrivial geodesic segment is virtually cyclic.*

The key ingredient then is *Rips Theory* which converts small actions (satisfying some mild restrictions which will hold in the case of groups $G$ which embed in hyperbolic groups) $G \curvearrowright T$, to decompositions of $G$ as an amalgam $G_1 \star_{G_3} G_2$ or HNN-extension $G = G_1 \star_{G_3}$, where the subgroup $G_3$ is again virtually cyclic. Thus, one obtains:

Theorem 9.154 (I. Rips, Z. Sela, [**RS94**]). *Suppose that $G$ does not split over a virtually cyclic subgroup. Then for every hyperbolic group $H$, $Hom_{inj}(G, H)/H$ is finite, where $Hom_{inj}$ consists of injective homomorphisms. In particular, if $G$ is itself hyperbolic, then $Out(G) = Aut(G)/G$ is finite.*

Some interesting and important groups $G$, like surface groups, do split over virtually cyclic subgroups. In this case, one cannot in general expect $Hom_{inj}(G, H)/H$ to be finite. However, it turns out that the only reason for lack of finiteness is the fact that one can precompose homomorphisms $G \to H$ with automorphisms of $G$ itself:

Theorem 9.155 (I. Rips, Z. Sela, [**RS94**]). *Suppose that $G$ is a 1-ended finitely-generated group. Then for every hyperbolic group $H$, the set*

$$Aut(G)\backslash Hom_{inj}(G, H)/H$$

*is finite. Here $Aut(G)$ acts on $Hom(G, H)$ by precomposition.*

## 9.20. Further properties of hyperbolic groups

1. Hyperbolic groups are ubiquitous:

Theorem 9.156 (See e.g. [**Del96**]). *Let $G$ be a non-elementary $\delta$-hyperbolic group. Then there exists $N$, so that for every collection $g_1, .., g_k \in G$ of elements of norm $\geq 1000\delta$, the following holds:*

*i. The subgroup generated by the elements $g_i^N$ and all their conjugates is free.*

*ii. Then the quotient group $G/\langle\langle g_1^n, ...g_k^n\rangle\rangle$ is again non-elementary hyperbolic for all sufficiently large $n$. In particular, infinite hyperbolic groups are never simple.*

Thus, by starting with, say, a nonabelian free group $F_n = G$, and adding to its presentation one relator of the form $w^n$ at a time (where $n$'s are large), one obtains non-elementary hyperbolic groups. Furthermore,

Theorem 9.157 (A. Ol'shanskii, [**Ol'91c**]). *Every non-elementary torsion-free hyperbolic group admits a quotient which is an infinite torsion group, where every nontrivial element has the same order.*

Theorem 9.158 (A. Ol'shanskii, [**Ol'95**], T. Delzant [**Del96**]). *Every non-elementary hyperbolic group $G$ is SQ-universal, i.e., every countable group embeds in a quotient of $G$.*

"Most" groups are hyperbolic:

Theorem 9.159 (A. Ol'shanskii [**Ol'92**]). *Fix $k \in \mathbb{N}$, $k \geq 2$ and let $A = \{a^{\pm 1}, a^{\pm 2}, ..., a_k^{\pm 1}\}$ be an alphabet. Fix $i \in \mathbb{N}$ and let $(n_1, ..., n_i)$ be a sequence of natural numbers. Let $N = N(k, i, n_1, ..., n_i)$ be the number of group presentations*

$$G = \langle a_1, ..., a_k | r_1, ..., r_i\rangle$$

*such that $r_1, ..., r_i$ are reduced words in the alphabet $A$ such that the length of $r_j$ is $n_j$, $j = 1, 2, , i$. If $N_h$ is the number of hyperbolic groups in this collection and if $n = min\{n_1, .., n_i\}$, then*

$$\lim_{n\to\infty} \frac{N_h}{N} = 1$$

*and convergence is exponentially fast.*

The model of randomness which appears in this theorem is by no means unique, we refer the reader to [**Gro03**], [**Ghy04**], [**Oll04**], [**KS08**] for further discussion of random groups.

Theorems 9.160, 9.161, 9.162 below first appeared in Gromov's paper [**Gro87**]; other proofs could be found for instance in [**Aea91**], [**BH99**], [**ECH**$^+$**92**], [**ECH**$^+$**92**], [**GdlH90**].

2. Hyperbolic groups have finite type:

Theorem 9.160. *Let $G$ be $\delta$-hyperbolic. Then there exists $D_0 = D_0(\delta)$ so that for all $D \geq D_0$ the Rips complex $\mathrm{Rips}_D(G)$ is contractible. In particular, $G$ has type $F_\infty$.*

3. Hyperbolic groups have controlled torsion:

Theorem 9.161. *Let $G$ be hyperbolic. Then $G$ contains only finitely many conjugacy classes of finite subgroups.*

4. Hyperbolic groups have solvable algorithmic problems:

Theorem 9.162. *Every $\delta$-hyperbolic group has solvable word and conjugacy problems.*

Furthermore:

Theorem 9.163 (I. Kapovich, [**Kap96**]). *Membership problem is solvable for quasiconvex subgroups of hyperbolic groups: Let $G$ be hyperbolic and $H < G$ be a quasiconvex subgroup of a $\delta$-hyperbolic group. Then the problem of membership in $H$ is solvable.*

Isomorphism problem is solvable:

Theorem 9.164 (Z. Sela, [**Sel95**]; F. Dahmani and V. Guirardel [**DG11**]). *Given two $\delta$-hyperbolic groups $G_1, G_2$, there is an algorithm to determine if $G_1, G_2$ are isomorphic.*

Note that Sela proved this theorem only for torsion-free 1-ended hyperbolic groups. This result was extended to all hyperbolic groups by Dahmani and Guirardel.

5. Hyperbolic groups are hopfian:

Theorem 9.165 (Z. Sela, [**Sel99**]). *For every hyperbolic group $G$ and every epimorphism $\phi : G \to G$, $Ker(\phi) = 1$.*

Note that every residually finite group is hopfian, but the converse, in general, is false. An outstanding open problem is to determine if all hyperbolic groups are residually finite (it is widely expected that the answer is negative). Every linear group is residually finite, but there are nonlinear hyperbolic groups, see [**Kap05**]. It is very likely that some (or even all) of the nonlinear hyperbolic groups described in [**Kap05**] are not residually finite.

6. Hyperbolic groups tend to be co-Hopfian:

Theorem 9.166 (Z. Sela, [**Sel97**]). *For every 1-ended hyperbolic group $G$, every monomorphism $\phi : G \to G$ is surjective, i.e., such $G$ is co-Hopf.*

7. All hyperbolic groups admit QI embeddings in the real-hyperbolic space $\mathbb{H}^n$:

Theorem 9.167 (M. Bonk, O. Schramm [**BS00**]). *For every hyperbolic group $G$ there exists $n$, such that $G$ admits a quasi-isometric embedding in $\mathbb{H}^n$.*

## 9.21. Relatively hyperbolic spaces and groups

Relatively hyperbolic groups were introduced by M. Gromov in the same paper as hyperbolic groups, namely in [**Gro87**]. While a model for hyperbolic groups were uniform lattices in negatively curved symmetric spaces, for relatively hyperbolic groups the model were non-uniform lattices in negatively curved spaces and, more generally, fundamental groups of complete Riemannian manifolds of finite volume and curvature $\leq -a^2 < 0$. A good picture is that of truncated hyperbolic spaces defined in Chapter 22 (see Figure 22.1). These are metric spaces hyperbolic relative to the boundary horospheres. In general, one considers a geodesic metric space $X$ and a collection $\mathcal{A}$ of subsets of it (called *peripheral subsets* when the relative hyperbolicity conditions are fulfilled).

The metric definition of relative hyperbolicity is consists of three conditions, the main one being very similar to the condition of thin triangles for hyperbolic spaces.

Definition 9.168. We say that $X$ is $(*)$–*relatively hyperbolic with respect to* $\mathcal{A}$ if for every $C \geqslant 0$ there exist two constants $\sigma$ and $\delta$ such for every triangle $T \subset X$ with $(1, C)$–quasi-geodesic edges, either there exists a point at distance

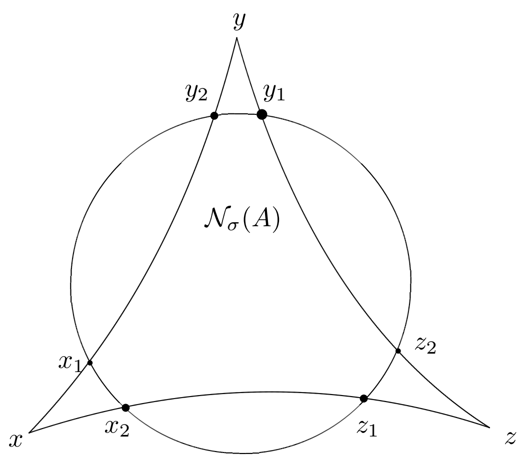



FIGURE 9.8. Second case of Definition 9.168.

at most $\sigma$ from each of the sides of $T$, or there exists a subset $A \in \mathcal{A}$ such that its $\sigma$-neighborhood $\mathcal{N}_\sigma(A)$ intersects each of the sides of the triangle; moreover ,for every vertex of the triangle, the two edges issuing from it enter $\mathcal{N}_\sigma(A)$ in two points at distance at most $\delta$ away from each other.

Clearly $(*)$–relative hyperbolicity is a rather weak condition. For instance every geodesic hyperbolic space is $(*)$–hyperbolic relative to every family of subsets covering it.

DEFINITION 9.169. The space $X$ is *hyperbolic relative* to $\mathcal{A}$ if it is $(*)$–hyperbolic relative to $\mathcal{A}$, and moreover, the following properties are satisfied:

($\alpha_1$) For every $r > 0$, the $r$–neighborhoods of two distinct subsets in $\mathcal{A}$ intersect in a set of diameter at most $D = D(r)$.

($\alpha_2$) Every geodesic of length $\ell$ with endpoints at distance at most $\frac{\ell}{3}$ from a set $A \in \mathcal{A}$, intersects the $M$-tubular neighborhood of $A$, for some universal constant $M$.

DEFINITION 9.170. A finitely generated group $G$ is *hyperbolic relative to* a finite set of subgroups $H_1, ..., H_n$ if, endowed with a word metric, $G$ is hyperbolic in the sense of Definition 9.169 relative to the collection $\mathcal{A}$ of left cosets $gH_i$ for $i \in \{1, 2, ..., n\}$ and $g$ in a set of representatives of $G/H_i$.

It follows from the definition that all $H_i$ are finitely generated, since the three metric conditions imply that the peripheral subsets are quasi-convex. The groups $H_i$ are called *peripheral subgroups*.

THEOREM 9.171 (C. Drutu, D. Osin, M. Sapir, [**DS05**],[**Osi06**],[**Dru09**]). *Relative hyperbolicity in the sense of Definition 9.170 is equivalent to (strong) relative hyperbolicity as defined in* [**Gro87**].

Other characterizations of (strong) relative hyperbolicity can be found in the papers [**Bow97**], [**Far98**], [**Dah03b**], [**DS05**], [**Osi06**]. Here we always mean *strong* relative hyperbolicity when we use the term. Also, in what follows, we will always assume that *every $H_i$ has infinite index in $G$.*

In the list of properties in Definition 9.169, one cannot drop property $(\alpha_1)$, as shown by the examples of groups in [**OOS09**] and in [**BDM09**, §7.1].

Many properties similar to those of hyperbolic groups are proved in the relatively hyperbolic case, in particular a Morse lemma, a characterization in terms of asymptotic cones [**DS05**], a relative linear filling [**Osi06**], action on the boundary as a convergence group [**Yam04**].

Hyperbolic groups are clearly relatively hyperbolic with peripheral subgroup $\{1\}$.

Other examples of relatively hyperbolic groups include:

(1) $G$ is hyperbolic and each $H_i$ is quasiconvex and *almost malnormal* in $G$ (see [**Far98**]). Almost malnormality of a subgroup $H \leqslant G$ means that for every $g \in G \setminus H$,
$$|gHg^{-1} \cap H| < \infty.$$
(2) $G$ is the fundamental group of a finite graph of groups with finite edge groups; then $G$ is hyperbolic relative to the vertex groups, see [**Bow97**].
(3) Fundamental groups of complete finite volume manifolds of pinched negative curvature; the peripheral subgroups are the fundamental groups of their cusps ([**Bow97**], [**Far98**]).
(4) Fully residually free groups, also known as limit groups of Sela; they have as peripheral subgroups a finite list of maximal abelian non-cyclic subgroups [**Dah03a**].

Similarly to hyperbolic groups, relatively hyperbolic groups were used to construct examples of infinite finitely generated groups with exotic properties. Denis Osin used in [**Osi10**] direct limits of relatively hyperbolic groups to construct torsion-free two-generated groups with exactly two conjugacy classes (i.e., all elements $\neq 1$ are conjugate to each other).

CHAPTER 10

# Abelian and nilpotent groups

## 10.1. Free abelian groups

DEFINITION 10.1. A group $G$ is called *free abelian* on a generating set $S$ if it is isomorphic to the direct sum

$$\oplus_{s\in S}\mathbb{Z}.$$

The minimal cardinality of $S$ is called *the rank of* $G$ and denoted $\mathrm{rank}\,(G)$, the set $S$ is called a *basis* of $G$.

Of course, if $|S| = n$, $G \cong \mathbb{Z}^n$. Given an abelian group $G$, we define its subgroup

$$G^2 = \{2x | x \in G\}.$$

Clearly, this subgroup is characteristic in $G$. Then, for the free abelian group $G = \oplus_{s\in S}\mathbb{Z}$, the quotient $G/G^2$ is isomorphic to

$$\oplus_{s\in S}\mathbb{Z}_2,$$

which has natural structure of a vector space over $\mathbb{Z}_2$ with basis $S$. Since every two bases of a vector space have the same cardinality, it follows that two bases of a free abelian group have the same cardinality, equal to $\mathrm{rank}\,(G)$.

EXERCISE 10.2. Every free abelian group is torsion-free.

Below is a characterization of free abelian groups by a *universality property*:

THEOREM 10.3. *Let $G$ be an abelian group generated by a set $X$. The group $G$ is free abelian with basis $X$ if and only if it satisfies the following* universality property*: For every abelian group $A$, every map $f : X \to A$ extends to a homomorphism $f : G \to A$.*

PROOF. Suppose that $G$ is free with the basis $X$. Every element $g \in G$ is uniquely represented as a sum

$$g = \sum_{x\in X} c_x \cdot x, c_x \in \mathbb{Z}$$

with only finitely many nonzero terms. Then, we extend $f$ to $G$ by

$$f(g) = \sum_{x\in X} c_x \cdot f(x).$$

Conversely, assume that $G, X$ satisfy the universality property. Let $A$ be free abelian with the basis $X$ and let $f : X \to X$ be the identity map and $f : G \to A$ be the extension. Then, by universality property of free abelian groups, the (identity) map $f^{-1}$ extends to a homomorphism $\bar{f} : A \to G$. Clearly, $\bar{f} = f^{-1}$. □

COROLLARY 10.4. *Let $0 \to A \to B \xrightarrow{r} C \to 0$ be a short exact sequence of abelian groups, where $C$ is free abelian. Then this sequence splits and $B \cong A \oplus C$.*

Proof. Let $c_i, i \in I$, denote a basis of $C$. Then, since $r$ is surjective, for every $c_i$ there exists $b_i \in B$ so that $r(b_i) = c_i$. By the universal property of free abelian groups, the map $s : c_i \to b_i$ extends to a homomorphism $s : C \to B$ so that $r \circ s = Id$. □

The following theorem is an abelian analogue of the Nielsen–Schreier theorem (Theorem 4.46), although, we are unaware of a topological or geometric proof:

Theorem 10.5. *1. Subgroups of free abelian groups are again free abelian.*
*2. If $G < F$ is a subgroup of a free abelian group $F$, then* $\mathrm{rank}\,(G) \leqslant \mathrm{rank}\,(F)$.

Proof. Let $X$ be a basis of a free abelian group $F = A_X$. For each subset $Y$ of $X$ let $A_Y$ be the free group with the basis $Y$, thus $A_Y$ embeds naturally as a free abelian subgroup $A_Y$ in $F_X$. For a subgroup $G < F$ let $G_Y$ denote the intersection $G \cap A_Y$.

Define a set $S$ consisting of triples $(G_Y, B, \phi)$, so that $G_Y$ is free with the basis $B$ and $\phi : B \hookrightarrow X$ is an embedding.

The set $S$ is nonempty, as we can take $Y$ consisting of a single element $x \in X$. Then $G_Y$ is either trivial or infinite cyclic (since $F$ is torsion-free). In both cases, $G_Y$ is free abelian and $B$ is either empty or consists of a single element.

We define a partial order $\leqslant$ on $S$ by:

$$(G_Y, B, \phi) \leqslant (G_Z, C, \psi) \iff Y \subset Z, B \subset C, \quad \phi = \psi|_B.$$

Suppose that $L$ is a chain in the above order indexed by an ordered set $M$:

$$\{(G_{Y_m}, B_m, \phi_m), m \in M\}, (G_{Y_m}, B_m, \phi_m) \leqslant (G_{Y_n}, B_n, \phi_n) \iff m \leqslant n.$$

Then the union

$$\bigcup_{m \in M} G_{Y_m}$$

is again a subgroup in $F$ and the set

$$C = \bigcup_{m \in M} B_m$$

is a basis in the above group. Furthermore, the maps $\phi_m$ determine an embedding $\psi : C \hookrightarrow X$. Thus,

$$(\bigcup_{m \in M} G_{Y_m}, C, \psi) \in S.$$

Therefore, by Zorn's Lemma, there exists a maximal element $(G_Y, B, \phi)$ of $S$. If $Y = X$ then $G_Y = G$ and we are done. Suppose that there exists $x \in X \setminus Y$. Set $Z := Y \cup \{x\}$. We will show that $G_Z$ is still free abelian with a basis $C$ containing $B$ and $\phi$ extends to an embedding $\psi : Z \to X$. If $G_Z = G_Y$, we take $C = B$, $\psi = \phi$. Otherwise, assume that $G_Z/G_Y \neq 0$. The quotient $A_Z/A_Y$ is isomorphic to $\mathbb{Z}$ and generated by the image $\bar{x}$ of $x$. The image of $G_Z$ in this quotient is isomorphic to $G_Z/G_Y$ and is generated by some $n \cdot \bar{x}$, $n \in \mathbb{Z} \setminus 0$. Let $g \in G_Z$ be an element which maps to $n \cdot \bar{x}$. The mapping $G_Z/G_Y \to \langle g \rangle$ splits the sequence

$$0 \to G_Y \to G_Z \to G_Z/G_Y = \mathbb{Z} \to 0$$

and, hence,

$$G_Z \cong G_Y \oplus \langle g \rangle \, .$$

This means that $C := B \cup \{g\}$ is a basis of $G_Z$; we extend $\phi$ to $C$ by $\psi(g) = x$. Thus, $(G_Z, C, \psi) \in S$. This contradicts maximality of $(G_Y, B, \phi)$.

We conclude that $G$ is free abelian and its basis embeds in the basis of $F$. $\square$

### 10.2. Classification of finitely generated abelian groups

THEOREM 10.6. *Every finitely generated abelian group $A$ is isomorphic to a finite direct sum of cyclic groups.*

PROOF. The proof below is taken from **[Mil12]**. The proof is induction on the number of generators of $A$.

If $A$ is 1-generated, the assertion is clear. Assume that the assertion holds for abelian groups with $\leqslant n-1$ generators and suppose that $A$ is an abelian group generated by $n$ elements. Consider all ordered generating sets $(a_1, ..., a_n)$ of $A$. Among such generating sets choose one, $S = (a_1, ..., a_n)$, so that the order of $a_1$ (denoted $|a_1|$) is the least possible. We claim that

$$A \cong \langle a_1 \rangle \oplus A' = \langle a_1 \rangle \oplus \langle a_2, ..., a_n \rangle .$$

(This claim will imply the assertion since, inductively, $A'$ splits as a direct sum of cyclic groups.) Indeed, if $A$ is not the direct sum as above, then we have a nontrivial relation

$$\sum_{i=1}^{n} r_i a_i = 0, r_i \in \mathbb{Z}, r_1 a_1 \neq 0. \qquad (10.1)$$

Without loss of generality, $0 < r_1 < |a_1|$ and $r_i \geqslant 0, i = 1, ...n$ (otherwise, we replace $a_i$'s with $-a_i$ whenever $r_i < 0$). Furthermore, let $d = gcd(r_1, ..., r_n)$ be the greatest common divisor of the numbers $r_i, i = 1, ..., n$. Set $q_i := \frac{r_i}{d}$.

LEMMA 10.7. *Suppose that $a_1, ..., a_n$ are generators of $A$ and $q_1, ..., q_n \in \mathbb{Z}_+$ are such that $gcd(q_1, ..., q_n) = 1$. Then there exists a new generating set $b_1, ..., b_n$ of $A$ so that*

$$b_1 = \sum_{i=1}^{n} q_i a_i.$$

PROOF. Proof of this lemma is a form of the Euclid's algorithm for computation of gcd. Note that $q := q_1 + ... + q_n \geqslant 1$. The proof of lemma is induction on $q$. If $q = 1$ then $b_1 \in \{a_1, ..., a_n\}$ and lemma follows. Suppose the assertion holds for all $q < m$, we will prove the claim for $q = m > 1$. After rearranging the indices, we can assume that $q_1 \geqslant q_2 > 0$.

Clearly, the set $\{a_1, a_1 + a_2, a_3, ..., a_n\}$ generates $A$. Furthermore,

$$gcd(q_1 - q_2, q_2, q_3, ..., q_n) = 1$$

and

$$q' := (q_1 - q_2) + q_2 + q_3 + ... + q_n < m$$

Thus, by the induction hypothesis, there exists a generating set $b'_1, ..., b'_n$ of $A$, where

$$b'_1 = (q_1 - q_2)a_1 + q_2(a_1 + a_2) + q_3 a_3 + ... + q_n a_n.$$

However, $b_1 = b'_1$. Lemma follows. $\square$

In view of this lemma, we get a new generating set $b_1, ..., b_n$ of $A$ so that

$$b_1 = \sum_{i=1}^{n} \frac{r_i}{d} a_i.$$

The equation (10.1) implies that $db_1 = 0$ and $d \leqslant r_1 < |a_1|$. Thus, the ordered generating set $(b_1, ..., b_n)$ of $A$ has the property that $|b_1| < |a_1|$, contradicting our choice of $S$. Theorem follows. $\square$

As the main corollary, we get:

THEOREM 10.8 (classification of abelian groups). *If $G$ is a finitely generated abelian group then there exist an integer $r \geqslant 0$, and $k$-tuples of prime numbers $(p_1, \ldots, p_k)$ and natural numbers $(m_1, \ldots, m_k)$, so that $p_1 \leqslant p_2 \leqslant \ldots \leqslant p_k$, $m_1 \geqslant \ldots \geqslant m_k$, and that*

$$G \simeq \mathbb{Z}^r \times \mathbb{Z}_{p_1^{m_1}} \times \cdots \times \mathbb{Z}_{p_k^{m_k}} . \tag{10.2}$$

*Furthermore, the number $r$, the $k$-tuples $(p_1, \ldots, p_k)$ and $(m_1, \ldots, m_k)$ are uniquely determined by $G$. The number $r$ is called the* rank *of $G$.*

PROOF. By Theorem 10.6, $G$ is isomorphic to the direct product of finitely many cyclic groups $C_1 \times \ldots C_r \times C_{r+1} \times \ldots \times C_n$, where $C_i$ is infinite cyclic for $i \leqslant r$ and finite cyclic for $i > r$.

EXERCISE 10.9. (Chinese remainder theorem) $\mathbb{Z}_s \times \mathbb{Z}_t \cong \mathbb{Z}_{st}$ if and only if the numbers $s, t$ are coprime.

In view of this exercise, we can split every finite cyclic group $C_i$ as a direct product of cyclic groups whose orders are prime powers. This proves existence of the decomposition (10.2).

We now consider the uniqueness part of the theorem.

EXERCISE 10.10. The number $r$ equals the rank of a maximal free abelian subgroup of $G$.

This exercise proves that $r$ is uniquely determined by $G$. Thus, in order to prove uniqueness of $p_i$'s and $m_i$'s it suffices to assume that $G$ is finite. Since the primes $p_i$ are the prime divisors of the order of $G$, the uniqueness question reduces to the case when $|G| = p^\ell$, for some prime number $p$, i.e., $G$ is an *abelian $p$-group*. Suppose that $G$ is an abelian $p$-group and

$$G \cong \mathbb{Z}_{p^{m_1}} \times \cdots \times \mathbb{Z}_{p^{m_k}}, \quad m_1 \geqslant \ldots \geqslant m_k.$$

Set $m = m_1$ and let $m_1 = m_2 = \ldots = m_d > m_{d+1}$. Clearly, the number $p^m$ is the largest order of an element of $G$. The subgroup $G_m$ of $G$ generated by elements of this order is clearly characteristic and equals the $d$-fold direct product of copies of $\mathbb{Z}_{p^m}$

$$\mathbb{Z}_{p^{m_1}} \times \cdots \times \mathbb{Z}_{p^{m_d}}$$

in the above factorization of $G$. Hence, the number $m_k$ and the number $d$ depend only on the group $G$. We then divide $G$ by $G_m$ and proceed by induction. $\square$

Another immediate corollary of Theorem 10.6 is

COROLLARY 10.11. *Every finitely generated abelian group $G$ is* polycyclic*, i.e., $G$ possesses a finite descending series*

$$G = N_0 \geqslant N_1 \geqslant \ldots \geqslant N_n \geqslant N_{n+1} = \{1\} , \tag{10.3}$$

*such that for every $i$ the factor $N_i/N_{i+1}$ is cyclic.*

COROLLARY 10.12. *A finitely generated abelian group is free abelian if and only if it is torsion-free.*

Exercise 10.13. Show that the torsion-free abelian group $\mathbb{Q}$ is not a free abelian group.

A group $T$ is said to be a *torsion group* if every element of $T$ has finite order.

Exercise 10.14. Every finitely generated abelian torsion group is finite.

Note that the for every abelian group $G$, the set $\mathrm{Tor}\,(G)$ of finite-order elements is a subgroup $T$ of $G$, called the *torsion subgroup* $T < G$. This subgroup of $G$ is *characteristic*, i.e., invariant under all automorphisms of $G$.

By applying Theorem 10.6, we get:

Corollary 10.15. *Every finitely generated abelian group $G$ is isomorphic to the direct sum $T \oplus F$, where $T$ is the torsion subgroup of $G$ and $F$ is a free abelian group.*

Corollary 10.16. *Let $G$ be an abelian group generated by $n$ elements. Then every subgroup $H$ of $G$ is finitely generated (with $\leqslant n$ generators).*

Proof. Theorem 10.3 implies that there exists an epimorphism $\phi : F = \mathbb{Z}^n \to G$. Let $A := \phi^{-1}(H)$. Then, by Theorem 10.5, the subgroup $A$ is free of rank $m \leqslant n$. Therefore, $H$ is also $m$-generated. □

Exercise 10.17. Construct an example of a finitely generated abelian group $G$ with the torsion subgroup $T$ and a subgroup $H \leqslant G$, so that there is no direct product decomposition $G = T \times F$ for which $H = (T \cap H) \times (F \cap H)$. Hint: Take $G = \mathbb{Z} \times \mathbb{Z}_2$ and $H$ infinite cyclic.

Exercise 10.18. Let $F$ be a free abelian group of rank $n$ and $B = \{x_1, ..., x_n\}$ be a generating set of $F$. Then $B$ is a basis of $F$.

Classification of finitely generated abelian groups allows one find a simple geometric model for such groups:

Lemma 10.19. *Every finitely generated abelian group $G$ of rank $n$ admits a geometric (in the sense of Definition 3.19) action on $\mathbb{R}^n$, so that every element of $G$ acts as a translation. In particular, $G$ is quasi-isometric to $\mathbb{R}^n$.*

Proof. Let $G = \mathbb{Z}^n \times \mathrm{Tor}\,(G)$. We let $\{e_1, \ldots, e_n\}$ denote a basis of $\mathbb{Z}^n$, we let $\mathbb{R}^n$ be the Euclidean vector space with the basis $e_1, \ldots, e_n$. Then every $g = \sum_{i=1}^n a_i e_n$ acts on $\mathbb{R}^n$ as the translation by the vector $(a_1, \ldots, a_n)$. This action of $\mathbb{Z}^n$ extends to $G$ by declaring that every $g \in \mathrm{Tor}\,(G)$ acts on $\mathbb{R}^n$ trivially. We leave it to the reader to check that this action is geometric and the quotient $\mathbb{R}^n/\mathbb{Z}^n$ is the $n$-torus $T^n$. □

## 10.3. Automorphisms of free abelian groups

Theorem 10.20. *The group of automorphisms of $\mathbb{Z}^n$ is isomorphic to $GL(n, \mathbb{Z})$.*

Proof. Consider the basis $\{e_1, \ldots, e_n\}$ of $\mathbb{Z}^n$, where

$$e_i = (\underbrace{0, \ldots, 0}_{i-1 \text{ times}}, 1, \underbrace{0, \ldots, 0}_{n-i \text{ times}}).$$

Let $\phi : \mathbb{Z}^n \to \mathbb{Z}^n$ be an automorphism. Set

$$\phi(e_i) = \sum_{j=1}^{n} m_{ij} e_j. \qquad (10.4)$$

We, thus, obtain a map $\mu : \phi \mapsto M_\phi = (m_{ij})$, where $M_\phi$ is a matrix with integer entries. We leave it to the reader to check that $\mu(\phi \circ \psi) = M_\phi M_\psi$. It follows that $\mu(\phi) \in GL(n, \mathbb{Z})$ for every $\phi \in Aut(\mathbb{Z}^n)$.

Given a matrix $M \in GL(n, \mathbb{Z})$, we define an endomorphism

$$\phi : \mathbb{Z}^n \to \mathbb{Z}^n,$$

using the equation (10.4). Since the map $\nu : M \mapsto \phi$ respects the composition, it follows that $\nu : GL(n, \mathbb{Z}) \to Aut(\mathbb{Z}^n)$ is a homomorphism and $\mu = \nu^{-1}$. □

Below we establish several properties of automorphisms of free abelian groups that are interesting by themselves and will also be useful later in the proof of the Milnor–Wolf Theorem.

LEMMA 10.21. *Let $v = (v_1, .., v_n) \in G = \mathbb{Z}^n$ be a vector with $\gcd(v_1, \ldots, v_n) = 1$. Then $H = G/\langle v \rangle$ is free abelian of rank $n - 1$. Moreover, there exists a basis $\{y_1, y_2, \ldots, y_{n-1}, v\}$ of $G$ such that $\{y_1 + \langle v \rangle, \ldots, y_{n-1} + \langle v \rangle\}$ is a basis of $H$.*

PROOF. First, let us show that the group $H$ is free abelian; since this group is finitely generated, it suffices to verify that the quotient group is torsion-free. We will use the notation $x \mapsto \bar{x}$ for the quotient map $G \to H$.

Let $\bar{u}$ be an element of finite order $k$ in $H$. Then $ku \in \langle v \rangle$, i.e., $ku = mv$ for some $m \in \mathbb{Z}$. Since $\gcd(v_1, \ldots, v_n) = 1$, it follows that $k|m$ and, hence, $u \in \langle v \rangle$, $\bar{u} = \bar{1}$.

Thus, $H = \mathbb{Z}^n/\langle v \rangle$ is torsion-free, and, hence, is free abelian of finite rank $m$. Next, the homomorphism $G \to H$ extends to a surjective linear map $\mathbb{R}^n \to \mathbb{R}^m$, whose kernel is the line spanned by $v$. Therefore, $m = n - 1$.

Let $\{\bar{x}_1, \ldots, \bar{x}_{n-1}\}$ be a basis on $H$. The map

$$\bar{x}_i \mapsto x_i, i = 1, \ldots, n-1,$$

extends to a group monomorphism $H \to G$; thus, the set $\{x_1, \ldots, x_{n-1}, v\}$ generates $\mathbb{Z}^n$. It follows that $\{x_1, \ldots, x_{n-1}, v\}$ is a basis of $G$. □

LEMMA 10.22. *If a matrix $M$ in $GL(n, \mathbb{Z})$ has all eigenvalues equal to 1 then there exists a finite ascending series of subgroups*

$$\{1\} = \Lambda_0 \leqslant \Lambda_1 \leqslant \cdots \leqslant \Lambda_{n-1} \leqslant \Lambda_n = \mathbb{Z}^n$$

*such that $\Lambda_i \simeq \mathbb{Z}^i$, $\Lambda_{i+1}/\Lambda_i \simeq \mathbb{Z}$ for all $i \geqslant 0$, $M(\Lambda_i) = \Lambda_i$ and $M$ acts on $\Lambda_{i+1}/\Lambda_i$ as identity.*

PROOF. Since $M$ has eigenvalue 1, there exists a vector $v = (v_1, .., v_n) \in \mathbb{Z}^n$ such that $\gcd(v_1, .., v_n) = 1$ and $Mv = v$. Then $M$ induces an automorphism of $H = \mathbb{Z}^n/\langle v \rangle \simeq \mathbb{Z}^{n-1}$ and the matrix $\bar{M}$ of this automorphism has only 1 as an eigenvalue. This follows immediately when writing the matrix of the automorphism $M$ with respect to a basis $\{x_1, x_2, \ldots, x_{n-1}, v\}$ of $\mathbb{Z}^n$ as in Lemma 10.21 and looking at the characteristic polynomial. Now, lemma follows from induction on $n$. □

LEMMA 10.23. *Let $M \in GL(n, \mathbb{Z})$ be a matrix such that each eigenvalue of $M$ has the absolute value 1. Then all eigenvalues of $M$ are roots of unity.*

PROOF. One can derive this lemma from [**BS66**, Theorem 2, p. 105] as follows. Take the number field $K \subset \mathbb{C}$ defined by the characteristic polynomial $p_M$ of $M$. Then each root of $p_M$ belongs to the ring of integers $\mathfrak{O} \subset K$. Since $det(M) = \pm 1$ it follows that each root of $p_M$ is a unit in $\mathfrak{O}$. Then the assertion immediately follows from [**BS66**, Theorem 2, p. 105]. Another proof along the same lines will be given in Section 13.6 using discreteness of the embedding of $F$ into its ring of adeles.

Below we will give an elementary proof, taken from [**Ros74**]. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $M$ listed with multiplicity. Then

$$tr(M^k) = \sum_{i=1}^{n} \lambda_i^k.$$

Since $M \in GL(n, \mathbb{Z})$, the above sums are integers for each $k \in \mathbb{Z}$. Consider the following elements

$$v_k := (\lambda_1^k, \ldots, \lambda_n^k) \in (S^1)^n$$

of the compact group $G = (S^1)^n$. Since the sequence $(v_k)$ contains a convergent subsequence $v_{k_l}$, we obtain

$$\lim_{l\to\infty} v_{k_{l+1}} v_{k_l}^{-1} = 1 \in G.$$

Setting $m_l := k_{l+1} - k_l > 0$, we get

$$\lim_{l\to\infty} \lambda_i^{m_l} = 1, i = 1, \ldots, n.$$

Thus the sequence

$$s_l = \sum_{i=1}^{n} \lambda_i^{m_l}$$

converges to $n$; since $tr(M^{m_l})$ is an integer, it follows $s_l$ is constant (and hence equals $n$) for all sufficiently large $l$. Therefore, for sufficiently large $l$,

$$\sum_{i=1}^{n} Re(\lambda_i^{m_l}) = n. \tag{10.5}$$

Since $|\lambda_i| = 1$,

$$Re(\lambda_i^{m_l}) \leqslant 1, \quad \forall l, i.$$

The equality (10.5) implies that $Re(\lambda_i^{m_l}) = 1$ for all $i$. Thus $\lambda_i^{m_l} = 1$ for all $i$ and all sufficiently large $l$; hence all eigenvalues of $M$ are roots of unity. ☐

LEMMA 10.24. *If a matrix $M$ in $GL(n, \mathbb{Z})$ has one eigenvalue $\lambda$ of absolute value at least 2 then there exists a vector $v \in \mathbb{Z}^n$ such that the following map is injective:*

$$\begin{array}{ccc} \bigoplus_{n\in\mathbb{Z}, n\geqslant 0} \mathbb{Z}_2 & \longrightarrow & \mathbb{Z}^n \\ (s_n)_n & \mapsto & s_0 v + s_1 M v + \ldots + s_n M^n v + \ldots . \end{array} \tag{10.6}$$

PROOF. The matrix $M$ defines an automorphism $\varphi : \mathbb{Z}^n \to \mathbb{Z}^n$, $\varphi(v) = Mv$. The dual map $\varphi^*$ has matrix $M^T$ in the dual canonical basis. Therefore it has the eigenvalue $\lambda$. It follows that there exists a linear form $f : \mathbb{C}^n \to \mathbb{C}$ such that $\varphi^*(f) = f \circ \varphi = \lambda f$.

Take $v \in \mathbb{Z}^n \setminus \operatorname{Ker} f$. Assume that the considered map is not injective. It follows that there exist some $(t_n)_n$, $t_n \in \{-1, 0, 1\}$, such that

$$t_0 v + t_1 M v + \ldots + t_n M^n v + \ldots = 0.$$

Let $N$ be the largest integer such that $t_N \neq 0$. Then

$$M^N v = r_0 v + r_1 M v + \ldots + r_{N-1} M^{N-1} v$$

where $r_i \in \{-1, 0, 1\}$. By applying $f$ to the equality we obtain

$$\left(r_0 + r_1\lambda + \cdots + r_{N-1}\lambda^{N-1}\right) f(v) = \lambda^N f(v),$$

whence

$$|\lambda|^N \leqslant \sum_{i=1}^{N-1} |\lambda|^i = \frac{|\lambda|^N - 1}{|\lambda| - 1} \leqslant |\lambda|^N - 1\,,$$

a contradiction. □

### 10.4. Nilpotent groups

We begin the discussion of nilpotent groups with some useful commutator identities:

LEMMA 10.25. *Let $(G, \cdot)$ be a group and $x, y, z$ elements in $G$. The following identities hold:*

(1) $[x, y]^{-1} = [y, x]$ ;

(2) $[x^{-1}, y] = [x^{-1}, [y, x]]\,[y, x]$ ;

(3) $[x, yz] = [x, y]\,[y, [x, z]]\,[x, z]$ ;

(4) $[xy, z] = [x, [y, z]]\,[y, z]\,[x, z]$ .

PROOF. (1) and (2) are immediate, (4) follows from (3) and (1). It remains to prove (3). Since $[y, [x, z]]\,[x, z] = y[x, z]y^{-1}$ we have that

$$[x, y]\,[y, [x, z]]\,[x, z] = xyx^{-1}[x, z]y^{-1} = xyzx^{-1}z^{-1}y^{-1} = [x, yz]\,.$$

□

NOTATION 10.26. For every $x_1, \ldots, x_n$ in a group $G$ we denote by $[x_1, \ldots, x_n]$ the *$n$-fold left-commutator*

$$[[[x_1, x_2], \ldots, x_{n-1}], x_n].$$

We declare that 1-fold left commutator $[x]$ is simply $x$.

Recall that for subsets $A, B$ in a group $G$, $[A, B]$ denotes the subgroup of $G$ generated by all commutators $[a, b], a \in A, b \in B$. In what follows we also use:

NOTATION 10.27. Given $n$ subgroups $H_1, H_2, \ldots, H_n$ in a group $G$ we denote by $[H_1, \ldots, H_n]$ the subgroup $[\ldots[H_1, H_2], \ldots, H_n] \leqslant G$.

We define *the lower central series* of a group $G$ by

$$C^1G \trianglerighteq C^2G \trianglerighteq \ldots \trianglerighteq C^nG \trianglerighteq \ldots$$

inductively by:

$$C^1G = G\,,\ C^{n+1}G = [G, C^nG]\,.$$

Note that each $C^kG$ is a *characteristic subgroup* of $G$.

DEFINITION 10.28. A group $G$ is called *$k$-step nilpotent* if $C^{k+1}G = \{1\}$. The minimal $k$ for which $G$ is $k$-step nilpotent is called the (nilpotence) *class* of $G$.

EXAMPLES 10.29. (1) Every abelian group is nilpotent of class 1.

(2) The group $\mathcal{U}_n$ of upper triangular $n \times n$ matrices with 1 on the diagonal, is nilpotent of class $n - 1$ (see Exercise 10.30).

(3) The *Heisenberg group*

$$H_{2n+1} = \left\{ \begin{pmatrix} 1 & x_1 & \dots & x_n & z \\ 0 & 1 & 0 & \dots & y_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & y_1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} ; \; x_1, \dots, x_n, y_1, \dots, y_n, z \in \mathbb{R} \right\}$$

is nilpotent of class 2.

If we restrict the parameters $x_1, \dots, x_n, y_1, \dots, y_n, z$ in $H_{2n+1}$ to integers, we obtain the *integer Heisenberg group*

$$H_{2n+1}(\mathbb{Z}) \leqslant H_{2n+1}.$$

The group $H_{2n+1}(\mathbb{Z})$ is finitely generated; we can take as generators the elementary matrices $N_{ij} = I + E_{ij}$ with

$$(i, j) \in \{(1, 2), \dots, (1, n + 1), (2, n), \dots, (n + 1, n)\}.$$

Both groups $H_{2n+1}(\mathbb{Z}), H_{2n+1}$ are nilpotent of class 2. Indeed $C^2 H_{2n+1}$ is the subgroup $x_i = y_i = 0, i = 1, \dots, n$.

Exercise 10.30. The goal of this exercise is to prove that the group $\mathcal{U}_n$ defined above is nilpotent of class $n - 1$.

Let $\mathcal{U}_{n,k}$ be the subset of $\mathcal{U}_n$ formed by matrices $(a_{ij})$ such that $a_{ij} = \delta_{ij}$ for $j < i + k$. Note that $\mathcal{U}_{n,1} = \mathcal{U}_n$ .

(1) Prove that for every $k \geqslant 1$ the map

$$\begin{array}{rcc} \varphi_k : \mathcal{U}_{n,k} & \to & \left(\mathbb{R}^{n-k}\, , +\right) \\ A = (a_{ij}) & \mapsto & (a_{1k+1}, a_{2k+2}, \dots, a_{n-kn}) \end{array}$$

is a homomorphism. Deduce that $(\mathcal{U}_{n,k})' \subset \mathcal{U}_{n,k+1}$ and that $\mathcal{U}_{n,k+1} \lhd \mathcal{U}_{n,k}$ for every $k \geqslant 1$.

(2) Let $E_{ij}$ be the matrix with all entries 0 except the $(i, j)$–entry, which is equal to 1. Consider the triangular matrix $T_{ij}(a) = I + aE_{ij}$.

Deduce from (1), using induction, that $\mathcal{U}_{n,k}$ is generated by the set

$$\{T_{ij}(a) \mid j \geqslant i + k, a \in \mathbb{R}\} .$$

(3) Prove that for every three distinct numbers $i, j, k$ in $\{1, 2, \dots, n\}$

$$[T_{ij}(a), T_{jk}(b)] = T_{ik}(ab) \, , \; [T_{ij}(a), T_{ki}(b)] = T_{kj}(-ab) \, ,$$

and that for all quadruples of distinct numbers $i, j, k, \ell$,

$$[T_{ij}(a), T_{k\ell}(b)] = I \, .$$

(4) Prove that $C^k \mathcal{U}_n \leqslant \mathcal{U}_{n,k+1}$ for every $k \geqslant 0$. Deduce that $\mathcal{U}_n$ is nilpotent.

*Remark.* All the arguments above work also when all matrices have integer entries. In this case (2) implies that $\mathcal{U}_n(\mathbb{Z})$ is generated by $\{T_{ij}(1) \mid j \geqslant i + 1\}$.

Lemma 10.31. *If $S$ be a generating set of a group $G$; then for every $k$ the subgroup $C^k G$ is generated by the $k$-fold left commutators in $S$ and their inverses, together with $C^{k+1} G$.*

Proof. We prove the assertion by induction on $k$. For $k = 0$ the statement is clear, since 1-fold commutators of elements of $S$ are just elements of $S$. Assume that the assertion holds for some $k \geqslant 0$ and consider $C^{k+1}G$.

By definition $C^{k+1}G$ is generated by all commutators $[c_k, g]$ with $c_k \in C^kG$ and $g \in G$. The induction hypothesis implies that $c_k = \ell_1^{\pm 1} \cdots \ell_m^{\pm 1} x$, where $m \in \mathbb{N}$, $\ell_i$ are $k$-fold left commutators in $S$ and $x \in C^{k+1}G$.

According to Lemma 10.25, (4),

$$[c_k, g] = [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1} x, g] = [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, [x, g]][x, g][\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, g] .$$

The first two factors are in $C^{k+2}G$, so it remains to deal with the third.

We write $g = s_1 \cdots s_r$, where $s_i \in S$, and we prove that $[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \cdots s_r]$ is a product of $(k+1)$–fold left commutators in $S$ and their inverses, and of elements in $C^{k+2}G$; our proof is another induction, this time on $m + r \geqslant 2$.

For the case $m + r = 2$ it suffices to note that $[\ell^{-1}, s] = [\ell^{-1}, [s, \ell]][s, \ell]$. The first factor is in $C^{k+2}G$, the second is the inverse of a $(k+1)$–fold left commutator.

Assume that the statement is true for $m + r = n \geqslant 2$. We now prove it for $m + r = k + 1$.

Assume that $m \geqslant 2$. We apply Lemma 10.25, (4), and obtain that

$$[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \dots s_r] = [\ell_1^{\pm 1} \cdots \ell_{m-1}^{\pm 1}, [\ell_m^{\pm 1}, g]]\, [\ell_m^{\pm 1}, s_1 \cdots s_r]\, [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \dots s_r] .$$

The first factor is in $C^{k+2}G$, and for the second and the third the induction hypothesis applies.

Likewise, if $r \geqslant 2$ then we apply Part 3 of Lemma 10.25, , and write

$$[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \cdots s_r] =$$
$$[\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_1 \cdots s_{r-1}]\, [s_1 \cdots s_{r-1}, [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_r]]\, [\ell_1^{\pm 1} \cdots \ell_m^{\pm 1}, s_r] . \quad \square$$

Corollary 10.32. *If $G$ is nilpotent, then $C^nG$ is generated by $k$-fold left commutators in $S$ and their inverses, where $k \geqslant n$.*

Proof. Suppose that $C^{m+1}G = \{1\}$. Then $C^mG$ is generated by the $m$-fold left commutators in $S$ and their inverses. By applying the reverse induction in $n$, each $C^nG$ is generated by the set of all $k$-fold left commutators of elements of $S$ and their inverses, $k \geqslant n$. $\square$

Definition 10.33. Given natural numbers $k$ and $m$, the *$k$-step* $m$-generated *free nilpotent* group is the quotient $N_{m,k}$ of the free group of rank $m$, $F_m$, by the normal subgroup $C^{k+1}F_m$.

Note that the free abelian group of rank $m$ is the 1-step $m$-generated free nilpotent group.

A consequence of Proposition 4.18 is the following.

Proposition 10.34 (Universal property of free nilpotent groups). *Every $k$–step nilpotent group $G$ with $m$ generators, is a quotient of the $m$-generated free $k$–step nilpotent group.*

Proof. Take a generating set $X$ of a $k$–step nilpotent group $G$, so that $X$ has cardinality $m$. The homomorphism $F_X = F_m \to G$ defined in Proposition 4.18 contains $C^{k+1}F(X)$ in its kernel. $\square$

Recall that the center of a group $H$ is denoted $Z(H)$. Given a group $G$, consider the sequence of normal subgroups $Z_i(G) \lhd G$ defined inductively by:

- $Z_0(G) = \{1\}$.
- If $Z_i(G) \lhd G$ is defined and $\pi_i : G \to G/Z_i(G)$ is the quotient map, then

$$Z_{i+1}(G) = \pi_i^{-1}\left(Z(G/Z_i(G))\right).$$

Note that $Z_{i+1}(G)$ is normal in $G$, as the preimage of a normal subgroup of a quotient of $G$. In particular,

$$Z_{i+1}(G)/Z_i(G) \cong Z(G/Z_i(G)).$$

Proposition 10.35. *The group $G$ is $k$-step nilpotent if and only if $Z_k(G) = G$.*

Proof. Assume that $G$ is nilpotent of class $k$. We prove by induction on $i \geqslant 0$ that $C^{k+1-i}G \leqslant Z_i(G)$. For $i = 0$ we have equality. Assume

$$C^{k+1-i}G \leqslant Z_i(G).$$

For every $g \in C^{k-i}G$ and every $x \in G$, $[g, x] \in C^{k+1-i}G \leqslant Z_i(G)$, whence $gZ_i(G)$ is in the center of $G/Z_i(G)$, i.e. $g \in Z_{i+1}(G)$. Thence, the inclusion follows by induction. For $i = k$ the inclusion becomes $C^1G = G \leqslant Z_k(G)$, hence, $Z_k(G) = G$.

Conversely, assume that there exists $k$ such that $Z_k(G) = G$. We prove by induction on $j \geqslant 1$ that $C^jG \leqslant Z_{k+1-j}(G)$. For $j = 1$ the two are equal. Assume that the inclusion is true for $j$. The subgroup $C^{j+1}G$ is generated by commutators $[c, g]$ with $c \in C^jG$ and $g \in G$. Since $c \in C^jG \leqslant Z_{k+1-j}(G)$, by the definition of $Z_{k+1-j}(G)$, the element $c$ commutes with $g$ modulo $Z_{k-j}(G)$, equivalently $[c, g] \in Z_{k-j}(G)$. This implies that $[c, g] \in Z_{k-j}(G)$. It follows that $C^{j+1}G \leqslant Z_{k-j}(G)$.

For $j = k+1$ this gives $C^{k+1}G \leqslant Z_0(G) = \{1\}$, hence $G$ is $k$-step nilpotent. □

Definition 10.36. The normal ascending series

$$Z_0(G) = \{1\} \leqslant Z_1(G) \leqslant \ldots \leqslant Z_i(G) \leqslant Z_{i+1}(G) \leqslant \ldots$$

is called the *upper central series* of the group $G$.

Thus, $G$ is nilpotent if and only if this series is finite, and its nilpotency class is the minimal $k$ such that $Z_k(G) = G$.

Remark 10.37. Yet another equivalent definition a nilpotent group, is to require that the group admits a finite normal series

$$\{1\} = \Gamma_0 \lhd \ldots \Gamma_i \lhd \Gamma_{i+1} \lhd \ldots \Gamma_{n-1} \lhd \Gamma_n = G,$$

so that $\Gamma_{i+1}/\Gamma_i \leqslant Z(G/\Gamma_i)$, or, equivalently, $[G, \Gamma_{i+1}] \leqslant \Gamma_i$. In particular, the quotients $\Gamma_{i+1}/\Gamma_i$ are abelian for each $i$. We will need only the fact that existence of such normal series implies that $G$ is $n$-step nilpotent. Indeed, the condition $\Gamma_{i+1}/\Gamma_i \leqslant Z(G/\Gamma_i)$ implies that $\Gamma_i \leqslant Z_i(G)$ for every $i$. In particular, $G = Z_n(G)$. Now, the assertion follows from Proposition 10.35 We refer to [**Hal76**, Theorem 10.2.2] for further details.

The following useful lemma is a converse to Corollary 10.32:

Lemma 10.38. *Let $S$ be a generating set of a group $G$. Suppose that all $N+1$-fold commutators $[s_1, \ldots, s_{N+1}]$ of elements of $S$ are trivial. Then $G$ is $N$-step nilpotent.*

Proof. Let $G_n$ be the subgroup of $\Gamma$ generated by the $n$-fold commutators $y_n = [s_1, \ldots, s_n]$ of generators $s_i \in S$ of the group $G$. For every generator $x$ of $G$ and every generator $y_n$ of $G_n$ we have:

$$[y_n, x] = y_n x y_n^{-1} x^{-1} \in G_{n+1} \subset G_n.$$

Since $y_n \in G_n$, it follows that $x y_n^{-1} x^{-1} \in G_n$ which implies that $G_n$ is a normal subgroup of $G$.

We claim that for every $n$, $G_{n-1}/G_n$ embeds (under the map induced by inclusion $G_{n-1} \hookrightarrow G$) in the center of $G/G_n$. To simplify the notation, we will regard $G_{n-1}/G_n$ as a subgroup of $G/G_n$. The proof of this statement is the reverse induction on $n$.

The subgroup $G_{N+1}$ is trivial, hence it is contained in the center of $G$. Suppose that the assertion holds for $n = k+1$, we will now prove it for $n = k$. To show that $G_{k-1}/G_k$ is in the center of $G/G_k$ it is enough to verify that for all elements $\bar{z}$ and $\bar{w}$ of generating sets of $G_{n-1}/G_n$ and $G/G_n$ respectively, the commutator $[\bar{z}, \bar{w}]$ is trivial.

The group $G$ is generated by the set $S$, the group $G_{n-1}$ is generated by the $n-1$-fold commutators $y_{n-1}$ of elements $x \in S$. Thus, the groups $G_{n-1}/G_n$ and $G/G_n$ are generated by the projections $\bar{x}, \bar{y}_{n-1}$ of the elements $x$, $y_{n-1}$. By definition of $G_n$ we have: $[y_{n-1}, x] \in G_n$, thus, dividing by $G_n$, we obtain $[\bar{y}_{n-1}, \bar{x}] = 1$. Thus, $G_{n-1}/G_n \leqslant Z(G/G_n)$ for every $n$ and Lemma follows from Remark 10.37. $\square$

Lemma 10.39. (1) *Every subgroup of a nilpotent group is nilpotent;*
(2) *If $G$ is nilpotent and $N \lhd G$ then $G/N$ is nilpotent;*
(3) *Direct product of a set of nilpotent groups is nilpotent.*

Proof. (1) Let $H$ be a subgroup in a nilpotent group $G$. Then $C^i H \leqslant C^i G$. Hence, if $G$ is $k$-step nilpotent then $C^{k+1} H = \{1\}$.

(2) If $\pi : G \to G/N$ is the quotient map, $\pi(C^i G) = C^i(G/N)$.

(3) The assertion follows from

$$C^j(\prod_{i \in I} G_i) = \prod_{i \in I} C^j G_i \ .$$

$\square$

Theorem 10.40. *Every subgroup of a finitely generated nilpotent group is finitely generated.*

Proof. We argue by induction on the class of nilpotency $k$. For $k = 1$ the group is abelian and the statement is already proven in Corollary 10.16. Assume that the assertion holds for $k$, let $G$ be a nilpotent group of class $k+1$ and let $H \subset G$ be a subgroup. By the induction hypothesis $H_1 = H \cap C^2 G$ and $H_2 = H/(H \cap C^2 G)$ are both finitely generated. Thus, $H$ fits in the short exact sequence

$$1 \to H_1 \to H \xrightarrow{\pi} H_2 \to 1,$$

where $H_1, H_2$ are finitely generated. Let $S_1, \bar{S}_2$ be finite generating sets of $H_1, H_2$. For each $\bar{s} \in \bar{S}_2$ pick $s \in \pi^{-1} S_2$. We leave it to the reader to check that the finite set

$$S_1 \cup \{s | \bar{s} \in S_2\}$$

is a generating set of $H$. $\square$

Our next goal is to prove some structural results for nilpotent groups. We begin with several lemmas concerning "calculus of commutators."

LEMMA 10.41. *If $A, B, C$ are three normal subgroups in a group $G$, then the subgroup $[A, B, C] \leqslant G$ is generated by the commutators $[a, b, c]$ with $a \in A, b \in B, c \in C$ and their inverses.*

PROOF. By definition, $[A, B, C]$ is generated by the commutators $[k, c]$ with $k \in [A, B]$ and $c \in C$. The element $k$ is a product $t_1 \cdots t_n$, where each $t_i$ is equal either to a commutator $[a, b]$ or to a commutator $[b, a]$, $a \in A, b \in B$.

We prove, by induction on $n$, that $[k, c]$ is a product of finitely many commutators $[a, b, c]$ and their inverses. For $n = 1$ we only need to consider the case $[t^{-1}, c]$, where $t = [a, b]$. By Lemma 10.25, (2),

$$[t^{-1}, c] = [c, t]^{t^{-1}} = [c^{t^{-1}}, t] = [c', t] = [a, b, c']^{-1}.$$

In the second equality above we applied the identity $\phi([x, y]) = [\phi(x), \phi(y)]$ for the inner automorphism $\phi(x) = x^{t^{-1}}$.

Assume that the statement is true when $k$ is a product of $n$ commutators $t_i$ and consider $k = k_1 t$, where $t$ is equal to either a commutator $[a, b]$ or a commutator $[b, a]$, and $k_1$ is a product of $n$ such commutators. According to Lemma 10.25, (4),

$$[k_1 t, c] = [t, c]^{k_1} [k_1, c].$$

Both factors are products of finitely many commutators $[a, b, c]$ and their inverses, by the induction hypothesis and the fact that $A, B, C$ are normal subgroups and so are invariant under conjugation. □

REMARK 10.42. The same result holds for $[H_1, \ldots, H_n]$ when all $H_i$ are normal subgroups.

LEMMA 10.43 (The Hall identity). *Given a group $G$ and three arbitrary elements $x, y, z$ in $G$, the following identity holds:*

$$\left[x^{-1}, y, z\right]^x \left[z^{-1}, x, y\right]^z \left[y^{-1}, z, x\right]^y = 1. \tag{10.7}$$

PROOF. The factor $\left[x^{-1}, y, z\right]^x$ equals $yxy^{-1}zyx^{-1}y^{-1}xz^{-1}x^{-1}$. The other two factors can be obtained by proper cyclic permutation and a direct calculation shows that all the terms reduce and the product is 1. □

COROLLARY 10.44. *Assume that $A, B, C$ are normal subgroups in $G$. Then*

$$[A, B, C] \leqslant [B, C, A][C, A, B]. \tag{10.8}$$

PROPOSITION 10.45. *Let $C^k G$ be the $k$-th group in the lower central series of a group $G$. Then for every $i, j \geqslant 1$*

$$\left[C^i G, C^j G\right] \leqslant C^{i+j} G. \tag{10.9}$$

PROOF. We prove by induction on $i \geqslant 1$ that for every $j \geqslant 1$,

$$\left[C^i G, C^j G\right] \leqslant C^{i+j} G.$$

For $i = 1$ this follows from the definition of $C^k G$. Assume that the statement is true for $i$. Consider $j \geqslant 1$ arbitrary.

$$[C^{i+1}G, C^j G] = [C^i G, G, C^j G] \leqslant [G, C^j G, C^i G][C^j G, C^i G, G] \leqslant$$
$$[C^{j+1}G, C^i G][C^{j+i}G, G] \leqslant C^{j+i+1}G.$$

□

We now prove that, as for abelian groups, all elements of finite order in a finitely generated nilpotent group form a subgroup, and that subgroup is finite.

LEMMA 10.46. *Let $G$ be a nilpotent group of class $k$. For every $x \in G$ the subgroup generated by $x$ and $C^2G$ is a normal subgroup, and it is nilpotent of class $\leqslant k-1$.*

PROOF. If $x \in C^2G$ then the statement is true.

Assume that $x \notin C^2G$ and let $H = \langle x, C^2G \rangle$. Then

$$H = \{x^m c \mid m \in \mathbb{Z},\, c \in C^2G\}.$$

For every $g \in G$, and $h \in H$, $h = x^m c$, $ghg^{-1} = x^m[x^{-m}, g]gcg^{-1}$, and, since the last two factors are in $C^2G$, the whole product is in $H$.

We now prove that $C^2H \leqslant C^3G$, which will end the proof.

Let $h, h'$ be two elements in $H$, $h = x^m c_1$, $h' = x^n c_2$. Then according to Lemma 10.25, (3),

$$[h, h'] = [h, x^n c_2] = [h, x^n]\,[x^n, [h, c_2]]\,[h, c_2].$$

The last term is in $C^3G$, hence the middle term is in $C^4G$.

For $[h, x^n] = [x^m c_1, x^n]$ we apply Lemma 10.25, (4), and obtain

$$[h, h'] = [x^m, [c_1, x^n]][c_1, x^n].$$

Since the last term is in $C^3G$ and the first in $C^4G$, lemma follows. □

THEOREM 10.47. *Let $G$ be a nilpotent group. The set of all finite order elements forms a characteristic subgroup of $G$, called the torsion subgroup of $G$ and denoted by* $\operatorname{Tor} G$.

PROOF. We argue by induction on the class of nilpotency $k$ of $G$. For $k = 1$ the $G$ group is abelian and the assertion is clear. Assume that the statement is true for all nilpotent groups of class $\leqslant k$, and consider a $(k+1)$–step nilpotent group $G$.

It suffices to prove that for two arbitrary elements $a, b$ of finite order in $G$, the product $ab$ is likewise of finite order. The subgroup $B = \langle b, C^2G \rangle$ is nilpotent of class $\leqslant k$, according to Lemma 10.46. By the induction hypothesis, all the set of finite order elements of $B$ is a characteristic subgroup $\operatorname{Tor} B \leqslant B$. Since $B$ is normal in $G$ it follows that $\operatorname{Tor} B$ is normal in $G$.

Assume that $a$ is of order $m$. Then

$$(ab)^m = aba^{-1}a^2ba^{-2}a^3b \cdots a^{-m+1}a^m ba^{-m}\,,$$

and right-hand side is a product of conjugates of $b$, hence it is in $\operatorname{Tor} B$. We conclude that $(ab)^m$ is of finite order. □

PROPOSITION 10.48. *A finitely generated nilpotent torsion group is finite.*

PROOF. We again argue by induction on the nilpotency class $n$ of the group $G$. For $n = 1$ we apply Exercise 10.14.

Assume that the property holds for all groups of class of nilpotency at most $n$ and consider $G$, a finitely generated torsion group that is $(n+1)$–step nilpotent. Then $C^2G$ and $G/C^2G$ are finite, by the induction hypothesis, whence $G$ is finite. □

COROLLARY 10.49. *Let $G$ be a finitely generated nilpotent group. Then the torsion subgroup* $\operatorname{Tor} G$ *is finite.*

Exercise 10.50. Let $D_\infty$ be the infinite dihedral group.

(1) Give an example of two elements $a, b$ of finite order in $D_\infty$ such that their product $ab$ is of infinite order.

(2) Is $D_\infty$ a nilpotent group ?

(3) Are any of the finite dihedral groups $D_{2n}$ nilpotent?

Lemma 10.51 (A. I. Malcev, [**Mal49a**]). *If $G$ is a nilpotent group with torsion-free center, then:*
*(a) Each quotient $Z_{i+1}(G)/Z_i(G)$ is torsion-free.*
*(b) $G$ is torsion-free.*

Proof. (a) We argue by induction on the nilpotence class $n$ of $G$. The assertion is clear for $n = 1$; assume it holds for all nilpotent groups of class $< n$. We first prove that the group $Z_{n-1}(G)/Z_n(G)$ is torsion-free.

We will show that for each nontrivial element $\bar{x} \in Z_2(G)/Z_1(G)$, there exists a homomorphism $\varphi \in \mathrm{Hom}(Z_2(G)/Z_1(G), Z_1(G))$ such that $\varphi(\bar{x}) \neq 1$. Since $Z_1(G)$ is torsion-free this would imply that $Z_2(G)/Z_1(G)$ is torsion-free. Let $x \in Z_2(G)$ be the element which projects to $\bar{x} \in Z_1(G)/Z_n(G)$. Thus $x \notin Z_1(G)$, therefore there exists an element $g \in G$ such that $[g, x] \in Z_1(G) - \{1\}$. Define the map $\tilde{\varphi} : Z_2(G) \to Z_1(G)$ by:

$$\tilde{\varphi}(y) := [y, g],$$

where $g \in G$ is an element above (so that $[g, x] \neq 1$). Clearly, $\tilde{\varphi}(x) \neq 1$; since $Z_1(G)$ is the center of $G$, the map $\tilde{\varphi}$ descends to a map $\varphi : Z_2(G)/Z_1(G) \to Z_1(G)$. It follows from Part 3 of Lemma 10.25 that $\tilde{\varphi}$ is a homomorphism. Hence, $\varphi$ is a homomorphism as well. Since $Z_n(G)$ is torsion-free, it follows that $Z_2(G)/Z_1(G)$ is torsion-free as well. Now, we replace $G$ by the group $\bar{G} = G/Z_1(G)$.

Since $Z_2(G)/Z_1(G)$ is torsion-free, the group $\bar{G}$ has torsion-free center. Hence, by the induction hypothesis, $Z_{i+1}(\bar{G})/Z_i(\bar{G})$ is torsion-free for every $i$. However,

$$Z_{i+1}(\bar{G})/Z_i(\bar{G}) \cong Z_i(G)/Z_{i-1}(G)$$

for every $i \geqslant 1$. Thus, every group $Z_i(G)/Z_{i-1}(G)$ is torsion-free, proving (a).

(b) In view of (a), for each $i$, $m \neq 0$ and each $x \in Z_i(G) \setminus Z_{i+1}(G)$ we have: $x^m \notin Z_{i+1}(G)$. Thus $x^m \neq 1$. Therefore, $G$ is torsion-free. $\square$

Corollary 10.52. *If $G$ is nilpotent then $G/\mathrm{Tor}\, G$ is torsion-free.*

## 10.5. Discreteness and nilpotence in Lie groups

The goal of this section is to prove theorems of Zassenhaus and Jordan. These theorems deal, respectively, with discrete and finite subgroups $\Gamma$ of Lie groups $G$ (with finitely many components). Theorem of Zassenhaus shows that, appropriately defined, "small elements" of $\Gamma$ generate a nilpotent subgroup of $G$. Jordan's theorem establishes that finite subgroups of $G$ are "almost abelian": Every finite group $\Gamma$ contains an abelian subgroup, whose index in $\Gamma$ is uniformly bounded. Historically, Jordan's theorem was proven first and then, Zassenhaus proved his theorem using similar ideas. We will prove things in the reverse order and we will be using Zassenhaus' results in order to prove Jordan's theorem.

**10.5.1. Zassenhaus neighborhoods.** We begin by defining 'smallness' in a Lie group: "Small" elements will be those which belong to a Zassenhaus neighborhood defined below.

DEFINITION 10.53. Let $G$ be a topological group. A *Zassenhaus neighborhood* in $G$ is an (open) neighborhood of the identity in $G$, denoted $U$ or $U_G$, which satisfies the following:

1. The commutator map sends $U \times U$ to $U$.
2. There exists a continuous function $\sigma : U \to \mathbb{R}$ so that $1 = \sigma^{-1}(0)$ is the point of minimum for $\sigma$ and

$$\sigma([A, B]) < \min(\sigma(A), \sigma(B))$$

for all $A \neq 1, B \neq 1$ in $U$.

Note that if $H < G$ is a topological subgroup and $U_G$ is a Zassenhaus neighborhood of $G$ then $U_H := U_G \cap H$ is a Zassenhaus neighborhood of $H$.

We will see that every Lie group has Zassenhaus neighborhoods. We start with some examples.

LEMMA 10.54. *Let $G = O(V)$ be the orthogonal group of a Hilbert space $V$ (the reader can think of finite-dimensional $V$ since this is the only case that we will need). We equip $End(V)$, the space of bounded linear operators in $V$, with the operator norm and set $\nu(A) := \|A - I\|$ for $A \in G$. Then the set $U$ given by the inequality $\nu(A) < 1/4$ is a Zassenhaus neighborhood in $G$.*

PROOF. We will use the function $\sigma = \nu$ in the definition of the Zassenhaus neighborhood. We will show that for all $A, B \in U \setminus 1$,

$$\nu([A, B]) < \min(\nu(A), \nu(B)),$$

which will also imply that $[\cdot, \cdot] : U \times U \to U$.

First, observe that multiplication by orthogonal transformations preserves the operator norm on $End(V)$. Applying this twice to operators $A, B$ such that $\nu(A) \leq \nu(B)$, we obtain:

$$\nu([A, B]) = \|AB - BA\| = \|(A - B)(A - I) - (A - I)(A - B)\| \leqslant$$
$$\|(A - B)(A - I)\| + \|(A - I)(A - B)\| \leq$$
$$2\nu(A)\|A - B\| = 2\nu(A)(\nu(A) + \nu(B)).$$

Since $\nu(A) \leqslant \nu(B) < 1/4$, we obtain

$$\nu([A, B]) < 2\nu(A)\left(\frac{1}{4} + \frac{1}{4}\right) = \nu(A). \quad \square$$

LEMMA 10.55. *Let $G = GL(V)$ be the general linear group of a Banach space $V$, i.e., group of invertible operators $A$ so that both $A$ and $A^{-1}$ are bounded. We again equip $End(V)$ with the operator norm and set $\sigma(A) := \max(\nu(A), \nu(A^{-1})$ for $A \in G$, where $\nu(A) := \|A - I\|$. Then the set $U$ given by the inequality $\sigma(A) < 1/8$ is a Zassenhaus neighborhood in $G$.*

PROOF. Our proof follows the same lines as in the orthogonal case. We will show that

$$\|ABA^{-1}B^{-1} - I\| < \min(\sigma(A), \sigma(B)).$$

The inequality

$$\|(ABA^{-1}B^{-1})^{-1} - I\| < \min(\sigma(A), \sigma(B))$$

will follow by interchanging $A$ and $B$. We again assume that $\sigma(A) \leq \sigma(B)$. Observe that $\|CD\|le\|C\| \cdot \|D\|$ for all $C, D \in End(V)$. Applying this twice, we get:

$$\|ABA^{-1}B^{-1} - I\| \leq \|B^{-1}\|\|ABA^{-1} - B\| \leq \|A^{-1}\|\|B^{-1}\|\|AB - BA\|.$$

If $\sigma(C) < c$ then $\|C^{-1}\| < 1 + c$ for every $C \in G$. Thus,

$$\|ABA^{-1}B^{-1} - I\| < (1+c)^2\|AB - BA\|$$

provided that $\sigma(A) < c, \sigma(B) < c$. The rest is the same as in the orthogonal case:

$$\|AB - BA\| \leq 2\sigma(A)(\sigma(A) + \sigma(B)) \leq 4\sigma(B)\sigma(A).$$

Putting it all together:

$$\|ABA^{-1}B^{-1} - I\| < 4c(1+c)^2\sigma(A).$$

Since for $c = 1/8$, $4c(1+c)^2 = \frac{1}{2}\left(\frac{9}{8}\right)^2 < 1$, we conclude that

$$\|ABA^{-1}B^{-1} - I\| < \sigma(A).$$

Thus,

$$\sigma([A, B]) < \min(\sigma(A), \sigma(B))$$

for all $A, B \in U$. ☐

REMARK 10.56. The above proofs, at first glance, look like trickery. What is really happening in the proof? Consider $G = GL(n, \mathbb{R})$. The point is then the commutator map has zero 1-st derivative at the point $(1, 1) \in G \times G$ (which one can easily see by using the Taylor expansion $A^{-1} = I - a + a^2$... for a matrix of the form $A = I + a$ where $a$ has small norm). Thus, by the basic calculus, $[A, B]$ will be "closer" to $I \in G$ than $A = I + a$ and $B = I + b$ if $a, b$ are sufficiently small. The above proofs provide explicit estimates for this argument.

We will say that a topological group $G$ *admits a basis of Zassenhaus neighborhoods* if $1 \in G$ admits a basis of topology consisting of Zassenhaus neighborhoods.

COROLLARY 10.57. *Suppose that $G$ is a linear Lie group. Then $1 \in G$ admits a basis of Zassenhaus neighborhoods.*

PROOF. First, suppose that $G = GL(V)$. Then the sets $U_t = \sigma^{-1}(t), t \in (0, \frac{1}{8})$ are Zassenhaus neighborhoods and their intersection is $1 \in G$. If $G$ is a Lie group which admits a continuous closed embedding $\phi : G \to GL(V)$, the sets $\phi^{-1}(U_t)$ will serve as a Zassenhaus basis. ☐

Note that being a subgroup of $GL(n, \mathbb{R})$ is not really necessary for this corollary since the conclusion is *local* at the identity in $G$. One says that a map $\phi : G_2 \to G_1$ is a *local embedding* of topological groups if it is continuous on its domain, is defined on some open neighborhood $U$ of $1 \in G_2$, $\phi(1) = 1$ and

$$\phi(g_1 g_2) = \phi(g_1)\phi(g_2),$$

whenever all three elements $g_1, g_2, g_1 g_2$ belong to $U$. Then, clearly, if $G_1$ admits a basis of Zassenhaus neighborhoods and $G_2$ is a locally compact group which locally embeds in $G_1$, then $G_2$ also admits a basis of Zassenhaus neighborhoods. We will use this trivial observation together with a deep theorem in Lie theory which is a combination of Lie's existence theorem on local homomorphisms of Lie groups induced by homomorphisms of their Lie algebras and Ado's theorem on linearity of finite-dimensional Lie algebras:

THEOREM 10.58. *Every Lie group $G$ admits a local embedding in $GL(V)$ for some finite-dimensional real vector space $V$.*

Note that if $G$ has discrete center, then the adjoint representation of $G$ is a local embedding of $G$ to $GL(\mathfrak{g})$, where $\mathfrak{g}$ is the Lie algebra of $G$. The difficulty is in the case of groups with nondiscrete center.

COROLLARY 10.59. *Every Lie group admits a basis of Zassenhaus neighborhoods.*

Why are Zassenhaus neighborhoods useful? We assume now that $G$ is a locally compact group which admits a basis of Zassenhaus neighborhoods and fix a Zassenhaus neighborhood $\Omega \subset G$ whose closure is compact and is contained in another Zassenhaus neighborhood $U \subset G$. Define inductively subsets $\Omega^{(i)}$ as $\Omega^{(i+1)} = [\Omega, \Omega^{(i)}]$, $\Omega^{(0)} := \Omega$. Since $\Omega$ is Zassenhaus,

$$\Omega^{(i+1)} \subset \Omega^{(i)}$$

for all $i$.

LEMMA 10.60. $E := \bigcap_i \overline{\Omega^{(i)}} = \{1\}$.

PROOF. Clearly, $E$ is compact and $[\Omega, E] = E$. Suppose that $E \neq \{1\}$ and take $f \in E$ with maximal $\sigma(f) > 0$, where $\sigma$ is the function in the definition of the Zassenhaus neighborhood. Then, $f = [g,h], g \in \Omega, h \in E$. Then,

$$\sigma(f) < \min(\sigma(g), \sigma(h)) \leqslant \sigma(h).$$

Contradiction. □

THEOREM 10.61 (Zassenhaus theorem). *Suppose that $G$ is a locally compact group which admits a relatively compact Zassenhaus neighborhood $\Omega$. Assume that $\Gamma < G$ is a discrete subgroup generated by the subset $X := \Gamma \cap \Omega$. Then $\Gamma$ is nilpotent. In particular, this property holds for all Lie groups.*

PROOF. In view of Lemma 10.31, it suffices to show that there exists $n$ so that all $n$-fold iterated commutators of elements of $X$ are trivial. By the definition of $\Omega^{(i)}$, all $i$-fold iterated commutators of $X$ are contained in $\Omega^{(i)}$. Since $\Gamma$ is discrete and $\Omega$ is relatively compact, we can have only finitely many distinct nontrivial iterated commutators of elements of $X$. Since $\bigcap_i \overline{\Omega^{(i)}} = \{1\}$, there exists $n$ so that $\Omega^{(n)}$ is disjoint from all these nontrivial commutators. Thus, all $n$-fold iterated commutators of the elements of $X$ are trivial. Hence, by Lemma 10.38, the group $\Gamma$ is nilpotent. □

In section 10.5.3 we will see how Zassenhaus theorem can be improved in the context of *finite subgroups* of Lie groups (Jordan's theorem).

**10.5.2. Some useful linear algebra.** In this section we discuss some basic linear algebra which will be used for the proof of Jordan's theorem.

Suppose that $V$ is a real Euclidean vector space (e.g., a Hilbert space, but we do not insist on completeness of the norm) with the inner product denoted $x \cdot y$

and the norm denoted $|x|$. We will endow the complexification $V^{\mathbb{C}}$ of $V$ with the Euclidean structure

$$(x + iy) \cdot (u + iv) = x \cdot u + y \cdot v.$$

Recall that the *operator norm* of a linear transformation $A \in End(V)$ is

$$\|A\| := \sup_{|x|=1} |Ax|.$$

Then $A$ extends naturally to a complex-linear transformation of $V^{\mathbb{C}}$ whose operator norm is $\leq \sqrt{2}\|A\|$. In what follows we will use the notation $\nu(A) = \|A - I\|$, the distance from $A$ to the identity in $End(V)$.

LEMMA 10.62. *Suppose that $A \in O(V)$ and $\nu(A) < \sqrt{2}$. Then $A$ is a* rotation *with rotation angles $< \pi/2$, i.e., for every nonzero vector $x \in V$,*

$$Ax \cdot x > 0.$$

PROOF. By assumption,

$$|Ax - x| < \sqrt{2}$$

for all unit vectors $x \in V$. Denoting $y$ the difference vector $Ax - x$, we obtain:

$$2 > y \cdot y = (Ax - x) \cdot (Ax - x) = 2 - 2(Ax \cdot x).$$

Hence, $Ax \cdot x > 0$. ☐

COROLLARY 10.63. *The same conclusion holds for the complexification of $A$.*

PROOF. Let $v = x + iy \in V^{\mathbb{C}}$. Then

$$Av \cdot v = (Ax + iAy) \cdot (x + iy) = Ax \cdot x + Ay \cdot y > 0$$

by the above lemma. ☐

LEMMA 10.64. *Suppose that $A, B \in O(V)$ and $\nu(B) < \sqrt{2}$. Then*

$$[A, BAB^{-1}] = 1 \iff [A, B] = 1.$$

PROOF. One implication is clear, so we assume that $[A, BAB^{-1}]$. Let $\lambda_j$'s be the (complex) eigenvalues of $A$. Then the complexification $V^{\mathbb{C}}$ splits as an $A$-invariant orthogonal sum

$$\oplus_j V^{\lambda_j},$$

where on each $V_j = V^{\lambda_j} = V_A^{\lambda_j}$ the orthogonal transformation $A$ acts *via* multiplication by $\lambda_j$. Here we assume that for $j \neq k$, $\lambda_j \neq \lambda_k$ for $j \neq k$. We refer to this orthogonal decomposition of $V^{\mathbb{C}}$ as $\mathbb{F}_A$. Then, clearly,

$$B(\mathbb{F}_A) = \mathbb{F}_{BAB^{-1}}$$

for any two orthogonal transformations $A, B \in O(V)$. Since $A$ commutes with $BAB^{-1}$, $A$ has to preserve the decomposition $\mathbb{F}_{BAB^{-1}}$ and, moreover, has to send each $W_j := V_{BAB^{-1}}^{\lambda_j} = B(V^{\lambda_j})$ to itself. What are the eigenvalues for this action of $A$ on $W_j$? They are $\lambda_k$'s for which $V^{\lambda_k}$ has nontrivial intersection with $W_j$. However, if $\lambda_j \neq \lambda_k$ then $V^{\lambda_j}$ is orthogonal to $V^{\lambda_k}$ and, hence, by Corollary 10.63, $B$ cannot send a (nonzero) vector from one space to the other. Therefore, in this case, $W_j \cap V_k = 0$. This leaves us with only one choice of the eigenvalue for the restriction $A|W_j$, namely $\lambda_j$. (Since the restriction has to have some eigenvalues!) Thus, $W_j \subset V_j$. However, $B$ sends $V_j$ to $W_j$ injectively, so $W_j = V_j$ and we conclude that $B(V_j) = V_j$. Since $A$ acts on $V_j$ *via* multiplication by $\lambda_j$, it follows that $B|V_j$ commutes with $|V_j$. This holds for all $j$, hence, $[A, B] = 1$. $\square$

Let $U$ denote a Zassenhaus neighborhood of the identity in $O(V)$ so that

$$U \subset \{A : \nu(A) < \sqrt{2}\}.$$

For instance, in view of Lemma 10.54, we can take $U$ given by the inequality $\nu(A) < 1/4$.

LEMMA 10.65. *Suppose that $G$ is a nilpotent subgroup of $O(V)$ generated by elements $A_j \in U$. Then $G$ is abelian.*

PROOF. Consider the lower central series of $G$:

$$G_1 = G, \ldots, G_i = [G, G_{i-1}], i = 1, ..., n,$$

where $G_{n+1} = 1$ and $G_n \neq 1$. We need to show that $n = 1$. Assume not and consider an $(n+1)$-fold iterated commutator of the generators $A_i$ of $G$, it has the form:

$$[[B, A], A] \in G_{n+1} = 1$$

where $A = A_j$ and $B \in G_{n-1}$ is an $n-1$-fold iterated commutator of the generators of $G$. Thus, $A$ commutes with $BAB^{-1}A^{-1}$. Since $A$ commutes with $A^{-1}$, we then conclude that $A$ commutes with $BAB^{-1}$. By the definition of a Zassenhaus neighborhood, if generators $A_i$ of $G$ are in $U$, then all their iterated commutators are also in $U$ and, hence, $B$ satisfies the inequality $\|B - I\| < \sqrt{2}$.

Thus, Lemma 10.64 implies that $A$ commutes with $B$ and, thus, every $n$-fold iterated commutator of generators in $G$ is trivial. Thus, $G_n = 1$ by Lemma 10.31. Contradiction. $\square$

**10.5.3. Jordan's theorem.** Notice that even the group $SO(2)$ contains finite subgroups of arbitrarily high order, but these subgroups, of course, all abelian. Considering the group $O(2)$ we find some non-cyclic subgroups of arbitrarily high order, but they all, of course, contain abelian subgroups of index 2. Jordan's theorem below shows that a similar statement holds for finite subgroups of other Lie groups as well:

THEOREM 10.66 (C. Jordan). *Let $L$ be a Lie group with finitely many connected components. Then there exists a number $q = q(L)$ such that each finite subgroup $F$ in $L$ contains an abelian subgroup of index $\leqslant q$.*

PROOF. If $L$ is not connected, we replace $L$ with $L_0$, the identity component of $L$ and $F$ with $F_0 := F \cap L_0$. Since $|F : F_0| \leqslant |L : L_0|$, it suffices to prove theorem only for subgroups of connected Lie groups. Thus, we assume in what follows that $L$ is connected.

1. We first consider the most interesting case, when the ambient Lie group is $K = O(V)$, the orthogonal group of some finite-dimensional Euclidean vector space.

Let $\Omega$ denote a relatively compact Zassenhaus neighborhood of $O(V)$ given by the inequality

$$\{A : \nu(A) < 1/4\}.$$

The finite subgroup $F \subset K$ is clearly discrete, therefore the subgroup $F' := \langle F \cap \Omega \rangle$ is nilpotent by Zassenhaus theorem. By Lemma 10.65, every nilpotent subgroup generated by elements of $\Omega$ is abelian. It therefore, follows that $F'$ is abelian.

It remains to estimate the index $|F : F'|$. Let $U \subset \Omega$ be a neighborhood of 1 in $K = O(V)$ such that $U \cdot U^{-1} \subset \Omega$ (i.e. products of pairs of elements $xy^{-1}$, $x, y \in U$, belong to $\Omega$). Let $q$ denote $Vol(K)/Vol(U)$, where $Vol$ is induced by the bi-invariant Riemannian metric on $K$.

LEMMA 10.67. *$|F : F'| \leqslant q$.*

PROOF. Let $x_1, \ldots, x_{q+1} \in F$ be distinct coset representatives for $F/F'$. Then

$$\sum_{i=1}^{q+1} Vol(x_i U) = (q+1)Vol(U) > Vol(K).$$

Hence there are $i \neq j$ such that $x_i U \cap x_j U \neq \emptyset$. Thus $x_j^{-1} x_i \in UU^{-1} \subset \Omega$. Hence $x_j^{-1} x_i \in F'$. Contradiction. $\square$

This proves Jordan's theorem for subgroups of $O(V)$.

2. Consider now the case when either $F$ or $L$ has trivial center. Consider the adjoint representation $L \to GL(V)$, where $V$ is the Lie algebra of $L$. This representation need not be faithful, but the kernel $C$ this representation is contained in the center of $L$, see Lemma 3.10. Hence, the kernel $C$ of the homomorphism

$$Ad : F \to \bar{F} \leqslant GL(V)$$

is contained in the center of $F$. Under our assumptions, $C$ is the trivial group and, hence, $F \cong \bar{F}$.

Next, we construct an $F$-invariant Euclidean metric on $V$: Start with an arbitrary positive-definite quadratic form $\mu_0$ on $V$ and then set

$$\mu := \sum_{g \in F} g^*(\mu_0).$$

It is clear that the quadratic form $\mu$ is again positive definite and invariant under $\bar{F}$. With respect to this quadratic form, $F \leqslant O(V)$. Thus, by Step 1, there exists an abelian subgroup $A := F' < F$ of index $\leqslant q(O(V))$.

3. We now consider the general case where we can no longer use elementary arguments. First, by Cartan–Iwasawa–Malcev theorem, see [**Hel01**], every connected locally compact contains unique, up to conjugation, maximal compact subgroup. We find such subgroup $K$ in $L$. By Chevalley's theorem [**Hel01**], every closed subgroup of a Lie group is again a Lie group. Hence, $K$ is also a Lie group. Since $F < L$ is finite, it is compact, and, thus, is conjugate to a subgroup of $K$. Next, every compact Lie group is linear according to a corollary of Peter-Weyl theorem, see e.g. [**OV90**, Theorem 10, page 245]. Thus, we can assume that $K$ is contained in $GL(V)$ for some finite-dimensional vector space $V$. Now, we proceed as in the Part 2. This proves Jordan's theorem. $\square$

REMARK 10.68. The above proof does not provide an explicit bound on $q(L)$. Boris Weisfeiler [**Wei84**] proved for $n > 63$, $q(GL(n, \mathbb{C})) \leqslant (n+2)!$, which is nearly optimal since $GL(n, \mathbb{C})$ contains the permutation group $S_n$ which has order $n!$. Weisfeiler did the work in 1984 shortly before he, tragically, disappeared in Chile in 1985. (On August 21 of 2012 a Chilean judge ordered the arrest of eight retired police and military officers in connection with the kidnapping and disappearance of Weisfeiler.)

CHAPTER 11

# Polycyclic and solvable groups

## 11.1. Polycyclic groups

DEFINITION 11.1. A group $G$ is *polycyclic* if it admits a subnormal descending series

$$(11.1)\quad G = N_0 \rhd N_1 \rhd \ldots \rhd N_n \rhd N_{n+1} = \{1\} \text{ such that } N_i/N_{i+1} \text{ is cyclic } \forall i \geqslant 0\,.$$

A series as in (11.1) is called a *cyclic series*, and its *length* is the number of non-trivial groups in this sequence, this number is $\leqslant n+1$ in (11.1).

If, moreover, $N_i/N_{i+1}$ is infinite cyclic $\forall i \geqslant 0$, then the group $G$ is called *poly-$C_\infty$* and the series is called a *$C_\infty$–series*.

We declare the trivial group to be poly-$C_\infty$ as well.

REMARK 11.2. If $G$ is poly-$C_\infty$ then Corollary 4.21 implies that $N_i \simeq N_{i+1} \rtimes \mathbb{Z}$ for every $i \geqslant 0$; thus, the group $G$ is obtained from $N_n \simeq \mathbb{Z}$ by *successive semi-direct products with $\mathbb{Z}$.*

For general polycyclic groups $G$ the above is no longer true, for instance, $G$ could be a finite group. However, the above property is *almost true* for $G$: *Every polycyclic group contains a normal subgroup of finite index which is poly-$C_\infty$* (see Proposition 11.8).

The following properties are immediate:

PROPOSITION 11.3. (1) *A polycyclic group has bounded generation property in the sense of Definition 4.13. More precisely, let $G$ be a group with a cyclic series* (11.1) *of length $n$ and let $t_i$ be such that $t_iN_{i+1}$ is a generator of $N_i/N_{i+1}$. Then every $g \in G$ can be written as $g = t_1^{k_1} \cdots t_n^{k_n}$, with $k_1, \ldots, k_n$ in $\mathbb{Z}$.*

(2) *A polycyclic torsion group is finite.*

(3) *Any subgroup of a polycyclic group is polycyclic, and, hence, finitely generated.*

(4) *If $N$ is a normal subgroup in a polycyclic group $G$, then $G/N$ is polycyclic.*

(5) *If $N \lhd G$ and both $N$ and $G/N$ are polycyclic then $G$ is polycyclic.*

(6) *Properties (3) and (5) hold with 'polycyclic' replaced by 'poly-$C_\infty$', but not (4).*

PROOF. (1) The statement follows by an easy induction on $n$.

(2) follows immediately from (1).

(3) Let $H$ be a subgroup in $G$. Given a cyclic series for $G$ as above, the intersections $H \cap N_i$ define a cyclic series for $H$.

(4) The proof is by induction on the minimal length $n$ of a cyclic series of $G$. For $n=1$, $G$ is cyclic and any quotient of $G$ is also cyclic.

Assume that the statement is true for all $k \leqslant n$, and consider $G$ for which the length of the shortest cyclic series is $n+1$. Let $N_1$ be the first term distinct from $G$ in this cyclic series. By the induction hypothesis, $N_1/(N_1 \cap N) \simeq N_1N/N$ is polycyclic. The subgroup $N_1N/N$ is normal in $G/N$ and $(G/N)/(N_1N/N) \simeq G/N_1N$ is cyclic, as it is a quotient of $G/N_1$. It follows that $G/N$ is polycyclic.

(5) Consider the cyclic series

$$G/N = Q_0 \geqslant Q_1 \geqslant \cdots Q_n = \{\bar{1}\}$$

and

$$N = N_0 \geqslant N_1 \geqslant \cdots N_k = \{1\}\,.$$

Given the quotient map $\pi : G \to G/N$ and $H_i := \pi^{-1}(Q_i)$, the following is a cyclic series for $G$:

$$G \geqslant H_1 \geqslant \ldots \geqslant H_n = N = N_0 \geqslant N_1 \geqslant \ldots N_k = \{1\}\,.$$

(6) The proofs of properties (3) and (5) with 'polycyclic' replaced by 'poly-$C_\infty$' are identical. A counter-example for (4) with 'polycyclic' replaced by 'poly-$C_\infty$' is $G = \mathbb{Z}$, $N = 2\mathbb{Z}$. □

REMARKS 11.4. (1) If $G$ is polycyclic then, in general, the subset $\operatorname{Tor} G \subset G$ of finite order elements in $G$ is neither a subgroup nor is a finite set.

Consider for instance the infinite dihedral group $D_\infty$. This group can be realized as the group of isometries of $\mathbb{R}$ generated by the symmetry $s : \mathbb{R} \to \mathbb{R}$, $s(x) = -x$ and the translation $t : \mathbb{R} \to \mathbb{R}$, $t(x) = x+1$, and as noted before (see Section 3.6) $D_\infty = \langle t \rangle \rtimes \langle s \rangle$. Therefore $D_\infty$ is polycyclic by Proposition 11.3, (5), but $\operatorname{Tor} D_\infty$ is the union of a left coset and the trivial subgroup:

$$\operatorname{Tor} G = s\,\langle t \rangle \cup \{1\}.$$

(2) Still, something almost as good holds for polycyclic groups: Every polycyclic group is virtually torsion-free (see Proposition 11.8).

PROPOSITION 11.5. *Every finitely generated nilpotent group is polycyclic.*

PROOF. This may be proved using Proposition 11.3, Part (5), and an induction on the nilpotency class or directly, by constructing a series as in (11.1) as follows: Consider the finite descending series with terms $C^kG$. For every $k \geqslant 1$, $C^kG/C^{k+1}G$ is finitely generated abelian. According to the classification of finitely generated abelian groups there exists a finite subnormal descending series

$$C^kG = A_0 \geqslant A_1 \geqslant \cdots \geqslant A_n \geqslant A_{n+1} = C^{k+1}G$$

such that every quotient $A_i/A_{i+1}$ is cyclic. By inserting all these finite descending series into the one defined by $C^kG$'s, we obtain a finite subnormal cyclic series for $G$. □

An edifying example of a polycyclic group is the following.

Proposition 11.6. *Let $m, n \geqslant 1$ be two integers, and let $\varphi : \mathbb{Z}^n \to \mathrm{Aut}(\mathbb{Z}^m)$ be a homomorphism.*

*The semidirect product $G = \mathbb{Z}^m \rtimes_\varphi \mathbb{Z}^n$ is a poly-$C_\infty$ group.*

Proof. The quotient $G/\mathbb{Z}^m$ is isomorphic to $\mathbb{Z}^n$. Therefore by Proposition 11.3, (6), the group $G$ is poly-$C_\infty$. □

Exercise 11.7. Let $\mathcal{T}_n$ be the group of invertible upper triangular $n \times n$ matrices with real entries.

(1) Prove that $\mathcal{T}_n$ is a semi-direct product of its nilpotent subgroup $\mathcal{U}_n$ introduced in Exercise 10.30, and of the subgroup of diagonal matrices.

(2) Prove that the subgroup of $\mathcal{T}_n$ generated by $I + E_{12}$ and by the diagonal matrix with $(-1, 1, \dots, 1)$ on the diagonal is isomorphic to the infinite dihedral group $D_\infty$. Deduce that $\mathcal{T}_n$ is not nilpotent.

Proposition 11.8. *A polycyclic group contains a normal subgroup of finite index which is poly-$C_\infty$.*

Proof. We argue by induction on the length $n$ of the shortest subnormal cyclic series as in (11.1). For $n = 1$ the group $G$ is cyclic and the statement obviously true. Assume that the assertion of Proposition is true for $n$ and consider a polycyclic group $G$ having a cyclic series (11.1).

The induction hypothesis implies that $N_1$ contains a normal subgroup $S$ of finite index which is poly-$C_\infty$. Lemma 3.38 implies that $S$ has a finite index subgroup $S_1$ which is normal in $G$. Proposition 11.3, Part (6), implies that $S_1$ is poly-$C_\infty$ as well.

If $G/N_1$ is finite then $S_1$ has finite index in $G$.

Assume that $G/N_1$ is infinite cyclic. Then the group $K = G/S_1$ contains the finite normal subgroup $F = N_1/S_1$ such that $K/F$ is isomorphic to $\mathbb{Z}$. Corollary 4.21 implies that $K$ is a semidirect product of $F$ and an infinite cyclic subgroup $\langle x \rangle$. The conjugation by $x$ defines an automorphism of $F$ and since $\mathrm{Aut}(F)$ is finite, there exists $r$ such that the conjugation by $x^r$ is the identity on $F$. Hence $F\langle x^r \rangle$ is a finite index subgroup in $K$ and it is a direct product of $F$ and $\langle x^r \rangle$. We conclude that $\langle x^r \rangle$ is a finite index normal subgroup of $K$. We have that $\langle x^r \rangle = G_1/S_1$, where $G_1$ is a finite index normal subgroup in $G$, and $G_1$ is poly-$C_\infty$ since $S_1$ is poly-$C_\infty$. □

Corollary 11.9. (a) *A poly-$C_\infty$ group is torsion-free.*

(b) *A polycyclic group is virtually torsion-free.*

Proof. In view of Proposition 11.8, it suffices to prove (a).We will consider be poly-$C_\infty$ groups $G$. We argue by induction on the minimal length of a cyclic series for a group $G$. For $n = 1$, the group $G$ is infinite cyclic and the statement holds. Assume that the statement is true for all groups with minimal length at most $n$ and consider a group $G$ for which the minimal length of a cyclic series (11.1) is $n+1$. Let $g$ be an element of finite order in $G$. Then its image in the infinite cyclic quotient $G/N_1$ is the identity, hence $g \in N_1$. The induction hypothesis implies that $g = 1$. □

Proposition 11.10. *Let $G$ be a finitely generated nilpotent group. The following are equivalent:*

1. *$G$ is poly-$C_\infty$;*
2. *$G$ is torsion-free;*
3. *the center of $G$ is torsion-free.*

Proof. Implication (1)$\Rightarrow$ (2) is Corollary 11.9, (a), while the implication (2)$\Rightarrow$ (3) is obvious. The implication (3)$\Rightarrow$ (1) follows from Lemma 10.51. $\square$

Remark 11.11. 1. Lemma 10.51 also implies that the upper central series of a torsion-free nilpotent group $G$ satisfies $Z_{i+1}(G)/Z_i(G)$ is torsion-free.

2. In contrast, the lower central series of a (finitely generated) nilpotent torsion-free group may have abelian quotients $C^{i+1}G/C^iG$ with non-trivial torsion. Indeed, given an integer $p \geqslant 2$, consider the following subgroup $G$ of the integer Heisenberg group $H_3(\mathbb{Z})$:

$$G = \left\{ \begin{pmatrix} 1 & k & n \\ 0 & 1 & pm \\ 0 & 0 & 1 \end{pmatrix} \; ; \; k, m, n \in \mathbb{Z} \right\} .$$

Since $H_3(\mathbb{Z})$ is poly-$C_\infty$, so is $G$. On the other hand, the commutator subgroup in $G$ is:

$$G' = \left\{ \begin{pmatrix} 1 & 0 & pn \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \; ; \; n \in \mathbb{Z} \right\} .$$

The quotient $G/G'$ is isomorphic to $\mathbb{Z}^2 \times \mathbb{Z}_p$.

Proposition 11.12. *Every polycyclic group is finitely presented.*

Proof. The proof is an easy induction on the minimal length of a cyclic series, combined with Proposition 4.27. $\square$

One parameter measuring the complexity of the "poly-$C_\infty$ part" of any polycyclic group is the *Hirsch number*, defined as follows.

Proposition 11.13. *The number of infinite factors in a cyclic series of a polycyclic group $G$ is the same for all series. This number is called the* Hirsch number *(or* Hirsch length*) of $G$.*

Proof. The proof will follow from the following observation on cyclic series:

Lemma 11.14. *Any refinement of a cyclic series is also cyclic. Moreover, the number of quotients isomorphic to $\mathbb{Z}$ is the same for both series.*

Proof. Consider a cyclic series

$$H_0 = G \geqslant H_1 \geqslant \ldots \geqslant H_n = \{1\} .$$

A refinement of this series is composed of the following sub-series

$$H_i = R_k \geqslant R_{k+1} \geqslant \ldots \geqslant R_{k+m} = H_{i+1} .$$

Each quotient $R_j/R_{j+1}$ embeds naturally as a subgroup in $H_i/R_{j+1}$, and the latter is a quotient of the cyclic group $H_i/H_{i+1}$; hence all quotients are cyclic. If $H_i/H_{i+1}$ is finite then all quotients $R_j/R_{j+1}$ are finite.

Assume now that $H_i/H_{i+1} \simeq \mathbb{Z}$. We prove by induction on $m \geqslant 1$ that exactly one among the quotients $R_j/R_{j+1}$ is isomorphic to $\mathbb{Z}$, and the other quotients are finite. For $m = 1$ the statement is clear. Assume that it is true for $m$ and consider the case of $m + 1$.

If $H_i/R_{k+m}$ is finite then all $R_j/R_{j+1}$ with $j \leqslant k + m - 1$ are finite. On the other, under this assumption, hand $R_{k+m}/R_{k+m+1}$ cannot be finite, otherwise $H_i/H_{i+1}$ would be finite.

Assume that $H_i/R_{k+m} \simeq \mathbb{Z}$. The induction hypothesis implies that exactly one quotient $R_j/R_{j+1}$ with $j \leqslant k + m - 1$ is isomorphic to $\mathbb{Z}$ and the others are finite. The quotient $R_{k+m}/R_{k+m+1}$ is a subgroup of $H_i/R_{k+m} \simeq \mathbb{Z}$ such that the quotient by this subgroup is also isomorphic to $\mathbb{Z}$. This can only happen when $R_{k+m}/R_{k+m+1}$ is trivial. □

Proposition 11.13 now follows from Lemmas 11.14 and 3.34. □

A natural question to ask is the following.

QUESTION 11.15. Since poly-$C_\infty$ groups are constructed by successive semi-direct products with $\mathbb{Z}$, is there a way to detect during this construction whether the group is nilpotent or not?

The answer to this question will be given in Section 12.6 and it has some interesting relation to the growth of groups.

### 11.2. Solvable groups: Definition and basic properties

Recall that $G'$ denotes the *derived subgroup* $[G, G]$ of the group $G$. Given a group $G$, we define its *iterated commutator subgroups* $G^{(k)}$ inductively by:

$$G^{(0)} = G, G^{(1)} = G', \dots, G^{(k+1)} = \left(G^{(k)}\right)', \dots$$

The descending series

$$G \trianglerighteq G' \trianglerighteq \dots \trianglerighteq G^{(k)} \trianglerighteq G^{(k+1)} \trianglerighteq \dots$$

is called the *derived series* of the group $G$.

Note that all subgroups $G^{(k)}$ are *characteristic* in $G$.

DEFINITION 11.16. A group $G$ is *solvable* if there exists $k$ such that $G^{(k)} = \{1\}$. The minimal $k$ such that $G^{(k)} = \{1\}$ is called the *derived length* of $G$. A solvable group of derived length at most two is called *metabelian*.

In particular, every solvable group $G$ of derived length $k$ satisfies the law:

$$\llbracket x_1, \dots, x_{2^k} \rrbracket = 1, \forall x_1, \dots x_{2^k} \in G. \tag{11.2}$$

Here and in what follows,

$$\llbracket x_1, \dots, x_{2^k} \rrbracket := [(x_1, \dots, x_{2^{k-1}}), (x_{2^{k-1}+1}, \dots, x_{2^k})] \tag{11.3}$$

and $\llbracket x_1, x_2 \rrbracket = [x_1, x_2]$.

EXERCISE 11.17. Find the values $n \in \mathbb{N}$ for which the symmetric group $S_n$ is solvable.

Exercise 11.18. Suppose that $G$ is the direct limit of a family of groups $G_i, i \in I$. Assume that there exist $k, m \in \mathbb{N}$ so that for every $i \in I$, the group $G_i$ contains a solvable subgroup $H_i$ of index $\leqslant k$ and derived length $\leqslant m$. Then $G$ is also virtually solvable: It contains a subgroup $H$ of index $\leqslant k$ and derived length $\leqslant m$. Hint: Use Exercise 1.5.

Proposition 11.19. (1) *If $N$ is a normal subgroup in $G$ and both $N$ and $G/N$ are solvable then $G$ is solvable.*

(2) *Every subgroup of a solvable group is solvable.*

(3) *If $G$ is solvable and $N \lhd G$ then $G/N$ is solvable.*

Note that the statement (1) is not true when 'solvable' is replaced by 'nilpotent'. This can be seen for instance from Proposition 12.26.

Proof. (1) Assume that $G/N$ is solvable of derived length $d$ and $N$ is solvable of derived length $m$. Since $(G/N)^{(d)} = \{\bar{1}\}$ it follows that $G^{(d)} \leqslant N$. Then, as $G^{(d+i)} \leqslant N^{(i)}$, it follows that $G^{(d+m)} = \{1\}$.

(2) Note that for every subgroup $H$ of a group $G$, $H' \leqslant G'$. Thus, by induction,

$$H^{(i)} \leqslant G^{(i)}.$$

If $G$ is solvable of derived length $k$ then $G^{(k)} = \{1\}$; thus $H^{(k)} = \{1\}$ as well and, hence, $H$ is also solvable.

(3) Consider the quotient map $\pi : G \to G/N$. It is immediate that $\pi\left(G^{(i)}\right) = (G/N)^{(i)}$, in particular if $G$ is solvable then $G/N$ is solvable. □

Exercise 11.20. (1) Prove that the group $\mathcal{T}_n$ of upper triangular $n \times n$ matrices in $GL(n, \mathbb{K})$, where $K$ is a field, is solvable. [Hint: you may use Exercise 10.30]

(2) Let $V$ be a $\mathbb{K}$-vector space of dimension $n$, and let

$$V_0 = 0 \subset V_1 \subset \cdots \subset V_{n-1} \subset V_n = V$$

be a *complete flag*, that is a flag such that each subspace $V_i$ has dimension $i$. Prove that the subgroup $G$ of $GL(V)$ consisting of elements $g$ such that $gV_i = V_i, \forall i$, is a solvable group.

(3) Let $V$ be a $\mathbb{K}$-vector space of dimension $n$, and let

$$V_0 = 0 \subset V_1 \subset \cdots \subset V_{k-1} \subset V_k = V$$

be a flag, not necessarily complete. Let $G$ be a subgroup of $GL(V)$ such that $gV_i = V_i$, for every $g \in G$ and every $i$. For every $i \in \{1, 2, \ldots, k-1\}$ let $\rho_i$ be the projection $G \to GL(V_{i+1}/V_i)$. Prove that if every $\rho_i(G)$ is solvable, then $G$ is also solvable.

Exercise 11.21. 1. Let $\mathbb{F}_k$ denote the field with $k$ elements. Use the 1-dimensional vector subspaces in $\mathbb{F}_k^2$ to construct a homomorphism $GL(2, \mathbb{F}_k) \to S_n$ for an appropriate $n$.

2. Prove that $GL(2, \mathbb{F}_2)$ and $GL(2, \mathbb{F}_3)$ are solvable.

## 11.3. Free solvable groups and Magnus embedding

As in the case of nilpotent groups, there exist universal objects in the class of solvable groups that we now describe.

DEFINITION 11.22. Given two integers $k, m \geqslant 1$, the *free solvable group of derived length $k$ with $m$ generators* is the quotient of the free group $F_m$ by the normal subgroup $F_m^{(k)}$.

When $k = 2$ we call the corresponding group *free metabelian group with $m$ generators*.

NOTATION 11.23. In what follows we use the notation $S_{m,k}$ for the free solvable group of derived length $k$ and with $m$ generators. Note that $S_{m,1}$ is $\mathbb{Z}^m$.

PROPOSITION 11.24 (Universal property of free solvable groups). *Every solvable group with $m$ generators and of derived length $k$, is a quotient of $S_{m,k}$.*

PROOF. Let $G$ be a solvable group of derived length $k$ and let $X$ be a generating set of $G$ of cardinality $m$. $X$ of cardinality $m$. The map defined in Proposition 4.18 contains $F(X)^{(k)}$ in its kernel, therefore it defines an epimorphism from the free solvable group $S_{m,k}$ to $G$. □

Our next goal is to define the *Magnus embedding* of the free solvable group $S_{r,k+1}$ into the wreath product $\mathbb{Z}^r \wr S_{r,k}$. Since $\mathbb{Z}^r \wr S_{r,k}$ is a semidirect product, Remark 3.76, (2), implies that in order to define a homomorphism

$$S_{r,k+1} \to \mathbb{Z}^r \wr S_{r,k}$$

one has to specify a homomorphism $\pi : S_{r,k+1} \to S_{r,k}$ and a derivation $d \in Der(S_{r,k+1}, \bigoplus_{S_{r,k}} \mathbb{Z}^r)$. Here we will use the following action of $S_{r,k+1}$ on $\bigoplus_{S_{r,k}} \mathbb{Z}^r$: We compose $\pi$ with the action of $S_{r,k}$ on itself *via* left multiplication.

To simplify the notation, we let $F = F_r$ denote the free group on $r$ generators $x_1, \ldots, x_r$. First, since $F/F^{(m)} = S_{r,m}$ for every $m$, and $F^{(k+1)} \leqslant F^{(k)}$, we have a natural quotient map

$$\pi : S_{r,k+1} \to S_{r,k}.$$

We now proceed to construct the derivation $d$. We will use definitions and results of Section 3.10. Note that $\bigoplus_{S_{r,k}} \mathbb{Z}^r$ is isomorphic (as a free abelian group) to

$$M_1 \oplus \ldots \oplus M_r$$

where for every $i$, $M_i = M = \mathbb{Z}S_{r,k}$, the group algebra of $S_{r,k}$. Since $S_{r,k}$ is the quotient of $F = F_r$, every derivation $\partial \in Der(\mathbb{Z}F, \mathbb{Z}F)$ projects to a derivation (denoted $\widehat{\partial}$) in $Der(\mathbb{Z}F, \mathbb{Z}S_{r,k})$. Thus, derivations $\partial_i \in Der(\mathbb{Z}F, \mathbb{Z}F)$ introduced in Section 3.10, projects to derivations $\widehat{\partial}_i \in Der(\mathbb{Z}F, M)$. Furthermore, every derivation $\widehat{\partial}_i \in Der(\mathbb{Z}F, M)$ extends to a derivation $d_i : \mathbb{Z}F \to \bigoplus_{S_{r,k}} \mathbb{Z}^r$ by

$$d_i : w \mapsto (0, \ldots, \widehat{\partial}_i(w), \ldots 0)$$

where we place $\widehat{\partial}_i(w)$ in the $i$-th slot. Since sum of derivations is again a derivation, we obtain a derivation

$$d = (\widehat{\partial}_1, \ldots, \widehat{\partial}_r) = d_1 + \ldots + d_r \in Der(\mathbb{Z}F, \bigoplus_{S_{r,k}} \mathbb{Z}^r).$$

For simplicity, in what follows, we denote $F^{(k)}$ by $N$ and, accordingly, $F^{(k+1)}$ by $N'$. Thus, $S_{r,k} = F/N$ and $S_{r,k+1} = F/N'$.

LEMMA 11.25. *The derivation $d$ projects to a derivation*

$$\bar{d} \in Der(\mathbb{Z}S_{r,k+1}, \bigoplus_{S_{r,k}} \mathbb{Z}^r).$$

PROOF. Let us check that $N'$ is in the kernel of $d$. Indeed, given a commutator $[x, y]$ with $x, y$ in $N$, property $(P_3)$ in Exercise 3.72 implies that (by computing in $\mathbb{Z}F$)

$$\partial_i[x, y] = (1 - xyx^{-1})\partial_i x + x(1 - yx^{-1}y^{-1})\partial_i y\,.$$

Since both $x, y \in N$ project to 1 in $S_{r,k}$, they act trivially on $M = \mathbb{Z}S_{r,k}$, it follows that

$$(1 - xyx^{-1}) \cdot \xi = 0 \text{ and } x(1 - yx^{-1}y^{-1}) \cdot \eta = 0, \quad \forall \xi, \eta \in M.$$

Hence, $d_i([x, y]) = 0$ for every $i$ and, thus, $d([x, y]) = 0$. Therefore, $d(N') = 0$ since the group $N'$ is generated by commutators $[x, y], x, y \in N$. For arbitrary $g \in F, h \in N'$, we have

$$d(gn) = d(g) + g \cdot d(n) = d(g).$$

Thus, the derivation $d$ projects to a derivation $\bar{d} \in Der((\mathbb{Z}S_{r,k+1}, \bigoplus_{S_{r,k}} \mathbb{Z}^r)$,

$$\bar{d}(gN') = d(g). \quad \square$$

Thus, according to Remark 3.76, the pair $(d, \pi)$ determines a homomorphism

$$\mathfrak{M} : S_{r,k+1} \to \mathbb{Z}^r \wr S_{r,k}\,.$$

THEOREM 11.26 (W. Magnus [**Mag39**]). *The homomorphism $\mathfrak{M}$ is injective; $\mathfrak{M}$ is called the* Magnus embedding.

We refer to [**Fox53**, Section (4.9)] for the proof of injectivity of $\mathfrak{M}$. Remarkably, the Magnus embedding also has nice geometric features:

THEOREM 11.27 (A. Sale [**Sal12**]). *The Magnus embedding is a quasi-isometric embedding.*

Clearly, the Magnus embedding is a useful tool for studying free solvable groups by induction on the derived length.

## 11.4. Solvable versus polycyclic

PROPOSITION 11.28. *Every polycyclic group is solvable.*

PROOF. This follows immediately by the induction argument on the length $n$ of a series as in (11.1), and Part (1) of Proposition 11.19. $\square$

DEFINITION 11.29. A group is said to be *Noetherian* if for every increasing sequence of subgroups

$$H_1 \leqslant H_2 \leqslant \cdots \leqslant H_n \leqslant \cdots \tag{11.4}$$

there exists $n_0$ such that $H_{n_0} = H_n$ for every $n \geqslant n_0$.

PROPOSITION 11.30. *A group $G$ is Noetherian if and only if every subgroup of $G$ is finitely generated.*

Proof. Assume that $G$ is a Noetherian group, and let $H \leqslant G$ be a subgroup which is not finitely generated. Pick $h_1 = H \setminus \{1\}$ and let $H_1 = \langle h_1 \rangle$. Inductively, assume that

$$H_1 < H_2 < ... < H_n$$

is a strictly increasing sequence of finitely generated subgroups of $H$, pick $h_{n+1} \in H \setminus H_n$, and set $H_{n+1} = \langle H_n, h_{n+1} \rangle$. We thus have a strictly increasing infinite sequence of subgroups of $G$, contradicting the assumption that $G$ is Noetherian.

Conversely, assume that all subgroups of $G$ are finitely generated, and consider an increasing sequence of subgroups as in (11.4). Then $H = \bigcup_{n \geqslant 1} H_n$ is a subgroup, hence generated by a finite set $S$. There exists $n_0$ such that $S \subseteq H_{n_0}$, hence $H_{n_0} = H = H_n$ for every $n \geqslant n_0$. □

Proposition 11.31. *A solvable group is polycyclic if and only if it is Noetherian.*

Proof. The 'only if' part follows immediately from Parts (1) and (3) of Proposition 11.3. Let $S$ be a Noetherian solvable group. We prove by induction on the derived length $k$ that $S$ is polycyclic.

For $k = 1$ the group is abelian, and since, by hypothesis, $S$ is finitely generated, it is polycyclic.

Assume that the statement is true for $k$ and consider a solvable group $G$ of derived length $k + 1$. The commutator subgroup $G' \leqslant G$ is also Noetherian and solvable of derived length $k$. Hence, by the induction hypothesis, $G'$ is polycyclic. The abelianization $G_{ab} = G/G'$ is finitely generated (because $S$ is, by hypothesis), hence it is polycyclic. It follows that $S$ is polycyclic by Proposition 11.3 (5). □

By Proposition 11.28 every nilpotent group is solvable. A natural question to ask is to find a relationship between nilpotency class and derived length.

Proposition 11.32. (1) *For every group $G$ and every $i \geqslant 0$, $G^{(i)} \leqslant C^{2^i} G$.*

(2) *If $G$ is a $k$-step nilpotent group then its derived length is at most*

$$[\log_2 k] + 1 .$$

Proof. (1) The statement is obviously true for $i = 0$. Assume that it is true for $i$. Then $G^{(i+1)} = \left[G^{(i)}, G^{(i)}\right] \leqslant \left[C^{2^i} G, C^{2^i} G\right] \leqslant C^{2^{i+1}} G$. In the last inclusion we applied Proposition 10.45.

(2) follows immediately from (1). □

Remark 11.33. The derived length can be much smaller than the nilpotency class: the dihedral subgroup $D_{2n}$ with $n = 2^k$ is $k$-step nilpotent and metabelian.

An instructive example of solvable group is the *lamplighter group*. This group is the wreath product $G = \mathbb{Z}_2 \wr \mathbb{Z}$ in the sense of Definition 3.65.

Exercise 11.34. Prove that if $K, H$ are solvable groups then $K \wr H$ is solvable. In particular, the lamplighter group $G$ is solvable (even metabelian).

We will see in the next section that if both $K$ and $H$ are finitely generated, then $K \wr H$ is also finitely generated. In particular, the lamplighter group is finitely generated. On the other hand:

(1) *Not all subgroups in the lamplighter group $G$ are finitely generated:* the subgroup $\bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$ of $G$ is not finitely generated.

(2) *The lamplighter group $G$ is not virtually torsion-free:* For any finite index subgroup $H \leqslant G$, $H \cap \bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$ has finite index in $\bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$; in particular this intersection is infinite and consists of elements of order at most 2.

Both (1) and (2) imply that the lamplighter group is *not polycyclic*.

(3) The commutator subgroup $G'$ of the lamplighter group $G$ coincides with the following subgroup of $\bigoplus_{n\in\mathbb{Z}} \mathbb{Z}_2$:

$$C = \{f : \mathbb{Z} \to \mathbb{Z}_2 \mid \mathrm{Supp}(f) \text{ has even cardinality}\}, \tag{11.5}$$

where $\mathrm{Supp}(f) = \{n \in \mathbb{Z} \mid f(n) = 1\}$.

[*NB. The notation here is additive, the identity element is* 0.]

In particular, $G'$ *is not finitely generated* and the group $G$ is metabelian (since $G'$ abelian).

We prove (3). First of all, $C$ is clearly a subgroup. Note also that

$$(f, m)^{-1} = (-\varphi(-m)f, -m),$$

where $\varphi$ is the action of $\mathbb{Z}$ on the space of functions $f : \mathbb{Z} \to \mathbb{Z}_2$. A straightforward calculation gives

$$[(f, m), (g, n)] = (f - g - \varphi(n)f + \varphi(m)g \,,\, 0).$$

Now, observe that either $\mathrm{Supp}(f)$ and $\mathrm{Supp}(\varphi(n)f)$ are disjoint, in which case $\mathrm{Supp}(f - \varphi(n)f)$ has cardinality twice the cardinality of $\mathrm{Supp}\, f$, or they overlap on a set of cardinality $k$; in the latter case, $\mathrm{Supp}(f - \varphi(n)f)$ has cardinality twice the cardinality of $\mathrm{Supp}\, f$ minus $2k$. The same holds for $\mathrm{Supp}(-g + \varphi(m)g)$. Since $C$ is a subgroup,

$$(f - g - \varphi(n)f + \varphi(m)g) = (f - \varphi(n)f - (g - \varphi(m)g)) \in C.$$

This shows that $G' \leqslant C$.

Consider the opposite inclusion. The subgroup $C$ is generated by functions $f : \mathbb{Z} \to \mathbb{Z}_2\,, \mathrm{Supp}\, f = \{a, b\}$, where $a, b$ are distinct integers; thus, it suffices to show that $(f, 0) \in G'$. Let $\delta_a : f : \mathbb{Z} \to \mathbb{Z}_2\,,\ \mathrm{Supp}\, \delta_a = \{a\}$. Then

$$[(\delta_a, 0), (0, b - a)] = (\delta_a - \varphi(b - a)\delta_a, 0) = (f, 0)$$

which implies that $(f, 0) \leqslant G'$.

We conclude this section by noting that, unlike polycyclic groups, solvable groups may not be finitely presented. An example of such group is the wreath product $\mathbb{Z} \wr \mathbb{Z}$ [**Bie79**]. We refer to the same paper for a survey on finitely presented solvable groups. Nevertheless, a solvable group may be finitely presented without being polycyclic; for instance the Baumslag-Solitar group

$$G = BS(1, p) = \left\langle a, b | aba^{-1} = b^p \right\rangle$$

is metabelian but not polycyclic (for $|p| \geqslant 2$). The derived subgroup $G'$ of $G$ is isomorphic to the additive group of $p$-adic rational numbers, i.e., rational numbers whose denominators are powers of $p$. In particular, $G'$ is not finitely generated. Hence, in view of Proposition 11.3, $G$ is not polycyclic.

EXERCISE 11.35. Show that $G = BS(1, p)$ is metabelian.

## 11.5. Failure of QI rigidity for solvable groups

THEOREM 11.36 (A. Dyubina (Erschler), [**Dyu00**]). *The class of finitely generated (virtually) solvable groups is not QI rigid.*

PROOF. The groups that will be used in the proof are wreath products. Recall that the *wreath product* $A \wr C$ of groups $A$ and $C$ is the semidirect product

$$\oplus_C A \rtimes C$$

where $C$ acts on the direct sum by precompositions: $f(x) \mapsto f(xc^{-1})$. Thus, elements of wreath products $A \wr C$ are pairs $(f, c)$, where $f : C \to A$ is a function with finite support and $c \in C$. The product structure on this set is given by the formula

$$(f_1(x), c_1) \cdot (f_2(x), c_2)) = (f_1(xc_2^{-1}) f_2(x), c_1 c_2).$$

Here and for the rest of the proof we use multiplicative notation when dealing with wreath products.

If $a_i, i \in I, c_j, j \in J$ are generators of $A$ and $C$ respectively, the elements $(1, c_j), j \in J$ and $(\delta_{a_i}, 1), i \in I$, will generate $G_A := A \wr C$. Here $\delta_a : C \to A$ is the function which sends $1 \in C$ to $a \in A$ and sends all other elements of $C$ to $1 \in A$. We will always equip $G_A$ with such generating sets. In particular, if $A$ and $C$ are finitely generated, so is $A \wr C$.

Suppose now that $A, B$ are finite groups of the same order, where $A$ is abelian, say, cyclic, and $B$ is a non-solvable group. For instance, we can take $A$ to be the alternating group $A_5$ (which is a simple nonabelian group of order 60) and $B = \mathbb{Z}_{60}$. We declare each nontrivial element of these groups to be a generator. Let $C$ be a finitely generated infinite abelian group, say, $\mathbb{Z}$, and consider the wreath products $G_A := A \wr C, G_B := B \wr C$. Let $\varphi : A \to B$ be a bijection sending 1 to 1. Extend this bijection to a map

$$\Phi : G_A \to G_B, \quad \Phi(f, c) = (\varphi \circ f, c).$$

LEMMA 11.37. *$\Phi$ extends to an isometry of Cayley graphs.*

PROOF. First, the inverse map $\Phi^{-1}$ is given by $\Phi(f, c) = (\varphi^{-1} \circ f, c)$. We now check that $\Phi$ preserves edges of the Cayley graphs. The group $G_A$ has two types of generators: $(1, c_j)$ and $(\delta_a, 1)$, where $c_j \in X$, a finite generating set of $C$ and $a \in A$ are all nontrivial elements of $A$. The same holds for the group $G_B$.

1. Consider the edges connecting $(f(x), c)$ to $(f(xc_j^{-1}), cc_j)$ in $Cayley(G_A)$. Applying $\Phi$ to the vertices of such edges we obtain

$$(\varphi \circ f(x), c), \quad (\varphi \circ f(xc_j^{-1}), cc_j).$$

Clearly, they are again within unit distance in $Cayley(G_B)$, since they differ by $(1, c_j)$.

2. Consider the edges connecting $(f(x), c), (f(x)\delta_a(x), c)$ in $Cayley(G_A)$. Applying $\Phi$ to the vertices we obtain

$$(\varphi \circ f(x), c), \quad (\varphi \circ f(x)\delta_b(x), c),$$

where $b = \varphi(a)$. Again, we obtain vertices which differ by $(\delta_b, 1)$, so they are within unit distance in $Cayley(G_B)$ as well. ☐

LEMMA 11.38. *The group $G_B$ is not virtually solvable.*

Proof. Let $\psi : G_B \to F$ be a homomorphism to a finite group. Then the kernel of the restriction $\psi|\oplus_C B$ is also solvable. The restriction of $\psi$ to the subgroup $B_c < \oplus_C B$ consisting of maps $f : C \to B$ supported at $\{c\}$, is determined by a homomorphism $\psi_c : B \to F$. There are only finitely many of such homomorphisms, while $C$ is an infinite group. Thus, we find $c_1 \neq c_2 \in C$ such that

$$\psi_{c_1} = \psi_{c_2} = \eta.$$

The kernel of the restriction $\psi|B_{c_1} \oplus B_{c_2}$ consists of pairs

$$(b_1, b_2) \in B_{c_1} \oplus B_{c_2} = B \times B, \quad \eta(b_1) = \eta(b_2)^{-1}$$

and contains the subgroup

$$\{(b, b^{-1}), b \in B\}.$$

However, this subgroup is isomorphic to $B$ (*via* projection to the first factor in the product $B \times B$). Thus, kernel of $\psi$ contains a subgroup isomorphic to $B$ and, hence is not solvable. □

Combination of these two lemmas implies the theorem. □

CHAPTER 12

# Growth of nilpotent and polycyclic groups

## 12.1. The growth function

Suppose that $X$ is a discrete metric space (see Definition 1.19) and $x \in X$ is a base-point. We define the *growth function*

$$\mathfrak{G}_{X,x}(R) := \text{card } \bar{B}(x,R),$$

the cardinality of the closed $R$-ball centered at $x$.

Given a simplicial complex $\mathcal{G}$ and a vertex $v$ as a base-point, the *growth function* of $\mathcal{G}$ is the growth function of its set of vertices with base-point $v$.

LEMMA 12.1 (Equivalence class of growth is QI invariant.). *Suppose that $f : (X,x) \to (Y,y)$ is a quasi-isometry (in the sense of Definition 5.6). Then $\mathfrak{G}_{X,x} \asymp \mathfrak{G}_{Y,y}$.*

PROOF. Let $\bar{f}$ be a coarse inverse to $f$ (in the sense of Definition 5.6), assume that $f, \bar{f}$ are $L$–Lipschitz. Then both $f$, $\bar{f}$ have multiplicity $\leqslant m$ (since $X$ and $Y$ are uniformly discrete). Then

$$f(\bar{B}(x,R)) \subset \bar{B}(y,LR).$$

It follows that card $\bar{B}(x,R) \leqslant m \, \text{card} \, \bar{B}(y,LR)$ and

$$\text{card } \bar{B}(y,R) \leqslant m \, \text{card } \bar{B}(x,LR). \quad \square$$

COROLLARY 12.2. *$\mathfrak{G}_{X,x} \asymp \mathfrak{G}_{X,x'}$ for all $x, x' \in X$ (see Notation 1.7 for the equivalence relation $\asymp$ between functions).*

Henceforth we will suppress the choice of the base-point in the notation for the growth function.

EXERCISE 12.3. Show that for each (uniformly discrete) space $X$, $\mathfrak{G}_X(R) \preceq e^R$.

For a group $G$ endowed with the word metric $\text{dist}_S$ corresponding to a finite generating set $S$ we sometimes will use the notation $\mathfrak{G}_S(R)$ for $\mathfrak{G}_G(R)$. Since $G$ acts transitively on itself, this definition does not depend on the choice of a base-point.

EXAMPLES 12.4.
(1) If $G = \mathbb{Z}^k$ then $\mathfrak{G}_S \asymp x^k$ for every finite generating set $S$.
(2) If $G = F_k$ is the free group of finite rank $k \geqslant 2$ and $S$ is the set of $2k$ generators then $\mathfrak{G}_S(n) \asymp (2k)^n$.

EXERCISE 12.5.
(1) Prove the two statements above.
(2) Prove that for every $n \geqslant 2$ the group $SL(n, \mathbb{Z})$ has exponential growth.

Proposition 12.6. (1) *If $S, S'$ are two finite generating sets of $G$ then $\mathfrak{G}_S \asymp \mathfrak{G}_{S'}$. Thus one can speak about the growth function $\mathfrak{G}_G$ of a group $G$, well defined up to the equivalence relation $\asymp$.*

(2) *If $G$ is infinite, $\mathfrak{G}_S|_{\mathbb{N}}$ is strictly increasing.*

(3) *The growth function is* sub-multiplicative*:*

$$\mathfrak{G}_S(r+t) \leqslant \mathfrak{G}_S(r)\mathfrak{G}_S(t)\,.$$

(4) *If* card $S = k$ *then* $\mathfrak{G}_S(r) \preceq k^r$ .

Proof. (1) follows immediately from Lemma 12.1 and Milnor–Schwartz lemma.

(2) Consider two integers $n < m$. As $G$ is infinite there exists $g \in G$ at distance $d \geqslant m$ from 1. The shortest path joining 1 and $g$ in Cayley$(G, S)$ can be parameterized as an isometric embedding $p : [0, d] \to$ Cayley$(G, S)$. The vertex $p(n + 1)$ is an element of $\bar{B}(1, m) \setminus \bar{B}(1, n)$.

(3) follows immediately from the fact that

$$\bar{B}(1, n+m) \subseteq \bigcup_{y\in\bar{B}(1,n)} \bar{B}(y, m)\,.$$

(4) follows from the existence of an epimorphism $\pi_S : F(S) \to G$. □

Property (3) implies that the function $\ln \mathfrak{G}_S(n)$ is sub-additive, hence by Fekete's Lemma, see e.g. [**HP74**, Theorem 7.6.1], there exists a (finite) limit

$$\lim_{n\to\infty} \frac{\ln \mathfrak{G}_S(n)}{n}.$$

Hence, we also get a finite limit

$$\gamma_S = \lim_{n\to\infty} \mathfrak{G}_S(n)^{\frac{1}{n}}\,,$$

called *growth constant*.

Property (2) implies that $\mathfrak{G}_S(n) \geqslant n$; whence, $\gamma_S \geqslant 1$.

Definition 12.7. If $\gamma_S > 1$ then $G$ is said to be of *exponential growth*. If $\gamma_S = 1$ then $G$ is said to be of *sub-exponential growth*.

Note that by Proposition 12.6, (1), if there exists a finite generating set $S$ such that $\gamma_S > 1$ then $\gamma_{S'} > 1$ for every other finite generating set $S'$. Likewise for equality with 1.

The notion of subexponential growth makes sense in general non-homogeneous setting.

Definition 12.8. Let $(X, \mathrm{dist})$ be a metric space for which the growth function is defined (e.g. a Riemannian metric, a discrete metric space, a simplicial complex). The space $X$ is said to be of *sub-exponential growth* if for some basepoint $x_0 \in X$

$$\limsup_{n\to\infty} \frac{\ln \mathfrak{G}_{x_0,X}(n)}{n} = 0\,.$$

Since for every other basepoint $y_0$, $\mathfrak{G}_{y_0,X}(n) \leqslant \mathfrak{G}_{x_0,X}\left(n + \mathrm{dist}(x_0, y_0)\right)$ , it follows that the definition is independent of the choice of basepoint.

According to Proposition 12.6, (1), the growth function $\mathfrak{G}_G$ of a finitely generated group $G$ is uniquely defined up to the equivalence relation $\asymp$.

Exercise 12.9. Use the growth function to prove that for $n \neq m$ the groups $\mathbb{Z}^n$ and $\mathbb{Z}^m$ are not quasi-isometric.

Proposition 12.10. (a) *If $H$ is a finitely generated subgroup in a finitely generated group $G$ then $\mathfrak{G}_H \preceq \mathfrak{G}_G$.*

(b) *If $H$ is a subgroup of finite index in $G$ then $\mathfrak{G}_H \asymp \mathfrak{G}_G$.*

(c) *If $N$ is a normal subgroup in $G$ then $\mathfrak{G}_{G/N} \preceq \mathfrak{G}_G$*

(d) *If $N$ is a finite normal subgroup in $G$ then $\mathfrak{G}_{G/N} \asymp \mathfrak{G}_G$.*

Proof. (a) If $X$ is a finite generating set of $H$ and $S$ is a finite generating set of $G$ containing $X$ then $\mathrm{Cayley}(H, X)$ is a subgraph of $\mathrm{Cayley}(G, S)$ and $\mathrm{dist}_X(1, h) \geqslant \mathrm{dist}_S(1, h)$ for every $h \in H$. In particular the closed ball of radius $r$ and center 1 in $\mathrm{Cayley}(H, X)$ is contained in the closed ball of radius $r$ and center 1 in $\mathrm{Cayley}(G, S)$.

(b) and (d) are immediate corollaries of Lemma 12.1 and Milnor–Schwartz lemma.

(c) Let $S$ be a finite generating set in $G$, and let $\bar{S} = \{sN \mid s \in S, s \notin N\}$ be the corresponding finite generating set in $G/N$. The epimorphism $\pi : G \to G/N$ maps the ball of center 1 and radius $r$ onto the ball of center 1 and radius $r$.

□

Exercise 12.11. Let $G$ and $H$ be two groups with finite generating sets $S$ and $X$ respectively. A homomorphism $\varphi : G \to H$ is called *expanding* if there exist constants $\lambda > 1$ and $C \geqslant 0$ such that for every $g \in G$ with $|g|_S \geqslant C$

$$|\varphi(g)|_X \geqslant \lambda |g|_S\,.$$

(1) Prove that given the integer Heisenberg group $G = H_3(\mathbb{Z})$ with the word metric described in Exercise 12.29, the endomorphism

$$\varphi_a : G \to G\,,\ \varphi_a(U_{kln}) = U_{(ak)(al)(a^2n)}\,,$$

is expanding if $a > 12$ is an integer, and that $\varphi_a(G)$ has finite index in $G$.

(2) Let $G$ be a group with a finite generating set $S$ and $H \leqslant G$ a finite index subgroup. We equip $G$ with the word metric $d_S$ and equip $H$ with the metric which is the restriction of $d_S$. Assume that there exists an expanding homomorphism $\varphi : H \to G$ such that $\varphi(H)$ has finite index in $G$. Prove that $G$ has polynomial growth.

More importantly, one has the following generalization of Efremovich's theorem **[Efr53]**:

Proposition 12.12 (Milnor–Efremovich–Schwartz). *Let $M$ be a connected complete Riemannian manifold with bounded geometry. If $M$ is quasi-isometric to a graph $\mathcal{G}$ with bounded geometry, then the growth function $\mathfrak{G}_{M,x_0}$ and the growth function of $\mathcal{G}$ with respect to an arbitrary vertex $v$, are equivalent in the sense of the equivalence relation $\asymp$.*

PROOF. The manifold $M$ has bounded geometry, therefore its sectional curvature is at least $a$ and at most $b$ for some constants $a \leqslant b$; moreover, there exists a uniform lower bound $2\rho > 0$ on the injectivity radius of $M$ at every point. Let $n$ denote the dimension of $M$. We let $V(x,r)$ denote volume of $r$-ball centered at the point $x \in M$ and let $V_a(r)$ denote the volume of the $r$-ball in the complete simply-connected $n$-dimensional manifold of constant curvature $a$.

The fact that the sectional curvature is at least $a$ implies, by Theorem 2.24, Part (1), that for every $r > 0$, $V(x,r) \leqslant V_a(r)$. Similarly, Theorem 2.24, Part (2), implies that the volume $V(x,\rho) \geqslant V_b(\rho)$.

Since $M$ and $\mathcal{G}$ are quasi-isometric, by Definition 5.1 it follows that there exist $L \geqslant 1$, $C \geqslant 0$, two $2C$–separated nets $A$ in $M$ and $B$ in $\mathcal{G}$ respectively, and a $L$–bi-Lipschitz bijection $\mathfrak{q} : A \to B$. Without loss of generality we may assume that $C \geqslant \rho$; otherwise we choose a maximal $2\rho$–separated subset $A'$ of $A$ and then restrict $\mathfrak{q}$ to $A'$.

According to Remark 2.17, (2), we may assume without loss of generality that the base-point $x_0$ in $M$ is contained in the net $A$, and that $\mathfrak{q}(x_0) = v$, the base vertex in $\mathcal{G}$.

For every $r > 0$ we have that

$$\mathfrak{G}_{M,x_0}(r) \geqslant \text{card}\, [A \cap B_M\,(x_0, r - C)]\, V_b(\rho) \geqslant \text{card}\, \left[B \cap B_{\mathcal{G}}\left(1, \frac{r-C}{L}\right)\right] V_b(\rho)$$

$$\geqslant \mathfrak{G}_{\mathcal{G}}\left(\frac{r-C}{L}\right) \frac{V_b(\rho)}{\mathfrak{G}_{\mathcal{G}}(2C)}\,.$$

Conversely,

$$\mathfrak{G}_{M,x_0}(r) \leqslant \text{card}\, [A \cap B_M\,(x_0, r + 2C)]\, V_a(2C) \leqslant$$

$$\text{card}\, [B \cap B_{\mathcal{G}}\,(1, L(r + 2C))]\, V_a(2C) \leqslant \mathfrak{G}_G\,(L(r+2C))\, V_a(2C)\,.$$

□

Thus, it follows from Theorem 4.32 that considering $\asymp$–equivalence classes of growth functions of universal covers of compact Riemannian manifolds is not different from considering equivalence classes of growth functions of finitely-presented groups.

REMARK 12.13. Note that in view of Theorem 5.41, every connected Riemannian manifold of bounded geometry is quasi-isometric to a graph of bounded geometry.

QUESTION 12.14. What is the set $Growth(groups)$ of the equivalence classes of growth functions of finitely generated groups?

QUESTION 12.15. What are the equivalence classes of growth functions for finitely presented groups ?

This question is equivalent to

QUESTION 12.16. What is the set $Growth(manifolds)$ of equivalence classes of growth functions for universal covers of compact connected Riemannian manifolds?

Clearly, $Growth(manifolds) \subset Growth(groups)$; it is unknown if this inclusion is proper.

We will see later on that:

$$\{\exp(t), t^n, n \in \mathbb{N}\} \subset Growth(manifolds) \subset Growth(groups)$$

One can refine Question 12.16 by defining $Growth_n(manifolds)$ as the set of equivalence classes of growth functions of universal covers of $n$-dimensional compact connected Riemannian manifolds. Since every finitely-presented group is the fundamental group of a closed smooth 4-dimensional manifold and growth function depends only on the fundamental group, we obtain:

$$Growth_4(manifolds) = Growth_n(manifolds), \quad \forall n \geqslant 4.$$

On the other hand:

THEOREM 12.17. *$Growth_2(manifolds) = \{1, t^2, e^t\}$, $Growth_3(manifolds) = \{1, t, t^3, t^4, e^t\}$.*

Below is an outline of the proof. Firstly, in view of classification of surfaces, for every closed connected oriented surface $S$ we have:

(1) If $\chi(S) = 2$ then $\pi_1(S) = \{1\}$ and growth function is trivial.
(2) If $\chi(S) = 0$ then $\pi_1(S) = \mathbb{Z}^2$ and growth function is equivalent to $t^2$.
(3) If $\chi(S) < 0$ then $\pi_1(S)$ contains a free nonabelian subgroup, so growth function is exponential.

In the case of 3-dimensional manifolds, one has to appeal to Perelman's solution of Thurston's geometrization conjecture. We refer to [**Kap01**] for the precise statement and definitions which appear below:

For every closed connected 3-dimensional manifold $M$ one of the following holds:

(1) $M$ admits a Riemannian metric of constant positive curvature, in which case $\pi_1(M)$ is finite and has trivial growth.
(2) $M$ admits a Riemannian metric locally isometric to the product metric $S^2 \times \mathbb{R}$. In this case growth function is linear.
(3) $M$ admits a flat Riemannian metric, so universal cover of $M$ is isometric to $\mathbb{R}^3$ and growth function is $t^3$.
(4) $M$ is homeomorphic to the quotient $H_3/\Gamma$, where $H_3$ is the 3-dimensional Heisenberg group and $\Gamma$ is a uniform lattice in $H_3$. In this case, in view of Exercise 12.29, growth function is $t^4$.
(5) Fundamental group of $M$ is solvable but not virtually nilpotent, thus, by Wolf's Theorem (theorem 12.52), growth function is exponential.
(6) In all other cases, $\pi_1(M)$ contains free nonabelian subgroup; hence, growth is exponential.

CONJECTURE 12.18 (J. Milnor [**Mil68**]). *The growth of a finitely generated group is either polynomial (i.e. $\mathfrak{G}_S(t) \preceq t^d$ for some integer $d$) or exponential (i.e. $\gamma_S > 1$).*

We will see below to which extent Milnor's conjecture holds, fails and to which extent it remains an open problem.

## 12.2. Isoperimetric inequalities

One can define, in the setting of graphs, the following concepts, inspired by, and closely connected to, notions introduced in Riemannian geometry (see Definitions 2.19 and 2.20). Recall that for a subset $F \subset V, F^c$ denotes its complement in $V$.

Definition 12.19. An *isoperimetric inequality* in a graph $\mathcal{G}$ of bounded geometry is an inequality satisfied by all finite subsets $F$ of vertices, of the form

$$\text{card}\,(F) \leq f(F) g\left(\text{card}\, E(F, F^c)\right) ,$$

where $f$ and $g$ are real-valued functions, $g$ defined on $\mathbb{R}_+$.

Definition 12.20. Let $\Gamma$ be a graph of bounded geometry, with the vertex set $V$ and edge set $E$. The *Cheeger constant* or the *Expansion Ratio* of the graph $\Gamma$ is defined as

$$h(\Gamma) = \inf\left\{ \frac{|E(F, F^c)|}{|F|} \;:\; F \text{ is a finite nonempty subset of } V,\, |F| \leqslant \frac{|V|}{2} \right\} .$$

Here $E(F, F^c)$ is edge boundary for both $F$ and $F^c$, i.e., the set of edges connecting $F$ to $F^c$ (see Definition 1.11). Thus, the condition $|F| \leqslant \frac{|V|}{2}$ insures that, in case $V$ is finite, one picks the smallest of the two sets $F$ and $F^c$ in the definition of the Cheeger constant. Intuitively, finite graphs with small Cheeger constant can be separated by vertex sets which are relatively small comparing to the size of (the smallest component of) their complements. In contrast, graphs with large Cheeger constant are "hard to separate."

Exercise 12.21. a. Let $\Gamma$ be a single circuit with $n$ vertices. Then $h(\Gamma) = \frac{2}{n}$.
b. Let $\Gamma = K_n$ be the complete graph on $n$ vertices, i.e., $\Gamma$ is the 1-dimensional skeleton of the $n-1$-dimensional simplex. Then

$$h(\Gamma) = \left\lfloor \frac{n}{2} \right\rfloor .$$

The inequalities in (1.1) imply that in every isoperimetric inequality, the edge-boundary can be replaced by the vertex boundary, if one replaces the function $g$ by an asymptotically equal function (respectively the Cheeger constants by bi-Lipschitz equivalent values). Therefore, in what follows we choose freely whether to work with the edge-boundary or with the vertex-boundary, depending on which one is more convenient.

There exists an isoperimetric inequality satisfied in every Cayley graph of an infinite group.

Proposition 12.22. *Let $\mathcal{G}$ be the Cayley graph of a finitely generated infinite group. For every finite set $F$ of vertices*

$$\text{card}\,(F) \leqslant [\text{diam}(F) + 1]\, \text{card}\,(\partial_V F) . \tag{12.1}$$

Proof. Assume that $\mathcal{G}$ is the Cayley graph of an infinite group $G$ with respect to a finite generating set $S$.

Let $d$ be the diameter of $F$ with respect to the word metric $\text{dist}_S$, and let $g$ be an element in $G$ such that $|g|_S = d+1$. Let $g_0 = 1, g_1, g_2, \ldots, g_d, g_{d+1} = g$ be the set of vertices on a geodesic joining 1 to $g$.

Given an arbitrary vertex $x \in F$, the element $xg$ is at distance $d+1$ from $x$; therefore, by the definition of $d$ it follows that $xg \in F^c$. In the finite sequence of vertices $x, xg_1, xg_2, \dots, xg_d, xg_{d+1} = xg$ consider the largest $i$ such that $xg_i \in F$. Then $i < d+1$ and $xg_{i+1} \in F^c$, whence $xg_{i+1} \in \partial_V F$, equivalently, $x \in [\partial_V F]\, g_{i+1}^{-1}$.

We have thus proved that $F \subseteq \bigcup_{i=1}^{d+1} [\partial_V F]\, g_i^{-1}$, which implies the inequality (12.1). □

An argument similar in spirit, but more elaborate, allows to relate isoperimetric inequalities and growth functions:

PROPOSITION 12.23 (Varopoulos inequality). *Let $\mathcal{G}$ be the Cayley graph of an infinite group $G$ with respect to a finite generating set $S$, and let $d$ be the cardinality of $S$. For every finite set $F$ of vertices*

$$\operatorname{card}(F) \leqslant 2d\, k \operatorname{card}(\partial_V F), \tag{12.2}$$

*where $k$ is the unique integer such that $\mathfrak{G}_S(k-1) \leqslant 2\operatorname{card}(F) < \mathfrak{G}_S(k)$.*

PROOF. Our goal is to show that with the given choice of $k$, there exists an element $g \in B_S(1,k)$ such that for a certain fraction of the vertices $x$ in $F$, the right-translates $xg$ are in $F^c$. In what follows we omit the subscript $S$ in our notation.

We consider the sum

$$\mathcal{S} = \frac{1}{\mathfrak{G}(k)} \sum_{g \in B(1,k)} \operatorname{card}\{x \in F \mid xg \in F^c\} = \frac{1}{\mathfrak{G}(k)} \sum_{g \in B(1,k)} \sum_{x \in F} \mathbf{1}_{F^c}(xg) =$$

$$\frac{1}{\mathfrak{G}(k)} \sum_{x \in F} \sum_{g \in B(1,k)} \mathbf{1}_{F^c}(xg) = \frac{1}{\mathfrak{G}(k)} \sum_{x \in F} \operatorname{card}\,[B(x,k) \setminus F]\,.$$

By the choice of $k$, the cardinality of each ball $B(x,k)$ is larger than $2\operatorname{card} F$, whence

$$\operatorname{card}\,[B(x,k) \setminus F] \geqslant \operatorname{card} F\,.$$

The denominator $\mathfrak{G}(k) \leqslant d\mathfrak{G}(k-1) \leqslant 2d\operatorname{card} F$. We, therefore, find as a lower bound for the sum $\mathcal{S}$, the value

$$\frac{1}{2d\operatorname{card} F} \sum_{x \in F} \operatorname{card} F = \frac{\operatorname{card} F}{2d}\,.$$

It follows that

$$\frac{1}{\mathfrak{G}(k)} \sum_{g \in B(1,k)} \operatorname{card}\{x \in F \mid xg \in F^c\} \geqslant \frac{\operatorname{card} F}{2d}\,.$$

The latter inequality implies that there exists $g \in B(1,k)$ such that

$$\operatorname{card}\{x \in F \mid xg \in F^c\} \geqslant \frac{\operatorname{card} F}{2d}\,.$$

We now argue as in the proof of Proposition 12.22, and for the element $g \in B(1,k)$ thus found, we consider $g_0 = 1, g_1, g_2, \dots, g_{m-1}, g_m = g$ to be the set of vertices on a geodesic joining 1 to $g$, where $m \leqslant k$. The set $\{x \in F \mid xg \in F^c\}$ is contained in the union $\bigcup_{i=1}^{m} [\partial_V F]\, g_i^{-1}$; therefore, we obtain

$$\frac{|F|}{2d} \leqslant k\, |\partial_V F|\,.$$

□

REMARKS 12.24. Proposition 12.23 was initially proved in **[VSCC92]** using random walks. The proof reproduced above follows **[CSC93]**.

Corollary 12.25. *Let $G$ be an infinite finitely generated group and let $F$ be an arbitrary set of elements in $G$.*

(1) *If $\mathfrak{G}_G \asymp x^n$ then*

$$\operatorname{card} F \leqslant K \left[\operatorname{card}\left(\partial_V F\right)\right]^{\frac{n}{n-1}} .$$

(2) *If $\mathfrak{G}_G \asymp \exp(x)$ then*

$$\frac{\operatorname{card} F}{\ln\left(\operatorname{card} F\right)} \leqslant K \operatorname{card}\left(\partial_V F\right) .$$

*In both inequalities above, the boundary $\partial_V F$ is considered in the Cayley graph of $G$ with respect to a finite generating set $S$, and $K$ depends on $S$.*

## 12.3. Wolf's Theorem for semidirect products $\mathbb{Z}^n \rtimes \mathbb{Z}$

In this section we explain how to prove Conjecture 12.18 in the case of semidirect products $\mathbb{Z}^n \rtimes \mathbb{Z}$. This easy example helps to understand the general case of polycyclic groups and the general Wolf's Theorem.

Note that the semidirect product is defined by a homomorphism $\varphi : \mathbb{Z} \to Aut(\mathbb{Z}^n) = GL(n, \mathbb{Z})$, and the latter is determined by $\theta = \varphi(1)$, which is represented by a matrix $M \in GL(n, \mathbb{Z})$. Therefore the same semidirect product is also denoted $\mathbb{Z}^n \rtimes_\theta \mathbb{Z} = \mathbb{Z}^n \rtimes_M \mathbb{Z}$.

Proposition 12.26. *A semidirect product $G = \mathbb{Z}^n \rtimes_M \mathbb{Z}$ is*

(1) *either virtually nilpotent (when $M$ has all eigenvalues of absolute value $1$);*
(2) *or of exponential growth (when $M$ has at least one eigenvalue of absolute value $\neq 1$).*

Remarks 12.27. (1) The group $G = \mathbb{Z}^n \rtimes_M \mathbb{Z}$ is nilpotent if $M$ has all eigenvalues equal to 1;

(2) The same is not in general true if $M$ has all eigenvalues of absolute value 1. The group $G = \mathbb{Z} \rtimes_M \mathbb{Z}$ with $M = (-1)$ is a counter-example. Indeed:
   (a) $G$ contains a finite index subgroup isomorphic to $\mathbb{Z}^2$. For example the subgroup $\mathbb{Z} \rtimes (2\mathbb{Z})$ is isomorphic to $\mathbb{Z} \rtimes_{M^2} \mathbb{Z} \simeq \mathbb{Z}^2$.
   (b) $G$ is not nilpotent. Let $s$ denote a generator of the first factor $\mathbb{Z}$ and $t$ a generator of the second factor $\mathbb{Z}$. Then $\left\langle t^2 \right\rangle$ is in the center of $G$, and $G/\left\langle t^2 \right\rangle$ is isomorphic to $\mathbb{Z} \rtimes_\varphi \mathbb{Z}_2$, where $\varphi : \mathbb{Z}_2 \to \mathbb{Z}$, $\varphi(k) =$ multiplication by $(-1)^k$. This latter subgroup is the infinite dihedral group $D_\infty$, which is solvable but not nilpotent (see Exercise 10.50).

   *Remark.* Thus $G$ is polycyclic, virtually nilpotent but not nilpotent. In particular statement (1) in Proposition 12.26 cannot be improved to '$G = \mathbb{Z}^n \rtimes_M \mathbb{Z}$ is nilpotent'.

Proof. Note that $\mathbb{Z}^n \rtimes_{\theta^N} \mathbb{Z}$ is a subgroup of finite index in $G = \mathbb{Z}^n \rtimes_\theta \mathbb{Z}$ (corresponding to the replacement of the second factor $\mathbb{Z}$ by $N\mathbb{Z}$). Thus we may replace $M$ by some power of $M$, and replace $G$ with a finite-index subgroup. We will retain the notation $G$ and $M$ for the finite-index subgroup and the power of $M$. Then, the matrix $M \in GL(n, \mathbb{Z})$ will have no nontrivial roots of unity as eigenvalues. In view of Lemma 10.23, this means that for every eigenvalue $\lambda \neq 1$ of $M$, $|\lambda| \neq 1$.

We have two cases to consider.

(1) The matrix $M$ has only eigenvalues equal 1. Lemma 10.22 then implies that there exists a finite series

$$\{1\} = H_n \leqslant H_{n-1} \leqslant \ldots \leqslant H_1 \leqslant H_0 = \mathbb{Z}^n$$

such that $H_i \simeq \mathbb{Z}^{n-i}$, each quotient $H_i/H_{i+1}$ is cyclic, the automorphism $\theta$ preserves each $H_i$ and induces the identity automorphism on $H_i/H_{i+1}$. Let $t$ denote the generator of the factor $\mathbb{Z}$ in the semidirect product $G = \mathbb{Z}^n \rtimes_M \mathbb{Z}$. For every $g \in \mathbb{Z}^n$, $tgt^{-1} = \theta(g)$. In particular, since $\theta$ projects to the identity automorphism of $H_i/H_{i+1}$, for every $h_i \in H_i$ we get:

$$t^k(h_iH_{i+1})t^{-k} = h_iH_{i+1}$$

that is $[t^k, h_i] \in H_{i+1}$.

We have an exact sequence $1 \to \mathbb{Z}^n \to G \to \mathbb{Z} \to 1$, and the projection of $C^2G = [G,G]$ onto $\mathbb{Z}$ is 0, hence $C^2G \leqslant \mathbb{Z}^n$. To finish the argument it suffices to prove that $[G, H_i] \subseteq H_{i+1}$ for every $i \geqslant 0$.

For every $h_i \in H_i$ and every element $zt^k \in G$ with $z \in \mathbb{Z}^n$ and $k \in \mathbb{Z}$,

$$[h_i, zt^k] = z[h_i, t^k]z^{-1} = zh_i\theta^k(h_i^{-1})z^{-1} =$$

$$zh_i\theta^k(h_i^{-1})z^{-1} = z[h_i, t^k]z^{-1} \in zH_{i+1}z^{-1} = H_{i+1},$$

since $H_0$ is abelian.

This proves that $[G, H_i] \subseteq H_{i+1}$ for every $i \geqslant 0$, which implies that $G$ is nilpotent.

(2) Assume that $M$ has an eigenvalue with absolute value strictly greater than 1. Up to replacing $\theta$ with $\theta^N$ we may assume that it has an eigenvalue with absolute value at least 2.

Lemma 10.24 applied to $M$ implies that there exists an element $v \in \mathbb{Z}^n$ such that distinct elements $s = (s_k) \in \bigoplus_{k\geqslant 0} \mathbb{Z}_2$ define distinct vectors

$$s_0v + s_1M_iv + \ldots + s_nM_i^nv + \ldots$$

in $\mathbb{Z}^n$. With the multiplicative notation for the binary operation in $G$, the above vectors correspond to pairwise distinct elements

$$g_s = v^{s_0}(tvt^{-1})^{s_1} \cdots (t^nvt^{-n})^{s_n} \cdots \in G$$

are also distinct. Now, consider the set $\Sigma_k$ of sequences $s = (s_k)$ for which $s_k = 0\,, \forall k \geqslant n+1$. Then the map

$$\Sigma_k \to G, \quad s \mapsto g_s$$

is injective and its image consists of $2^{n+1}$ elements $g_s$. Assume that the generating set of $G$ contains the elements $t$ and $v$. With respect to this generating set, the word-length $|g_s|$ is at most $3n+1$ for every $s \in \Sigma_k$. Thus, for every $n$ we obtain $2^{n+1}$ distinct elements of $G$ of length at most $3n+1$, whence $G$ has exponential growth. □

Remark 12.28. What remains to be proven is that the two cases in Proposition 12.26 are mutually exclusive, i.e., that a nilpotent group cannot have exponential growth. We shall prove in Section 12.5 that nilpotent (hence virtually nilpotent) groups have in fact polynomial growth. The following exercise contains a proof of this statement in the case of the integer Heisenberg group.

Exercise 12.29. Consider the *integer Heisenberg group*

$$G = H_3(\mathbb{Z}) = \left\{ U_{kln} = \begin{pmatrix} 1 & k & n \\ 0 & 1 & l \\ 0 & 0 & 1 \end{pmatrix} \; ; \; k, l, n \in \mathbb{Z} \right\},$$

with the finite set of generators $S = \{u^{\pm 1}, v^{\pm 1}, z^{\pm 1}\}$, where $u = I + E_{12}$, $v = I + E_{23}$, $z = I + E_{13}$.

For every $g \in G$ denote by $|g|$ the distance $\mathrm{dist}_S(1, g)$.

(1) Prove that $U_{kln} = u^k v^l z^{n-kl}$ for every $k, l, n \in \mathbb{Z}$.

This in particular shows that every element of $G$ can be written as $u^k v^l z^m$ with $k, l, m \in \mathbb{Z}$, and that this decomposition is unique for every element (since it is entirely determined by its entries).

(2) Prove that $[u^k, v^l] = z^{kl}$. Deduce that $|z^m| \leqslant 6\sqrt{|m|}$ and that

$$|u^k v^l z^m| \leqslant |k| + |l| + 6\sqrt{|m|}\,. \tag{12.3}$$

(3) Prove that $|u^k v^l z^m| \leqslant r$ implies $|k| + |l| \leqslant r$ and $|m| \leqslant r^2$.

(4) Deduce that

$$|u^k v^l z^m| \geqslant \frac{1}{2}\left(|k| + |l| + \sqrt{|m|}\right) \tag{12.4}$$

(5) Deduce that there exist constants $c_2 > c_1 > 0$ such that the growth function $\mathfrak{G}_S$ of $G$ satisfies

$$c_1 n^4 \leqslant \mathfrak{G}_S(n) \leqslant c_2 n^4\,,\ \forall n \geqslant 1\,.$$

## 12.4. Distortion of a subgroup in a group

**12.4.1. Definition, properties, examples.** If $H$ is a finitely generated subgroup of a finitely generated group $G$, we can choose finite generating sets $X$ of $H$ and $S$ of $G$, so that $X \subset G$. In this case $\mathrm{Cayley}(H, X)$ is a subgraph of $\mathrm{Cayley}(G, S)$, hence for every $h, h' \in H$, $\mathrm{dist}_X(h, h') \geqslant \mathrm{dist}_S(h, h')$.

We want to measure how much larger the distances $\mathrm{dist}_X(h, h')$ can be, compared to the distances $\mathrm{dist}_S(h, h')$, for $h, h' \in H$. Since all word metrics are left invariant it suffices to compare the two metrics when $h' = 1$.

Definition 12.30. The *distortion function of the subgroup $H$ in the group $G$* with respect to some finite generating sets $X$ of $H$ and $S$ of $G$, is the function $\Delta_G^H : \mathbb{N} \to \mathbb{N}$ defined by

$$\Delta_G^H(n) = \max\left\{\mathrm{dist}_X(1, h) \mid h \in H\,,\ \mathrm{dist}_S(1, h) \leqslant n\right\}\,.$$

The subgroup $H$ is called *undistorted* (in $G$) if $\Delta_G^H(n) \asymp n$.

Exercise 12.31. Show that $H$ is undistorted if and only if the embedding $\iota : H \to G$ is a quasi-isometric embedding.

In general, the distortion function measures failure of the embedding $H \to G$ to be quasi-isometric.

CONVENTION 12.32. In what follows, when discussing distortion we always assume that both the ambient group $G$ and the subgroup $H$ are infinite.

Below are the basic properties of the distortion function:

PROPOSITION 12.33. (1) *If $\widetilde{X}$ and $\widetilde{S}$ are finite generating sets of $H$ and $G$ respectively, and $\widetilde{\Delta}_G^H$ is the distortion function with respect to these generating sets, then $\widetilde{\Delta}_G^H \asymp \Delta_G^H$. Thus up to the equivalence relation $\asymp$, the distortion function of the subgroup $H$ in the group $G$ is uniquely defined by $H$ and $G$.*

(2) *For every finitely generated subgroup $H$ in a finitely generated group $G$, $\Delta_G^H(n) \succeq n$.*

(3) *If $H$ has finite index in $G$ then $\Delta_G^H(n) \asymp n$.*

(4) *Let $K \lhd G$ is a finite normal subgroup and let $H \leqslant G$ be a finitely generated subgroup; set $\bar{G} := G/K, \bar{H} := H/K$. Then*

$$\Delta_G^H \asymp \Delta_{\bar{G}}^{\bar{H}}.$$

(5) *If $K \leqslant H \leqslant G$ then*

$$\Delta_G^K \preceq \Delta_H^K \circ \Delta_G^H .$$

(6) *If $A$ is a finitely generated abelian group then for every subgroup $H$ in $A$, $\Delta_G^H(n) \asymp n$.*

PROOF. (1) Let $\lambda = \max_{s \in S} |s|_{\widetilde{S}}$ and $\mu = \max_{x \in X} |x|_{\widetilde{X}}$. Then

$$\max\{|h|_X : h \in H\,,\, |h|_S \leqslant n\} \leqslant \max\{|h|_X : h \in H\,,\, |h|_{\widetilde{S}} \leqslant \lambda n\} \leqslant$$
$$\mu \max\{|h|_{\widetilde{X}} : h \in H\,,\, |h|_{\widetilde{S}} \leqslant \lambda n\}\,.$$

Thus $\Delta_G^H(n) \leqslant \mu \widetilde{\Delta}_G^H(\lambda n)$. The opposite inequality is proved by interchanging the roles of the generating sets.

(2) If we take finite generating sets $S$ and $X$ of $G$ and $H$ respectively so that $S \subset X$, then the embedding $H \to G$ is 1-Lipschitz with respect to the resulting word metrics. Whence $\Delta_G^H(n) \geqslant n$.

(3) The statement follows immediately from the fact that the inclusion map $H \to G$ is a quasi-isometry.

(4) This equivalence follows from the fact that the projections $G \to \bar{G}$ and $H \to \bar{H}$ are quasi-isometries.

(5) Consider $\mathrm{dist}_K$, $\mathrm{dist}_H$ and $\mathrm{dist}_G$ three word metrics, and an arbitrary element $k \in K$ such that $\mathrm{dist}_G(1, k) \leqslant n$. Then $\mathrm{dist}_H(1, k) \leqslant \Delta_G^H(n)$ whence

$$\mathrm{dist}_K(1, k) \leqslant \Delta_H^K\left(\Delta_G^H(n)\right)\,.$$

(6) By the classification theorem of finitely generated abelian groups (Theorem 10.6), every such group is the direct product of a finite group and free abelian group. In particular, every finitely generated abelian group is virtually torsion-free. Therefore, by combining (3) and (5), it suffices to consider the case where $G$ is torsion-free of rank $n$. Then $G$ acts by translations geometrically on $\mathbb{R}^n$; its rank $m$ subgroup $H$ also acts geometrically on a subspace $\mathbb{R}^m \subset \mathbb{R}^n$. Since $\mathbb{R}^m$

is isometrically embedded in $\mathbb{R}^n$, it follows that the embedding $H \to G$ is quasi-isometric. Hence, $H$ is undistorted in $G$ and $\Delta_G^H(n) \asymp n$. □

Below is an example of a subgroup with non-linear (and in fact exponential) distortion:

LEMMA 12.34. *Let $G = \mathbb{Z}^m \rtimes_M \mathbb{Z}$, where $M \in GL(m, \mathbb{Z})$.*
*If $M$ has an eigenvalue with absolute value different from 1 then*

$$\Delta_G^{\mathbb{Z}^m}(n) \asymp e^n\,. \tag{12.5}$$

PROOF. Note that (12.5) is equivalent to the existence of constants $b \geqslant a > 1$ and $c_i > 0$, $i = 1, 2$, such that for every $n \in \mathbb{N}$,

$$c_1 a^n \leqslant \Delta_G^{\mathbb{Z}^m}(n) \leqslant c_2 b^n\,. \tag{12.6}$$

*Lower bound.* There exists $N$ such that $M^N$ has an eigenvalue with absolute value at least 2. According to Proposition 12.33, we may replace in our arguments the group $G$ by the finite index subgroup $\mathbb{Z}^m \rtimes (N\mathbb{Z})$. Thus, without loss of generality, we may assume that $M$ has an eigenvalue with absolute value at least 2.

Lemma 10.24 implies that there exists a vector $v \in \mathbb{Z}^m$ such that the map

$$\begin{array}{ccc} \mathbb{Z}_2^{k+1} & \to & \mathbb{Z}^m \\ s = (s_n) & \mapsto & s_0 v + s_1 M v + \ldots + s_k M^k v \end{array}$$

is injective. If we denote by $t$ the generator of the factor $\mathbb{Z}$ and we use the multiplicative notation for the operation in the group $G$, then the element

$$w_s = s_0 v + s_1 M v + \ldots + s_k M^k v \in \mathbb{Z}^m$$

can be rewritten as

$$w_s = v^{s_0}(tvt^{-1})^{s_1} \cdots (t^k v t^{-k})^{s_k}\,.$$

Thus we obtain $2^{k+1}$ elements of $\mathbb{Z}^m$ of the form $w_s$, and if we assume that $t$ and $v$ are in the generating set defining the metric, the length of all these elements is at most $3k+1$.

In the subgroup $\mathbb{Z}^m$ we consider the generating set $X = \{e_i \mid 1 \leqslant i \leqslant m\}$, where $e_i$ is the $i$-th element in the canonical basis. Then for every $w \in \mathbb{Z}^m$, $|w|_X = |w_1| + \cdots + |w_m|$, i.e. $|w|_X = \|w\|_1$, where $\|\ \|_1$ denotes the $\ell_1$–norm on $\mathbb{R}^m$.

Let $r = \max\{\|w_s\|_1 : s = (s_n) \in \mathbb{Z}_2^{k+1}\}$. The ball in $(\mathbb{Z}^m, \|\ \|_1)$ with center 0 and radius $r$ contains all the products $w_s$, i.e. $2^{k+1}$ elements, whence $r^m \succeq 2^{k+1}$, and $r \succeq a_1^k$, where $a_1 = 2^{\frac{1}{m}}$.

We have thus obtained that $\Delta_G^{\mathbb{Z}^m}(3k+1) \succeq a_1^k$, whence $\Delta_G^{\mathbb{Z}^m}(n) \succeq a^n$, where $a = a_1^{\frac{1}{3}}$.

*Upper bound.* Consider the generating set $X = \{e_i \mid 1 \leqslant i \leqslant m\}$ in $\mathbb{Z}^m$ and the generating set $S = X \cup \{t\}$ in $G$. Let $w$ be an element of $\mathbb{Z}^m$ such that $|w|_S \leqslant n$. It follows that

$$w = t^{k_0} v_1 t^{k_1} v_2 \cdots t^{k_{\ell-1}} v_\ell t^{k_\ell}\,, \tag{12.7}$$

where $k_j \in \mathbb{Z}$, $k_0$ and $k_\ell$ possibly equal to 0 but all the other exponents of $t$ are non-zero, $v_j \in \mathbb{Z}^m$, and

$$\sum_{j=0}^{\ell} |k_j| + \sum_{j=1}^{\ell} \|v_j\|_1 \leqslant n\,.$$

We may rewrite (12.7) as

(12.8)
$$w = \left(t^{k_0} v_1 t^{-k_0}\right)\left(t^{k_0+k_1} v_2 t^{-k_0-k_1}\right)\cdots\left(t^{k_0+..+k_{\ell-1}} v_\ell t^{-k_0-..-k_{\ell-1}}\right) t^{k_0+..+k_{\ell-1}+k_\ell}\,.$$

The uniqueness of the decomposition of every element in $G$ as $wt^q$ with $w \in \mathbb{Z}^m$ and $q \in \mathbb{Z}$, implies that $k_0 + .. + k_{\ell-1} + k_\ell = 0$. With this correction, the decomposition in (12.8), rewritten with the additive notation and using the fact that $t^k v t^{-k} = M^k v$ for every $v \in \mathbb{Z}^m$, is as follows

$$w = M^{k_0} v_1 + M^{k_0+k_1} v_2 + \cdots + M^{k_0+..+k_{\ell-1}} v_\ell\,.$$

Let $\alpha_+$ be the maximum among absolute values of the eigenvalues of $M$, $\alpha_-$ be the maximum of absolute values of eigenvalue of $M^{-1}$; set $\alpha = \max(\alpha_+, \alpha_-)$.

In $GL(m, \mathbb{C})$ the matrix $M$ can be written as $PDUP^{-1}$, where $D$ is diagonal, $U$ is upper triangular with entries 1 on the diagonal and $DU = UD$ (the multiplicative Jordan decomposition of $M$).

Then $M^k = PD^kU^kP^{-1}$, and $\|M^k\| \leqslant \lambda \|D^k\|\|U^k\| \leqslant \lambda' \alpha^{|k|} k^m \leqslant \mu \beta^{|k|}$, for an arbitrary $\beta > \alpha$ and all sufficiently large values of $k$. Therefore,

$$\|w\|_1 \preceq \|M^{k_0}\|\,\|v_1\|_1 + \|M^{k_0+k_1}\|\,\|v_2\|_1 + \cdots + \|M^{k_0+..+k_{\ell-1}}\|\,\|v_\ell\|_1 \preceq$$

$$\beta^{|k_0|}\,\|v_1\|_1 + \beta^{|k_0|+|k_1|}\,\|v_2\|_1 + \cdots + \beta^{|k_0|+..+|k_{\ell-1}|}\,\|v_\ell\|_1 \preceq \beta^n\, n \preceq \beta^{2n}\,.$$

We thus conclude that $\Delta_G^{\mathbb{Z}^m}(n) \preceq \beta^{2n}$. ☐

Example 12.35. Let $G := \left\langle a, b : aba^{-1} = b^p \right\rangle$, $p \geqslant 2$. Then the subgroup $H = \langle b \rangle$ is exponentially distorted in $G$.

Proof. To establish the lower exponential bound note that:

$$g_n := a^n b a^{-n} = b^{p^n},$$

hence $d_G(1, g_n) = 2n + 1$, $d_H(1, g_n) = p^n$, hence

$$\Delta_G^H(R) \geqslant p^{[(R-1)/2]}.$$

We will leave the upper exponential bound as an exercise. ☐

Exercise 12.36. Consider the group

$$G = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} ;\ a = 2^n\,,\ b = \frac{m}{2^k}\,,\ n, m, k \in \mathbb{Z} \right\}.$$

Note that $G$ has a finite generating set consisting of matrices $d = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and $u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

(1) Prove that the group $G$ has exponential growth.

(2) Prove that the cyclic subgroup generated by $u$ has exponential distortion.

In general, distortion functions for subgroups can be as bad as one can imagine, for instance, nonrecursive.

Example 12.37. [Mikhailova's construction.] Let $Q$ be a finitely-presented group with Dehn function $\delta(n)$. Let $a_1, \ldots, a_m$ be generators of $Q$ and $\phi : F_m \to Q$ be the epimorphism from the free group of rank $m$ sending free generators of $F_m$ to the elements $a_i$, $i = 1, \ldots, m$. Consider the group $G = F_m \times F_m$ and its subgroup

$$H = \langle (g_1, g_2) \in G | \phi(g_1) = \phi(g_2) \rangle .$$

This construction of $H$ is called *Mikhailova's construction*, it is a source of many pathological examples in group theory. The subgroup $H$ is finitely generated and its distortion in $G$ is $\asymp \delta(n)$. In particular, if $Q$ has unsolvable word problem then its distortion in $G$ is nonrecursive. We refer the reader to [**OS01**, Theorem 2] for further details.

**12.4.2. Distortion of subgroups in nilpotent groups.** The goal of this section is to estimate the distortion function for subgroups of nilpotent groups.

Lemma 12.38. *Let $G$ be a finitely generated nilpotent group of class $k$ and let $C^kG$ be the last non-trivial element in its lower central series. If $S$ is a finite set of generators for $G$ and $g$ is an arbitrary element in $C^kG$ then there exists a constant $\lambda = \lambda(S, g)$ such that*

$$|g^n|_S \leqslant \lambda n^{\frac{1}{k}} \text{ for every } n \in \mathbb{N} .$$

Proof. We argue by induction on $k$. The statement is clearly true for $k = 1$. Assume that it is true for $k$ and consider $G$, a $(k+1)$–step nilpotent group.

Note that $C^{k+1}G$ is central in $G$, in particular it is abelian. The subgroup $C^{k+1}G$ also has a finite set of generators of the form $[s, c]$, with $s \in S$ and $c \in C^kG$ (e.g., we can take as the generators of $C^{k+1}G$ the inverses of $(k+1)$–fold left commutators of generators of $G$, see Lemma 10.31). Since $C^{k+1}G$ is abelian suffices to prove the statement of lemma for $g$ equal to one of these generators $[s, c]$.

The formulas (3) and (4) in Lemma 10.25, imply that for every $x, x' \in G$ and $y, y' \in C^kG$ we have

$$[x, yy'] = [x, y][x, y'] \text{ and } [xx', y] = [x, y][x', y] . \tag{12.9}$$

Here we used the fact that $C^{k+1}G$ is central to deduce that $[y, [x, y']] = 1$ and $[x, [x', y]] = 1$, and to swap $[x, y]$ and $[x', y]$.

In particular

$$[x, y^a] = [x, y]^a \text{ and } [x^b, y] = [x, y]^b . \tag{12.10}$$

Given $n$ we let $q$ denote the smallest integer such that $q > n^{\frac{1}{k+1}}$. Note that our goal is to show that $|[s, c]^n|_S$ is bounded by $\lambda\, q$ for a suitable choice of $\lambda$.

There exist two positive integers $a, b$ such that $n = aq^k + b$ and $0 \leqslant b < q^k$; moreover, $n < q^{k+1}$ implies that $a < q$. The formulas in (12.10) then imply that

$$[s, c]^n = \left[s^a, c^{q^k}\right] \left[s, c^b\right] .$$

The induction hypothesis applied to the group $G/C^{k+1}G$ (where finite generating set of the quotient is image of $S$), and to the element $c\, C^{k+1}G$ implies that $c^{q^k} = k_1 z_1$ and $c^b = k_2 z_2$, where $|k_i|_S \leqslant \mu q$, for a constant $\mu = \mu(S, c)$, and $z_i \in C^{k+1}G$, for $i = 1, 2$.

The formulas (12.9) imply that for every $x \in G$, $[x, k_i z_i] = [x, k_i]$. Therefore

$$[s, c]^n = [s^a, k_1]\, [s, k_2] \ ,$$

whence $[s,c]^n$ has $S$-length at most

$$2(a+\mu q)+2(1+\mu q)\leqslant 4(1+\mu)q\leqslant 8(1+\mu)n^{\frac{1}{k+1}}.$$

Thus, we can take $\lambda = 8(1+\mu)$. □

Corollary 12.39. *Let $G$ be a finitely generated nilpotent group of class $k$ and let $H := C^kG$ be the last non-trivial element in its lower central series. Then:*

(1) *The restriction of the distance function to free abelian factor of $H$ satisfies $\text{dist}_G(1,g) \preceq \text{dist}_H(1,g)^{\frac{1}{k}}$.*

(2) *If $H$ is infinite then its distortion function satisfies $\Delta_G^H(n) \succeq n^k$.*

Proof. The group $H = C^kG$ is abelian, hence isomorphic to $\mathbb{Z}^m \times F$ for some $m \in \mathbb{N}$ and a finite abelian group $F$. Let $\{t_1, \ldots, t_m\}$ be a basis for $\mathbb{Z}^m$ and $\tau_1, \ldots, \tau_q$ respective generators of the cyclic factors of $F$. We consider the word metric in $H$ corresponding to the generating set $\{t_1, \ldots, t_m, \tau_1, \ldots, \tau_q\}$. Consider the shortest word in this generating set representing $g$,

$$g = t_1^{\alpha_1} \cdots t_m^{\alpha_m} \tau_1^{\beta_1} \cdots \tau_q^{\beta_q}.$$

Then

$$\text{dist}_H(1,g) = \sum_{i=1}^{m} |\alpha_i| + \sum_{j=1}^{q} |\beta_j|. \tag{12.11}$$

Let $D$ denote the diameter of the finite group $F$ with respect to $\text{dist}_S$ and let

$$\lambda := \max_i \lambda(S, t_i),$$

where $\lambda(S,t_i)$ is as in Lemma 12.38. Then:

$$|g|_S \leqslant \sum_{i=1}^{m} |t_i^{\alpha_i}|_S + |\tau_1^{\beta_1} \cdots \tau_q^{\beta_q}|_S \leqslant \lambda \sum_{i=1}^{m} |\alpha_i|^{\frac{1}{k}} + D. \tag{12.12}$$

Now, the statement (1) follows from (12.11) and (12.12). Statement (2) is an immediate consequence of (1). □

Our next goal is to prove the inequalities opposite to those in Corollary 12.39. Since the distortion function is preserved by taking quotients by finite subgroups and in view of Corollary 10.52, it suffices to consider the case when the group $G$ is torsion-free, in particular, all abelian quotients $C^iG/C^{i+1}G$ are torsion-free.

Lemma 12.40. *If $X$ is a finite generating set for a nilpotent group $G$, then there exists a finite generating set $\widehat{X}$ containing $X$ and such that for every $x, y \in \widehat{X}$, $[x,y] \in \widehat{X}$.*

Proof. We define inductively certain finite subsets of $G$. Let $T_1 = X$ and set $T_{i+1} := \{[a,b] \mid a, b \in T_i\}$ for $i \geqslant 2$. We next claim that for every $m$,

$$T_m \subseteq C^mG.$$

Indeed, $T_2 \subseteq C^2G$. Assume inductively that $T_i \subseteq C^iG$. Then $T_{i+1}$ consists of commutators $[a,b]$ with $a, b \in T_i \subset C^i$ For such $a, b$, clearly,

$$[a,b] \in [C^iG, G] = C^{i+1}G.$$

In particular, for $i \geqslant k$, $T_i = \{1\}$. We then take

$$\widehat{X} := \bigcup_{i \geqslant 1} T_i.$$

□

Definition 12.41. Let $G$ be a nilpotent finitely generated group. We call a finite generating set $S$ of $G$ an *lcs-generating set* (where lcs stands for 'lower central series') if for every $i \geqslant 1$, $S_i := S \cap C^i G$ generates $C^i G$. For such a generating set we denote by $S_i'$ the complement $S_i \setminus S_{i+1}$.

Note that for any generating set $X$, the set $\widehat{X}$ is lcs-generating, according to Lemma 10.31. Observe also that the projection of an lcs-generating set to every quotient $G/C^{i+1}G$ is again an lcs-generating set.

We say that an lcs-generating set $T$ of $G$ is *closed* if $T = \widehat{T}$, i.e., $T$ is closed with respect to the operation of taking commutators. Given a closed lcs-generating set we filter $T$ by

$$T^i := T \cap C^i G$$

and set $T_i := T^i \setminus T^{i+1}$.

Definition 12.42. If $G$ is a finitely generated nilpotent group and $S$ is an *lcs-generating set* of $G$, then for any word $w$ in $S \cup S^{-1}$ we define its length $|w|_S$ as usual and its *i-length* $|w|_i$ as the number of occurrences of letters from $S_i' \cup S_i'^{-1}$ in the word $w$.

The *lcs-length of* a word $w$ is the finite sequence $(|w|_0, |w|_1, \ldots, |w|_k, \ldots)$. An element $g$ in $G$ is said to have *lcs-length at most* $(r_0, r_1, .., r_k, \ldots)$,

$$lcs_S(g) \leqslant (r_0, r_1, \ldots, r_k, \ldots),$$

if $g$ can be expressed as a word in $S \cup S^{-1}$ of lcs-length $(n_0, n_1, \ldots, n_k, \ldots)$, with $n_i \leqslant r_i$ for all $i \geqslant 0$.

We now are ready to prove the following:

Proposition 12.43. *Let $G$ be a finitely generated nilpotent group of class $k+1$ and let $H := C^i G$, $i \leqslant k$. Then:*

(1) *For $g \in H$,*

$$\operatorname{dist}_H(1, g) \asymp \operatorname{dist}_G(1, g)^i .$$

(2) *The distortion function $\Delta_G^H(n) \asymp n^i$.*

Statement (2) is an immediate consequence of (1). We now prove (1). The relation $\operatorname{dist}_G(1, g)^i \preceq \operatorname{dist}_H(1, g)$ is proven in Corollary 12.39, (1). In what follows we prove $\operatorname{dist}_H(1, g) \preceq \operatorname{dist}_G(1, g)^i$ . As we observed before, it suffices to consider the case when $G$ is torsion-free, which we will assume from now on. The main step in the proof is the following result.

Lemma 12.44. *Let $G$ be a finitely generated torsion-free nilpotent group of class $k$ with an lcs generating set $S$. Then there exists a sequence of closed lcs generating sets $S^{(i)}$ of subgroups $C^i G$, $i = 1, \ldots, k$, so that the following holds:*

*1. For every pair of numbers $\lambda > 0$, $r \geqslant 1$ and every element $g \in C^i G$ with $lcs_S(g) \leqslant (\lambda r, \lambda r^2, ..., \lambda r^k)$, we have*

$$lcs_{S^{(i)}}(g) \leqslant (\lambda_i r^i, \lambda r^{i+1}, \ldots, \lambda_i r^k)$$

*where $\lambda_i$ depends only on $S$ and on $\lambda$.*

*2. Furthermore,*

$$|g|_{S^{(i)}} \leqslant 2\lambda_i r^i$$

PROOF. The proof is induction on $i$. The assertion is clear for $i = 1$ since we can simply take $S^{(1)} = \widehat{S}$: The new lcs generating set of $G$ is closed and the word length (as well as the lcs–length) increases only by a constant factor depending only on $S$, see Exercise 4.72 or Theorem 5.29.

We will describe only the induction step $1 \to 2$ since the general induction $i \to i+1$ is identical (replacing $G = C^1G$ with $C^iG$).

In what follows we fix the constants $\lambda > 0$ and $r \geqslant 1$ and consider elements $g \in G$ such that $lcs_S(g) \leqslant (\lambda r, \lambda r^2, ..., \lambda r^k)$. Our goal is to construct a new lcs generating set $T$ for $G$ so that whenever $g$ as above satisfies $g \in C^2G$, we also have

$$lcs_T(g) \leqslant (0, \lambda_2 r^2, \ldots, \lambda_2 r^k).$$

We then will take $S^{(2)} := T^2$.

We first modify the generating set $S$ by replacing it with another closed lcs generating set $T$ so that $T_1$ projects to a free generating set of the abelianization $G_{ab} = G/G' = G/C^2G$ (recall that $G_{ab}$ is torsion-free since $G$ is). This again increases lengths of all words by a uniformly bounded factor, we retain the notation $\lambda$ for the constant so that $lcs_T(g) \leqslant (\lambda r, \lambda r^2, \ldots, \lambda r^k)$. We let $t_1, \ldots, t_m$ denote the elements of $T_1$.

Let $w = w_0$ be a word in $T$ representing $g$, so that $lcs_T(w) \leqslant (\lambda r, \lambda r^2, ..., \lambda r^k)$. Let $\ell = \ell_1$ denote $|w_0|_1$, i.e., the number of times the letters $t^{\pm 1}$, $t \in T_1$, appear in $w_0$. We next construct, by induction on $j \leqslant \ell$, a sequence of words $(w_j)_{0 \leqslant j \leqslant \ell}$ all representing $g$, such that for every $j$, $w_j = v_{j-1} t_{r_j}^{\pm 1} u_j$ for some $t = t_{r_j} \in T_1$, and:

(a) $t^{\pm 1}$ occurs at most $\ell - j$ times in $u_j$.

(b) For every $i \geqslant 1$,

$$|w_j|_i \leqslant |w_{j-1}|_i + |w_{j-1}|_{i-1}$$

where, we set $|x|_0 := 0$ for every word $x$.

(c) $v_j = v_{j-1} t_{r_j}^{\pm 1}$ with $v_0$ being the empty word.

(d) In the words $v_j$, letters $t_p^{\pm 1}$ always in the increasing order with respect to $p$, i.e., $t_p^{\pm 1}$ is always to the left of $t_q^{\pm 1}$, whenever $p < q$.

Namely, $w_1$ is obtained from $w_0$ by considering the left-most occurrence of $t_1^{\pm 1}, t_1 \in T_1$, and moving $t_1^{\pm 1}$ to the front of the word *via* the relations

$$x t_1^{\pm 1} = t_1^{\pm 1} x \left[x^{-1}, t_1^{\mp 1}\right]$$

Likewise $w_{j+1}$ is obtained from $w_j$ by considering the left-most occurrence of $t_{r_{j+1}}^{\pm 1}$ in $u_j$ and moving it to the left in the similar fashion, where $r_j$ is a weakly increasing sequence. Note that for each $t = t_{r_j}$, $\left[x^{-1}, t^{\mp 1}\right]$ is in the set $T^2$ and the number of occurrences of $t^{\pm 1}$ does not increase in this process. Then properties (a)—(d) are immediate.

In the end of the induction process, we convert $w$ to a word $w_\ell$ which has the form

$$t_1^{d_1} \ldots t_m^{d_m} w',$$

where $w'$ is a word in the alphabet $T^2 \cup (T^2)^{-1}$. Since the set $T_1$ projects to a free basis of $G_{ab}$ and $g \in C^2G$, it follows that

$$t_1^{d_1} \dots t_m^{d_m} = 1$$

in $G$; thus, the element $g \in G$ is represented by the word $w'$. Our next goal is to estimate the lcs–length of the word $w'$.

Using (b) we obtain by induction on $j$ that for every $i = 1, \dots, k$ and $j = 1, \dots \ell$,

$$|w_j|_i \leqslant \sum_{s=0}^{a} \binom{j}{s} |w_0|_{i-s} \tag{12.13}$$

where $a = \min(i-1, j)$. The induction step follows from the formula

$$\binom{j}{s} + \binom{j}{s-1} = \binom{j+1}{s}.$$

Since $|w_0|_{i-s} \leqslant \lambda r^{i-s}$ and

$$\binom{j}{s} \leqslant j^s \leqslant \ell^s \leqslant \lambda r^s,$$

we obtain:

$$|w_j|_i \leqslant \sum_{s=0}^{a} \ell^s \lambda r^{i-s} \leqslant \sum_{s=0}^{a} \lambda r^s \lambda r^{i-s} \leqslant i\lambda^i r^i.$$

In particular,

$$|w'|_i \leqslant |w_\ell|_i \leqslant i\lambda^i r^i \leqslant \lambda_2 r^i,$$

where $\lambda_2 = k\lambda^k$. This proves Part 1 of Lemma. In order to prove Part 2, we observe that, for $r \geqslant 2$,

$$|g|_{S^{(i)}} \leqslant \sum_{j=i}^{k} \lambda_i r^j = \lambda_i \frac{r^{i+1} - r^i}{r-1} \leqslant 2\lambda_i r^i.$$

□

Now, we can conclude the proof of Proposition 12.43, Part (2). We take a redundant lcs generating set $S$ of $G$, so that the subset $S_1$ already generates $G$. Again, it suffices to prove the inequality

$$|g|_{S'} \leqslant \mu |g|_{S_1}^i$$

for some finite generating set $S'$ of $C^iG$ and a constant $\mu$ independent of $g \in C^iG$.

Applying Lemma 12.44, we obtain a new lcs generating set $T := S^{(i)}$ of $G$. Let $w$ be a shortest word in the alphabet $S_1 \cup S_1^{-1}$ representing an element $g \in C^iG$, let $r$ denote the length of $w$. Then

$$lcs_S(w) \leqslant (r, 0, \dots, 0) \leqslant (r, r^2, \dots, r^k).$$

According to Lemma 12.44, $g$ is represented by a word $w'$ in the alphabet $T_i \cup T_i^{-1}$ so that

$$|w'|_{T_i} \leqslant \mu r^i,$$

where $\mu = 2\lambda_i$ and, thus, $\mu$ depends only on $S$. Therefore, for the generating set $S' = T_i$ of the group $C^iG$,

$$|g|_{S'} \leqslant |w'|_{T_i} \leqslant \mu r^i = \mu |g|_{S_1}^i\ .$$

□

Another interesting consequence of Lemma 12.44 is a control on the exponents of the bounded generation property for nilpotent groups.

Proposition 12.45 (Controlled bounded generation for nilpotent groups). *Let $G$ be a finitely generated nilpotent group of class $k$. Let $S$ be an lcs generating set of $G$, and $S_i \subset S$ be a subset*

*For every $i \in \{1,..,k\}$, let $S_i = \{t_{i1}, \ldots, t_{iq_i}\}$ be a set of elements in $C^iG$ projecting onto a set of generators of the abelian group $C^iG/C^{i+1}G$, and let*

$$S = \bigcup_{i=1}^{k} (S_i)$$

*(so $S$ is an lcs-generating set for $G$). Then every element $g \in G$ can be written as*

$$g = \prod_{i=1}^{k} t_{i1}^{m_{i1}} \cdots t_{iq_i}^{m_{iq_i}},$$

*with $m_{ij} \in \mathbb{Z}$, such that $|m_{i1}| + \ldots + |m_{iq_i}| \leqslant C|g|_S^i$, for every $i \in \{1,..,k\}$, , where $C$ is a constant depending only on $G$ and on $S$.*

Proof. We argue by induction on the class $k$. For $k = 1$ the group is abelian and the statement is obvious. Assume that the statement is true for the class $k-1$ and let $G$ be a nilpotent group of class $k \geqslant 2$. Let $S_i$ and $S$ be as in the statement of the proposition, and let $g$ be an arbitrary element in $G$. The induction hypothesis implies that $g = pc$, where $c \in C^kG$ and

$$p = \prod_{i=1}^{k-1} t_{i1}^{m_{i1}} \cdots t_{iq_i}^{m_{iq_i}},$$

where $m_{ij} \in \mathbb{Z}$ are such that

$$|m_{i1}| + \ldots + |m_{iq_i}| \leqslant C|g|_S^i, \tag{12.14}$$

for every $i \in \{1,..,k\}$, where $C$ is a constant depending only on $G$ and $S$.

Then, by the inequalities (12.14), the element $c = p^{-1}g$ in $C^kG$ has lcs-length with respect to $S$ at most $(\lambda r, \lambda r^2, \ldots, \lambda r^k)$, where $r = |g|_S$ and $\lambda = C+1$. Lemma 12.44 then implies that there exists a new generating set $T$ of $C^kG$, (determined by $S$), such that $|c|_T \leqslant \mu r^k$, where $\mu$ only depends on $T$. Then $c = t_{k1}^{m_{k1}} \cdots t_{kq_k}^{m_{kq_k}}$, where $|m_{k1}| + \ldots + |m_{kq_k}| \leqslant \eta |c|_S \leqslant \eta\mu r^k$, where $\eta$ depends only on $S$. Now, the assertion follows by combining product decompositions of $p$ and $c$. $\square$

Proposition 12.43 generalizes to all subgroups:

Proposition 12.46. *Let $G$ be a finitely generated nilpotent group of class $k$. Then, for every subgroup $H$ in a $G$, $\Delta_G^H(n) \preceq n^\gamma$, where $\gamma = 2^{k-1}$.*

Proof. We prove the statement by induction on the nilpotency class $k$ of $G$. The statement is obviously true for $k = 1$ since subgroups of abelian groups are undistorted. Assume that proposition holds for $k$ and consider a subgroup $H$ in a group $G$ of nilpotency class $k+1$. Set $G' = C^2G$ and $H' := H \cap G'$. Consider a finite generating set $X'$ of $H'$ contained in a finite generating set $X$ of $H$ and let $S$ be a finite generating set of $G$ containing $X$ and containing a generating set $S'$ of $G'$.

Let $h$ be an arbitrary element in $H$. Note that $G/G'$ is abelian, $H/H'$ embeds naturally in $G/G'$ as a subgroup with linear distortion function. Thus, there exists a word $w$ in $X$ of length at most $\lambda_1|h|_S$ such that $h = wh'$, where $h' \in H'$, and $\lambda_1$ is independent of $h$. By Proposition 12.43 we have that

$$|h'|_{S'} \leqslant \lambda_2 |h'|_S^2,$$

for some $\lambda_2$ independent of $h$. Note that $h' = w^{-1}h$ satisfies

$$|h'|_S \leqslant |w|_S + |h|_S \leqslant (\lambda_1 + 1)|h|_S.$$

By the induction hypothesis, $|h'|_{X'} \leqslant \lambda_3 |h'|_{S'}^{\beta}$ where $\beta = 2^{k-1}$ and $\lambda_3$ independent of $h'$. It follows that

$$|h'|_X \leqslant |h'|_{X'} \leqslant \lambda_2 \lambda_3^2 |h'|_S^{2\beta}.$$

We conclude that

$$|h|_X \leqslant |w|_X + |h'|_X \leqslant \lambda_1 |h|_S + \lambda_2 \lambda_3^2 |h'|_S^{2\beta} \leqslant$$
$$\lambda_1 |h|_S + \lambda_2 \lambda_3^2 (\lambda_1 + 1)^{2\mathfrak{G}} |h|_S^{2\beta} \leqslant \lambda_5 |h|_S^{\alpha},$$

where $\alpha = 2\beta = 2^k$. □

In fact a stronger statement holds:

Theorem 12.47. *For every infinite subgroup $H$ in a finitely generated nilpotent group $G$ there exists a rational positive number $\alpha$ such that*

$$\Delta_G^H(n) \asymp n^{\alpha}.$$

Theorem 12.47 was originally proven by M. Gromov in **[Gro93]** (see also **[Var99]**); later on, an explicit computation of the possible exponents $\alpha$ was established by D. Osin in **[Osi01]**. More precisely, given an element of infinite order $h$ in a nilpotent group $G$, its *weight* in $G$, $\nu_G(h)$, is the defined as the maximal $i$ such that $\langle h \rangle \cap C^i G \neq \{1\}$. The exponent $\alpha$ in Theorem 12.47 is the maximum of the fractions $\frac{\nu_G(h)}{\nu_H(h)}$ over all the elements $h$ of infinite order in $H$.

## 12.5. Polynomial growth of nilpotent groups

Theorem 12.48 (Bass–Guivarc'h Theorem). *Let $G$ be a finitely generated nilpotent group. Assume that $G$ is of class $k$ and, for every $i \geqslant 1$, let $m_i$ be the rank of the abelian group*

$$C^i G / C^{i+1} G.$$

*Define the number $d = d(G) = \sum_{i=1}^k i m_i$. Then the growth function of $G$ satisfies*

$$\mathfrak{G}_G(n) \asymp n^d. \qquad (12.15)$$

Proof. In the proof below, $\lambda_i$'s are constants depending only on the generating set of the group $G$. We will use the notation $B_G(1, r)$ to denote the $r$-ball in the group $G$ centered at $1 \in G$, with the word metric given by suitable finite generating set of $G$.

We argue by induction on the class $k$. For $k = 1$ the group $G$ is abelian and the statement is obvious.

Assume that the statement holds for $k - 1$ and consider $G$ of class $k \geqslant 2$; let $H = C^k G$ be the last non-trivial subgroup in the lower central series of $G$. Let $d_1 = d - k\, m_k$. If $H$ is finite then $m_k = 0$, we apply the induction hypothesis for $G/H$; since $G$ and $G/H$ have equivalent growth functions the result follows.

We now assume that $H$ is infinite, i.e. $m_k \geqslant 1$. Fix a finite generating set $S$ of $G$ and use its projection as the generating set of $G/H$.

*Lower bound.* By our choice of generating sets, the ball $B_G(1, r)$ maps onto the ball $B_{G/H}(1, r)$ under the projection $G \to G/H$. The induction hypothesis applied to $G/H$ implies that

$$N = card(B_{G/H}(1, n)) \geqslant \lambda_1 n^{d_1}.$$

Let $\{g_1, \ldots, g_N\} \subset B_G(1, n)$ denote the preimage of $B_{G/H}(1, n)$. Since the abelian group $H$ has growth function $t^{m_k}$, Part (1) of Corollary 12.39 implies that

$$card(B_G(1, n) \cap H) \geqslant \lambda_2 n^{km_k}$$

Therefore, the ball $B_G(1, 2n)$ contains the set

$$\bigcup_{i=1}^{N} g_i(B_G(1, n) \cap H)$$

of cardinality at least

$$N\lambda_2 n^{km_k} \geqslant \lambda_1 \lambda_2 n^{d_1 + km_k} = \lambda_3 n^d = \lambda_3 2^{-d} (2n)^d.$$

Thus, for even $t = 2n$,

$$\mathfrak{G}_G(t) \geqslant \lambda_4 t^d.$$

The case of odd $t$ is left as an exercise to the reader.

*Upper bound.* The proof is analogous to the lower bound. Recall that the image of $B_G(1, n)$ in $G/H$ is the ball $B_{G/H}(1, n)$. By the induction hypothesis there exist at most $\lambda_5 n^{d_1}$ elements

$$\bar{g}_1, \ldots, \bar{g}_N \in B_{G/H}(1, n),$$

which are projections of elements $g_i \in B_G(1, n)$, $i = 1, \ldots, N$. Assume that $g = g_i h \in g_i H$. Then $|h|_S \leqslant |g|_S + |g_i|_S \leqslant 2n$. By Proposition 12.43 there are at most $\lambda_6 n^{km_k}$ elements $h \in H$ satisfying this inequality. It then follows that there are at most $\lambda_5 \lambda_6 n^{d_1 + km_k} = \lambda_7 n^d$ distinct elements $g \in B_G(1, n)$. □

### 12.6. Wolf's Theorem

NOTATION 12.49. If $G$ is a group, a semidirect product $G \rtimes_\Phi \mathbb{Z}$ is defined by a homomorphism $\Phi : \mathbb{Z} \to \mathrm{Aut}\,(G)$. The latter homomorphism is entirely determined by $\Phi(1) = \varphi$. Following the notation in Section 12.3, we set

$$G \rtimes_\varphi \mathbb{Z} := G \rtimes_\Phi \mathbb{Z}$$

PROPOSITION 12.50. *Let $G$ be a finitely generated nilpotent group and let $\varphi \in Aut(G)$. Then the polycyclic group $S = G \rtimes_\varphi \mathbb{Z}$ is*

(1) *either virtually nilpotent;*
(2) *or has exponential growth.*

REMARK 12.51. The statement (1) in Proposition 12.50 cannot be improved to '$G$ nilpotent', see Remark 12.27, Part (2).

PROOF. The automorphism $\varphi$ preserves the lower central series of $G$; let $\theta_i$ denote the restriction of $\varphi$ to $C^iG$. Then $\theta_i$ projects to an automorphism $\varphi_i$ of the finitely generated abelian group $B_i := C^iG/C^{i+1}G$. Therefore $\varphi_i$ induces an automorphism $\psi_i$ of $\mathrm{Tor}\, B_i$ and an automorphism $\overline{\varphi}_i$ of $B_i/\mathrm{Tor}\, B_i \simeq \mathbb{Z}^{m_i}$. Each automorphism $\overline{\varphi}_i$ is determined by a matrix $M_i$ in $GL(m_i, \mathbb{Z})$. Analogously to the proof of Proposition 12.26, we have two case to consider:

(1) All matrices $M_i$ only have eigenvalues of absolute value 1, hence (by Lemma 10.23) all eigenvalues are roots of unity. Then there exists $N$ such that for the automorphism $\varphi^N$ the corresponding matrices $M_i$ have only eigenvalues equal to 1 and the corresponding automorphisms of finite abelian groups $\psi_i : \mathrm{Tor}\, B_i \to \mathrm{Tor}\, B_i$ are equal to the identity. Without loss of generality we may therefore assume that all matrices $M_i$ have all eigenvalues 1 and all the $\psi_i$ are the identity automorphisms, since this can be achieved by replacing $S$ with its finite index subgroup

$$G \rtimes_{\varphi} (N\mathbb{Z}) \simeq G \rtimes_{\varphi^N} \mathbb{Z}.$$

Lemma 10.22 applied to each $\overline{\varphi}_i$ and the similar statement applied to each $\psi_i = \mathrm{id}_{\mathrm{Tor}\, B_i}$, imply that the lower central series of $G$ is a sub-series of a series

$$\{1\} = H_n \leqslant H_{n-1} \leqslant \ldots \leqslant H_1 \leqslant H_0 = G$$

such that each $H_i/H_{i+1}$ is cyclic, $\varphi$ preserves each $H_i$ and induces the identity map on $H_i/H_{i+1}$. We denote by $t$ the generator of the semidirect factor $\mathbb{Z}$ in the decomposition $S = G \rtimes \mathbb{Z}$. Then, by the definition of the semidirect product, for every $g \in G$, $tgt^{-1} = \varphi(g)$. The fact that $\varphi$ acts as the identity on each $H_i/H_{i+1}$ implies that $t^k(hH_{i+1})t^{-k} = hH_{i+1}$ for every $h$ in $H_i$; equivalently

$$[t^k, h] \in H_{i+1} \tag{12.16}$$

for every such $h$.

Since the group $S$ is the middle term of the exact sequence

$$1 \to G \to S \to \mathbb{Z} \to 1$$

and the projection of $C^2S = [S, S]$ to $\mathbb{Z}$ is $\{1\}$, it follows that $C^2S \leqslant G$.

We claim that for every $i \geqslant 0$, $[S, H_i] \subseteq H_{i+1}$. Indeed, every element $s \in S$ has the form $s = gt^k$, with $g \in G$ and $k \in \mathbb{Z}$; consider an arbitrary element $h \in H_i$. In view of the commutator identity (3) in Lemma 10.25,

$$[h, gt^k] = [h, g][g, [h, t^k]][h, t^k] \ .$$

According to (12.16), $[h, t^k] \in H_{i+1}$. Also, since the lower central series of $G$ is a subseries of $(H_i)$, there exists $r \geqslant 1$ such that $C^rG \geqslant H_i \geqslant H_{i+1} \geqslant C^{r+1}G$. Then, $h \in H_i \leqslant C^rG$ and $[h, g] \in C^{r+1}G \leqslant H_{i+1}$. Likewise, as $[h, t^k] \in H_{i+1} \leqslant C^rG$, the commutator $[g, [h, t^k]] \in C^{r+1}G \leqslant H_{i+1}$. By putting it all together, we conclude that $[h, s] \in H_{i+1}$ and, hence, $[S, H_i] \subseteq H_{i+1}$

The easy induction now shows that $C^{i+2}S \leqslant H_i$ for every $i \geqslant 1$; in particular, $C^{n+2}S \leqslant H_n = \{1\}$. Therefore, $S$ is virtually nilpotent.

(2) Assume that at least one matrix $M_i$ has an eigenvalue with absolute value strictly greater than 1, in particular, $m_i \geqslant 2$. The group $S$ contains the subgroup

$$S_i := C^iG \rtimes_{\theta_i} \mathbb{Z}.$$

Furthermore, the subgroup $C^{i+1}G$ is normal in $S_i$ and $S_i/C^{i+1}G \cong B_i \rtimes_{\varphi_i} \mathbb{Z}$. Lastly,

$$B_i \rtimes_{\varphi_i} \mathbb{Z}/\mathrm{Tor}\, B_i \cong \mathbb{Z}^{m_i} \rtimes_{M_i} \mathbb{Z}.$$

According to Proposition 12.26, the group $\mathbb{Z}^{m_i} \rtimes_{M_i} \mathbb{Z}$ has exponential growth. Therefore, in view of Proposition 12.10, parts (a) and (c), the groups $B_i \rtimes_{\varphi_i} \mathbb{Z}$,

$S_i/C^{i+1}G$, $S_i$, and, hence, $S$ all have exponential growth. Thus, in the case (2), $S$ has exponential growth. □

Proposition 12.50 combined with Proposition 3.39 on subgroups of finite index in finitely generated groups will be used to prove Wolf's Theorem.

THEOREM 12.52 (Wolf's Theorem). *A polycyclic group is either virtually nilpotent or has exponential growth.*

PROOF. According to Proposition 11.8, it suffices to prove the statement for poly-$C_\infty$ groups. Let $G$ be a poly-$C_\infty$ group, and consider a finite subnormal descending series

$$G = N_0 \geqslant N_1 \geqslant \ldots \geqslant N_n \geqslant N_{n+1} = \{1\}$$

such that $N_i/N_{i+1} \simeq \mathbb{Z}$ for every $i \geqslant 0$. We argue by induction on $n$. For $n=0$ the group $G$ is infinite cyclic and the statement is obvious. Assume that the assertion of theorem holds for $n$ and consider the case of $n+1$. By the induction hypothesis, the subgroup $N_1 \leqslant G$ is either virtually nilpotent or has exponential growth. In the second case the group $G$ itself has exponential growth.

Assume that $N_1$ is virtually nilpotent. Corollary 4.21 implies that $G$ decomposes as a semidirect product $N_1 \rtimes_\psi \mathbb{Z}$, corresponding to a homomorphism $\Psi : \mathbb{Z} \to \mathrm{Aut}\,(N_1)$, $\theta = \Psi(1)$.

By hypothesis $N_1$ contains a nilpotent subgroup $H$ of finite index. According to Proposition 3.39, (2), we may moreover assume that $H$ is characteristic in $N_1$. In particular $H$ is invariant under the automorphisms $\psi$. We retain the notation $\psi$ for the restriction $\theta|H$. Therefore, $H \rtimes_\theta \mathbb{Z}$ is a normal subgroup of $G$. Moreover, $H \rtimes_\theta \mathbb{Z}$ has finite index in $G$, since $G/(H \rtimes_\theta \mathbb{Z})$ is the quotient of the finite group $N_1/H$.

By Proposition 12.50, $H \rtimes_\theta \mathbb{Z}$ is either virtually nilpotent or of exponential growth. Therefore, the same alternative then holds for $N_1 \rtimes_\theta \mathbb{Z} = G$. □

## 12.7. Milnor's theorem

THEOREM 12.53. *A finitely generated solvable group is either polycyclic or has exponential growth.*

We begin the proof by establishing a property on conjugates implied by sub-exponential growth:

LEMMA 12.54. *If a finitely generated group $G$ has sub-exponential growth then for every $\beta_1, \ldots, \beta_m, g \in G$, the set of conjugates*

$$\{g^k \beta_i g^{-k} \mid k \in \mathbb{Z}, i = 1, \ldots, m\}$$

*generates a finitely generated subgroup $N \leqslant G$.*

PROOF. By induction on $i$, it suffices to prove lemma for $m = 1$.

NOTATION 12.55. We set $\alpha := \beta_1$ and let $\alpha_k$ denote $g^k \alpha g^{-k}$ for every $k \in \mathbb{Z}$.

The goal is to prove that finitely many elements in the set $\{\alpha_k \mid k \in \mathbb{Z}\}$ generate $N$.

Identify $\mathbb{Z}_2$ with the set $\{0, 1\}$ and consider the map

$$\begin{array}{rcc} \mu = \mu_m : \prod_{i=0}^m \mathbb{Z}_2 & \to & G \\ \mu : (s_i) & \mapsto & g\alpha^{s_0} g\alpha^{s_1} \cdots g\alpha^{s_m} . \end{array}$$

If for every $m \in \mathbb{N}$ the map $\mu$ is injective then we have $2^{m+1}$ products as above, and if $g, g\alpha$ are in the set of generators of $G$, all these products are in $B_G(1, m+1)$. This contradicts the hypothesis that $G$ has sub-exponential growth. It follows that there exists some $m$ and two distinct sequences $(s_i), (t_i)$ in $\prod_{i=0}^m \mathbb{Z}_2$ such that

$$g\alpha^{s_0} g\alpha^{s_1} \cdots g\alpha^{s_m} = g\alpha^{t_0} g\alpha^{t_1} \cdots g\alpha^{t_m} . \tag{12.17}$$

Assume that $m$ is minimal with this property. This, in particular, implies that $s_0 \neq t_0$ and $s_m \neq t_m$.

With the notation of 12.55,

$$g\alpha^{s_0} g\alpha^{s_1} \cdots g\alpha^{s_m} = \alpha_1^{s_0}\alpha_2^{s_1} \cdots \alpha_{m+1}^{s_m} g^{m+1} .$$

The equality (12.17) the becomes

$$\alpha_1^{s_0}\alpha_2^{s_1} \cdots \alpha_{m+1}^{s_m} = \alpha_1^{t_0}\alpha_2^{t_1} \cdots \alpha_{m+1}^{t_m} .$$

Since $s_m \neq t_m$ it follows that $s_m - t_m = \pm 1$. Then

$$\alpha_{m+1}^{\pm 1} = \alpha_m^{-s_{m-1}} \cdots \alpha_2^{-s_1} \alpha_1^{t_0 - s_0} \alpha_2^{t_1} \cdots \alpha_m^{t_{m-1}} . \tag{12.18}$$

If in (12.18) we conjugate by $g$, we obtain that

$$\alpha_{m+2}^{\pm 1} = \alpha_{m+1}^{-s_{m-1}} \cdots \alpha_3^{-s_1} \alpha_2^{t_0 - s_0} \alpha_3^{t_1} \cdots \alpha_{m+1}^{t_{m-1}} .$$

This and (12.18) imply that $\alpha_{m+2}$ is a product of powers of $\alpha_1, \dots, \alpha_m$. Then, by induction, every $\alpha_n$ with $n \in \mathbb{N}$ is a product of powers of $\alpha_1, \dots, \alpha_m$, and the same is true for $\alpha_n$ with $n \in \mathbb{Z}$ by considering inverses. Therefore, every generator $\alpha_n$ of $N$ belongs to the subgroup of $N$ generated by the elements $\alpha_1, \dots, \alpha_m$. Hence, the elements $\alpha_1, \dots, \alpha_m$ generate $N$. □

Exercise 12.56. Use Lemma 12.54 to prove that the finitely generated group $H$ described in Example 4.7 has exponential growth.

We now are ready to prove Theorem 12.53; our proof by induction on the derived length $d$. For $d = 1$ the group $G$ is finitely generated abelian and the statement is true. Assume that the alternative holds for finitely generated solvable groups of derived length $\leqslant d$ and consider $G$ of derived length $d+1$. Then $H = G/G^{(d)}$ is finitely generated solvable of derived length $d$. By the induction hypothesis, either $H$ has exponential growth or $H$ is polycyclic. If $H$ has exponential growth then $G$ has exponential growth too (see statement (c) in Proposition 12.10).

Assume therefore that $H$ is polycyclic. In particular, $H$ is finitely presented by Proposition 11.12. Theorem 12.53 will follow from:

Lemma 12.57. *Consider a short exact sequence*

$$1 \to A \to G \xrightarrow{\pi} H \to 1 , \quad \text{with } A \text{ abelian and } G \text{ finitely generated.} \tag{12.19}$$

*If $H$ is polycyclic then $G$ is either polycyclic or has exponential growth.*

Proof. We assume that $G$ has sub-exponential growth and will prove that $G$ is polycyclic. This is equivalent to the fact that $A$ is finitely generated. Since $H$ is polycyclic, it has bounded generation property (see Proposition 11.3); hence, there exist finitely many elements $h_1, \dots, h_q$ in $H$ such that every element $h \in H$ can be written as

$$h = h_1^{m_1} h_2^{m_2} \cdots h_q^{m_q} , \text{ with } m_1, m_2, \dots, m_q \in \mathbb{Z} .$$

Choose $g_i \in G$ such that $\pi(g_i) = h_i$ for every $i \in \{1, 2, \ldots, q\}$. Then every element $g \in G$ can be written as

$$g = g_1^{m_1} g_2^{m_2} \cdots g_q^{m_q} a\,, \text{ with } m_1, m_2, \ldots, m_q \in \mathbb{Z} \text{ and } a \in A\,. \tag{12.20}$$

Since $H$ is finitely presented, by Lemma 4.26 there exist finitely many elements $a_1, \ldots, a_k$ in $A$ such that every element in $A$ is a product of $G$-conjugates of $a_1, \ldots, a_k$. According to (12.20), all the conjugates of $a_j$ are of the form

$$g_1^{m_1} g_2^{m_2} \cdots g_q^{m_q} a_j \left(g_1^{m_1} g_2^{m_2} \cdots g_q^{m_q}\right)^{-1}\,. \tag{12.21}$$

By Lemma 12.54, the group $A_q$ generated by all conjugates $g_q^m a_j g_q^{-m}$ with $m \in \mathbb{Z}$ and $j \in \{1, \ldots, k\}$ is finitely generated. Let $S_q$ be its finite generating set. Then the conjugates $g_{q-1}^n g_q^m a_j g_q^{-m} g_{q-1}^{-n}$ with $m, n \in \mathbb{Z}$ and $j \in \{1, \ldots, k\}$ are in the group generated by $g_{q-1}^n s g_{q-1}^{-n}$ with $n \in \mathbb{Z}$ and $s \in S_q$. Again Lemma 12.54 implies that such a group is finitely generated. Continuing by induction we conclude that the group $A$ generated by all the conjugates in (12.21), is finitely generated. Hence, $G$ is polycyclic. □

This also concludes the proof of Milnor's theorem, Theorem 12.53. □

By combining theorems of Milnor and Wolf one obtains:

THEOREM 12.58. *Every finitely generated solvable group either is virtually nilpotent or it has exponential growth.*

Note that the above was later strengthened by Rosenblatt as follows:

THEOREM 12.59 ([**Ros74**]). *Every finitely generated solvable group either is virtually nilpotent or it contains a free non-abelian subsemigroup.*

Another application of Lemma 12.54 is the following proposition which will be used in the proof of Gromov's theorem on groups of polynomial growth:

PROPOSITION 12.60. *Suppose that $G$ is a finitely generated group of sub-exponential growth, which fits in a short exact sequence*

$$1 \to K \to G \xrightarrow{\pi} \mathbb{Z} \to 1.$$

*Then $K$ is finitely generated. Moreover, if $\mathfrak{G}_G(R) \preceq R^d$ then $\mathfrak{G}_K(R) \preceq R^{d-1}$.*

PROOF. Let $\gamma \in G$ be an element which projects to the generator 1 of $\mathbb{Z}$. Let $\{f_1, \ldots, f_k\}$ denote a set of generators of $G$. Then for each $i$ there exists $s_i \in \mathbb{Z}$ such that $\pi(f_i \gamma^{s_i}) = 0 \in \mathbb{Z}$. Define elements $g_i := f_i \gamma^{s_i}$, $i = 1, \ldots, k$. Clearly, the set $\{g_1, \ldots, g_k, \gamma\}$ generates $G$. Without loss of generality we may assume that each generator $g_i$ is nontrivial. Define

$$S := \{\gamma_{m,i} := \gamma^m g_i \gamma^{-m}\ ;\ m \in \mathbb{Z}, i = 1, \ldots, k\}.$$

We claim that the (infinite) set $S$ generates $K$. Indeed, clearly, $S \subset K$. Every $g \in K$ can be written as a word $w = w(g_1, \ldots, g_k, \gamma)$. We then move all entries of powers of $\gamma$ in the word $w$ to the end of $w$ by using the relations

$$\gamma^m g_i = \gamma_{m,i} \gamma^m.$$

As the result, we obtain a word $w' = u\gamma^a$ in the alphabet $S \cup S^{-1} \cup \{\gamma, \gamma^{-1}\}$, where $u$ contains only the letters in $S \cup S^{-1}$ and $a \in \mathbb{Z}$. Since $g$ projects to $0 \in \mathbb{Z}$, $a = 0$. Claim follows.

Lemma 12.54 implies that there exists $M(i)$ so that the subgroup $K$ is generated by the finite set

$$\{\gamma_{l,i}\ ;\ |l| \leqslant M(i), i = 1, \dots, k\}.$$

This proves the first assertion of the Proposition.

Now let us prove the second assertion which estimates the growth function of $K$. Take a finite generating set $Y$ of the subgroup $K$ and set $X := Y \cup \{\gamma\}$, where $\gamma$ is as above. Then $X$ is a generating set of $G$. Given $n \in \mathbb{N}$ let $N := \mathfrak{G}_Y(n)$, where $\mathfrak{G}_Y$ is the growth function of $K$ with respect to the generating set $Y$. Thus there exists a subset

$$H := \{h_1, \dots, h_N\} \subset K$$

where $|h_i|_Y \leqslant n$ and $h_i \neq h_j$ for all $i \neq j$. Then we obtain a set $T$ of $(2n+1) \cdot N$ pairwise distinct elements

$$h_i \gamma^j, \quad -n \leqslant j \leqslant n, \quad i = 1, \dots, N.$$

It is clear that $\|h_i \gamma^j\|_X \leqslant 2n$ for each $h_j \gamma^j \in T$. Therefore

$$n\mathfrak{G}_Y(n) \leqslant (2n+1)\mathfrak{G}_Y(n) = (2n+1)N \leqslant \mathfrak{G}_X(2n) \leqslant C(2n)^d = 2^d C \cdot n^d,$$

for some constant $C$ depending only on $X$. It follows that

$$\mathfrak{G}_Y(n) \leqslant 2^d C \cdot n^{d-1} \preceq n^{d-1}. \quad \square$$

R. Grigorchuk [**Gri83, Gri84a, Gri84b**] constructed finitely generated groups of *intermediate growth*, i.e. their growth is superpolynomial but subexponential. More precisely, Grigorchuk proved that for every sub-exponential function $f$ there exists a group of intermediate growth whose growth function is larger than $f(n)$ for infinitely many $n$. A. Erschler adapted his arguments to show that for every such function $f$, a direct sum of two Grigorchuk groups has growth function larger than $f(n)$ for all but finitely many $n$ [**Ers04**].

The first examples of groups of intermediate growth for which the function is known (up to the equivalence relation $\asymp$), were constructed by L. Bartholdi and A. Erschler in [**BE12**]. For every $k \in \mathbb{N}$, they provided examples of torsion groups $G_k$ and of torsion-free groups $H_k$ such that their growth functions satisfy

$$\mathfrak{G}_{G_k}(x) \asymp \exp\left(x^{1-(1-\alpha)^k}\right),$$

and

$$\mathfrak{G}_{H_k}(x) \asymp \exp\left(\log x \, (x^{1-(1-\alpha)^k}\right),$$

Here, $\alpha$ is the number satisfying $2^{3-\frac{3}{\alpha}} + 2^{2-\frac{2}{\alpha}} + 2^{1-\frac{1}{\alpha}} = 2$.

We note that all currently known groups of intermediate growth have growth larger than $2^{\sqrt{n}}$.

Existence of finitely presented groups of intermediate growth is unknown. In particular the Grigorchuk groups do not answer Question 12.16.

CHAPTER 13

# Tits' Alternative

In this chapter we will prove

THEOREM 13.1 (Tits' Alternative, [**Tit72**]). *Let $L$ be a Lie group with finitely many connected components and $\Gamma \subset L$ be a finitely generated subgroup. Then either $\Gamma$ is virtually solvable or $\Gamma$ contains a free nonabelian subgroup.*

REMARK 13.2. In the above one cannot replace 'virtually solvable' by 'solvable'. Indeed consider the Heisenberg group $H_3 \leqslant GL(3,\mathbb{R})$ and $A_5 \leqslant GL(5,\mathbb{R})$. The group $\Gamma = H_3 \times A_5 \leqslant GL(8,\mathbb{R})$ is not solvable (because $A_5$ is simple) and does not contain a free nonabelian subgroup (because it has polynomial growth).

COROLLARY 13.3. *Suppose that $\Gamma$ is a finitely generated subgroup of $GL(n,\mathbb{R})$. Then $\Gamma$ has either polynomial or exponential growth.*

PROOF. By Tits' Alternative, either $\Gamma$ contains a nonabelian free subgroup (and hence $\Gamma$ has exponential growth) or $\Gamma$ is virtually solvable. For virtually solvable groups the assertion follows from Theorem 12.58. □

### 13.1. Zariski topology and algebraic groups

The proof of Tits' theorem relies in part on some basic results from theory of affine algebraic groups. We recall some terminology and results needed in the argument. For a more thorough presentation see [**Hum75**] and [**OV90**].

The proof of the following general lemma is straightforward, and left as an exercise to the reader.

LEMMA 13.4. *For every commutative ring $A$ the following two statements are equivalent:*

(1) *every ideal in $A$ is finitely generated;*
(2) *the set of ideals satisfies the ascending chain condition (ACC), that is, every ascending chain of ideals*
$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq \cdots$$
*stabilizes, i.e. there exists an integer $N$ such that $I_n = I_N$ for every $n \geqslant N$.*

DEFINITION 13.5. An commutative ring is called *noetherian* if it satisfies one (hence both) statements in Lemma 13.4.

Note that a field seen as a ring is always noetherian. Other examples of noetherian rings come from the following

THEOREM 13.6 (Hilbert's ideal basis theorem, see [**Hum75**]). *If $A$ is a noetherian ring then the ring of multivariable polynomials $A[X_1, ..., X_n]$ is also noetherian.*

From now on, we fix a field $\mathbb{K}$.

Definition 13.7. An *affine algebraic set* in $\mathbb{K}^n$ is a subset $Z$ in $\mathbb{K}^n$ that is the solution set of a system of multivariable polynomial equations $p_j = 0\,,\ \forall j \in J$, with coefficients in $\mathbb{K}$:

$$Z = \{(x_1, .., x_n) \in \mathbb{K}^n\ ;\ p_j(x_1, .., x_n) = 0, j \in J\}.$$

We will frequently say "algebraic subset" when referring to affine algebraic set.

For instance, the algebraic subsets in the affine line (1-dimensional vector space $V$) are finite subsets and the entire of $V$, since every nonzero polynomial in one variable has at most finitely many zeroes.

There is a one-to-one map associating to every algebraic subset in $\mathbb{K}^n$ an ideal in $K[X_1, ..., X_n]$:

$$Z \mapsto I_Z = \{p \in K[X_1, ..., X_n]\ ;\ p|_Z \equiv 0\}\,.$$

Note that $I_Z$ is the kernel of the homomorphism $p \mapsto p|_Z$ from $K[X_1, ..., X_n]$ to the ring of functions on $Z$. Thus, the ring $K[X_1, ..., X_n]/I_Z$ may be seen as a ring of functions on $Z$; this quotient ring is called the *coordinate ring of* $Z$ or the *ring of polynomials on* $Z$, and denoted $\mathbb{K}[Z]$.

Theorem 13.6 and Lemma 13.4 imply the following.

Lemma 13.8.
(1) *Every algebraic set is defined by finitely many equations.*
(2) *The set of algebraic subsets of $\mathbb{K}^n$ satisfies the descending chain condition (DCC): every descending chain of algebraic subsets*

$$Z_1 \supseteq Z_2 \supseteq \cdots \supseteq Z_i \supseteq \cdots$$

*stabilizes, i.e., for some integer $N \geqslant 1$, $Z_i = Z_N$ for every $i \geqslant N$.*

The pair $(Z, \mathbb{K}[Z])$ (a *ringed space*) is an *affine algebraic variety* or simply an *affine variety*, or, by abusing the terminology, just a *(sub)variety*. We will frequently conflate affine varieties and the corresponding algebraic subsets.

Definition 13.9. A *morphism* between two affine varieties $Y$ in $\mathbb{K}^n$ and $Z$ in $\mathbb{K}^m$ is a map of the form $\varphi : Y \to Z$, $\varphi = (\varphi_1, ..., \varphi_m)$, such that $\varphi_i$ is in $\mathbb{K}[Y]$ for every $i \in \{1, 2, ..., m\}$.

Note that every morphism is induced by a morphism $\tilde{\varphi} : \mathbb{K}^n \to \mathbb{K}^m$, $\tilde{\varphi} = (\tilde{\varphi}_1, ..., \tilde{\varphi}_m)$, with $\tilde{\varphi}_i : \mathbb{K}^n \to \mathbb{K}$ a polynomial function for every $i \in \{1, 2, ..., m\}$.

An *isomorphism* between two affine varieties $Y$ and $Z$ is an invertible map $\varphi : Y \to Z$ such that both $\varphi$ and $\varphi^{-1}$ are morphisms. When $Y = Z$, an isomorphism is called an *automorphism*.

Exercise 13.10. 1. If $f : Y \to Z$ is a morphism of affine varieties and $W \subset Z$ is a subvariety, then $f^{-1}(W)$ is a subvariety in $Y$. In particular, every linear automorphism of $V = \mathbb{K}^n$ sends subvarieties to subvarieties and, hence, the notion of a subvariety is independent of the choice of a basis in $V$.

2. Show that the projection map $f : \mathbb{C}^2 \to \mathbb{C}$, $f(x, y) = x$, does not map subvarieties to subvarieties.

Let $V$ be an $n$-dimensional vector space over a field $\mathbb{K}$. The *Zariski topology* on $V$ is the topology having as closed sets all the algebraic subsets in $V$. It is clear that the intersection of algebraic subsets is again an algebraic subset. Let

$Z = Z_1 \cup ... \cup Z_\ell$ be a finite union of algebraic subsets, $Z_i$ defined by a set of polynomials $P_i$ , $i \in \{1, ..., \ell\}$. According to Lemma 13.8, (1), we may take each $P_i$ to be finite. Define the new set of polynomials

$$P := \left\{ p = \prod_{i=1}^{\ell} p_i ; \;\; p_i \in P_i \text{ for every } i \in \{1, ..., \ell\} \right\} .$$

The solution set of the system of equations $p = 0$ , $p \in P$, is $Z$.

The induced topology on a subvariety $Z \subseteq V$ is also called the Zariski topology. Note that this topology can also be defined directly using polynomial functions in $\mathbb{K}[Z]$. According to Exercise 13.10, morphisms between affine varieties are continuous with respect to the Zariski topologies.

The *Zariski closure* of a subset $E \subset V$ can also be defined by means of the set $P_E$ of all polynomials which vanish on $E$, i.e. it coincides with

$$\{x \in V \mid p(x) = 0, \, \forall p \in P_E\} .$$

A subset $Y \subset Z$ in an affine variety is called *Zariski-dense* if its Zariski closure is the entire of $Z$.

Lemma 13.8, Part (2), implies that the closed sets in Zariski topology satisfy the descending chain condition (DCC).

Definition 13.11. A topological space such that the closed sets satisfy the DCC (or, equivalently, with the property that the open sets satisfy the ACC) is called *noetherian.*

Lemma 13.12. *Every subspace of a noetherian topological space (with the subspace topology) is noetherian.*

Proof. Let $X$ be a space with topology $\mathcal{T}$ such that $(X, \mathcal{T})$ is noetherian, and let $Y$ be an arbitrary subset in $X$. Consider a descending chain of closed subsets in $Y$:

$$Z_1 \supseteq Z_2 \supseteq \cdots \supseteq Z_n \supseteq \ldots$$

Every $Z_i$ is equal to $Y \cap C_i$ for some closed set $C_i$ in $X$. We leave it to the reader to check that $C_i$ can be taken equal to the closure $\overline{Z}_i$ of $Z_i$ in $X$.

The descending chain of closed subsets in $X$,

$$\overline{Z}_1 \supseteq \overline{Z}_2 \supseteq \cdots \supseteq \overline{Z}_n \supseteq \ldots$$

stabilizes, hence, so does the chain of the subsets $Z_i$. □

Proposition 13.13. *Every noetherian topological space $X$ is compact.*

Proof. Compactness of $X$ is equivalent to the condition that for every family $\{Z_i \; : \; i \in I\}$ of closed subsets in $X$, if $\bigcap_{i \in I} Z_i = \emptyset$ then there exists a finite subset $J$ of $I$ such that $\bigcap_{j \in J} Z_j = \emptyset$ . Assume that all finite intersections of a family as above are non-empty. Then we construct inductively a descending sequence of closed sets that never stabilizes. The initial step consists of picking an arbitrary set $Z_{i_1}$, with $i_1 \in I$. At the $n$th step we have a non-empty intersection $Z_{i_1} \cap Z_{i_2} \cap ... \cap Z_{i_n}$; hence, there exists $Z_{i_{n+1}}$ with $i_{n+1} \in I$ such that $Z_{i_1} \cap Z_{i_2} \cap ... \cap Z_{i_n} \cap Z_{i_{n+1}}$ is a non-empty proper closed subset of $Z_{i_1} \cap Z_{i_2} \cap ... \cap Z_{i_n}$. □

We now discuss a strong version of connectedness, relevant in the setting of noetherian spaces.

Lemma 13.14. *For a topological space $X$ the following properties are equivalent:*

(1) *every open non-empty subset of $X$ is dense in $X$;*

(2) *two open non-empty subsets have non-empty intersection;*

(3) *$X$ cannot be written as a finite union of proper closed subsets.*

We leave the proof of this lemma as an exercise to the reader.

Definition 13.15. A topological space is called *irreducible* if it is non-empty and one of (hence all) the properties in Lemma 13.14 is (are) satisfied. A subset of a topological space is *irreducible* if, when endowed with the subset topology, it is an irreducible space.

Exercise 13.16. (1) Prove that $\mathbb{K}^n$ with Zariski topology is irreducible.

(2) Prove that an algebraic variety $Z$ is irreducible if and only if $\mathbb{K}[Z]$ does not contain zero divisors.

The following properties are straightforward and their proof is left as an exercise to the reader.

Lemma 13.17. (1) *The image of an irreducible space under a continuous map is irreducible.*

(2) *The cartesian product of two irreducible spaces is an irreducible space, when endowed with the product topology.*

Note that the Zariski topology on $\mathbb{K}^{n+m} = \mathbb{K}^n \times \mathbb{K}^m$ is *not* the product topology. Nevertheless, one has:

Lemma 13.18. *Let $V_1, V_2$ be finite-dimensional vector spaces over $\mathbb{K}$ and $Z_i \subset V_i, i = 1, 2$, be irreducible subvarieties. Then the product $Z := Z_1 \times Z_2 \subset V = V_1 \times V_2$ is an irreducible subvariety in the vector space $V$.*

Proof. Let $Z = W_1 \cup W_2$ be a union of two proper subvarieties. For every $z \in Z_1$ the product $\{z\} \times Z_2$ is isomorphic to $Z_2$ (*via* projection to the second factor) and, hence, irreducible. On the other hand,

$$\{z\} \times Z_2 = ((\{z\} \times Z_2) \cap W_1) \cup ((\{z\} \times Z_2) \cap W_2)$$

is a union of two subvarieties. Thus, for every $z \in Z_1$, one of these subvarieties has to be the entire $\{z\} \times Z_2$. In other words, either $\{z\} \times Z_2 \subset W_1$ or $\{z\} \times Z_2 \subset W_2$. We then partition $Z_1$ in two subsets $A_1, A_2$:

$$A_i = \{z \in Z_1 : \{z\} \times Z_2 \subset W_i\}, i = 1, 2.$$

Since each $W_1, W_2$ is a proper subvariety, both $A_1, A_2$ are proper subsets of $Z_1$. We will now prove that both $A_1, A_2$ are subvarieties in $Z_1$. We will consider the case of $A_1$ since the other case is obtained by relabeling. Let $f_1, \ldots, f_k$ denote generators of the ideal of $W_1$. We will think of each $f_i$ as a function of two variables $f = f(X_1, X_2)$, where $X_k$ stands for the tuple of coordinates in $V_k, k = 1, 2$. Then

$$A_1 = \{z \in Z_2 : f_i(z, z_2) = 0, \forall z \in Z_1, i = 1, \ldots, k\}.$$

However, for every fixed $z \in Z_1$, the function $f_i(z, \cdot)$ is a polynomial function $f_{i,z}$ on $Z_2$. Therefore, $A_1$ is the solution set of the system of polynomial equations on $Z_1$:

$$\{f_{i,z} = 0 : i = 1, \ldots, k, z \in Z_1\}.$$

Therefore, $A_1$ is a subvariety. This contradicts irreducibility of $Z_2$. □

Lemma 13.19. *Let $(X, \mathcal{T})$ be a topological space.*

(1) *A subset $Y$ of $X$ is irreducible if and only if its closure $\overline{Y}$ in $X$ is irreducible.*
(2) *If $Y$ is irreducible and $Y \subseteq A \subseteq \overline{Y}$ then $A$ is irreducible.*
(3) *Every irreducible subset $Y$ of $X$ is contained in a maximal irreducible subset.*
(4) *The maximal irreducible subsets of $X$ are closed and they cover $X$.*

Proof. (1) For every open subset $U \subset X$, $U \cap Y \neq \emptyset$ if and only if $U \cap \overline{Y} \neq \emptyset$. This and Lemma 13.14, (2), imply the equivalence.

(2) Now let $U, V$ be two open sets in $A$. Then $U = A \cap U_1$ and $V = A \cap V_1$, where $U_1, V_1$ are open in $X$. Since $U_1 \cap \overline{Y} \neq \emptyset$ and $V_1 \cap \overline{Y} \neq \emptyset$ it follows that both $U_1$ and $V_1$ have non-empty intersections with $Y$. Then irreducibility of $Y$ implies that $U_1 \cap V_1 \cap Y$ is non-empty, whence $U \cap V \neq \emptyset$.

(3) The family $\mathcal{I}_Y$ of irreducible subsets containing $Y$ has the property that every ascending chain has a maximal element, which is the union. It can be easily verified that the union is again irreducible, using Lemma 13.14, (2).

It follows by Zorn's Lemma that $\mathcal{I}_Y$ has a maximal element.

(4) follows from (1) and (3). □

Theorem 13.20. *A noetherian topological space $X$ is a union of finitely many distinct maximal irreducible subsets $X_1, X_2, ..., X_n$ such that for every $i$, $X_i$ is not contained in $\bigcup_{j \neq i} X_j$. Moreover, every maximal irreducible subset in $X$ coincides with one of the subsets $X_1, X_2, ..., X_n$. This decomposition of $X$ is unique up to a renumbering of the $X_i$'s.*

Proof. Let $\mathcal{F}$ be the collection of closed subsets of $X$ that cannot be written as a finite union of maximal irreducible subsets. Assume that $\mathcal{F}$ is non-empty. Since $X$ is noetherian, $\mathcal{F}$ satisfies the DCC, hence by Zorn's Lemma it contains a minimal element $Y$. As $Y$ is not irreducible, it can be decomposed as $Y = Y_1 \cup Y_2$, where $Y_i$ are closed and, by the minimality of $Y$, both $Y_i$ decompose as finite unions of irreducible subsets (maximal in $Y_i$). According to Lemma 13.19, (3), $Y$ itself can be written as union of finitely many maximal irreducible subsets, a contradiction. It follows that $\mathcal{F}$ is empty.

If $X_i \subseteq \bigcup_{j \neq i} X_j$ then $X_i = \bigcup_{j \neq i} (X_j \cap X_i)$. As $X_i$ is irreducible it follows that $X_i \subseteq X_j$ for some $j \neq i$, hence by maximality $X_i = X_j$, contradicting the fact that we took only distinct maximal irreducible subsets. A similar argument is used to prove that every maximal irreducible subset of $X$ must coincide with one of the sets $X_i$.

Now assume that $X$ can be also written as a union of distinct maximal irreducible subsets $Y_1, Y_2, ..., Y_m$ such that for every $i$, $Y_i$ is not contained in $\bigcup_{j \neq i} Y_j$. For every $i \in \{1, 2, ..., m\}$ there exists a unique $j_i \in \{1, 2, ..., n\}$ such that $Y_i = X_{j_i}$. The map $i \mapsto j_i$ is injective, and if some $k \in \{1, 2, ..., n\}$ is not in the image of this map then it follows that $X_k \subseteq \bigcup_{i=1}^m Y_i \subseteq \bigcup_{j \neq k} X_j$, a contradiction. □

Definition 13.21. The subsets $X_i$ defined in Theorem 13.20 are called *the irreducible components* of $X$.

Note that we can equip every Zariski-open subset $U$ of a (finite-dimensional) vector space $V$ with the Zariski topology, which is the subset topology with respect to the Zariski topology on $V$. Then $U$ is also Noetherian. We will be using the Zariski topology primarily in the context of the group $GL(V)$, which we identify with the Zariski open subset of $V \otimes V^*$, the space of $n \times n$ matrices with nonzero determinant.

DEFINITION 13.22. An *algebraic subgroup* of $GL(V)$ is a Zariski-closed subgroup of $GL(V)$.

Given an algebraic subgroup $G$ of $GL(V)$, the binary operation $G \times G \to G$, $(g, h) \mapsto gh$ is a morphism. The inversion map $g \mapsto g^{-1}$, as well as the left-multiplication and right-multiplication maps $g \mapsto ag$ and $g \mapsto ga$, by a fixed element $a \in G$, are automorphisms of $G$.

EXAMPLE 13.23. (1) The subgroup $SL(V)$ of $GL(V)$ is algebraic, defined by the equation $\det(g) = 1$.

(2) The group $GL(n, \mathbb{K})$ can be identified to an algebraic subgroup of $SL(n+1, \mathbb{K})$ by mapping every matrix $A \in GL(n, \mathbb{K})$ to the matrix

$$\begin{pmatrix} A & 0 \\ 0 & \frac{1}{\det(A)} \end{pmatrix}.$$

Therefore, in what follows, it will not matter if we consider algebraic subgroups of $GL(n, \mathbb{K})$ or of $SL(n, \mathbb{K})$.

(3) The group $O(V)$ is an algebraic subgroup, as it is given by the system of equations $M^T M = \mathrm{Id}_V$.

(4) More generally, given an arbitrary quadratic form $q$ on $V$, its stabilizer $O(q)$ is obviously algebraic. A special instance of this is the *symplectic group* $Sp(2k, \mathbb{K})$, preserving the form with the following matrix (given with respect to the standard basis in $V = \mathbb{K}^{2n}$)

$$J = \begin{pmatrix} 0 & K \\ -K & 0 \end{pmatrix}, \text{ where } K = \begin{pmatrix} 0 & \dots & 1 \\ 0 & \iddots & 0 \\ 1 & \dots & 0 \end{pmatrix}.$$

LEMMA 13.24. *If $\Gamma$ is a subgroup of $SL(V)$ then its Zariski closure $\bar{\Gamma}$ in $SL(V)$ is also a subgroup.*

PROOF. Consider the map $f : SL(V) \to SL(V)$ given by $f(\gamma) = \gamma^{-1}$. Then $f$ is a polynomial isomorphism and, hence, $f(\bar{\Gamma})$ is Zariski closed in $SL(V)$. Since $\Gamma$ is a subgroup, $f(\bar{\Gamma})$ contains $\Gamma$. Thus, $\bar{\Gamma} \cap f(\bar{\Gamma})$ is a Zariski closed set containing $\Gamma$. It follows that $\bar{\Gamma} = f(\bar{\Gamma})$ and hence $\bar{\Gamma}$ is stable under the inversion. The argument for the multiplication is similar. □

If $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, then $V = \mathbb{K}^n$ also has the *standard* or *classical* topology, given by the Euclidean metric on $V$. We use the terminology *classical topology* for the induced topology on subsets of $V$. Classical topology, of course, is stronger than Zariski topology.

THEOREM 13.25 (See for instance Chapter 3, §2, in [**OV90**]). (1) *An algebraic subgroup of $GL(n, \mathbb{C})$ is irreducible in the Zariski topology if and only if it is connected in the classical topology.*

(2) *A connected (in classical topology) algebraic subgroup of $GL(n, \mathbb{R})$ is irreducible in the Zariski topology.*

PROPOSITION 13.26. *Let $G$ be an algebraic subgroup in $GL(V)$.*

(1) *Only one irreducible component of $G$ contains the identity element. This component is called the* identity component *and is denoted by $G_0$.*
(2) *The subset $G_0$ is a normal subgroup of finite index in $G$ whose cosets are the irreducible components of $G$.*

REMARK 13.27. Proposition 13.26, (2), implies that for algebraic groups the irreducible components are disjoint. This is not true in general for algebraic varieties, consider, for instance, the subvariety $\{xy = 0\} \subset \mathbb{K}^2$.

PROOF. (1) Let $X_1, ..., X_k$ be irreducible components of $G$ containing the identity. According to Lemma 13.18, the product set $X_1 \times \ldots \times X_k$ is irreducible. Since the product map is a morphism, the subset $X_1 \cdots X_k \subset G$ is irreducible as well; hence by Lemma 13.19, (3), and by Theorem 13.20 this subset is contained in some $X_j$. The fact that every $X_i$ with $i \in \{1, ..., k\}$ is contained in $X_1 \cdots X_k$, hence in $X_j$, implies that $k = 1$.

(2) Since the inversion map $g \mapsto g^{-1}$ is an algebraic automorphism of $G$ (but not a group automorphism, of course) it follows that $G_0$ is stable with respect to the inversion. Hence for every $g \in G_0$, $gG_0$ contains the identity element, and is an irreducible component. Therefore, $gG_0 = G_0$. Likewise, for every $x \in G$, $xG_0x^{-1}$ is an irreducible component containing the identity element, hence it equals $G_0$. The cosets of $G_0$ (left or right) are images of $G_0$ under automorphisms, therefore also irreducible components. Thus there can only be finitely many of them. □

In what follows we list some useful properties of algebraic groups. We refer the reader to [**OV90**] for the details:

1. A complex or real algebraic group is a complex, respectively real, Lie group.

2. Every Lie group $G$ (resp. algebraic group over a field $\mathbb{K}$), contains a *radical* Rad$G$, which is the largest connected (resp. irreducible) solvable normal Lie (resp. algebraic) subgroup of $G$. The radical is the same if the group is considered with its real or its complex Lie structure. A group with trivial radical is called *semisimple.*

3. The quotient of an algebraic group by its radical is an algebraic semisimple group.

4. The commutator subgroup of an irreducible algebraic group is an irreducible algebraic subgroup. An irreducible algebraic semisimple group coincides with its commutator subgroup.

5. One of the most remarkable properties of algebraic semisimple groups is the following: given such a group $G$ and its representation as a linear group $G \hookrightarrow GL(V)$, the space $V$ decomposes into a direct sum of $G$-invariant subspaces so that the restriction of the action of $G$ to any of these subspaces is irreducible, i.e. there are no proper $G$-invariant subspaces.

6. From the classification of normal subgroups in a semisimple connected Lie group (see for instance [**OV90**, Theorem 4, Chapter 4, §3]) it follows that the image

of an algebraic irreducible semisimple group under an algebraic homomorphism is an algebraic irreducible semisimple group.

As an application of the formalism of algebraic groups, we will now give a "cheap" proof of the fact that the group $SU(2)$ contains a subgroup isomorphic to $F_2$, the free group on two generators:

LEMMA 13.28. *The subset of monomorphisms* $F_2 \to SU(2)$ *is dense in the variety* $Hom(F_2, SU(2))$.

PROOF. Consider the space $V = Hom(F_2, SL(2,\mathbb{C})) = SL(2,\mathbb{C}) \times SL(2,\mathbb{C})$; every element $w \in F_2$ defines a polynomial function

$$f_w : V \to SL(2,\mathbb{C}), \quad f_w(\rho) = \rho(w).$$

Since $SL(2,\mathbb{R}) \leqslant SL(2,\mathbb{C})$ contains a subgroup isomorphic to $F_2$ (see Example 4.38), it follows that for every $w \neq 1$, the function $f_w$ takes values different from 1. In particular, the subset $E_w := f_w^{-1}(1)$ is a proper (complex) subvariety in $V$. Since $SL(2,\mathbb{C})$ is a connected complex manifold, the variety $SL(2,\mathbb{C})$ is irreducible; hence, $V$ is irreducible as well. It follows that for every $w \neq 1$, $E_w$ has empty interior (in the classical topology) in $V$. Suppose that for some $w \neq 1$, the intersection

$$E'_w := E_w \cap SU(2) \times SU(2)$$

contains a nonempty open subset $U$. In view of Exercise 3.8, $SU(2)$ is Zariski dense (over $\mathbb{C}$) in $SL(2,\mathbb{C})$; hence, $U$ (and, thus, $E_w$) is Zariski dense in $V$. It then follows that $E_w = V$, which is false. Therefore, for every $w \neq 1$, the closed (in the classical topology) subset $E'_w \subset Hom(F_2, SU(2))$ has empty interior. Since $F_2$ is countable, by Baire category theorem, the union

$$E := \bigcup_{w \neq 1} E'_w$$

has empty interior in $Hom(F_2, SU(2))$. Since every $\rho \notin E$ is injective, lemma follows. ☐

Since $SU(2)/\pm I$ is isomorphic to $SO(3)$, we obtain

LEMMA 13.29. *The subset of monomorphisms* $F_2 \to SO(3)$ *is dense in the variety* $Hom(F_2, SO(3))$.

## 13.2. Virtually solvable subgroups of $GL(n,\mathbb{C})$

This and the following section deal with virtually solvable subgroups of the general linear group and limits of sequences of such groups. This material (namely, Theorem 13.45 or the weaker Proposition 13.44 that will also suffice) will be needed in the proof of the Tits' Alternative.

Let $G$ be a subgroup of $GL(V)$, where $V \cong \mathbb{C}^n$. We will think of $V$ as a $G$-module. In particular we will talk about $G$-submodules and quotient modules: The former are $G$-invariant subspaces $W$ of $V$, the latter are quotients $V/W$, where $W$ is a $G$-submodule. The $G$-module $V$ is *reducible* if there exists a proper $G$-submodule $W \subset V$. We say that $G$ is *upper-triangular* (or the $G$-module $V$ is *upper-triangular*) if it is conjugate to a subgroup of the group $B$ of upper-triangular matrices in $GL(V)$. In other words, there exists a *complete flag* $0 \subset V_1 \subset ... \subset V_n = V$ of $G$-submodules in $V$, where $\dim(V_i) = i$ for each $i$. Of course, reducibility makes sense only for modules of dimension $> 1$; however, by abusing the terminology, we will regard modules of dimension $\leqslant 1$ as reducible by default.

The group $B$ (and its conjugates in $GL(V)$) is called the *Borel subgroup* of $GL(V)$.

LEMMA 13.30. *Suppose that $G$ is an abstract group so that every $G$-module $V \cong \mathbb{C}^k$ with $2 \leqslant k \leqslant n$ is reducible. Then every $n$-dimensional $G$-module $V$ is upper-triangular.*

PROOF. Since $G \curvearrowright V$ is reducible, there exists a proper submodule $W \subset V$. Thus $\dim(W) < n$ and $\dim(V/W) < n$. Now, the assertion follows by induction on the dimension. $\square$

For a vector space $V$ over $\mathbb{K}$ we let $P(V)$ denote the corresponding projective space:

$$P(V) = (V \setminus \{0\})/\mathbb{K}^*.$$

LEMMA 13.31. *Let $G < GL(V)$ be upper-triangular. Then the fixed-point set $\mathrm{Fix}(G)$ of the action of $G$ on the projective space $P(V)$ is nonempty and consists of a disjoint union of projective subspaces $P(V_\ell), \ell = 1, ..., k$, so that the subspaces $V_i \subset V$ are* linearly independent*, i.e.:*

$$\mathrm{Span}(\{V_1, ..., V_k\}) = \bigoplus_{\ell=1}^{k} V_\ell.$$

*In particular, $k \leqslant \dim(V)$.*

PROOF. For $g \in GL(V)$ we let $a_{ij}(g)$ denote the $i, j$ matrix coefficient of $g$. Then, since $G$ is upper-triangular, the maps $\chi_i : g \to a_{ii}(g)$ are homomorphisms $\chi : G \to \mathbb{C}^*$, called *characters* of $G$. The (multiplicative) group of characters of $G$ is denoted $X(G)$. We let $J \subset \{1, ..., n\}$ be the set of all indices $j$ such that

$$a_{ij}(g) = a_{ji}(g) = 0, \forall g \in G, \forall i \neq j.$$

We then break the set $J$ into disjoint subsets $J_1, ..., J_m$ which are preimages of points $\chi \in X(G)$ under the map

$$j \in J \mapsto \chi_j \in X(G).$$

Set $V_\ell := \mathrm{Span}(\{e_i, i \in J_\ell\})$, where $e_1, ..., e_n$ form the standard basis in $V$. It is clear that $G$ fixes each $P(V_\ell)$ pointwise since each $g \in G$ acts on $V_\ell$ *via* the scalar multiplication by $\chi_\ell(g)$. We leave it to the reader to check that

$$\bigcup_{\ell=1}^{m} P(V_\ell)$$

is the entire fixed-point set $\mathrm{Fix}(G)$. $\square$

In what follows, the topology on subgroups of $GL(V)$ is always the Zariski topology, in particular, connectedness always means Zariski–connectedness.

THEOREM 13.32 (A. Borel). *Let $G$ be a connected solvable Lie group. Then every $G$-module $V$ (where $V$ is a finite-dimensional complex vector space) is upper-triangular.*

Proof. In view of Lemma 13.30, it suffices to prove that every such module $V$ is reducible. The proof is an induction on the derived length $d$ of $G$.

We first recall a few facts about eigenvalues of the elements of $GL(V)$. Let $Z_{GL(V)}$ denote the center of $GL(V)$, i.e. the group of matrices of the form $\mu \cdot I, \mu \in \mathbb{C}^*$, where $I$ is the unit matrix.

Let $g \in GL(V) \setminus Z_{GL(V)}$. Then $g$ has linearly independent eigenspaces $E_{\lambda_j}(g), j = 1, ..., k$, labeled by the corresponding eigenvalues $\lambda_j, 1 \leqslant j \leqslant k$, where $2 \leqslant k \leqslant n$. We let $\mathcal{E}(g)$ denote the set of (unlabeled) eigenspaces

$$\{E_{\lambda_j}(g), j = 1, ..., k\}.$$

Let $B_g$ denote the abelian subgroup of $GL(V)$ generated by $g$ and the center $Z_{GL(V)}$. Then for every $g' \in B_g$, $\mathcal{E}(g') = \mathcal{E}(g)$ (with the new eigenvalues, of course). Therefore, if $N(B_g)$ denotes the normalizer of $B_g$ in $G$, then $N(B_g)$ preserves the set $\mathcal{E}(g)$, however, elements of $N(B_g)$ can permute the elements of $\mathcal{E}(g)$. (Note that $N(B_g)$ is, in general, larger than $N(\langle g \rangle)$, the normalizer of $\langle g \rangle$ in $G$.) Since $\mathcal{E}(g)$ has cardinality $\leqslant n$, there is a subgroup $N^o = N^o(B_g) < N(B_g)$ of index $\neq n!$ that fixes the set $\mathcal{E}(g)$ element-wise, i.e., every $h \in N^o$ will preserve each $E_\lambda(g)$, where $\lambda \in Sp(g)$, the spectrum of $g$. Of course, $h$ need not act trivially on $E_\lambda(g)$. Since $g \notin Z_G$, this means that there exists a proper $N^o$-invariant subspace $E_\lambda(g) \subset V$.

We next prove several needed for the proof of Borel's theorem.

Lemma 13.33. *Let $A$ be an abelian subgroup of $GL(V)$. Then the $A$-module $V$ is reducible.*

Proof. If $A \leqslant Z_{GL(V)}$, there is nothing to prove. Assume, therefore, that $A$ contains an element $g \notin Z_{GL(V)}$. Since $A \leqslant N(B_g)$, it follows that $A$ preserves the collection of subspaces $\mathcal{E}(g)$. Since $A$ is abelian, it cannot permute these subspaces. Therefore, $A$ preserves the proper subspace $E_{\lambda_1}(g) \subset V$ and hence $A \curvearrowright V$ is reducible. ☐

Lemma 13.34. *Suppose that $G < GL(V)$ is a connected metabelian group, so that $G' = [G, G] \leqslant Z_{GL(V)}$. Then the $G$-module $V$ is reducible.*

Proof. The proof is analogous to the proof of the previous lemma. If $G < Z_{GL(V)}$ there is nothing to prove. Pick, therefore some $g \in G \setminus Z_{GL(V)}$. Since the image of $G$ in $PGL(V)$ is abelian, the group $G$ is contained in $N(B_g)$. Since $G$ is connected, it cannot permute the elements of $\mathcal{E}(g)$. Hence $G$ preserves each $E_{\lambda_i}(g)$. Since every subspace $E_{\lambda_i}(g)$ is proper, it follows that the $G$-module $V$ is reducible. ☐

Similarly, we have:

Lemma 13.35. *Let $G < GL(V)$ be a metabelian group whose projection to $PGL(V)$ is abelian. Then $G$ contains a reducible subgroup of index $\leqslant n!$.*

Proof. We argue as in the proof of the previous lemma, except $G$ may permute the elements of $\mathcal{E}(g)$. However, it will contain an index $\leqslant n!$ subgroup which preserves each $E_{\lambda_j}(g)$ and the assertion follows. ☐

We can now prove Theorem 13.32. Lemma 13.33 proves the theorem for abelian groups, i.e., solvable groups of derived length 1. Suppose the assertion holds for all connected groups of derived length $< d$ and let $G < GL(V)$ be a connected solvable group of derived length $d$. Then $G' = [G, G]$ has derived length $< d$. Thus by the induction hypothesis, $G'$ is upper-triangular. By Lemma 13.31, $\mathrm{Fix}(G') \subset P(V)$ is a nonempty disjoint union of independent projective subspaces $P(V_i), i = 1, ..., \ell$. Since $G'$ is normal in $G$, $\mathrm{Fix}(G')$ is invariant under $G$. Since $G$ is connected, it cannot interchange the components $P(V_i)$ of $\mathrm{Fix}(G)$. Therefore, it has to preserve each $P(V_i)$. If one of the $P(V_i)$'s is a proper projective subspace in $P(V)$, then $V_i$ is $G$-invariant and hence the $G$-module $V$ is reducible. Therefore, we assume that $\ell = 1$ and $V_1 = V$, i.e., $G'$ acts trivially on $P(V)$. This means that $G' < Z_{GL(V)}$ is abelian and hence $G$ is 2-step nilpotent. Now, the assertion follows from Lemma 13.34. This concludes the proof of Theorem 13.32. □

The following is a converse to Theorem 13.32:

Proposition 13.36. *For $V = \mathbb{C}^n$ the Borel subgroup $B < GL(V)$ is solvable of derived length $n$. Thus, a connected subgroup of $GL(V)$ is solvable if and only if it is conjugate to a subgroup of $B$, i.e., Borel subgroups are the maximal solvable connected subgroups of $GL(V)$. In particular, the derived length of every connected subgroup of $GL_n(\mathbb{C})$ is at most $n$.*

Proof. The proof is induction on $n$. The assertion is clear for $n = 1$ as $GL_1(\mathbb{C}) \cong \mathbb{C}^*$ is abelian. Suppose it holds for $n' = n - 1$, we will prove it for $n$. Let $B^{(i)} := [B^{(i-1)}, B^{(i-1)}], B^{(0)} = B$ be the derived series of $B$.

Let $0 = V_0 \subset V_1 \subset ... \subset V_n$ be the complete flag invariant under $B$. Set $W := V/V_1$, let $B_W$ be the image of $B$ in $GL(W)$. The kernel $K$ of the homomorphism $B \to B_W$ is isomorphic to $\mathbb{C}^*$. The group $B_W$ preserves the complete flag

$$0 = W_0 := V_1/V_1 \subset W_1 := V_2/V_1 \subset ... \subset W = V/V_1.$$

Therefore, by the induction assumption it has derived length $n-1$. Thus $B^{(n)} := [B^{(n-1)}, B^{(n-1)}] \subset K \cong \mathbb{C}^*$. Since $\mathbb{C}^*$ is abelian $[B^{(n)}, B^{(n)}] = 0$, i.e., $B$ has derived length $n$. □

Remark 13.37. Theorem 13.32 is false for non-connected solvable subgroups of $GL(V)$. Take $n = 2$, let $A$ be the group of diagonal matrices in $SL(2, \mathbb{C})$ and let

$$s = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then $s$ normalizes $A$ and $s^2 \in A$. We let $G$ be the group generated by $A$ and $s$ which is isomorphic to the semidirect product of $A$ and $\mathbb{Z}_2$. In particular, $G$ is solvable of derived length 2. On the other hand, it is clear that the $G$-module $\mathbb{C}^2$ is irreducible.

Theorem 13.38. *There exist functions $\nu(n), \delta(n)$ so that every virtually solvable subgroup $\Gamma \leqslant GL(V)$ contains a solvable subgroup $\Lambda$ of index $\leqslant \nu(n)$ and derived length $\leqslant \delta(n)$.*

Proof. Let $d$ denote the derived length of a finite index solvable subgroup of $\Gamma$. Let $\overline{\Gamma}$ denote the Zariski closure of $\Gamma$ in $GL(V)$. Then $\overline{\Gamma}$ has only finitely many (Zariski) connected components (see Theorem 13.20).

Lemma 13.39. *The group $\overline{\Gamma}$ is contains a finite index subgroup which is a solvable group of derived length $d$.*

Proof. We will use $k$-fold iterated commutators

$$[\![g_1, \ldots, g_{2^k}]\!]$$

defined in the equation (11.3). Let $\Gamma^o < \Gamma$ denote a solvable subgroup of derived length $d$ and finite index $m$ in $\Gamma$; thus

$$\Gamma = \gamma_1\Gamma^o \cup ... \cup \gamma_m\Gamma^o.$$

The group $\Gamma^o$ satisfies polynomial equations of the form $(g_1, ..., g_{2^d}) = 1$. Therefore, $\Gamma$ satisfies the polynomial equations in the variables $g_j$:

$$\gamma_i[\![g_1, \ldots, g_{2^d}]\!] = 1, i = 1, ..., m.$$

Hence, the Zariski closure $\overline{\Gamma}$ of $\Gamma$ satisfies the same set of polynomial equations. It follows that $\overline{\Gamma}$ contains a subgroup of index $m$ which is solvable of derived length $d$. $\square$

Let $G$ be the (Zariski) connected component of the identity of $\overline{\Gamma}$, which implies that $G \lhd \overline{\Gamma}$.

LEMMA 13.40. *The group $G$ is solvable of derived length $\leqslant n$.*

PROOF. Let $H \lhd G$ be the maximal solvable subgroup of derived length $d$ of finite index. Thus as above, $H$ is given by imposing polynomial equations of the form $[\![g_1, \ldots, g_{2^d}]\!] = 1$ on tuples of the elements of $G$, i.e., $H$ is Zariski closed. Since $H$ has finite index in $G$, it is also open. Since $G$ is connected, it follows that $G = H$, i.e., $G$ is solvable and has derived length $\leqslant n$ by Proposition 13.36. $\square$

It is clear that $\Gamma \cap G$ is a finite index subgroup of $\Gamma$ whose index is at most $|\overline{\Gamma} : G|$. Unfortunately, the index $|\overline{\Gamma} : G|$ could be arbitrarily large. We will see, however, that we can enlarge $G$ to a (possibly disconnected) subgroup $\widehat{G} \leqslant \overline{\Gamma}$ which is still solvable but has a uniform upper bound on $|\overline{\Gamma} : \widehat{G}|$ and a uniform bound on the derived length.

We will get a bound on the index and the derived length by the dimension induction. The base case where $n = 1$ is clear, so we assume that for each $n' < n$ and each virtually solvable subgroup $\Gamma' \leqslant GL_{n'}(\mathbb{C})$ there exists a solvable group $\widehat{G}'$

$$G' \leqslant \widehat{G}' \leqslant \overline{\Gamma}'$$

as required, with a uniform bound $\nu(n')$ on the index $|\overline{\Gamma}' : \widehat{G}'|$ and so that the derived length of $\widehat{G}'$ is at most $\delta(n') \leqslant \delta(n-1)$.

Let $\mathcal{V} := \{V_1, \ldots, V_\ell\}$ denote the maximal collection of (independent) subspaces in $V$ so that $G$ fixes each $P(V_i)$ pointwise (see Theorem 13.32 and Lemma 13.31). In particular, $\ell \leqslant n$. Since $G$ is normal in $\overline{\Gamma}$, the collection $\mathcal{V}$ is invariant under $\overline{\Gamma}$. Let $K \leqslant \overline{\Gamma}$ denote the kernel of the action of $\overline{\Gamma}$ on the set $\mathcal{V}$. Clearly, $G \leqslant K$ and $|\overline{\Gamma} : K| \leqslant \ell! \leqslant n!$. We will, therefore, study the pair $G \leqslant K$.

REMARK 13.41. Note that we just proved that every virtually solvable subgroup $\Gamma \leqslant GL(n, \mathbb{C})$ contains a reducible subgroup of index $\leqslant n!c(n)$, where $c(n) := q(PGL(n, \mathbb{C}))$ is the function from Jordan's Theorem 10.66. Indeed, if $\ell > 1$, the subgroup $K \cap \Gamma$ (of index $\neq n!$) preserves a proper subspace $V_1$. If $\ell = 1$, then $G$ is contained in $Z_{GL(V)}$ and hence $\Gamma$ projects to a finite subgroup $\Phi < PGL(V)$. After replacing $\Phi$ with an abelian subgroup $A$ of index $\neq q(PGL(V))$ (see Jordan's Theorem 10.66), we obtain a metabelian group $\tilde{A} < \Gamma$ whose center is contained in $Z_{GL(V)}$. Now the assertion follows from Lemma 13.35.

The group $K$ preserves each $V_i$ and, by construction, the group $G$ acts trivially on each $P(V_i)$. Therefore, the image $Q_i$ of $K/G$ in $PGL(V_i)$ is finite. (The finite group $K/G$ need not act faithfully on $P(V_i)$.) By Jordan's Theorem 10.66, the group $Q_i$ contains an abelian subgroup of index $\leqslant c(\dim(V_i)) \leqslant c(n)$. Hence, $K$ contains a subgroup $N \lhd K$ of index at most

$$\prod_{i=1}^{\ell} c(\dim(V_i)) \leqslant c(n)^n$$

which acts as an abelian group on

$$\prod_{i=1}^{\ell} P(V_i).$$

We again note that $G \leqslant N$. The image of the restriction homomorphism $\phi : N \to GL(U)$,

$$U := V_1 \oplus \ldots \oplus V_\ell$$

is therefore a *metabelian* group $M$.

We also have the homomorphism $\psi : N \to GL(W)$, $W = V/U$ with the image $N_W$. This group contains the connected solvable subgroup $G_W := \psi(G)$ of finite index. To identify the intersection $\mathrm{Ker}(\phi) \cap \mathrm{Ker}(\psi)$ we observe that $V = U \oplus W$ and the group $N$ acts by matrices of the block-triangular form:

$$\begin{bmatrix} x & y \\ 0 & z \end{bmatrix}$$

where $x \in M$, $z \in N_W$. Then the kernel of the homomorphism $\phi \times \psi : N \to M \times N_W$ consists of matrices of the upper-triangular form

$$\begin{bmatrix} 1 & y \\ 0 & 1 \end{bmatrix}.$$

Thus by Proposition 13.36, $L = \mathrm{Ker}(\phi \times \psi)$ is solvable of derived length $\leqslant n$.

By the induction hypothesis, there exists a solvable group $\widehat{G_W}$ of derived length $\leqslant \delta(n-1)$, so that

$$G_W \leqslant \widehat{G_W} \leqslant N_W$$

and $|N_W : \widehat{G_W}| \leqslant \nu(n-1)$. Therefore, for $\widehat{G} := (\phi \times \psi)^{-1}(M \times \widehat{G_W})$, we obtain a commutative diagram

$$\begin{array}{ccccccc} 1 \to & L & \longrightarrow & N & \overset{\phi\times\psi}{\longrightarrow} & M \times N_W & \longrightarrow \\ & || & & \uparrow \iota' & & \uparrow \iota & \\ 1 \to & L & \longrightarrow & \widehat{G} & \overset{\phi\times\psi}{\longrightarrow} & M \times \widehat{G_W} & \longrightarrow \end{array}$$

where $\iota$ is the inclusion of index $i \leqslant \nu(n-1)$ subgroup and, hence, $\iota'$ is also the inclusion of index $i$ subgroup. Furthermore, $L$ is solvable of derived length $\leqslant n$, $M \times \widehat{G_W}$ is solvable of derived length$\leqslant \max(2, \delta(n-1))$. Putting it all together, we get

$$|\overline{\Gamma} : \widehat{G}| \leqslant \nu(n) := \nu(n-1)n!(c(n))^n,$$

where $\widehat{G}$ is solvable of derived length $\leqslant \delta(n) := \max(2, \delta(n-1)) + n$. Intersecting $\widehat{G}$ with $\Gamma$ we obtain $\Lambda < \Gamma$ of index at most $\nu(n)$ and derived length $\leqslant \delta(n)$. Theorem 13.38 follows. □

## 13.3. Limits of sequences of virtually solvable subgroups of $GL(n, \mathbb{C})$

Throughout this section, all vector spaces under consideration will be complex and finite-dimensional.

We say that a subgroup $G < GL(V)$, $V \cong \mathbb{C}^n$, is *virtually reducible* if $G$ contains a finite index subgroup $H$ which has reducible action on $V$. A subgroup which is not virtually reducible is called *virtually irreducible.* Recall that modules of dimension 1 are regarded as reducible by default.

Remark 13.42. In order to distinguish this notion of irreducibility from the irreducibility in the context of algebraic groups, we will refer to the later as *Zariski–irreducibility.*

Lemma 13.43. *Let $G \leqslant GL(V)$ be a subgroup which is not virtually solvable. Then $G$ contains a finite index subgroup $H$ which admits an $H$-module $W$, which is either a submodule or quotient module of $H \curvearrowright V$, such that $H \curvearrowright W$ is virtually irreducible.*

Proof. The proof is by induction on the dimension of $V$. The statement is clear if $V$ is 1-dimensional. Suppose it holds in all dimensions $< n$. If $G$ itself is virtually irreducible, we are done. Otherwise, we take a finite index subgroup $G_1 < G$ so that the $G_1 \curvearrowright V$ is reducible. Let $W \subset V$ be a $G_1$-invariant subspace. If the images of $G_1$ in $GL(W)$ and $GL(V/W)$ are both virtually solvable, then $G$ is itself virtually solvable. If one of these images is not virtually solvable, the statement follows from the induction hypothesis. □

Proposition 13.44. *Let $\Gamma \leqslant GL(n, \mathbb{C})$ be a finitely-generated virtually irreducible subgroup. Then there exists a neighborhood $\Xi$ of $id$ in $\mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$ so that every $\rho \in \Xi$ has image which is not virtually solvable.*

Proof. Suppose to the contrary that there exists a sequence

$$\rho_j \in \mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$$

converging to $id$, so that each $\Gamma_j := \rho_j(\Gamma)$ is virtually solvable. Since each $\Gamma_j$ is virtually solvable, by Remark 13.41 it contains a reducible subgroup of index $\leqslant n!c(n)$. Let $\Phi < \Gamma$ denote the intersection of the preimages of these subgroups under $\rho_j$'s. Clearly, $|\Gamma : \Phi| < \infty$. After passing to a subsequence, we may assume that each $\Gamma_j$ preserves a proper projective subspace $P_j \subset \mathbb{CP}^{n-1}$ of a fixed dimension $k$. By passing to a further subsequence, we may assume that the subspaces $P_j$ converge to a proper projective subspace $P \subset \mathbb{CP}^{n-1}$. Since each $\Gamma_j$ preserves $P_j$, the group $\Phi$ also preserves $P$. Hence, $\Gamma \curvearrowright V$ is virtually reducible, contradicting our assumptions. □

Although the above proposition will suffice for the proof of the Tits' Alternative, we will prove a slightly stronger assertion:

Theorem 13.45. *Let $\Gamma \subset GL(n, \mathbb{C})$ be a finitely-generated subgroup which is not virtually solvable. Then there exists a neighborhood $\Sigma$ of $id$ in $\mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$ so that every $\rho \in \Sigma$ has image which is not virtually solvable.*

Proof. We argue analogously to the proof of Proposition 13.44. Suppose to the contrary that there exists a sequence $\rho_j \in \mathrm{Hom}(\Gamma, G)$ converging to $id$, so that each $\Gamma_j := \rho_j(\Gamma)$ is virtually solvable. By Theorem 13.38, for each $j$ there exists a subgroup $\Lambda_j \leqslant \Gamma_j$ of index $\leqslant \nu(n)$ which is solvable of derived length $\leqslant d = \delta(n)$.

Let $\Lambda \leqslant \Gamma$ denote the intersection of $\rho_j^{-1}(\Lambda_j)$. Again, $|\Gamma : \Lambda| < \infty$. Each group $\Gamma_j$ satisfies the law:

$$[\![g_1, ..., g_{2^d}]\!] = 1$$

where $[\![g_1, ..., g_{2^d}]\!]$ is the $d$-fold iterated commutator as in (11.3). Therefore, for every $2^d$-tuple of elements $\gamma_i$ of $\Lambda$ we have

$$[\![\gamma_1, ..., \gamma_{2^d}]\!] = \lim_{j\to\infty} [\![\rho_j(\gamma_1), ..., \rho_j(\gamma_{2^d})]\!] = 1.$$

Hence, $\Lambda$ is solvable of derived length $\leqslant d$. □

### 13.4. Reduction to the case of linear subgroups

PROPOSITION 13.46. *It suffices to prove Theorem 13.1 for subgroups $\Gamma \leqslant GL(V)$, where $V$ is a finite-dimensional real vector space, and the Zariski closure of $\Gamma$ in $GL(V)$ is a Zariski–irreducible semisimple algebraic group, acting irreducibly on $V$.*

PROOF. The first step is to reduce the problem from subgroups in Lie groups with finitely many connected components to subgroups of some $GL(V)$.

Let $L$ be a Lie group with finitely many components. The connected component of the identity $L_0 \subset L$ is then a finite index normal subgroup. Thus $\Gamma \cap L_0$ has finite index in $\Gamma$. Therefore, we can assume that $L$ is connected.

LEMMA 13.47. *There exists a homomorphism $\phi : \Gamma \to GL_n(\mathbb{R})$, $n = \dim(G)$, whose kernel is contained in the center of $\Gamma$.*

PROOF. Since $L$ is connected, kernel of the adjoint representation $Ad : L \to GL(T_eL)$ is contained in the center of $L$, see Lemma 3.10. Now, take $\phi := Ad|\Gamma$. □

Observe that

1. $\Gamma$ is virtually solvable if and only if $\phi(\Gamma)$ is virtually solvable.
2. $\Gamma$ contains a free subgroup if and only if $\phi(\Gamma)$ contains a free subgroup.

Therefore, we can assume that $\Gamma$ is a linear group, $\Gamma \subset GL(n, \mathbb{R})$.

Let $G$ be the Zariski-closure of $\Gamma$ in $GL(V)$. Although $G$ need not be Zariski–irreducible, by Proposition 13.26 it has only finitely many irreducible components. Thus, after passing to a finite index subgroup in $\Gamma$, we may assume that $G$ is Zariski–irreducible.

According to the results mentioned in the end of Section 13.1, $G$ contains a normal algebraic Zariski–irreducible subgroup which is solvable, $\mathrm{Rad}(G)$, and the quotient $G/\mathrm{Rad}(G)$ is a semisimple algebraic Zariski–irreducible subgroup. Clearly the image of $\Gamma$ by the algebraic projection $\pi : G \to G/\mathrm{Rad}(G)$ is Zariski dense in $G/\mathrm{Rad}(G)$, and it suffices to prove the alternative for $\pi(\Gamma)$. Thus we may assume that the Zariski closure $G$ of $\Gamma$ is Zariski–irreducible and semisimple.

If the action $G \curvearrowright V$ is reducible then we take the direct sum decomposition

$$V = \bigoplus_{i=1}^{s} V_i$$

in $G$-invariant subspaces, so that the action of $G$ on each $V_i$ is irreducible. If we denote by $\rho_i$ the homomorphism $G \to GL(V_i)$ then it suffices to prove the Tits' Theorem for each $\rho_i(\Gamma)$. Indeed, if it is proved, then either all $\rho_i(\Gamma)$ are solvable, in which case $\Gamma$ itself is solvable (see Exercise 11.20), or some $\rho_i(\Gamma)$ contains a free non-abelian subgroup, in which case $\Gamma$ itself does, as $\rho_i(\Gamma)$ is a quotient of $\Gamma$.

Note that when we replace in our problem $\Gamma$ by $\rho_i(\Gamma)$, we have to replace $G$ by the Zariski closure $G_i$ of $\rho_i(\Gamma)$ in $GL(V_i)$. Note also that

$$\rho_i(\Gamma) \leqslant \rho_i(\overline{\Gamma}) = \rho_i(G) \leqslant \overline{\rho_i(\Gamma)} = G_i \leqslant \overline{\rho_i(G)}\,.$$

According to the considerations in the end of Section 13.1, $\rho_i(G)$ is an algebraic Zariski-irreducible semisimple group. In particular it coincides with its closure, hence $G_i = \rho_i(G)$. Thus $G_i$ acts irreducibly on $V_i$ because $G$ does, and $G_i$ is Zariski–irreducible and semisimple because $\rho_i(G)$ is. This concludes the proof of Proposition 13.46. □

## 13.5. Tits' Alternative for unbounded subgroups of $SL(n)$

In this section we prove Tits' Alternative for subgroups $\Gamma$ of $SL(n,\mathbb{K})$ that are unbounded with respect to the standard norm, where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. For technical reasons, one should also consider the case of other *local fields* $\mathbb{K}$. Recall that a local field is a field with a norm $|\cdot|$ which determines a locally compact topology on $\mathbb{K}$. The most relevant examples for us are when $\mathbb{K}=\mathbb{R}, \mathbb{K}=\mathbb{C}$, $\mathbb{K}=\mathbb{Q}_p$ and, more generally, $\mathbb{K}$ is the completion of a finite extension of $\mathbb{Q}$.

In what follows, $V$ is an $n$-dimensional vector space over a local field $\mathbb{K}$, $n=\dim(V)>1$. We fix a basis $e_1,\ldots,e_n$ in $V$. Then the norm $|\cdot|$ on $\mathbb{K}$ determines the Euclidean norms $\|\cdot\|$ on $V$ and on its exterior powers.

NOTATION 13.48. We will use the notation $E^c$ to denote the complement $X\setminus E$ of a subset $E\subset X$.

We shall prove the following.

THEOREM 13.49. *Let $\Gamma \leqslant GL(V)$ be a finitely-generated group which is not relatively compact, and such that the Zariski closure of $\Gamma$ in $GL(V)$ is a Zariski-irreducible semisimple algebraic group acting irreducibly on $V$. Then $\Gamma$ contains a free non-abelian subgroup.*

PROOF. In the argument, the free subgroups will be constructed using the Ping-pong Lemma 4.37. The role of the space $X$ in that lemma will be played by the projective space.

NOTATION 13.50. We let $P(V)$ denote the projective space of $V$. When there is no possibility of confusion we do not mention the vector space anymore, and simply denote the projective space by $P$.

The ideal situation would be to find a pair of elements $g,h$ in $\Gamma$ with properties as in Chapter 4, Section 4.5. Since such elements may not exist in $\Gamma$ in general, we try to 'approximate' the situation in Lemma 4.42.

Recall that, according to the Cartan decomposition (see Section 4.5), every element $g\in GL(V)$ can be written as $g=kdh$, where $k$ and $h$ are in the compact subgroup $K$ of $GL(V)$ and $d$ is a diagonal matrix with entries on the diagonal such that $|a_1|\geqslant|a_2|\geqslant\ldots\geqslant|a_n|>0$.

DEFINITION 13.51. We call a sequence of elements $(g_i)$ in $GL(V)$ a *diverging sequence* if their matrix norms diverge to infinity.

It is immediate from the compactness of $K$ that the elements $g_i$ of a diverging sequence have Cartan decomposition $g_i = k_i d_i h_i$ such that $|a_1(g_i)| \to \infty$ as $i \to \infty$.

For every diverging sequence, there exists a maximal $m \in \{1, \ldots, n-1\}$ with the property that

$$\limsup_{i\to\infty} \frac{|a_m(g_i)|}{|a_1(g_i)|} > 0\,.$$

By passing to a subsequence we may assume that

$$\lim_{i\to\infty} \frac{|a_m(g_i)|}{|a_1(g_i)|} = 2\ell > 0$$

and also that $k_i$ and $h_i$ converge to some $k \in K$ and $h \in K$ respectively. We formalize these observations as follows:

Definition 13.52. We call a sequence $(g_i)$ *$m$-contracting*, for $m < \dim V$, if its elements have Cartan decompositions $g_i = k_i d_i h_i$ satisfying the following convergence properties:

(1) $k_i$ and $h_i$ converge to some $k$ and $h$ in $K$;

(2) $d_i$ are diagonal matrices with diagonal entries $a_1(g_i), \ldots, a_n(g_i)$ such that

$$|a_1(g_i)| \geqslant |a_2(g_i)| \geqslant \ldots \geqslant |a_n(g_i)|, \quad |a_1(g_i)| \to \infty$$

and

$$\lim_{i\to\infty} \frac{|a_m(g_i)|}{|a_1(g_i)|} > 0\,.$$

(3) The number $m$ is maximal with the above properties.

Observe now that *since $\Gamma$ is unbounded, it contains an $m$-contracting sequence $(g_i)$, for some $1 \leqslant m < \dim V$.*

In what follows we analyze the dynamics of an $m$-contracting sequence $\sigma = (g_i)$. We use the following notation and terminology, consistent to that in Definition 13.52 and the notation used in §4.5:

Notation 13.53.

$$A(g_i) = k_i\,[\mathrm{Span}(e_1, \ldots, e_m)] \quad \text{and } A(\sigma) = k\,[\mathrm{Span}(e_1, \ldots, e_m)]\,.$$

$$E(g_i) = h_i^{-1}\,[\mathrm{Span}(e_{m+1}, \ldots, e_n)] \quad \text{and } E(\sigma) = h^{-1}\,[\mathrm{Span}(e_{m+1}, \ldots, e_n)]\,.$$

Here the bracket stands for the projection to $P(V)$. We call $A(\sigma)$ the *attracting subspace* of the sequence $\sigma$ and $E(\sigma)$ the *repelling subspace* of the sequence $\sigma$.

When $m = 1$ we call $A(\sigma)$ the *attracting point* and $E(\sigma)$ (sometimes also denoted $H(\sigma)$) the *repelling hyperplane* of the sequence $\sigma$.

Note that since $k_i \to k$ and $h_i \to h$, they converge in the compact-open topology as transformations of $P(V)$; hence $A(g_i)$ converge to $A(\sigma)$, and $E(g_i)$ converge to $E(\sigma)$ with respect to the Hausdorff metric.

Example 13.54. To make things more concrete, consider the case $\dim V = 2$ and $\mathbb{K} = \mathbb{R}$. Then $P(V) = \mathbb{P}^1$ is the circle on which the group $PSL(2, \mathbb{R})$ acts by linear-fractional transformations. Since $0 < m < 2$, it follows that $m = 1$ and, hence, every diverging sequence contains a 1-contracting subsequence. It is easy to see that, for a 1-contracting sequence, the sequence of inverses has to be 1-contracting as well. Moreover, the repelling hyperplanes in $P(V)$ are again points.

Thus, each diverging sequence $g_i \in PSL(2,\mathbb{R})$ contains a subsequence $g_{i_n}$ for which there exists a pair of points $A$ and $H$ in $P(V)$ such that

$$\lim_{n\to\infty} g_{i_n}|_{P(V)\setminus\{H\}} = A \quad \text{and} \quad \lim_{n\to\infty} g_{i_n}^{-1}|_{P(V)\setminus\{A\}} = H\,,$$

uniformly on compact sets. For instance, if $g_{i_n} = g^n$, and $g$ is parabolic, then $A = H$ is the fixed point of $g$. If $g$ is hyperbolic then $A$ is the attractive and $H$ is the repelling fixed point of $g$. Thus, in general (unlike in the diagonal case), $A(g_i)$ may belong to $E(g_i)$.

The following is a uniform version of Lemma 4.41 for $m$-contracting sequences:

LEMMA 13.55. *Let $\sigma = (g_i)$ be an $m$-contracting sequence. For each compact $K \subset E(\sigma)^c$ there exist $L$ and $i_0$ so that $g_i$ is $L$–Lipschitz on $K$, for every $i \geqslant i_0$.*

PROOF. Assume that $g_i$'s satisfy (for all sufficiently large $i$) the following:

$$|a_1(g_i)| \geqslant |a_2(g_i)| \geqslant \ldots \geqslant |a_m(g_i)| \geqslant \ell |a_1(g_i)|\,,$$

where $\ell > 0$ is a constant independent of $i$.

By the assumption, $hK$ is disjoint of $[\mathrm{Span}(e_{m+1}, \ldots, e_n)]$, so the Hausdorff distance between these two compact sets is $2\varepsilon > 0$. Since the sets $h_i K$ converge to $hK$ in the Hausdorff metric, as $i \to \infty$, we may assume that for large $i$, the set $h_i K$ is contained in $K_\varepsilon$, where

$$K_\varepsilon = \overline{\mathcal{N}}_\varepsilon(hK) = \{p \in P(V) \mid \mathrm{dist}(p, hK) \leqslant \varepsilon\}\,.$$

Since $k_i$ act as isometries on $P(V)$, it suffices to prove that $d_i$'s are $L$-Lipschitz maps, for some uniform $L$ and $i$ large enough. In what follows, we consider an arbitrary diagonal matrix $d = d_i$ with eigenvalues $a_1, \ldots, a_n$.

Then every point $[u]$ of $K_\varepsilon$ is at distance $\gg \varepsilon$ from $[\mathrm{Span}(e_{m+1}, \ldots, e_n)]$. Without loss of generality, we may assume that $u = (u_1, \ldots, u_n)$ is a unit vector. Set

$$u' = (u_1, \ldots, u_m, 0, \ldots, 0), \quad u'' = (0, \ldots, 0, u_{m+1}, \ldots, u_n)$$

Suppose that $0 < \delta \leqslant \frac{1}{2\sqrt{n}}$ and the vector $u$ (as above) is such that

$$|u_i| \leqslant \delta, \forall i = 1, \ldots, m.$$

Then,

$$|u - u''|_{max} = |u'|_{max} \leqslant \delta.$$

Lemma 1.74 then implies that

$$|u \wedge u''| \leqslant 2n\delta,$$

while

$$|u''| \gg 1 - \sqrt{n}\delta \geqslant \frac{1}{2}.$$

Combining these inequalities,we obtain

$$d([u],[u'']) \leqslant 4n\delta.$$

Since, by assumption, $\varepsilon \leqslant d([u],[u''])$, we see that

$$\delta \geqslant \frac{\varepsilon}{4n}.$$

Therefore, for every unit vector $u$ so that $[u] \in K_\varepsilon$,

$$\max_{k=1,\ldots,m} |u_k| \geqslant \delta = \delta(\varepsilon) = \min\left(\frac{\varepsilon}{4n}, \frac{1}{2\sqrt{n}}\right). \tag{13.1}$$

In particular, for such $u$, there exists $k \in \{1, \dots, m\}$, so that

$$|d(u)|^2 \geqslant |a_k|^2 |u_k|^2 \geqslant \ell^2 |a_1|^2 \delta^2 .$$

Let $[v]$ and $[w]$ be two points in $K_\epsilon$. Then, in the archimedean case,

$$|d(v) \wedge d(w)|^2 = \sum_{p<q} |a_p v_p a_q w_q - a_q v_q a_p w_p|^2 = \sum_{p<q} |a_p a_q|^2 |v_p w_q - v_q w_p|^2 \leqslant$$

$$|a_1|^4 \sum_{p<q} |v_p w_q - v_q w_p|^2 = |a_1|^4 |v \wedge w|^2,$$

while in the nonarchimedean case we also get:

$$|d(v) \wedge d(w)| = \max_{p,q} |a_p v_p a_q w_q - a_q v_q a_p w_p| \leqslant |a_1|^2 |v \wedge w|.$$

By combining these inequalities, for unit vectors $u, v$ satisfying $[u], [v] \in K_\epsilon$, we obtain

$$d(g(v), g(w)) = \frac{|g(v) \wedge g(w)|}{|g(v)| \cdot |g(w)|} \leqslant \frac{|v \wedge w|}{\ell^2 \delta} = \frac{d(v,w)}{\ell^2 \delta}. \quad \square$$

LEMMA 13.56. *Let $g$ be an element in $GL(V)$ with Cartan decomposition $g = kdh$, where $d$ is a diagonal matrix with entries $a_1, \dots, a_n$ on the diagonal such that $|a_1| > |a_2| \geqslant \dots \geqslant |a_n| > 0$. If $\frac{|a_2|}{|a_1|} < \varepsilon^2/\sqrt{n}$, then $g$ maps the complement of the $\varepsilon$–neighborhood of the hyperplane $H = h^{-1}[\mathrm{Span}(e_2, \dots, e_n)]$ into the ball with center $k[e_1]$ and radius $\varepsilon$ .*

PROOF. Since $k$ and $h$ are isometries of $P(V)$, it clearly suffices to prove the statement for $g = d$, $k = h = 1$. Let $[v]$ be a point in $P(V)$ such that $\mathrm{dist}([v], [\mathrm{Span}(e_2, ..., e_n)]) \geqslant \varepsilon$. Then, as in the proof of Lemma 4.42,

$$d([dv], [e_1]) = \frac{|dv \wedge e_1|}{|dv|} \leqslant \sqrt{n} \frac{|a_2|}{\varepsilon |a_1|} < \varepsilon .$$

$\square$

LEMMA 13.57. *If $\sigma = (g_i)$ is a 1-contracting sequence with attracting point $p = A(\sigma)$ and repelling hyperplane $H(\sigma)$, then for every closed ball $B \subseteq H(\sigma)^c$, the maps $g_i|_B$ converge uniformly to the constant function on $B$ which maps everything to the point $p$.*

PROOF. Consider an arbitrary closed ball $B$ in $H(\sigma)^c$. Then $hB$ is a closed ball in the complement of $[\mathrm{Span}(e_{m+1}, ..., e_n)]$. By compactness on $P(V)$, there exists $\varepsilon > 0$ so that the minimal distance from $hB$ to $[\mathrm{Span}(e_{m+1}, ..., e_n)]$ is $\geqslant 2\varepsilon$. Consider

$$B_\varepsilon = \{x \in P(V) \mid \mathrm{dist}(x, B) \leqslant \varepsilon\} ,$$

which is also a closed ball, at minimal distance $\geqslant \varepsilon$ from $[\mathrm{Span}(e_{m+1}, ..., e_n)]$. For all sufficiently large $i$, the ball $h_i B$ is contained in $B_\varepsilon$. Therefore, it suffices to prove that the maps $k_i d_i|_{B_\varepsilon}$ converge uniformly to the constant function on $B_\varepsilon$ which maps everything to the point $p$

Consider $\delta = \varepsilon/2$. For all sufficiently large $i$, according to Lemma 13.56, $d_i(B_\delta)$ is contained in $B([e_1], \delta)$. On the other hand, for all large $i$, the point $k_i[e_1]$ belongs to the ball $B(p, \delta)$. Whence,

$$k_i d_i(B_\varepsilon) \subset B(p, \varepsilon).$$

$\square$

Lemma 13.58. *Let $(g_i)$ be a diverging sequence of elements in $GL(V)$.*

(1) *If there exists a closed ball $B$ with non-empty interior and a point $p$ such that $g_i|_B$ converge uniformly to the constant function on $B$ which maps everything to the point $p$, then $(g_i)$ contains a 1-contracting subsequence with attracting point $p$.*

(2) *If, moreover, there exists a hyperplane $H$ such that for every closed ball $B \subseteq H^c$, $g_i|_B$ converge uniformly to the constant function on $B$ which maps everything to the point $p$, then $(g_i)$ contains a 1-contracting subsequence with the attracting point $p$ and the repelling hyperplane $H$.*

Proof. (1) Since $(g_i)$ is diverging, it contains a subsequence $\sigma$ (whose elements we again denote $g_i$) which is $m$-contracting for some $m$. By replacing $B$ with a smaller ball, we may assume that $B$ is in $E(\sigma)^c$.

Let $g_i = k_i d_i h_i$ denote the Cartan decomposition of $g_i$. By the above observations, for all sufficiently large $i$, the balls $h_i B$ are disjoint from $[\mathrm{Span}(e_{m+1}, ..., e_n)]$. The sequence of closed metric balls $h_i B$ Hausdorff–converges to the closed metric ball $hB$. Therefore, there exists $i_0$ and a closed ball $B'$ contained in the intersection

$$\bigcap_{i \geqslant i_0} h_i B.$$

By the hypothesis, the closed sets $k_i d_i(B')$ Hausdorff–converge to the point $p$.

For every point $[v] \in B'$ represented by a vector $v$, we have:

$$[d_i v] = \left[ v_1 e_1 + \frac{a_2(g_i)}{a_1(g_i)} v_2 e_2 + \ldots + \frac{a_n(g_i)}{a_1(g_i)} v_n e_n \right].$$

After passing to a subsequence, we may assume that

$$\lim_{i\to\infty} \frac{a_k(g_i)}{a_1(g_i)} = \lambda_k, k = 1, \ldots, m.$$

Since our sequence is $m$-contracting,

$$|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_m| > 0.$$

If $m \geqslant 2$ then we may find two distinct points $[v], [v']$ in $B'$ represented by two unit vectors $v = (v_1, \ldots, v_n), v' = (v'_1, \ldots, v')$ so that

$$\lim_{i\to\infty} [d_i v] = [w], \quad \lim_{i\to\infty} [d_i v'] = [w'],$$

$[w] \neq [w'], \quad w = v_1 e_1 + \lambda_2 v_2 e_2 + \ldots + \lambda_m v_m e_m, w' = v'_1 e_1 + \lambda_2 v'_2 e_2 + \ldots + \lambda_m v'_m e_m.$ Assume that $d([w], [w']) = \epsilon > 0$. As

$$[u] = \lim_{i\to\infty} [k_i d_i v] = \lim_{i\to\infty} [k_i w_i], \quad [u'] = \lim_{i\to\infty} [k_i d_i v'] = \lim_{i\to\infty} [k_i w'_i],$$

it follows that the $d([u], [u']) = \epsilon > 0$. This contradicts the assumption that the sequence of sets $k_i d_i(B')$ Hausdorff–converges to a point. It follows that $m = 1$, i.e., $\sigma = (g_i)$ is 1-contracting. If $A(\sigma) \neq p$ then a contradiction easily follows from Lemma 13.57.

(2) According to (1), the sequence $(g_i)$ contains a subsequence $\sigma$ which is 1-contracting, with $A(\sigma) = p$. We continue with the notation introduced in the proof of (1). If $H(\sigma) \neq H$ then at least one of the points $h^{-1}[e_2], \ldots, h^{-1}[e_n]$ is not in $H$. Assume that it is $h^{-1}[e_2]$, and that its distance to $H$ is $2\epsilon > 0$. For sufficiently large all $i$'s, the points $h_i^{-1}[e_2]$ belong to the ball $B(h^{-1}[e_2], \epsilon)$, disjoint

from $H$. It follows that the sequence $k_i d_i[e_2] = k_i[e_2]$ must converge to $p = k[e_1]$ by the assumption, and also to $k[e_2]$, since $\lim_{i\to\infty} k_i = k$. Contradiction. $\square$

The following lemma is an easy consequence of Lemma 13.58, and it is left as an exercise to the reader.

LEMMA 13.59. *Let $(g_i)$ be a 1-contracting sequence in $PGL(V)$, and $f, h \in PGL(V)$. Then the sequence $(fg_ih)$ contains a 1-contracting subsequence $\sigma = (g_i')$ such that*

$$A(\sigma) = f(A(\sigma)), \quad E(\sigma) = h^{-1}E(\sigma).$$

LEMMA 13.60. *Let $(g_i)$ be a diverging sequence in $PGL(V)$. Then there exists a vector space $W$ and an embedding $\rho : PGL(V) \hookrightarrow PGL(W)$ so that a subsequence in $(\rho(g_i))$ is 1-contracting in $PGL(W)$.*

PROOF. After passing to a subsequence, we may assume that the sequence $\sigma = (g_i)$ is $m$-contracting for some $m$, $0 < m < n$. We consider the $m$-th exterior power of $V$,

$$W := \Lambda^m V.$$

The action of $GL(V)$ on $V$ extends naturally to its action on $W$ we obtain the embedding $\rho : GL(V) \hookrightarrow GL(W)$. Clearly, for a matrix $g \in GL(V)$, the norms of the singular values of $\rho(g) \in GL(W)$ are the products

$$\prod_{j_1 < \ldots < j_m} |a_{j_1} \cdots a_{j_m}(g)|.$$

where $a_j(g)$ is the $j$-th singular value of $g$. Then, $|a_1(\rho(g_i))| = |a_1 \cdots a_m(g_i)|$ and it is immediate that

$$\lim_{i\to\infty} \frac{a_l(\rho(g_i))}{a_1(\rho(g_i))} = 0, \forall l > 1. \quad \square$$

We now return to the proof of the Tits alternative for the subgroup $\Gamma < GL(V)$. Recall that we are working under the assumption that the Zariski closure $G = \overline{\Gamma}$ of $\Gamma$ in $GL(V)$ satisfies certain conditions, namely $G$ is Zariski–irreducible, semisimple and it acts irreducibly on $V$.

After replacing $V$ with $W$ as above, since

$$\rho(\Gamma) \leqslant \rho(G) = \rho(\overline{\Gamma}) \leqslant \overline{\rho(\Gamma)} \leqslant \overline{\rho(G)}$$

and $\rho(G)$ is still an algebraic Zariski–irreducible semisimple subgroup (see the end of Section 13.1), it follows that $\overline{\rho(\Gamma)} = \rho(G)$. In what follows, we let $\Gamma$ and $G$ denote $\rho(\Gamma)$ and $\rho(G)$, and we denote the sequence $(\rho(g_i))$ by $(g_i)$.

If the action $G \curvearrowright W$ is reducible, we take a direct sum decomposition

$$W = \bigoplus_{i=1}^{s} W_i$$

into $G$-invariant subspaces, so that the restriction of the $G$-action to each is irreducible. This defines homomorphisms $\rho_i : G \to GL(W_i)$, and all $G_i = \rho_i(G)$ are algebraic Zariski–irreducible semisimple subgroups. In particular, $G_i = [G_i, G_i]$, hence every $G_i$ is, in fact, contained in $SL(W_i)$. In particular for the 1-dimensional spaces $W_i$, the group $G_i$ is trivial. Without loss of generality, we can, therefore, assume that each subspace $W_i$ has dimension $> 1$.

Lemma 13.61. *For some $s$, the sequence $\sigma = (g_i)$ restricted to $W_s$ is 1-contracting.*

Proof. Let $p = A(\sigma) \in P(W)$ and $H = H(\sigma) \subset P(W)$ be the attracting point and, respectively, the repelling hyperplane of the sequence $\sigma = (g_i)$. Since the subspaces $W_t$ are $G$-invariant, for each $t$ either $p \in P(W_t)$ or $P(W_t) \subset H$. Since $H$ is a hyperplane in $P(W)$, it follows that $p \in P(W_s)$ for some $s$. The restriction of $(g_i)$ to $P(W_s)$ converges to $p$ away from $H \cap P(W_s)$. Since $\dim(W_s) > 1$, we are done. $\square$

Let $\rho_s$ be the representation $G \to SL(W_s)$. Our goal will be to prove that $\rho_s(\Gamma)$ contains a free non-abelian group, whence it will follow that $\Gamma$ contains such a group, which will conclude the proof. For simplicity of notation, in what follows, we denote $\rho_s(\Gamma)$ by $\Gamma$, its Zariski closure by $G$ and the vector space $W_s$ by $V$. As before, the Zariski closure of $\rho_s(\Gamma)$ is Zariski–irreducible and semisimple.

Theorem 13.62. *Let $\Gamma$ be a subgroup in $SL(V)$ containing a 1-contracting sequence of elements, and such that the Zariski closure $\overline{\Gamma}$ of $\Gamma$ is Zariski–irreducible and that $\Gamma$ acts irreducibly on $V$. Then $\Gamma$ contains a free non-abelian subgroup.*

Before beginning the proof, we note that the 1-contracting sequence that we now have at our disposal in the group $\Gamma$ does not suffice yet, not even to construct one of the two elements in a ping-pong pair "modeled" after the one in Lemma 4.42. Indeed, for every $i \in \mathbb{N}$ the action of the element $g_i \in \Gamma$ on the projective space $P = P(V)$ is, as represented in Figure 13.1 (where we picture projective space as a sphere). According to Lemma 13.56, for every $\epsilon > 0$ and all sufficiently large $i$, the transformation $g_i$ (with the Cartan decomposition $k_i d_i h_i$) maps the complement of the $\epsilon$-neighborhood of $H(\sigma) = h_i^{-1}[\mathrm{Span}(e_2, \ldots, e_n)]$ into the $\epsilon$-neighborhood of the point $A(\sigma) = k_i [e_1]$, with the notation of 13.53.

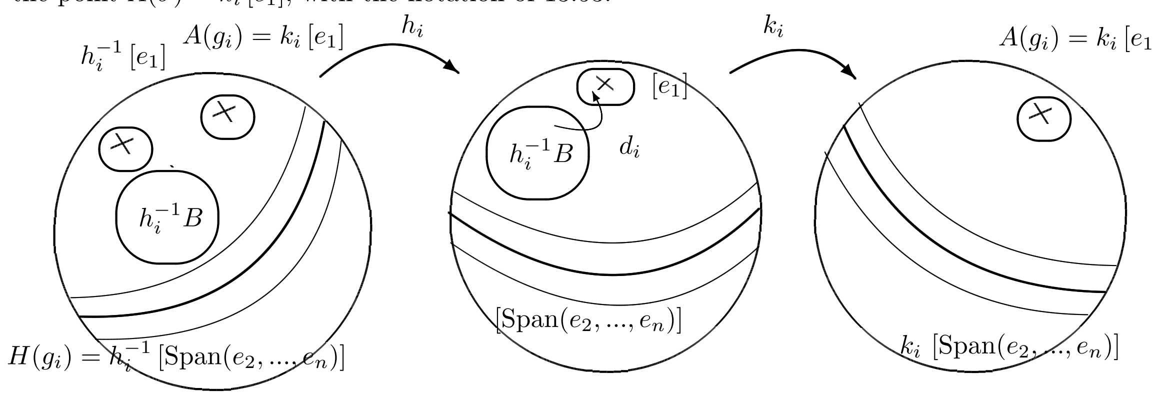



Figure 13.1. The action of $g_i$.

The first problem occurs when one iterates $g_i$, i.e. one considers $g_i^2, g_i^3$, etc. Nothing guarantees that $g_i^2$ would also map the complement of the $\epsilon$-neighborhood of $H(g_i)$ into the $\epsilon$-neighborhood of $A(g_i)$, for large $i$. This only happens when the $\epsilon$-neighborhood of $A(g_i)$ is disjoint from the $\epsilon$-neighborhood of $H(g_i)$. Our hypothesis does not ensure this, since no conditions can be imposed on $h_i, k_i$ and their limits (see comments in Example 13.54). We will use Lemma 13.59 and the notion of a *separating set* developed in the sequel to circumvent this difficulty.

**Separating sets.**

DEFINITION 13.63. A subset $F \subset PGL(V)$ is called *m-separating* if for every choice of points $p_1, \ldots, p_m \in P = P(V)$ and hyperplanes $H_1, \ldots, H_m \subset P$, there exists $f \in F$ so that

$$f^{\pm 1}(p_i) \notin H_j, \forall i, j = 1, \ldots, m.$$

It will now become apparent why we endeavored to ensure the two irreducibility properties (for the Zariski topology, and for the action) for the Zariski closure of $\Gamma$.

PROPOSITION 13.64. *Let $\Gamma \subset SL(V)$ be a subgroup with the property that its Zariski closure is Zariski–irreducible and it acts irreducibly on $V$. For every $m$, $\Gamma$ contains a finite $m$-separating subset $F$.*

PROOF. Let $G$ be the Zariski closure of $\Gamma$. Let $P^\vee$ denote the space of hyperplanes in $P$ (i.e. the projective space of the dual of $V$). For each $g \in G$ let $M_g \subset P^m \times (P^\vee)^m$ denote the collection of $2m$-tuples

$$(p_1, \ldots, p_m, H_1, \ldots, H_m)$$

so that

$$g(p_i) \in H_j \text{ or } g^{-1}(p_i) \in H_j$$

for some $i, j = 1, \ldots, m$.

LEMMA 13.65. *If $\Gamma$ is as in Proposition 13.64 then*

$$\bigcap_{g \in \Gamma} M_g = \emptyset\,.$$

PROOF. Suppose to the contrary that the intersection is nonempty. Then there exists a $2m$-tuple $(p_1, \ldots, p_m, H_1, \ldots, H_m)$ so that for every $g \in \Gamma$,

$$\exists\ i, j \text{ so that } g(p_i) \in H_j \text{ or } g^{-1}(p_i) \in H_j\,. \tag{13.2}$$

The set of elements $g \in SL(V)$ such that (13.2) holds for the given $2m$-tuple is Zariski–closed, and $G$ is the Zariski closure of $\Gamma$, hence all $g \in G$ also satisfy (13.2).

Let $G^{\pm}_{p_i, H_j}$ denote the set of $g \in G$ so that

$$g^{\pm 1}(p_i) \in H_j.$$

Clearly, these subsets are Zariski–closed and cover the group $G$. Since $G$ Zariski–irreducible, it follows that one of these sets, say $G^{+}_{p_i, H_j}$, is the entire of $G$. Therefore, for every $g \in G$, $g(p_i) \in H_j$. Thus, projectivization of the vector subspace $L$ spanned by the $G$–orbit (of lines) $G \cdot p_i$ is contained in $H_j$. The subspace $L$ is proper and $G$-invariant. This contradicts the hypothesis that $G$ acts irreducibly on $V$. ☐

We now finish the proof of Proposition 13.64. Let $M_g^c$ denote the complement of $M_g$ in $P^m \times (P^\vee)^m$. This set is Zariski open. By Lemma 13.65, the sets $M_g^c$ $(g \in \Gamma)$ cover the space $P^m \times (P^\vee)^m$. Since $\mathbb{K}$ is a local field, the product $P^m \times (P^\vee)^m$ is compact and, thus, the above open cover contains a finite subcover. Hence, there exists a finite set $F \subset \Gamma$ so that

$$\bigcup_{f\in F} M_f^c = P^m \times (P^\vee)^m.$$

This set satisfies the assertion of Proposition 13.64. □

REMARK 13.66. The above proposition holds even if the field $\mathbb{K}$ is not local. Then the point is that by Hilbert's Nullstellensatz, there exists a finite subset $F \subset \Gamma$ so that

$$\bigcap_{f\in F} M_f = \bigcap_{g\in\Gamma} M_g = \emptyset.$$

With this modification, the above proof goes through.

**Ping-pong sequences.** We now begin the proof of Theorem 13.62, which will be split in several lemmas.

In what follows we fix a 4-separating finite subset $F \subset \Gamma \subset PGL(V)$. We will use the notation $f$ for the elements of $F$.

LEMMA 13.67. *There exists $f \in F$ so that (after passing to a subsequence in $(g_i)$) both sequences $h_i := g_i f g_i^{-1}$ and $g_i f^{-1} g_i^{-1}$ are 1-contracting.*

PROOF. After passing to a subsequence $\sigma = (g_i)$, we can assume that the sequence $\sigma^- = (g_i^{-1})$ is $m$-contracting, with attracting subspace $A(\sigma^-)$ and repelling subspace $E(\sigma^-)$. Pick a point $q$ in the complement of the subspace $E(\sigma^-)$. After passing to a subsequence in $(g_i)$ again, we can assume that $\lim_i g_i^{-1}(q) = u \in A(\sigma^-)$. Let $A(\sigma)$ and $H(\sigma)$ be the attracting point and the repelling hyperplane of the sequence $\sigma$.

Since $F$ is a separating subset, there exists $f \in F$ so that $f^{\pm 1}(u) \notin H(\sigma)$.

Take a small closed ball $B(q,\epsilon) \subset P$ centered at $q$ and disjoint from $E(\sigma^-)$. According to Lemma 13.55, $g_i^{-1}(B(q,\epsilon)) \subset B\left(g_i^{-1}(q), L\epsilon\right)$ for al large $i$ and $L$ independent of $i$. It follows that for all large $i$

$$g_i^{-1}(B(q,\epsilon)) \subset B(u, 2L\epsilon).$$

By Lemma 4.41, $f g_i^{-1}(B(q,\epsilon)) \subset B(f(u), L'\epsilon)$ for all large $i$ and $L'$ independent of $i$. Note that if we reduce $\epsilon$, the constants $L$ and $L'$ will not change. We take $\epsilon$ small enough so that the sets $B(f(u), L'\epsilon)$ and $\mathcal{N}_\epsilon(H(\sigma))$ are disjoint. Since the sequence $(g_i)$ restricted to the complement of $\mathcal{N}_\epsilon(H(\sigma))$ converges uniformly to the point $A(\sigma)$ it follows that the sequence $g_i f g_i^{-1}|_{B(q,\epsilon)}$ converges uniformly to the point $A(\sigma)$. Lemma 13.58, (1), now implies that $(g_i)$ contains a 1-contracting subsequence.

The same argument for $f^{-1}$ concludes the proof. □

Thus, we have found a 1-contracting sequence $\tau = (h_i)$ in $\Gamma$ such that the sequence $\tau^- = (h_i^{-1})$ is also 1-contracting.

LEMMA 13.68. *There exists $f \in F$ such that, for a subsequence $\eta = (y_i)$ of the sequence $(fh_i)$, both $\eta$ and $\eta^- = \left(y_i^{-1}\right)$ are 1-contracting. Moreover,*

$$A(\eta) \notin H(\eta) \quad and \quad A(\eta^-) \notin H(\eta^-). \qquad (13.3)$$

PROOF. By Lemma 13.59, for any choice $f \in F$, the sequence $(fh_i)$ contains a 1-contracting subsequence $\eta = (y_i)$, with $\eta^- = \left(y_i^{-1}\right)$ likewise 1-contracting, and

$$A(\eta) = f(A(\tau)),\ H(\eta) = H(\tau),$$
$$A\left(\eta^-\right) = A\left(\tau^-\right),\ H\left(\eta^-\right) = fH\left(\tau^-\right).$$

Now, the assertion follows from the fact that $F$ is a 4-separating set. ☐

DEFINITION 13.69. [Ping-pong pair] A pair of sequences $\eta = (y_i)$ and $\zeta = (z_i)$ is called a *ping-pong pair* if both sequences are as in Lemma 13.68 and, furthermore, $A(\eta^\pm) \notin H(\zeta^\pm)$ and $A(\zeta^\pm) \notin H(\eta^\pm)$.

Let $\eta = (y_i)$ be the sequence from Lemma 13.68.

LEMMA 13.70. *There exists $f \in F$ so that the sequences $(y_i), (z_i) = (fy_if^{-1})$ contain subsequences that form a ping-pong pair.*

PROOF. By Lemma 13.59, after replacing $\eta = (y_i)$ with a subsequence, we may assume that $\zeta = (z_i)$ and $\zeta^- = \left(z_i^{-1}\right)$ are 1-contracting and $A\left(\zeta^{\pm 1}\right) = fA\left(\eta^{\pm 1}\right)$, while $H\left(\zeta^{\pm 1}\right) = fH\left(\eta^{\pm 1}\right)$. Now, the assertion follows from the fact that $F$ is 4-separating. ☐

*End of proof of Theorem 13.62.* Lemma 13.70 implies that $\Gamma$ contains a ping-pong pair of sequences $\eta = (y_i),\ \zeta = (z_i)$. For every small $\epsilon$ and all large $i$, we have:

$$\mathcal{N}_\epsilon(H(\eta))^c \xrightarrow{y_i} B(A(\eta),\epsilon)$$
$$\mathcal{N}_\epsilon\left(H\left(\eta^-\right)\right)^c \xrightarrow{y_i^{-1}} B\left(A\left(\eta^-\right),\epsilon\right)$$
$$\mathcal{N}_\epsilon(H(\zeta))^c \xrightarrow{z_i} B(A(\zeta),\epsilon)$$
$$\mathcal{N}_\epsilon\left(H\left(\zeta^-\right)\right)^c \xrightarrow{z_i^{-1}} B\left(A\left(\zeta^-\right),\epsilon\right)$$

Moreover, for $\epsilon$ sufficiently small, the balls on the right-hand side are contained in the complements of tubular neighborhoods on the left-hand side. Therefore, the above statements also hold with transformations $y_i, y_i^{-1}, z_i, z_i^{-1}$ replaced by their $k$-th iterations for all $k > 0$.

We choose $\epsilon$ small enough so that

$$B(A(\eta),\epsilon) \cap \mathcal{N}_\epsilon\left(H(\eta) \cup H(\zeta) \cup H\left(\zeta^-\right)\right) = \emptyset,$$
$$B\left(A\left(\eta^-\right),\epsilon\right) \cap \mathcal{N}_\epsilon\left(H\left(\eta^-\right) \cup H(\zeta) \cup H\left(\zeta^-\right)\right) = \emptyset,$$
$$B(A(\zeta),\epsilon) \cap \mathcal{N}_\epsilon\left(H(\zeta) \cup H(\eta) \cup H\left(\eta^-\right)\right) = \emptyset,$$
$$B\left(A\left(\zeta^-\right),\epsilon\right) \cap \mathcal{N}_\epsilon\left(H\left(\zeta^-\right) \cup H(\eta) \cup H\left(\eta^-\right)\right) = \emptyset.$$

For $\epsilon$ small as above, we consider the sets

$$\widetilde{A} = B(A(\eta),\epsilon) \cup B\left(A\left(\eta^-\right),\epsilon\right)$$

and

$$\widetilde{B} = B(A(\zeta),\epsilon) \cup B\left(A\left(\zeta^-\right),\epsilon\right).$$

Since $A(\eta) \in H\left(\eta^-\right)$, $A\left(\eta^-\right) \in H(\eta)$ and $A(\zeta) \in H\left(\zeta^-\right)$, $A\left(\zeta^-\right) \in H(\zeta)$, our hypotheses imply that $\widetilde{A} \cap \widetilde{B} = \emptyset$. Moreover for all large $i$, for every $k \in \mathbb{Z} \setminus \{0\}$,

$$y_i^k\left(\widetilde{B}\right) \subseteq \widetilde{A} \text{ and } z_i^k\left(\widetilde{A}\right) \subseteq \widetilde{B}.$$

Lemma 4.37 now implies that for all large $i$, the group $\langle y_i, z_i\rangle$ is a free group of rank 2. ☐

### 13.6. Free subgroups in compact Lie groups

The compact case is more complicated. Let $\Gamma$ be a relatively compact finitely-generated subgroup of $G = SL(n, \mathbb{C})$. According to Proposition 13.46, we may assume that the Zariski closure of $\Gamma$ in $SL(n, \mathbb{C})$ is Zariski–irreducible, semisimple, and that it acts irreducibly, i.e., it does not preserve a proper subspace of $\mathbb{C}^n$. Note that in this section, unlike in the previous one, $G$ denotes $SL(n, \mathbb{C})$, not the Zariski closure of $\Gamma$.

Let $\gamma_1, \ldots, \gamma_m$ denote generators of $\Gamma$ and consider the subfield $F$ in $\mathbb{C}$ generated by the matrix entries of these matrices.

**Reduction to a number field case.** Consider the *representation variety* $R(\Gamma, G) = \mathrm{Hom}(\Gamma, G)$. This space can be described as follows. Let

$$\langle \gamma_1, \ldots, \gamma_m | r_1, \ldots \rangle$$

be a presentation of $\Gamma$ (the number of relators could be infinite). Each homomorphism $\rho : \Gamma \to G$ is determined by the images of the generators of $\Gamma$. Hence $R(\Gamma, G)$ is a subset of $G^m$. A map $\rho : \gamma_i \mapsto G, i = 1, \ldots, m$ extends to a homomorphism of $\Gamma$ if and only if

$$\forall j, \rho(r_j) = 1. \tag{13.4}$$

Since the relators $r_j$ are words in $\gamma_1^{\pm 1}, \ldots, \gamma_m^{\pm 1}$, the equations (13.4) amount to polynomial equations on $G^m$. Hence, $R(\Gamma, G)$ is given by a system of polynomial equations and has a natural structure of an affine algebraic variety. Since the formula for the inverse in $SL(n)$ involves only integer linear combinations of products of matrix entries, it follows that the above equations have integer (in particular, rational) coefficients. In other words, the representation variety $R(\Gamma, G)$ is *defined over* $\mathbb{Q}$.

Proposition 13.71. *Let $Z$ be an affine variety in $\mathbb{C}^N$ defined by polynomial equations with rational coefficients and let $\overline{\mathbb{Q}}$ be the field of algebraic numbers, the algebraic closure of $\mathbb{Q}$. Then the set $Z \cap \overline{\mathbb{Q}}^N$ is dense in $Z$ with respect to the classical topology on $\mathbb{C}^N$.*

Proof. The proof is by induction on $N$. The assertion is clear for $N = 1$. Indeed, in this case either $Z = \mathbb{C}$ or $Z$ is a finite set of roots of a polynomial with rational coefficients: These roots are algebraic numbers. Suppose the assertion holds for subvarieties in $\mathbb{C}^{N-1}$. Pick a point $x = (x_1, \ldots, x_N) \in Z$ and let $q_i$ be a sequence of rational numbers converging to the first coordinate $x_1$. For each rational number $q_i$, the intersection $Z \cap \{x_1 = q_i\}$ is again an affine variety defined over $\mathbb{Q}$ which sits inside $\mathbb{C}^{N-1}$. Now the claim follows from the induction hypothesis by taking a diagonal sequence. □

Corollary 13.72. *Algebraic points are dense in $R = R(\Gamma, G)$ with respect to the classical topology. In other words, for every homomorphism $\rho : \Gamma \to G$, there exists a sequence of homomorphisms $\rho_j : \Gamma \to G$ converging to $\rho$ so that the matrix entries of the images of generators $\rho_j(\gamma_i)$ are in $\overline{\mathbb{Q}}$.*

We now let $\rho_i \in R(\Gamma, G)$ be a sequence which converges to the identity representation $\rho : \Gamma \to \Gamma \subset G$. Recall that in section 13.3, we proved that for every finitely-generated subgroup $\Gamma \subset GL(n, \mathbb{C})$ which is not virtually solvable, there exists a neighborhood $\Sigma$ of $\rho = id$ in $\mathrm{Hom}(\Gamma, GL(n, \mathbb{C}))$ so that every $\rho' \in \Xi$ has image which is not virtually solvable. Therefore, without loss of generality, we may assume that each $\rho_j(\Gamma)$ constructed above is not virtually solvable.

LEMMA 13.73. *If* $\Gamma_j := \rho_j(\Gamma)$ *contains a free subgroup* $\Lambda_j$ *of rank 2 then so does* $\Gamma$.

PROOF. Let $g_1, g_2 \in \Gamma$ be the elements which map to the free generators $h_1, h_2$ of $\Lambda_j$ under $\rho_j$. Let $\Lambda$ be the subgroup of $\Gamma$ generated by $g_1, g_2$. We claim that $\Lambda$ is free of rank 2. Indeed, since $\Lambda_j$ is free of rank 2, there exists a homomorphism $\phi_j : \Lambda_j \to \Lambda$ sending $h_k$ to $g_k$, $k = 1, 2$. The composition $\phi_j \circ \rho_j$ is the identity since it sends each $h_k$ to itself. Hence, $\phi_j : \Lambda_j \to \Lambda$ is an isomorphism. □

Thus, it suffices to consider the case when the field $F$ (generated by matrix entries of generators of $\Gamma$) is a number field, i.e., is contained in $\overline{\mathbb{Q}}$. The absolute Galois group $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on $F$ and hence on $SL(n, F)$:

$$\forall \sigma \in Gal(\overline{\mathbb{Q}}/\mathbb{Q}), \quad A = (a_{ij}) \in SL(n, F), \quad \sigma(A) := (a_{ij}^{\sigma}).$$

Every $\sigma \in Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ extends (by identity) to the set of transcendental numbers and, hence, extends to an automorphism $\sigma$ of $\mathbb{C}$. Therefore, $\sigma$ determines an automorphism $\sigma$ of $SL(n, \mathbb{C})$ (which, typically, it discontinuous in the classical topology). Therefore, $\sigma$ will send the subgroup $\Gamma \subset SL(n, F)$ to $\sigma(\Gamma) \subset G(\sigma(F)) \subset SL(n, \mathbb{C})$. The homomorphism $\sigma : \Gamma \to \Gamma' := \sigma(\Gamma)$ is $1-1$ and, therefore, if for some $\sigma$ the group $SL(n, \sigma(F))$ happens to be a non-relatively compact subgroup of $SL(n, \mathbb{C})$ we are done by Theorem 13.49.

However, it could happen that for each $\sigma$ the group $G(\sigma(F))$ is relatively compact and, thus, we seemingly gained nothing. Nevertheless, there is a remarkable construction which saves the proof.

*Adeles.* (See [**Lan64**, Chapter 6].) The ring of adeles was introduced by A. Weil in 1936. For a number field $F$ consider various norms $|\cdot| : F \to \mathbb{R}_+$, see §1.7.

Suppose that $F$ is a finitely–generated number field. Then $F$ is a finite extension of $\mathbb{Q}$. Let $\mathrm{Nor}(F)$ denote the set of all norms on $F$ which restrict to either the absolute value or to one of the $p$-adic norms on $\mathbb{Q} \subset F$. We will use the notation $F_\nu$, $\mathbb{Q}_\nu$ to denote the completion of $F$ with respect to the norm $\nu$, we let $O_\nu \subset F_\nu$ denote the ring of integers:

$$O_\nu = \{x \in F_\nu : \nu(x) \leqslant 1\}.$$

Note that for each $x \in \mathbb{Q}$, $x \in O_p$ for all but finitely many $p$'s, since $x$ has only finitely many primes in its denominator. The same is true for elements of $F$: For all but finitely many $\nu \in \mathrm{Nor}(F)$, $\nu(x) \leqslant 1$. We will use the notation $\nu_p$ for the $p$-adic norm on $\mathbb{Q}$.

*Product formula*: For each $x \in \mathbb{Q} \setminus \{0\}$

$$\prod_{\nu \in \mathrm{Nor}(\mathbb{Q})} \nu(x) = 1.$$

Indeed, if $x = p$ is prime, then $|p| = p$ for the archimedean norm, $\nu(p) = 1$ if $\nu \neq \nu_p$ is a nonarchimedean norm and $\nu_p(p) = 1/p$. Thus, the product formula holds for prime numbers $x$. Since norms are multiplicative functions from $\mathbb{Q}^\times$ to $\mathbb{R}_+$, the product formula holds for arbitrary $x \neq 0$. A similar product formula is true for an arbitrary algebraic number field $F$:

$$\prod_{\nu \in \mathrm{Nor}(F)} (\nu(x))^{N_\nu} = 1,$$

where $N_\nu = [F_\nu : \mathbb{Q}_\nu]$, see [**Lan64**, Chapter 6].

Definition 13.74. The ring of *adeles* is the *restricted product*

$$\mathbb{A}(F) := \prod_{\nu \in \mathrm{Nor}(F)}' F_\nu,$$

i.e. the subset of the direct product

$$\prod_{\nu \in \mathrm{Nor}(F)} F_\nu \tag{13.5}$$

which consists of sequences whose projection to $F_\nu$ belongs to $O_\nu$ for all but finitely many $\nu$'s.

We topologize $\mathbb{A}(F)$ *via* the subset topology induced from the product (13.5), which, in turn, is equipped with the product topology. Note that the ring operations are continuous with respect to this topology.

For instance, if $F = \mathbb{Q}$ then $\mathbb{A}(\mathbb{Q})$ is the restricted product

$$\mathbb{R} \times \prod_{p \text{ is prime}}' \mathbb{Q}_p.$$

Remark 13.75. Actually, it suffices to use the ring of adeles $\mathbb{A}(\mathbb{Q})$. This is done *via* the following procedure called *the restriction of scalars*: The field $F$ is an $m$-dimensional vector space over $\mathbb{Q}$. This determines an embedding

$$\Gamma \subset GL(n, F) \hookrightarrow \prod_{i=1}^{m} GL(n, \mathbb{Q}) \subset GL(n + m, \mathbb{Q})$$

and reduces our discussion to the case $\Gamma \subset GL(n + m, \mathbb{Q})$.

*Now, a miracle happens:*

Theorem 13.76 (See e.g. Chapter 6, Theorem 1 of [**Lan64**]). *The image of the diagonal embedding $F \hookrightarrow \mathbb{A}(F)$ is a discrete subset in $\mathbb{A}(F)$.*

Proof. It suffices to verify that 0 is an isolated point. Take the archimedean norms $\nu_1, \ldots, \nu_m \in Nor(F)$ (there are only finitely many of them since the Galois group $Gal(F/\mathbb{Q})$ is finite) and consider the open subset

$$U = \prod_{i=1}^{m} \{x \in F_{\nu_i} : \nu_i(x) < 1/2\} \times \prod_{\mu \in \mathrm{Nor}(F) \setminus \{\nu_1, \ldots, \nu_m\}} O_\mu$$

of $\mathbb{A}(F)$. Then for each $(x_\nu) \in U$,

$$\prod_{\nu \in \mathrm{Nor}(F)} \nu(x_\nu) < 1/2 < 1.$$

Hence, by the product formula, the intersection of $U$ with the image of $F$ in $\mathbb{A}(F)$ consists only of $\{0\}$. □

In order to appreciate this theorem, the reader should consider the case $F = \mathbb{Q}$ which is dense in the completion of $\mathbb{Q}$ with respect to every norm.

Recall that $\Gamma$ is a subgroup in $SL(n, F)$. The diagonal embedding above defines an embedding

$$\Gamma \subset SL(n, F) \hookrightarrow SL(n, \mathbb{A}(F)) \subset \prod_{\nu \in \mathrm{Nor}(F)} SL(n, F_\nu)$$

with discrete image.

For each norm $\nu \in \mathrm{Nor}(F)$ we have the projection $p_\nu : \Gamma \to SL(n, F_\nu)$. If the image $p_\nu(\Gamma)$ is relatively compact for each $\nu$ then $\Gamma$ is relatively compact in $\prod_{\nu \in \mathrm{Nor}(F)} SL(n, F_\nu)$, by Tychonoff's Theorem. As $\Gamma$ is also discrete, this implies that $\Gamma$ is finite, a contradiction.

Thus, there exists a norm $\nu \in \mathrm{Nor}(F)$ such that the image of $\Gamma$ in $SL(n, F_\nu)$ is not relatively compact. If $\nu$ happens to be archimedean we are done as before. The more interesting case occurs if $\nu$ is nonarchimedean. Then the field $F_\nu = \mathbf{k}$ is a local field (just like the $p$-adic completion of the rational numbers) and we appeal to Theorem 13.49 to conclude that $\Gamma$ contains a free subgroup in this case as well. This concludes the proof of the Tits' Alternative (Theorem 13.1). □

REMARK 13.77. 1. The above proof works only if $\Gamma$ is finitely generated. The general case will be treated below.

2. Tits' proof also works for algebraic groups over fields of positive characteristic, see [**Tit72**]. However, in the case of infinitely-generated groups one has to modify the assertion, since $GL(n, F)$, where $F$ is an infinite algebraic extension of a finite field, provides a counter-example otherwise.

3. The arguments in the above proof mostly follow the ones of Breuillard and Gelander in [**BG03**].

Note that a consequence of the previous arguments is the following.

THEOREM 13.78. *Let $\Gamma$ be a finitely generated group that does not contain a free non-abelian subgroup. Then:*

(1) *If $\Gamma$ is a subgroup of an algebraic group $L$ then its Zariski closure $G$ is virtually solvable.*
(2) *If $\Gamma$ is a subgroup of a Lie group $L$ with finitely many connected components, then the closure $G$ of $\Gamma$ in the Lie group $L$ is virtually solvable.*

*Furthermore, in both cases above, the solvable subgroup $S$ of $G$ has derived length at most $\delta = \delta(L)$ and the index $|G : S|$ is at most $\nu = \nu(L)$.*

PROOF. The arguments in the proof of Theorem 13.1 imply the statement (1). The statement (2) follows in a similar manner. Indeed, as in Section 13.4, using the adjoint representation one can reduce the problem to the setting of linear subgroups, and there the closure in the standard topology is contained in the Zariski closure. □

**Tits Alternative without finite generation assumption.**

We will need

LEMMA 13.79. *Every countable field $F$ of zero characteristic embeds in $\mathbb{C}$.*

Proof. Since $F$ has characteristic zero, its prime subfield $P$ is isomorphic to $\mathbb{Q}$. Then $F$ is an extension of the form

$$P \subset E \subset F,$$

where $P \subset E$ is an algebraic extension and $E \subset F$ is a purely transcendental extension (see [Chapter VI.1][**Hun80**]). The algebraic number field $E$ embeds in $\bar{Q} \subset \mathbb{C}$. Since $F$ is countable, $F/E$ has countable dimension and, therefore,

$$F = E(u_1, \ldots, u_m)$$

or

$$F = E(u_1, \ldots, u_m, \ldots).$$

Sending variables $u_j$ to independent transcendental numbers $z_j \in \mathbb{C}$, we then obtain an embedding $F \hookrightarrow \mathbb{C}$. □

Theorem 13.80 (Tits Alternative). *Let $F$ be a field of zero characteristic and $\Gamma$ be a subgroup of $GL(n, F)$. Then either $\Gamma$ is virtually solvable or $\Gamma$ contains a free nonabelian subgroup.*

Proof. The group $\Gamma$ is the direct limit of the direct system of its finitely-generated subgroups $\Gamma_i$. Let $F_i \subset F$ denote the subfield generated by the matrix entries of the generators of $\Gamma_i$. Then $\Gamma_i \leqslant GL(n, F_i)$. Since $F$ (and, hence, every $F_i$) has zero characteristic, the field $F_i$ embeds in $\mathbb{C}$ (see Lemma 13.79).

If one of the groups $\Gamma_i$ contains a free nonabelian subgroup, then so does $\Gamma$. Assume, therefore, that this does not happen. Then, in view of the Tits Alternative (for finitely generated linear groups), each $\Gamma_i$ is virtually solvable. For $\nu = \nu(GL(n, \mathbb{C}))$ and $\delta = \delta(GL(n, \mathbb{C}))$, every $i$ there exists a subgroup $\Lambda_i \leqslant \Gamma_i$ of index $\leqslant \delta$, so that $\Lambda_i$ has derived length $\leqslant \delta$ (see Theorem 13.38). In view of Exercise 11.18, the group $\Gamma$ is also virtually solvable. □

CHAPTER 14

# Gromov's Theorem

The main objective of this chapter is to prove the converse of Bass–Guivarc'h Theorem 12.48.

THEOREM 14.1 (M. Gromov, [**Gro81**]). *If* $\Gamma$ *is a finitely generated group of polynomial growth then* $\Gamma$ *is virtually nilpotent.*

COROLLARY 14.2. *Suppose that* $\Gamma$ *is a finitely generated group which is quasi-isometric to a nilpotent group. Then* $\Gamma$ *is virtually nilpotent.*

PROOF. Follows directly from Gromov's theorem since polynomial growth is a QI invariant. □

REMARK 14.3. An alternative proof of the above corollary (which does not use Gromov's theorem) was given by Y. Shalom [**Sha04**].

We will actually prove a slightly stronger version (Theorem 14.5 below) of Theorem 14.1, which is due to van der Dries and Wilkie [**dDW84**] (our proof mainly follows [**dDW84**]).

DEFINITION 14.4. A finitely generated group $\Gamma$ has *weakly polynomial growth* of degree $\leqslant a$ if there exists a sequence of positive numbers $R_n$ diverging to infinity and a pair of numbers $C$ and $a$, for which

$$\mathfrak{G}(R_n) \leqslant CR_n^a, \forall n \in \mathbb{N}.$$

THEOREM 14.5. *If* $\Gamma$ *has weakly polynomial growth then it is virtually nilpotent.*

Gromov's proof of polynomial growth theorem relies heavily upon the work of Montgomery and Zippin on Hilbert's 5-th problem (characterization of Lie groups as topological groups).[1] Therefore in the following section we collect several *elementary* facts in point-set topology and review *highly nontrivial* results of Montgomery and Zippin.

## 14.1. Montgomery-Zippin Theorems

Recall that a *topological group* is a group $G$ which is given topology so that the group operations (multiplication and inversion) are continuous. A continuous group action of a topological group $G$ on a topological space $X$ is a continuous map

$$\mu : G \times X \to X$$

such that $\mu(e, x) = x$ for each $x \in X$ and for each $g, h \in G$

$$\mu(gh, x) = \mu(g) \circ \mu(h)(x).$$

[1]In 1990 J. Hirschfeld [**Hir90**] gave an alternative solution of Hilbert's problem, based on nonstandard analysis. However, his proof does not seem to apply in the context of more general topological group actions needed for Gromov's proof.

(Here and in what follows $e$ denotes the identity element in a group.) In particular, for each $g \in G$ the map $x \mapsto \mu(g)(x)$ is a homeomorphism $X \to X$. Thus, each action $\mu$ defines a homomorphism $G \to \mathrm{Homeo}(X)$. The action $\mu$ is *effective* if this homomorphism is injective.

Throughout this section we will consider only *metrizable* topological spaces $X$. We will topologize the group of homeomorphisms $\mathrm{Homeo}(X)$ *via* the compact-open topology; thus, we obtain a continuous action $\mathrm{Homeo}(X) \times X \to X$.

Definition 14.6. [Property A, section 6.2 of [**MZ74**]] Suppose that $H$ is a separable, locally compact topological group. Then $H$ is said to satisfy Property A if for each neighborhood $V$ of the identity $e \in H$ there exists a compact subgroup $K \subset H$ so that $K \subset V$ and $H/K$ (equipped with the quotient topology) is a Lie group.

In other words, the group $H$ can be *approximated by* the Lie groups $H/K$. Here is an example to keep in mind. Let $H$ be the additive group $\mathbb{Q}_p$ of $p$-adic numbers. The sets

$$H_{i,p} := \{x \in \mathbb{Q}_p : |x|_p \leqslant p^{-i}\}, i \in \mathbb{N},$$

are open and form a basis of topology at the identity element $0 \in \mathbb{Q}_p = H$. For instance, for $i = 0$, $H_{0,p}$ is the group of $p$-adic integers $O_p$. Now, the fact that the $p$-adic norm $|x|_p$ is nonarchimedean implies that $H_{i,p}$ is a subgroup of $H$. Furthermore, this subgroup is clearly closed; we leave it to the reader to verify that $H_{i,p}$ is also compact. The quotient $H/H_{i,p}$ has discrete quotient topology since $H_{i,p}$ is both open and closed in $H$. Hence, $G_{i,p} = H/H_{i,p}$ is a Lie group.

Exercise 14.7. The groups $G_{i,p}$ is obviously abelian; show that $G_{i,p}$ is finitely generated and compute its free and torsion parts. Hint: First compute $\mathbb{Q}_p/O_p$.

Theorem 14.8 (D. Montgomery and L. Zippin, [**MZ74**], Chapter IV). *Each separable locally compact group $H$ contains an open and closed subgroup $\hat{H} \leqslant H$ such that $\hat{H}$ satisfies Property A.*

The following theorem is proven in [**MZ74**], section 6.3, Corollary on page 243.

Theorem 14.9 (D. Montgomery and L. Zippin). *Suppose that $X$ is a topological space which is connected, locally connected, finite-dimensional and locally compact. Suppose that $H$ is a separable locally compact group satisfying Property A, $H \times X \to X$ is a topological action which is effective and transitive. Then $H$ is a Lie group.*

We are mainly interested in the following corollary for metric spaces.

Theorem 14.10. *Let $X$ be a metric space which is proper, connected, locally connected and finite-dimensional. Let $H$ be a closed subgroup in* $\mathrm{Homeo}(X)$ *with the compact-open topology, such that $H \curvearrowright X$ is transitive. If there exists $L \in \mathbb{R}$ such that each $h \in H$ is $L$–Lipschitz, then the group $H$ is a Lie group with finitely many connected components.*

Proof. It is clear that $H \times X \to X$ is a continuous effective action. It follows from the Arzela-Ascoli theorem that $H$ is locally compact.

Lemma 14.11. *The group $H$ is separable.*

Proof. Pick a point $x \in X$. Given $r \in \mathbb{R}_+$, consider the subset

$$H_r = \{h \in H : \mathrm{dist}(x, h(x)) \leqslant r\}.$$

By Arzela–Ascoli theorem, each $H_r$ is a compact set. Therefore

$$H = \bigcup_{r \in \mathbb{N}} H_r$$

is a countable union of compact subsets. Thus, it suffices to prove separability of each $H_r$. Recall that $\bar{B}(x, R)$ denotes the closed $R$-ball in $X$ centered at the point $x$. For each $R \in \mathbb{R}_+$ define the map

$$\phi_R : H \to C_L(\bar{B}(x, R), X)$$

given by the restriction $h \mapsto h|\bar{B}(x, R)$. Here $C_L(\bar{B}(x, R), X)$ is the space of $L$-Lipschitz maps from $\bar{B}(x, R)$ to $X$. Observe that $C_L(\bar{B}(x, R), X)$ is metrizable *via*

$$\text{dist}(f, g) = \max_{y \in \bar{B}(x,R)} \text{dist}(f(y), g(y)).$$

Thus, the image of $H_r$ in each $C_L(\bar{B}(x, R), X)$ is a compact metrizable space. We claim now that each $\phi_R(H_r)$ is separable. Indeed, for each $i \in \mathbb{N}$ take $\mathcal{E}_i \subset \phi_R(H_r)$ to be an $\frac{1}{i}$–net. The union

$$\bigcup_{i \in \mathbb{N}} \mathcal{E}_i$$

is a dense countable subset of $\phi_R(H_r)$. On the other hand, the group $H$ (as a topological space) is homeomorphic to the *inverse limit*

$$\varprojlim_{R \in \mathbb{N}} \phi_R(H),$$

i.e., the subset of the product $\prod_i \phi_i(H)$ (given the product topology) which consists of sequences $(g_i)$ such that

$$\phi_j(g_i) = g_j, j \leqslant i.$$

Let $E_i \subset \phi_i(H_r)$ be a dense countable subset. For each element $h_i \in E_i$ consider a sequence $(g_j) = \tilde{h}_i$ in the above inverse limit such that $g_i = h_i$. Let $\tilde{h}_i \in H$ be the element corresponding to this sequence $(g_j)$. It is clear now that

$$\bigcup_{i \in \mathbb{N}} \{\tilde{h}_i \in H \; ; \; h_i \in E_i\}$$

is a dense countable subset of $H_r$. □

Corollary 14.12. *For each open subgroup $U \subset H$, the quotient $H/U$ is a countable set.*

Proof. Let $I \subset H$ be a dense countable set. The sets

$$hU, h \in H$$

are open subsets of $H$ so that $hU = gU$ or $hU \cap gU = \emptyset$ for all $g, h \in H$. The countable set $I$ intersects every $hU, h \in H$. Therefore, the above collection of open subsets of $H$ consists of countably many elements. □

Thus, we now know that the topological group $H$ is locally compact and separable. Therefore, by Theorem 14.8, $H$ contains an open and closed subgroup $\hat{H} \leqslant H$ satisfying Property A.

Lemma 14.13. *For every $x \in X$, the orbit $Y := \hat{H}x \subset X$ is open in $X$.*

Proof. If $Y$ is not open then it has empty interior (since $\hat{H}$ acts transitively on $Y$). Since $\hat{H} \subset H$ is closed, Arzela–Ascoli theorem implies that $Y$ is closed as well. Since $\hat{H}$ is open in $H$, by Corollary 14.12 the coset $S := H/\hat{H}$ is countable. Choose representatives $g_i$ of $S$, $i \in I$, where $I$ is countable. Then

$$\bigcup_{i\in I} g_i Y = X.$$

Therefore, the space $X$ is a countable union of closed subsets with empty interior. However, by Baire's theorem, each first category subset in the locally compact metric space $X$ has empty interior. Contradiction. □

We now can conclude the proof of Theorem 14.10. Let $Z \subset Y$ be the connected component of $Y := \hat{H}x$ as above, containing the point $x$. The stabilizer $F \subset \hat{H}$ of $Z$ is both closed and open in $\hat{H}$. Therefore, $F$ again has the Property A and the assumptions of Theorem 14.9 are satisfied by the action $F \curvearrowright Z$. It follows $F$ is a Lie group. Since $F \subset H$ is an open subgroup, the group $H$ is a Lie group as well. Let $K$ be the stabilizer of $x$ in $H$. The subgroup $K$ is a compact Lie group and, therefore, has only finitely many connected components. Since the action $H \curvearrowright X$ is transitive, $X$ is homeomorphic to $H/K$. Connectedness of $X$ now implies that $H$ has only finitely many connected components. □

### 14.2. Regular Growth Theorem

We now proceed to construct, for a group $\Gamma$ of weakly polynomial growth, a representation $\rho : \Gamma \to \mathrm{Isom}(X)$, where $X$ is a metric space as in Theorem 14.10, and $H = \mathrm{Isom}(X)$ acts transitively on $X$.

The first naive attempt would be to take $X$ to be a Cayley graph $\mathrm{Cayley}(\Gamma, S)$ of $\Gamma$. But in that case $\mathrm{Isom}(X)$ does not act transitively on $X$. If we replace the Cayley graph with its set of vertices on the other hand then we achieve homogeneity but loose connectedness. The ingenious idea of Gromov is to take $X$ to be a limit of rescaled Cayley graphs $(\mathrm{Cayley}(\Gamma, S), \lambda_n \mathrm{dist})$, where $\lambda_n$ is a sequence of positive numbers converging to 0. Gromov originally used Gromov-Hausdorff convergence to define the limit; we will take $X$ to be an asymptotic cone of $\mathrm{Cayley}(\Gamma, S)$ instead; equivalently $X$ is an asymptotic cone of $\Gamma$ with the word metric. Such an asymptotic cone inherits both the homogeneity from $\Gamma$ (see Proposition 7.58) and the property of being geodesic from $\mathrm{Cayley}(\Gamma, S)$ (see (3) in Proposition 7.54). In particular it is connected and locally connected. The asymptotic cone $X$ is also complete, by Proposition 7.56. These properties and the Hopf-Rinow Theorem 1.29 imply that in order to prove properness it suffices to prove local compactness.

To sum up, if we wish to apply Theorem 14.10 to an asymptotic cone, it remains to use the hypothesis of polynomial growth to find an asymptotic cone that is locally compact and finite dimensional. In what follows we explain how to choose a scaling sequence $\boldsymbol{\lambda}$ so that $\mathrm{Cone}_\omega(\Gamma, \mathbf{1}, \boldsymbol{\lambda})$ has both properties.

A metric space $X$ is called *p-doubling* if each $R$-ball in $X$ is covered by $p$ balls of radius $R/2$. One way to show that a metric space $X$ is doubling is to estimate its the *packing number* of $R$-balls in $X$. The *packing number* $p(B)$ of a ball $B = B(x, R) \subset X$ is the supremum of cardinalities of $R/2$-separated subsets $\mathcal{N}$ of $B$. If $\mathcal{N}$ is a maximal subset as above, then

$$\forall x \in B \; \exists y \in \mathcal{N} \text{ so that } d(x, y) < R/2.$$

(This condition is slightly stronger than the one of being an $R/2$-net.) In other words, the collection of open balls $\{B(x, R/2) : x \in \mathcal{N}\}$ is a covering of $B$. Thus, there exist a covering of $B$ by $p(B)$ balls of radius $R/2$. If $p(\bar{B}(x, R)) \leqslant p$ for every $x$ and $R$, then $X$ *has packing number* $\leqslant p$; such $X$ is necessarily $p$-doubling. The reader should compare this (trivial) statement with the statement of the Regular Growth Theorem below.

EXERCISE 14.14. Show that doubling implies polynomial growth for uniformly discrete spaces.

Note that, being scale-invariant, doubling property passes to asymptotic cones. The following lemma, although logically unnecessary for the proof of Gromov's theorem, motivates its arguments.

LEMMA 14.15. *If $X$ is $p$-doubling then the Hausdorff dimension of $X$ is at most $\log_2(p)$.*

PROOF. Consider a metric ball $B = B(o, R) \subset X$. We first cover $B$ by balls $B(x_i, R/2)$, $i = 1, \ldots, p$. We then cover each of the new balls by balls of radius $R/4$ and proceed inductively. On $n$-th step of induction we have a covering of $B$ by $p^n$ balls of radius $2^{-n}R$. The $n$-th sum of radii in the definition of the $\alpha$-Hausdorff measure of $B$ (1.9) then equals

$$\sum_{i=1}^{p^n} 2^{-n\alpha} R^\alpha = R^\alpha \left(\frac{p}{2^\alpha}\right)^n .$$

This quantity is converges to 0 as $n \to \infty$ provided that $p < 2^\alpha$, i.e., $\alpha > \log_2(p)$. Thus, $\mu_\alpha(B) = 0$ for every metric ball in $X$. Representing $X$ as a countable union of concentric metric balls, we conclude that $\mu_\alpha(X) = 0$ for every $\alpha > \log_2(p)$. $\square$

Thus, asymptotic cone of every doubling metric space has finite Hausdorff and, hence, covering, dimension.

Although there are spaces of polynomial growth which are not doubling, the Regular Growth Theorem below shows that groups of polynomial growth exhibit doubling-like behavior, which suffices for proving that the asymptotic cone is finite-dimensional.

Our discussion below follows the paper of Van den Dries and Wilkie, [**VdDW84**], Gromov's original statement and proof of the Regular Growth Theorem were different (although, some key arguments were quite similar).

THEOREM 14.16 (Regular growth theorem). *Let $\Gamma$ be a finitely generated group. Assume that there exists an infinitely large number $R = (R_n)^\omega$ in the ultrapower $\mathbb{R}_+^\omega$ such that the growth function satisfies:*

$$\mathfrak{G}_\Gamma(R_n) = \operatorname{card} B_\Gamma(1, R_n) \leqslant C R_n^a\,, \ \forall n \in \mathbb{N}\,, \tag{14.1}$$

*where $C > 0$ and $a \in \mathbb{N}$ are constants independent of $n$. Let $\epsilon > 0$.*

*Then there exists $\eta \in [\log R\,,\, R] \subset \mathbb{R}_+^\omega$ such that the ball $B\left(1, \frac{\eta}{4}\right)$ in the ultrapower $\Gamma^\omega$ endowed with the metric defined in (7.2) satisfies the following:*

*For every $i \in \mathbb{N}$, $i \geqslant 4$, all the sets of $\frac{\eta}{i}$-separated points in the ball $B\left(1, \frac{\eta}{4}\right)$ have cardinality at most $i^{a+\epsilon}$.*

*In particular, taking $i = 8$, we see that every $\frac{\eta}{4}$-ball in $\Gamma^\omega$ has packing number $\leqslant 8^{a+\epsilon}$ (with respect to the nonstandard metric).*

Recall that a metric space $M$ is $\varepsilon$-separated if for all distinct points $m_1, m_2 \in M$, $\mathrm{dist}(m_1, m_2) \geqslant \varepsilon$. For infinitely large numbers, see Definition 7.29 and Exercise 7.30. The difference between the assertion of this theorem and the statement that $\Gamma^\omega$ has finite packing number is that *we are not estimating packing numbers of all metric balls, but only of metric balls of certain radii.*

PROOF. Suppose to the contrary that for every $\eta \in [\log R\,,\, R] \subset \mathbb{R}^\omega_+$ there exists $i \in \mathbb{N}$, $i \geqslant 4$, such that the ball $B\left(1, \frac{\eta}{4}\right)$ contains at least $i^{a+\epsilon}$ points that are $\frac{\eta}{i}$-separated.

Then we define the function

$$\iota : [\log R, R] \to \mathbb{N}^\omega,\ \iota(\eta) \text{ is the smallest } i \in \mathbb{N} \text{ for which the above holds.}$$

The image of $\iota$ is contained in $\mathbb{N}$, identified to $\widehat{\mathbb{N}} \subset \mathbb{N}^\omega$.

It is easy to check that $\iota$ is an internal map defined by the sequence of maps: $\iota_n : [\log R_n\,,\, R_n] \to \mathbb{N}$, $\iota_n(r) =$ the minimal $i \in \mathbb{N}$, $i \geqslant 4$, such that $B_\Gamma\left(1\,,\, \frac{r}{4}\right)$ contains at least $i^{a+\epsilon}$ points that are $\frac{r}{i}$-separated.

The image of $\iota$ is therefore internal, and contained in $\widehat{\mathbb{N}} \subset \mathbb{N}^\omega$. According to Lemma 7.33, the image of $\iota$ has to be finite. Thus, there exists $K \in \mathbb{N}$ such that

$$\iota(\eta) \in [4, K], \quad \forall \eta \in [\log R, R].$$

This means that for every $\eta \in [\log R, R]$ there exists $i = \iota(\eta) \in \{4, \ldots, K\}$ such that the ball $B\left(1, \frac{\eta}{2}\right)$ contains at least $i^{a+\epsilon}$ disjoint balls of radii $\frac{\eta}{2i}$.

In particular: For $R$ there exists $i_1 = \iota(R) \in \{4, \ldots, K\}$ such that the ball $B\left(1, \frac{R}{2}\right) \subset \Gamma^\omega$ contains at least $i_1^{a+\epsilon}$ disjoint balls

$$B\left(x_1(1), \frac{R}{2i_1}\right), B\left(x_2(1), \frac{R}{2i_1}\right), \ldots, B\left(x_{t_1}(1), \frac{R}{2i_1}\right) \text{ with } t_1 \geqslant i_1^{a+\epsilon}.$$

Since $\Gamma^\omega$ is a group which acts on itself isometrically and transitively, all the balls in the list above are isometric to $B\left(1, \frac{R}{2i_1}\right)$.

EXERCISE 14.17. For every natural number $k$, $k\log(R) < R$.

Thus, $\frac{R}{i_1} \in [\log R\,,\, R]$; hence there exists $i_2 = \iota\left(\frac{R}{i_1}\right)$ such that the ball $B\left(1, \frac{R}{2i_1}\right)$ contains at least $i_2^{a+\epsilon}$ disjoint balls of radii $\frac{R}{2i_1 i_2}$.

It follows that $B\left(1, \frac{R}{2}\right)$ contains a family of disjoint balls

$$B\left(x_1(2), \frac{R}{2i_1i_2}\right), B\left(x_2(2), \frac{R}{2i_1i_2}\right), \ldots, B\left(x_{t_2}(2), \frac{R}{2i_1i_2}\right) \text{ with } t_2 \geqslant i_1^{a+\epsilon} i_2^{a+\epsilon}.$$

We continue *via* the nonstandard induction. Consider $u \in \mathbb{N}^\omega$ such that $B\left(1, \frac{R}{2}\right)$ contains a family of disjoint balls

$$B\left(x_1(u), \frac{R}{2i_1i_2\cdots i_u}\right), B\left(x_2(u), \frac{R}{2i_1i_2\cdots i_u}\right), \ldots, B\left(x_{t_u}(u), \frac{R}{2i_1i_2\cdots i_u}\right),$$

with $t_u \geqslant (i_1 i_2 \cdots i_u)^{a+\epsilon}$.

We construct the next generation of points

$$x_1(u+1), \ldots, x_{t_{u+1}}(u+1)$$

by considering, within each ball

$$B\left(x_2(u), \frac{R}{2i_1i_2\cdots i_u}\right)$$

the centers of $i_{u+1}^{a+\epsilon}$ disjoint balls of radii

$$\frac{R}{2i_1i_2\cdots i_ui_{u+1}},$$

where

$$i_{u+1} = \iota\left(\frac{R}{i_1i_2\cdots i_u}\right)$$

Here and below the product $i_1\cdots i_{u+1}$ is defined *via* the nonstandard induction as in the end of Section 7.3.

This induction process continues as long as $R/(i_1\cdots i_u) \geqslant \log R$. Recall that $i_j \geqslant 2$, hence

$$\frac{R}{i_1\cdots i_u} \leqslant 2^{-u}R.$$

Therefore, if $u > \log R - \log\log R$ then

$$\frac{R}{i_1\cdots i_u} < \log R.$$

Thus, there exists $u \in \mathbb{N}^\omega$ such that

$$\frac{R}{i_1i_2\cdots i_{u+1}} < \log R \leqslant \frac{R}{i_1i_2\cdots i_u} \leqslant \frac{KR}{i_1i_2\cdots i_{u+1}} \Leftrightarrow$$

$$\frac{R}{\log R} < i_1i_2\cdots i_{u+1} \leqslant \frac{KR}{\log R}\,. \tag{14.2}$$

Let's "count" the "number" (nonstandard of course!) of points $x_i(k)$ we constructed between the first step of the induction and the $u$-th step of the induction:

We get $i_1^{a+\epsilon}i_2^{a+\epsilon}\cdots i_{u+1}^{a+\epsilon}$ points; from (14.2) we get:

$$\left(\frac{R}{\log R}\right)^{a+\epsilon} \leqslant (i_1i_2\cdots i_u)^{a+\epsilon}\,.$$

What does this inequality actually mean? Recall that $R$ and $u$ are represented by sequences of real and natural numbers $R_n, u_n$ respectively. The above inequality thus implies that for $\omega$–all $n \in \mathbb{N}$, one has:

$$\left(\frac{R_n}{\log R_n}\right)^{a+\epsilon} \leqslant \operatorname{card} B(1, R_n)\,.$$

Since $\operatorname{card} B(e, R_n) \leqslant CR_n^a$, we get:

$$R_n^\epsilon \leqslant C(\log(R_n))^{a+\epsilon},$$

for $\omega$–all $n \in \mathbb{N}$. This contradicts the assumption on polynomial growth of $\Gamma$. $\square$

## 14.3. Corollaries of regular growth.

Proposition 14.18. *Let $\Gamma$ be a finitely generated group for which there exists an infinitely large number $R = (R_n)^\omega$ in the ultrapower $\mathbb{R}_+^\omega$ such that the growth function satisfies (14.1). Fix real numbers $a$ and $\epsilon > 0$ as in Theorem 14.16 and let $\eta = (\eta_n)$ be a sequence provided by the conclusion of Regular Growth Theorem 14.16; let $\lambda = (\lambda_n)$ with $\lambda_n = \frac{1}{\eta_n}$.*

*Then the asymptotic cone $X_\omega = \mathrm{Cone}_\omega(\Gamma; 1, \lambda)$ is*

(a) *locally compact;*

(b) *has Hausdorff dimension at most $a+\epsilon$. In particular, in view of Theorem 1.53, $X_\omega$ has finite covering dimension.*

Proof. (a) Since $X_\omega$ is homogeneous, it suffices to prove that the closed ball $C = \bar{B}\left(1, \frac{1}{4}\right) \subset X_\omega$ is compact. Since $C$ is complete, it suffices to show that it is *totally bounded*, i.e., for every $\delta > 0$ there exists a finite cover of $C$ by $\delta$-balls (see [**Nag85**]).

Let *dist* denote the word metric on $\Gamma$. By Theorem 14.16, the ball $B(1, \frac{1}{4}) \subset (\Gamma, \lambda_n dist)$ satisfies the property that for every integer $i \geqslant 4$, every $\frac{1}{i}$-separated subset $E \subset B(1, \frac{1}{4})$ contains at most $i^{a+\epsilon}$ points. The same assertion clearly holds for the ultralimit $X_\omega$. Therefore, we pick some $i \in \mathbb{N}$ such that $\frac{1}{i} < \delta$ and choose (by Zorn's lemma) a maximal $\frac{1}{i}$-separated subset $E \subset C$. Then, by maximality (see Lemma 1.36),

$$C \subset \bigcup_{x \in E} B(x, \frac{1}{i}) \subset \bigcup_{x \in E} B(x, \delta).$$

We, thus, have a finite cover of $C$ by $\delta$-balls and, therefore, $C$ is compact.

(b) We first verify that the Hausdorff dimension of the ball $B(1, 1/4)$ is at most $a+\epsilon$. Pick $\alpha > a+\epsilon$. For each $i$ consider a maximal $\frac{1}{i}$-separated set $x_{1\omega}, x_{2\omega}, \ldots, x_{t\omega}$ in $B(1, 1/4)$, with $t \leqslant i^{a+\epsilon}$.

Then $B(1, 1/4)$ is covered by the balls

$$B(x_{j\omega}, 1/i), j = 1, \ldots, t\,.$$

We get:

$$\sum_{j=1}^{t} (1/i)^\alpha \leqslant i^{a+\epsilon}/i^\alpha = i^{a+\epsilon-\alpha}.$$

Since $\alpha > a+\epsilon$, $\lim_{i\to\infty} i^{a+\epsilon-\alpha} = 0$. Hence $\mu_\alpha(B(1, 1/4)) = 0$.

Thus by homogeneity of $X_\omega$, $\dim_{Haus}(B(x, 1/4)) \leqslant a+\epsilon$ for each $x \in X_\omega$.

By (a) and Theorem 1.29 $X_\omega$ is proper, hence it is covered by countably many balls $B(x_n, 1/4)$, $n \in \mathbb{N}$. For every $\alpha > a+\epsilon$, additivity of $\mu_\alpha$ implies that

$$\mu_\alpha(X_\omega) \leqslant \sum_{n=1}^{\infty} \mu_\alpha\left(B(x_n, 1/4)\right) = 0.$$

Therefore $\dim_H(X_\omega) \leqslant a+\epsilon$. □

Exercise 14.19. 1. Use local compactness of $X_\omega$ to show that $G$ is doubling.

2. Suppose that $G$ is virtually nilpotent. Show that $G$ satisfies doubling condition using Theorem 12.48.

## 14.4. Weak polynomial growth

Here we prove several elementary properties of groups of weakly polynomial growth (Definition 14.4) that will be used in the next section.

LEMMA 14.20. *If $G$ has weak polynomial growth then for every normal subgroup $N \lhd G$, the quotient $G/N$ also has weak polynomial growth.*

PROOF. We equip $H = G/N$ with generating set which is the image of the finite generating set of $G$. Then $B_H(1, R)$ is the image of $B_G(1, G)$. Hence,

$$\operatorname{card} B_H(1, R) \leqslant \operatorname{card} B_G(1, R).$$

Thus, $H$ also has weak polynomial growth. □

LEMMA 14.21. *If $G$ has exponential growth then it cannot have weak polynomial growth.*

PROOF. Since $G$ has exponential growth,

$$\lim_{r\to\infty} \frac{1}{r} \log(\mathfrak{G}(r)) > 0.$$

Suppose that $G$ has weak polynomial growth. This means that growth function of $G$ satisfies

$$\mathfrak{G}(R_n) = \operatorname{card} B_G(1, R_n) \leqslant C R_n^a$$

for a certain sequence $(R_n)$ diverging to infinity and constants $C$ and $a$. Hence,

$$\frac{1}{R_n} \log(\mathfrak{G}(R_n)) \leqslant \frac{\log(C)}{R_n} + \frac{a}{R_n} \log(R_n).$$

However,

$$\lim_{R\to\infty} \left( \frac{\log(C)}{R} + \frac{a}{R} \log(R) \right) = 0.$$

Contradiction. □

LEMMA 14.22. *Let $\Gamma$ be a finitely generated subgroup of a Lie group $G$ with finitely many components. If $\Gamma$ has weakly polynomial growth then $\Gamma$ is virtually nilpotent.*

PROOF. According to Tits' alternative, either $\rho(\Gamma)$ contains a free nonabelian subgroup or is virtually solvable. In the former case, $\rho(\Gamma)$ cannot have weak polynomial growth (see Lemma 14.21). Thus $\rho(\Gamma)$ is virtually solvable. Similarly, by applying Theorem 12.58, since $\Gamma$ has weakly polynomial growth, $\Gamma$ has to be is virtually nilpotent. □

## 14.5. Displacement function

In this section we discuss certain metric properties of action of a finitely generated group $\Gamma$ on itself by left translations. These properties will be used to prove Gromov's theorem. We fix a finite generating set $S$ of $\Gamma$, Cayley graph $\operatorname{Cayley}(\Gamma, S)$ and the corresponding word metric on $\Gamma$.

We define certain *displacement functions* $\Delta$ for the action $\Gamma \curvearrowright \Gamma$ by left multiplication. For every $\gamma \in \Gamma$, $x \in \operatorname{Cayley}(\Gamma, S)$ and $r > 0$ we define the function measuring the maximal displacement by $\gamma$ on the ball $B(x, r) \subset \operatorname{Cayley}(\Gamma, S)$:

$$\Delta(\gamma, x, r) = \max\{\operatorname{dist}(y, \gamma y)\ ;\ y \in B(x, r)\}\,.$$

When $x = 1$ we use the notation $\Delta(\gamma, r)$ for the displacement function.

For a subset of $F \subset \Gamma$, define

$$\Delta(F, x, r) = \sup_{\gamma \in F} \Delta(\gamma, x, r) .$$

Likewise, we write $\Delta(F, r)$ when $x = 1$.

Clearly, for every $g \in \Gamma$,

$$\Delta(F, g, r) = \Delta(g^{-1} F g, r) .$$

LEMMA 14.23. *Fix $r > 0$ and a finite subset $F$ in $\Gamma$. Then the function* $\mathrm{Cayley}(\Gamma, S) \to \mathbb{R}$, $x \mapsto \Delta(F, x, r)$ *is 2–Lipschitz.*

PROOF. Let $x, y$ be two points in $\mathrm{Cayley}(\Gamma, S)$. Let $p$ be an arbitrary point in $B(x, r) \subset \mathrm{Cayley}(\Gamma, S)$. A geodesic in $\mathrm{Cayley}(\Gamma, S)$ connecting $p$ to $y$ has length at most $r + \mathrm{dist}(x, y)$, hence it contains a point $q \in B(y, r)$ with $\mathrm{dist}(p, q) \leqslant \mathrm{dist}(x, y)$. For an arbitrary $\gamma \in F$,

$$\mathrm{dist}(p, \gamma p) \leqslant \mathrm{dist}(q, \gamma q) + 2\mathrm{dist}(x, y) \leqslant \Delta(F, y, r) + 2\mathrm{dist}(x, y) .$$

It follows that $\Delta(F, x, r) \leqslant \Delta(F, y, r) + 2\mathrm{dist}(x, y)$. The inequality $\Delta(F, y, r) \leqslant \Delta(F, x, r) + 2\mathrm{dist}(x, y)$ is proved similarly. $\square$

LEMMA 14.24. *Suppose that $\Delta(S, r)$ is bounded as a function of $r$. Then $\Gamma$ is virtually abelian.*

PROOF. Suppose that $\mathrm{dist}(sx, x) \leqslant C$ for all $x \in \Gamma$ and $s \in S$. Then

$$\mathrm{dist}(x^{-1} s x, 1) \leqslant C,$$

and, therefore, the conjugacy class of $s$ in $\Gamma$ has cardinality $\leqslant \mathfrak{G}_\Gamma(C) = N$. We claim that the centralizer $Z_\Gamma(s)$ of $s$ in $\Gamma$ has finite index in $\Gamma$: Indeed, if $x_0, \ldots, x_N \in \Gamma$ then there are $i, k$, $0 \leqslant i \neq k \leqslant N$, such that

$$x_i^{-1} s x_i = x_k^{-1} s x_k \Rightarrow [x_k x_i^{-1}, s] = 1 \Rightarrow x_k x_i^{-1} \in Z_\Gamma(s).$$

Thus, the intersection

$$A := \bigcap_{s \in S} Z_\Gamma(s)$$

has finite index in $\Gamma$. Therefore, $A$ is an abelian subgroup of finite index in $\Gamma$. $\square$

## 14.6. Proof of Gromov's theorem

In this section we prove Theorem 14.5 and, hence, Theorem 14.1 as well.

Let $\Gamma$ be a group satisfying the assumptions of Theorem 14.5 and $a, \epsilon, R \in \mathbb{R}^*, \eta \in \mathbb{R}^*$ be the quantities appearing in Theorem 14.5. In what follows we fix a finite generating set $S$ of $\Gamma$ and the corresponding Cayley graph $\mathrm{Cayley}(\Gamma, S)$.

Suppose that $\Gamma$ has weakly polynomial growth with respect to a sequence $(R_n)$ diverging to infinity. Take the diverging sequence $(\eta_n)$ given by the Regular Growth Theorem applied to the group $\Gamma$. Let $\lambda = (\lambda_n)$ with $\lambda_n = \frac{1}{\eta_n}$. Construct the asymptotic cone $X_\omega = \mathrm{Cone}_\omega(\Gamma; 1, \lambda)$ of the Cayley graph of $\Gamma$ *via* rescaling by the sequence $\lambda_n$ and considering the sequence $e$ of base-points constant equal to the identity in $\Gamma$. By Proposition 14.18, the metric space $X_\omega$ is connected, locally connected, finite-dimensional and proper.

According to Proposition 7.58, we have a homomorphism

$$\alpha : \Gamma_e^\omega \to L := \mathrm{Isom}(X_\omega)$$

such that $\alpha(\Gamma_e^\omega)$ acts on $X_\omega$ transitively. We also get a homomorphism

$$\rho : \Gamma \to L, \rho = \iota \circ \alpha,$$

where $\iota : \Gamma \hookrightarrow \Gamma_e^\omega$ is the diagonal embedding $\iota(\gamma) = (\gamma)^\omega$. Since the isometric action $L \curvearrowright X_\omega$ is effective and transitive, according to Theorem 14.10, the group $L$ is a Lie group with finitely many components.

REMARK 14.25. Observe that the point-stabilizer $L_y$ for $y \in X_\omega$ is a compact subgroup in $L$. Therefore $X_\omega = L/L_x$ can be given a left-invariant Riemannian metric $ds^2$. Hence, since $X_\omega$ is connected, by using the exponential map with respect to $ds^2$ we see that if $g \in L$ fixes an open ball in $X_\omega$ pointwise, then $g = id$.

The subgroup $\rho(\Gamma) \leqslant L$ has weak polynomial growth because $\Gamma$ has weak polynomial growth (see Lemma 14.20). By Lemma 14.22, $\rho(\Gamma)$ is virtually nilpotent.

The main problem is that $\rho$ may have a large kernel. Indeed, if $\Gamma$ is abelian then the homomorphism $\rho$ is actually trivial. An induction on the degree $d$ of weak polynomial growth allows to get around this problem and prove Gromov's Theorem. In the induction step, we shall use $\rho$ to construct an epimorphism $\Gamma \to \mathbb{Z}$, and then apply Proposition 12.60.

If $d = 0$, then $\mathfrak{G}_\Gamma(R_n)$ is bounded. Since growth function is monotonic, it follows that $\Gamma$ is finite and there is nothing to prove.

Suppose that each group $\Gamma$ of weak polynomial growth of degree $\leqslant d - 1$ is virtually nilpotent. Let $\Gamma$ be a group of weak polynomial growth of degree $\leqslant d$, i.e.,

$$\mathfrak{G}_\Gamma(R_n) \leqslant C_\Gamma R_n^d,$$

for some sequence $R_n$ diverging to infinity. There are two cases to consider:

(a) The image of the homomorphism $\rho$ above is infinite. Then, after passing to a finite index subgroup in $\Gamma$ (we retain the notation $\Gamma$ for the subgroup), we get a homomorphism from $\Gamma$ to a torsion-free infinite nilpotent group. The latter has infinite abelianization, hence, we get an epimorphism $\phi : \Gamma \to \mathbb{Z}$. If $K = \mathrm{Ker}(\phi)$ is not finitely generated, then $\Gamma$ has exponential growth (see Proposition 12.60), which is a contradiction. Therefore, $K$ is finitely generated. Repeating arguments in the proof of Proposition 12.60 verbatim we see that $K$ has weakly polynomial growth of degree $\leqslant d - 1$. Thus, by the induction hypothesis, $K$ is a virtually nilpotent group. Therefore, $\Gamma$ is solvable. Applying Lemma 14.21, we conclude that $\Gamma$ is virtually nilpotent as well.

(b) $\rho(\Gamma)$ is finite.

First we note that we can reduce to the case when $\rho(\Gamma) = \{1\}$. Indeed, consider the subgroup of finite index $\Gamma' := \mathrm{Ker}(\rho) \subset \Gamma$. For every $\gamma \in \Gamma'$, we have that $\mathrm{dist}_\Gamma(x_n, \gamma x_n) = o(\eta_n)$, for every sequence $(x_n) \in \Gamma^\mathbb{N}$ with $\mathrm{dist}_\Gamma(1, x_n) = O(\eta_n)$. Since $\Gamma'$ is quasi-isometric to $\Gamma$, the same is true for sequences $(x_n)$ in $\Gamma'$ and $\mathrm{dist}_{\Gamma'}$. Thus, $\Gamma'$ acts trivially on its own asymptotic cone $\mathrm{Cone}_\omega(\Gamma'; 1, \lambda)$, and it clearly suffices to prove that $\Gamma'$ is virtually nilpotent.

Hence, from now on we assume that $\rho(\Gamma) = \{1\}$, equivalently that $\mathrm{Ker}\, \rho = \Gamma$. What is the metric significance of this condition?

EXERCISE 14.26. Let $\Delta$ denote the displacement function for the action of $\Gamma$ on itself *via* left multiplication introduced in Section 14.5. Show that the condition

$\operatorname{Ker}\rho = \Gamma$ is equivalent to the fact that

$$\omega\text{-lim}\,\frac{\Delta(S, R\eta_n)}{\eta_n} = 0\,, \quad \text{for every } R > 0\,. \tag{14.3}$$

In other words, all generators of $\Gamma$ act on $\Gamma$ with *sublinear* (with respect to $(\eta_n)$) displacement.

If the function $\Delta(S, r)$ were bounded then $\Gamma$ would be virtually abelian (Lemma 14.24), which would conclude the proof. Thus, we assume that $\Delta(S, r)$ diverges to infinity as $r \to \infty$.

LEMMA 14.27. *For every $\epsilon \in (0, 1]$ there exists a sequence $(x_n)$ in $\Gamma$ such that*

$$\omega\text{-}lim\,\frac{\Delta(x_n^{-1} S x_n, \eta_n)}{\eta_n} = \epsilon.$$

PROOF. By (14.3), for $\omega$–all $n \in \mathbb{N}$ we have $\Delta(S, \eta_n) \leqslant \epsilon\eta_n/2$. Thus, for $\omega$-all $n$, there exists $p_n \in \Gamma$ so that $\Delta(S, p_n, \eta_n) \leqslant \eta_n/2$. Fix $n$ in the above set with $\omega$-measure 1. Since the function $\Delta(S, r)$ diverges to infinity, there exists $q_n \in \Gamma$ such that

$$\Delta(S, q_n, \eta_n) \geqslant \max_{s \in S} \operatorname{dist}(q_n, s q_n) > 2\eta_n.$$

The Cayley graph $\mathrm{Cayley}(\Gamma, S)$ is connected and the function $\mathrm{Cayley}(\Gamma, S) \to \mathbb{R}$, $p \mapsto \Delta(S, p, \eta_n)$ is continuous by Lemma 14.23. Hence, for $\omega$-all $n$, there exists $y_n \in \mathrm{Cayley}(\Gamma, S)$ such that

$$\Delta(S, y_n, \eta_n) = \epsilon\eta_n.$$

The point $y_n$ is not necessarily in the vertex set of the Cayley graph $\mathrm{Cayley}(\Gamma, S)$. Pick a point $x_n \in \Gamma$ within the distance $\frac{1}{2}$ from $y_n$. Again by Lemma 14.23

$$|\Delta(S, x_n, \eta_n) - \epsilon\eta_n| \leqslant 1.$$

It follows that $|\Delta(x_n^{-1} S x_n, \eta_n) - \epsilon\eta_n| \leqslant 1$ and, therefore,

$$\omega\text{-lim}\,\frac{\Delta(x_n^{-1} S x_n, \eta_n)}{\eta_n} = \epsilon. \quad \square$$

For every $0 < \epsilon \leqslant 1$ we consider a sequence $(x_n)$ as in Lemma 14.27 and define the homomorphism

$$\rho_\epsilon : \Gamma \to \Gamma_e^\omega,\ \rho_\epsilon(g) = \left(x_n^{-1} g x_n\right)^\omega \in \Gamma^\omega.$$

Note that since $\Delta(x_n^{-1} S x_n, \eta_n) = O(\epsilon\eta_n)$, the elements $\rho_\epsilon(g)$ belong to $\Gamma_e^\omega$.

Clearly, the image of $\rho_\epsilon$ is non-trivial. If for some $\epsilon > 0$, $\rho_\epsilon(\Gamma)$ is infinite we are done as in (a). Hence we assume that $\rho_\epsilon(\Gamma)$ is finite for all $\epsilon \in (0, 1]$.

Next, we reduce the problem to the case when all groups $\rho_\epsilon(\Gamma)$ are finite abelian. According to Jordan's theorem 10.66, there exists $q = q(L)$ so that each finite group $\rho_\epsilon(\Gamma)$ contains an abelian subgroup of index $\leqslant q$. For each $\epsilon$ consider the preimage $\Gamma_\epsilon$ in $\Gamma$ of the abelian subgroup in $\rho_\epsilon(\Gamma)$ which is given by Jordan's theorem applied to $L = \Gamma_e^\omega$. The index of $\Gamma_\epsilon$ in $\Gamma$ is at most $q$. Let $\Gamma'$ be the intersection of all the subgroups $\Gamma_\epsilon, \epsilon > 0$. Then, $\Gamma'$ has finite index in $\Gamma$ and $\rho_\epsilon(\Gamma')$ is finite abelian.

The topology of the group $L$ is the compact-open topology with respect to its action on $X_\omega$, thus an $\epsilon$–neighborhood of the identity in $L$ contains all isometries $h \in L$ such that

$$\Delta(h, 1) \leqslant \epsilon,$$

where $\Delta(h,1)$ is the displacement of $h$ on the unit ball $B(e_\omega,1) \subset X_\omega$. By our choice of $x_n$, for every generator $s \in S$ of $\Gamma$, $\Delta(\rho_\epsilon(s),1) \leqslant \epsilon$, and for one of the generators the inequality becomes an equality. After multiplying $\epsilon$ if necessary by some factor depending on $\Gamma'$, we may assume that the same is true for a set $S'$ of generators of $\Gamma'$.

Assume there exists an $M \in \mathbb{N}$ such that the order $|\rho_\epsilon(\Gamma')|$ is at most $M$ for all $\epsilon$ (or for all $\epsilon_i$ in a sequence $(\epsilon_i)$ converging to 0). The above implies that for every $g \in \Gamma'$, $\delta(\rho_\epsilon(g),1) \leqslant M\epsilon$. It follows that $L$ contains arbitrarily small finite cyclic subgroups, which is impossible since $L$ is a Lie group. Therefore,

$$\lim_{\epsilon \to 0} |\rho_\epsilon(\Gamma')| = \infty\,.$$

Then $\Gamma'$ admits epimorphisms to finite abelian groups of arbitrarily large order. Since all such homomorphisms have to factor through the abelianization $(\Gamma')_{ab}$, the group $(\Gamma')^{ab}$ has to be infinite. Since $(\Gamma')_{ab}$ is finitely generated it follows that it has nontrivial free factor, hence $\Gamma'$ again admits an epimorphism to $\mathbb{Z}$. We apply Proposition 12.60 and the induction hypothesis, and conclude that $\Gamma'$ is virtually nilpotent. Thus $\Gamma$ is also virtually nilpotent, and we are done. This concludes the proof of Theorem 14.5 and, hence, of Theorem 14.1. □

The following version of Gromov's theorem was proved by F. Point:

THEOREM 14.28 ([**Poi95**]). *If* $(\Gamma, \mathrm{dist})$ *is a finitely generated group with a word metric such that for a fixed sequence* $\epsilon_n \to 0$ *the cone* $\mathrm{Cone}_\omega(\Gamma,(1),(\epsilon_n))$ *is proper and has finite Minkowski dimension for every ultrafilter* $\omega$*, then* $\Gamma$ *is virtually nilpotent.*

Below we review some properties of asymptotic cones of nilpotent groups.

Let $(\Gamma, \mathrm{dist})$ be a finitely generated nilpotent group endowed with a word metric, let $\mathrm{Tor}\,(\Gamma)$ be the torsion subgroup of $\Gamma$ and let $H$ be the torsion-free nilpotent group $\Gamma/\mathrm{Tor}\,(\Gamma)$.

THEOREM 14.29 (A. I. Mal'cev [**Mal49b**]). *Every finitely generated torsion-free nilpotent group* $H$ *is isomorphic to a uniform lattice in a connected nilpotent Lie group* $N$.

With every $k$-step nilpotent Lie group $N$ with Lie algebra $\mathfrak{n}$ one associates the *graded Lie algebra* $\bar{\mathfrak{n}}$ obtained as the direct sum

$$\oplus_{i=1}^{k} \mathfrak{c}^i\mathfrak{n}/\mathfrak{c}^{i+1}\mathfrak{n},$$

where $\mathfrak{c}^i\mathfrak{n}$ is the Lie algebra of $C^iN$. Every finite-dimensional Lie algebra is the Lie algebra of a connected Lie group; thus, consider the connected nilpotent Lie group $\overline{N}$ with the Lie algebra $\bar{\mathfrak{n}}$. The group $\overline{N}$ is called the *graded Lie group* of the group $\Gamma$ and of the Lie group $N$. We refer to Pansu's paper [**Pan83**] for the definition of Carnot-Caratheodory metric appearing in the following theorem:

THEOREM 14.30 (P. Pansu, [**Pan83**]). **(a)** *All the asymptotic cones of the finitely generated nilpotent group* $\Gamma$ *are isometric to the graded Lie group* $\overline{N}$ *endowed with a* Carnot-Caratheodory metric $\mathrm{dist}_{CC}$. *In particular, the Lie group* $\overline{N}$ *(treated as a topological space[2]) is a quasi-isometry invariant of* $\Gamma$.

[2]Actually, $\overline{N}$ treated as a Lie group is also a quasi-isometry invariant of $\Gamma$.

**(b)** *For every sequence* $\varepsilon_j > 0$ *converging to* 0 *and every word metric* dist *on* $\Gamma$*, the sequence of metric spaces* $(\Gamma, \varepsilon_j \cdot \mathrm{dist})$ *converges in the modified Hausdorff metric to* $(\overline{N}, \mathrm{dist}_{CC})$.

**(c)** *The sub-bundle in* $\overline{N}$ *defining the Carnot-Caratheodory metric is independent of the word metric on* $\Gamma$*, only the norm on this subbundle depends on the word metric.*

**(d)** *The dimension of* $\overline{N}$ *equals the cohomological dimension of* $\Gamma$*, which, in turn, equals*

$$\dim(\Gamma) = \sum_{i=1}^{k} m_i,$$

*where* $m_i$ *is the rank of the abelian quotient* $C^i\Gamma/C^{i+1}\Gamma$.

**e)** *The Hausdorff dimension of* $(\overline{N}, \mathrm{dist}_{CC})$ *equals to the degree of polynomial growth of* $\Gamma$*, that is to*

$$d(\Gamma) = \sum_{i=1}^{k} i\ m_i$$

Note that, according to Theorem 7.60, (a) implies (b) in Pansu's theorem.

Remark 14.31. One says that two metric spaces are *asymptotically bi-Lipschitz* if their asymptotic cones are bi-Lipschitz homeomorphic. Pansu's theorem above allows one to construct an example of two asymptotically bi-Lipschitz nilpotent groups, which are not quasi-isometric. Indeed, by Pansu's theorem, every nilpotent Lie group $\Gamma$ is asymptotically bi-Lipschitz to its associated graded Lie group $\overline{N}$ On the other hand, in [**Sha04**, p. 151-152] Y. Shalom gives an example, which he attributes to Y. Benoist, of a nilpotent Lie group not quasi-isometric to the corresponding associated graded Lie group.

### 14.7. Quasi-isometric rigidity of nilpotent and abelian groups

Theorem 14.32 (M. Gromov). *Suppose that* $\Gamma_1, \Gamma_2$ *are quasi-isometric finitely generated groups and* $\Gamma_1$ *is nilpotent. Then* $\Gamma_2$ *is virtually nilpotent.*

Proof. Being nilpotent, $\Gamma_1$ has polynomial growth of degree $d$. Since growth is quasi-isometric invariant, $\Gamma_2$ also has polynomial growth of degree $d$. By Theorem 14.1, $\Gamma_2$ is virtually nilpotent. ☐

Theorem 14.33 (P. Pansu). *Suppose that* $\Gamma_1, \Gamma_2$ *are quasi-isometric finitely generated groups and* $\Gamma_1$ *is abelian. Then* $\Gamma_2$ *is virtually abelian.*

Proof. Let $d$ denote the rank of $\Gamma_1$. Then $\mathfrak{G}_{\Gamma_1}(t) \asymp t^d$. Furthermore, $d$ is the rational cohomological dimension of $\Gamma_1$. Then $\Gamma_2$ also growth $\asymp t^d$. As we observed above, $\Gamma_2$ is virtually nilpotent. Let $\Gamma := \Gamma_2/\mathrm{Tor}\,\Gamma_2$. By Bass–Guivarc'h Theorem (Theorem 12.48),

$$d = \sum_{i=1}^{k} i\ m_i,$$

where $m_i$ is the rank of $C^i\Gamma/C^{i+1}\Gamma$. By part (d) of Pansu's theorem, rational cohomological dimension is a quasi-isometry invariant of a finitely generated nilpotent group. (Actually, R. Sauer later proved in [**Sau06**] that rational cohomological dimension is a quasi-isometry invariant of *every* finitely generated group.) Therefore,

$$d = dim(\Gamma) = \sum_{i=1}^{k} m_i,$$

and

$$\sum_{i=1}^{k} i\ m_i = \sum_{i=1}^{k} m_i.$$

The latter implies that $k = 1$, i.e., $\Gamma$ is abelian. □

CHAPTER 15

# The Banach-Tarski paradox

## 15.1. Paradoxical decompositions

DEFINITION 15.1. Two subsets $A, B$ in a metric space $(X, \mathrm{dist})$ are *congruent* if there exists an isometry $\phi : X \to X$ such that $\phi(A) = B$.

DEFINITION 15.2. Two sets $A, B$ in a metric space $X$ are *piecewise congruent* (or *equidecomposable*) if, for some $k \in \mathbb{N}$, they admit partitions $A = A_1 \sqcup ... \sqcup A_k$, $B = B_1 \sqcup ... \sqcup B_k$ such that for each $i \in \{1, ..., k\}$, the sets $A_i$ and $B_i$ are congruent.

Two subsets $A, B$ in a metric space $X$ are *countably piecewise congruent* (or *countably equidecomposable*) if they admit partitions $A = \bigsqcup_{n\in\mathbb{N}} A_n$, $B = \bigsqcup_{n\in\mathbb{N}} B_n$ such that for every $n \in \mathbb{N}$, the sets $A_n$ and $B_n$ are congruent.

REMARK 15.3. Thus, by using empty sets for some $A_n, B_n$, we see that piecewise congruence as a stronger form of countably piecewise congruence.

EXERCISE 15.4. Prove that (countably) piecewise congruence is an equivalence relation.

DEFINITION 15.5. A set $E$ in a metric space $X$ is *paradoxical* if there exists a partition

$$E = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

and isometries $\varphi_1, ..., \varphi_k, \psi_1, ..., \psi_m$ of $X$, so that

$$\varphi_1(X_1) \sqcup ... \sqcup \varphi_k(X_k) = E$$

and

$$\psi_1(Y_1) \sqcup ... \sqcup \psi_m(Y_m) = E\ .$$

A set $E$ in a metric space $X$ is *countably paradoxical* if there exists a partition

$$E = \bigsqcup_{n\in\mathbb{N}} X_n \sqcup \bigsqcup_{m\in\mathbb{N}} Y_m$$

and two sequences of isometries $(\varphi_n)_{n\in\mathbb{N}}, (\psi_m)_{m\in\mathbb{N}}$ of $X$, so that

$$\bigsqcup_{n\in\mathbb{N}} \varphi_n(X_n) = E, \text{ and } \bigsqcup_{m\in\mathbb{N}} \psi_m(Y_m) = E.$$

EXERCISE 15.6. 1. If $E, E' \subset X$ are piecewise-congruent and $E$ is paradoxical, then so is $E'$.

2. If $E, E' \subset X$ are countably piecewise-congruent and $E$ is countably paradoxical, then so is $E'$.

Using earlier work of Vitali and Hausdorff, Banach and Tarski proved the following:

Theorem 15.7 (Banach-Tarski paradox [**BT24**]). (1) *Any two bounded subsets with non-empty interior in* $\mathbb{R}^n$ *(for* $n \gg 3$*) are piecewise congruent.*

(2) *Any two bounded subsets with non-empty interior in* $\mathbb{R}^n$ *(for* $n \in \{1, 2\}$*) are countably piecewise congruent.*

Corollary 15.8. (1) *Every Euclidean ball is paradoxical in* $\mathbb{R}^n, n \geqslant 3$*, and countably paradoxical in* $\mathbb{R}^n, n \in \{1, 2\}$*.*

(2) *For every* $n \geqslant 3$ *and every* $m \in \mathbb{N}$*, every ball in* $\mathbb{R}^n$ *is piecewise congruent to* $m$ *copies of this ball (one can "double" the ball).*

(3) *A pea and the sun are piecewise congruent (any two Euclidean* $n$*-balls are piecewise-congruent for* $n \geqslant 3$*).*

Remark 15.9. The Banach-Tarski paradox emphasizes that it is impossible to find a finitely-additive measure defined on *all* subsets of the Euclidean space of dimension at least 3 that is invariant with respect to isometries and takes the value one on a unit cube. The main point in their theorem is that the congruent pieces $A_i, B_i$ are not *Lebesgue measurable.*

Remark 15.10 (Banach-Tarski paradox and axiom of choice). The Banach-Tarski paradox is neither provable nor disprovable with Zermelo-Fraenkel axioms (ZF) only: It is impossible to prove that a unit ball in $\mathbb{R}^3$ is paradoxical in ZF, it is also impossible to prove it is not paradoxical. An extra axiom is needed, e.g., the axiom of choice (AC). In fact, work of M. Foreman & F. Wehrung [**FW91**] and J. Pawlikowski [**Paw91**] shows that the Banach–Tarski paradox can be proved assuming ZF and the Hahn–Banach theorem (which is a weaker axiom than AC, see Section 7.1).

## 15.2. Step 1 of the proof of the Banach–Tarski theorem

We will prove only Corollary 15.8, Parts 1 and 2 and only in the case $n \leqslant 3$. The general statement of Theorem 15.7 for two bounded subset with non-empty interiors is derived from the doubling of a ball by using the Banach–Bernstein-Schroeder theorem (see [**Wag85**]). The general statement in $\mathbb{R}^n, n \geqslant 3$, can be easily either derived from the statement for $n = 3$, or proved directly by adapting the proof in dimension 3.

The first step in the proof is common to all dimensions.

**Step 1: The unit sphere $\mathbb{S}^n$ is piecewise congruent to $\mathbb{S}^n \setminus C$, where $C$ is any countable set, and $n = 1, 2$.**

We first prove that there exists a rotation $\rho$ around the origin such that for any integer $n \geqslant 1$, $\rho^n(C) \cap C = \emptyset$. This is obvious in the plane (only a countable set of rotations do not satisfy this).

In the space we first select a line $\ell$ through the origin such that its intersection with $\mathbb{S}^2$ is not in $C$. Such a line exists because the set of lines through the origin containing points in $C$ is countable. Then we look for a rotation $\rho_\theta$ of angle $\theta$ around $\ell$ such that for any integer $n \geqslant 1$, $\rho_\theta^n(C) \cap C = \emptyset$. Indeed take $A$ the set of angles $\alpha$ such that the rotation of angle $\alpha$ around $\ell$ sends a point in $C$ to another point in $C$. There are countably many such angles, therefore the set $A' = \bigcup_{n \geqslant 1} \frac{1}{n} A$ is also countable. Thus, we may choose an angle $\theta \notin A'$.

Take $\mathcal{O} = \bigsqcup_{n \geqslant 0} \rho_\theta^n(C)$ and decompose $\mathbb{S}^2$ as $\mathbb{S}^2 = \mathcal{O} \sqcup (\mathbb{S}^2 \setminus \mathcal{O})$. Then $(\mathcal{O} \setminus C) \sqcup (\mathbb{S}^2 \setminus \mathcal{O}) = \mathbb{S}^2 \setminus C$. We, thus, have a piecewise congruence of $\mathbb{S}^2$ to $\mathbb{S}^2 \setminus C$ which sends $\mathcal{O}$ to $\mathcal{O} \setminus C$ by $\rho_\theta$ and is the identity on $\mathbb{S}^2 \setminus \mathcal{O}$.

## 15.3. Proof of the Banach–Tarski theorem in the plane

**Step 2 (using the axiom of choice): The unit circle $\mathbb{S}^1$ is countably paradoxical.**

Let $\alpha$ be an irrational number and let $R \in SO(2)$ be the counter-clockwise rotation of angle $2\pi\alpha$. Then the map $m \mapsto R^m$ is an injective homomorphism $\mathbb{Z} \to SO(2)$. *Via* this homomorphism, $\mathbb{Z}$ acts on the unit circle $\mathbb{S}^1$. *According to the axiom of choice* there exists a subset $D \subset \mathbb{S}^1$ which intersects every $\mathbb{Z}$-orbit in exactly one point.

Since $\mathbb{Z}$ decomposes as $2\mathbb{Z} \sqcup (2\mathbb{Z}+1)$, the unit circle decomposes as

$$2\mathbb{Z} \cdot D \sqcup (2\mathbb{Z}+1) \cdot D.$$

Now, for each $X_n = R^{2n} \cdot D$ consider the isometry $\varphi_n = R^{-n}$, and for each $Y_n = R^{2n+1} \cdot D$ consider the isometry $\psi_n = R^{-n-1}$. Clearly $\mathbb{S}^1 = \bigsqcup_{n \in \mathbb{Z}} \varphi_n(X_n)$ and $\mathbb{S}^1 = \bigsqcup_{n \in \mathbb{Z}} \psi_n(Y_n)$.

**Step 3: The unit disk $\mathbb{D}^2$ is countably paradoxical.**

Let $\mathbb{D}^2$ be the closed unit disk in $\mathbb{R}^2$ centered at a point $O$. Step 1 and the fact that $\mathbb{D}^2 \setminus \{O\}$ can be written as the set

$$\{\lambda x \;;\; \lambda \in (0,1]\,,\, x \in \mathbb{S}^1\},$$

imply that $\mathbb{D}^2 \setminus \{O\}$ is countably paradoxical. Thus, it suffices to prove that $\mathbb{D}^2 \setminus \{O\}$ is piecewise congruent to $\mathbb{D}^2$. Take $\mathbb{S}^1\left(\left(\frac{1}{2}, 0\right), \frac{1}{2}\right)$, the unit circle with center $\left(\frac{1}{2}, 0\right)$ and radius $\frac{1}{2}$. For simplicity, we denote this circle $\mathbb{S}_{1/2}$. Then

$$\mathbb{D}^2 \setminus \{O\} = \mathbb{D}^2 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} \setminus \{O\}.$$

According to Step 1, $\mathbb{S}_{1/2} \setminus \{O\}$ is piecewise congruent to $\mathbb{S}_{1/2}$, hence $\mathbb{D}^2 \setminus \{O\}$ is piecewise congruent to

$$\mathbb{D}^2 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} = \mathbb{D}^2. \quad \square$$

Remark 15.11 (Stronger result). Instead of the splitting $\mathbb{Z} = 2\mathbb{Z} \sqcup (2\mathbb{Z}+1)$ of $\mathbb{Z}$ into two 'copies' of itself, we might consider a splitting of $\mathbb{Z}$ into infinitely countably many 'copies' of itself. Indeed the subsets $\mathbb{Z}^{(k)} = 2^k\mathbb{Z} + 2^{k-1}$, $k \in \mathbb{N}$, form a partition of $\mathbb{Z}$. This allows to prove, following the same proof as above, that a unit disk is countably piecewise congruent to countably many copies of itself.

Proof. As in Step 2, we write $\mathbb{S}^1 = \mathbb{Z}D = \bigsqcup_{k \in \mathbb{N}} \mathbb{Z}^{(k)} D$. The idea is to move by isometries the copies of $D$ in $\mathbb{Z}^{(k)} D$ so as to form the $k$-th copy of the unit circle. Indeed, if for the set $X_{k,m} = R^{2^k m + 2^{k-1}} D$ we consider the isometry

$$\phi_{k,m} = T_{(2k,0)} \circ R^{-2^k m - 2^{k-1} + m},$$

then

$$\bigsqcup_{m \in \mathbb{Z}} \phi_{k,m}(X_{k,m})$$

is equal to $\mathbb{S}^1((2k, 0), 1)$.

Thus, $\mathbb{S}^1$ is countably piecewise congruent to

$$\bigsqcup_{k\in\mathbb{N}} \mathbb{S}^1((2k,0),1).$$

This extends to the corresponding disks with their centers removed. In Step 3 we proved that a punctured disk is piecewise congruent to the full disk. This allows to finish the argument. $\square$

### 15.4. Proof of the Banach–Tarski theorem in the space

We now explain prove Banach–Tarski theorem for $A$, the unit ball in $\mathbb{R}^3$ and $B$, the disjoint union of two unit balls in $\mathbb{R}^3$.

**Step 2: a paradoxical decomposition for the free group of rank 2.**

Let $F_2$ be the free group of rank 2 with generators $a, b$. Given $u$, a reduced word in $a, b, a^{-1}, b^{-1}$, we denote by $\mathcal{W}_u$ the set of reduced words in $a, b, a^{-1}, b^{-1}$ with the prefix $u$. Every $x \in F_2$ defines a map $L_x : F_2 \to F_2$, $L_x(y) = xy$ (left translation by $x$).

Then

$$F_2 = \{1\} \sqcup \mathcal{W}_a \sqcup \mathcal{W}_{a^{-1}} \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}} \tag{15.1}$$

but also $F_2 = L_a\mathcal{W}_{a^{-1}} \sqcup \mathcal{W}_a$ , and $F_2 = L_b\mathcal{W}_{b^{-1}} \sqcup \mathcal{W}_b$. We slightly modify the above partition in order to include $\{1\}$ into one of the other four subsets. Consider the following modifications of $\mathcal{W}_a$ and $\mathcal{W}_{a^{-1}}$:

$$\mathcal{W}'_a = \mathcal{W}_a \setminus \{a^n \,;\, n \geqslant 1\} \text{ and } \mathcal{W}'_{a^{-1}} = \mathcal{W}_{a^{-1}} \sqcup \{a^n \,;\, n \geqslant 0\}\,.$$

Then

$$F_2 = \mathcal{W}'_a \sqcup \mathcal{W}'_{a^{-1}} \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}} \tag{15.2}$$

and

$$F_2 = L_a\mathcal{W}'_{a^{-1}} \cup \mathcal{W}'_a\,.$$

**Step 3: A paradoxical decomposition for the unit sphere (using the axiom of choice).**

According to the Tits Alternative (see also Example 13.29), the free group $F_2$ embeds as a subgroup in the orthogonal group $SO(3)$. For every $w \in F_2$ we denote by $R_w$ the rotation of $\mathbb{R}^3$ given by this embedding.

Let $C$ be the (countable) set of intersections of $\mathbb{S}^2$ with the union of axes of the rotations $R_w$, $w \in F_2 \setminus \{1\}$. Since $C$ is countable, by Step 1, $\mathbb{S}^2$ is piecewise congruent to $\mathbb{S}^2 \setminus C$. The set $\mathbb{S}^2 \setminus C$ is a disjoint union of orbits of $F_2$. *According to the axiom of choice* there exists a subset $D \subset \mathbb{S}^2 \setminus C$ which intersects every $F_2$-orbit in $\mathbb{S}^2 \setminus C$ exactly once. (The removal of the set $C$ ensures that the action of $F_2$ is free, i.e., no nontrivial element of $F_2$ fixes a point, that is all orbits are copies of $F_2$.)

By Step 2,

$$F_2 = \mathcal{W}'_a \sqcup \mathcal{W}'_{a^{-1}} \sqcup \mathcal{W}_b \sqcup \mathcal{W}_{b^{-1}}.$$

This defines a decomposition

$$\mathbb{S}^2 \setminus C = F_2 \cdot D = \mathcal{W}'_a \cdot D \sqcup \mathcal{W}'_{a^{-1}} \cdot D \sqcup \mathcal{W}_b \cdot D \sqcup \mathcal{W}_{b^{-1}} \cdot D\,. \tag{15.3}$$

The fact that the subsets in the union (15.3) are pairwise disjoint reflects the fact that the action of $F_2$ on $\mathbb{S}^2 \setminus C$ is free. Since $F_2$ admits a paradoxical decomposition, so does $\mathbb{S}^2 \setminus C$. Since the latter is piecewise-congruent to $\mathbb{S}^2$, it follows that $\mathbb{S}^2$ also admits a paradoxical decomposition.

We will now show that $\mathbb{S}^2$ is piecewise congruent to a disjoint union of two copies of $\mathbb{S}^2$. Let $v$ denote the vector $(3, 0, 0)$ in $\mathbb{R}^3$ and let $T_v$ denote the isometry of $\mathbb{R}^3$ which is the translation by $v$.

In view of the decomposition (15.3), the set $\mathbb{S}^2 \setminus C$ is piecewise congruent to

$$\mathcal{W}'_a \cdot D \sqcup R_a \mathcal{W}'_{a^{-1}} \cdot D \sqcup T_v\left(\mathcal{W}_b D\right) \sqcup T_v \circ R_b\left(\mathcal{W}_{b^{-1}} D\right) = \mathbb{S}^2 \setminus C \sqcup T_v\left(\mathbb{S}^2 \setminus C\right) .$$

This and Step 1 imply that $\mathbb{S}^2$ is piecewise congruent to $\mathbb{S}^2 \sqcup T_v \mathbb{S}^2$, i.e., one can "double" the ball. Part 2 of Corollary 15.8 now follows by induction.

**Step 4: A paradoxical decomposition for the unit ball.**

The argument is very similar to the last step in the 2-dimensional case. Let $\mathbb{B}^3$ denote the closed unit ball in $\mathbb{R}^3$ centered at $O$. Step 3 and the fact that the unit ball $\mathbb{B}^3 \setminus \{O\}$ can be written as the set

$$\{\lambda x\ ;\ \lambda \in (0, 1]\, ,\, x \in \mathbb{S}^2\},$$

imply that $\mathbb{B}^3 \setminus \{O\}$ is piecewise congruent to

$$\mathbb{B}^3 \setminus \{O\} \sqcup T_v\left(\mathbb{B}^3 \setminus \{O\}\right).$$

Thus, it remains to prove that $\mathbb{B}^3 \setminus \{O\}$ is piecewise congruent to $\mathbb{B}^3$. We denote by $\mathbb{S}_{1/2}$ the sphere with the center $\left(\frac{1}{2}, 0, 0\right)$ and radius $\frac{1}{2}$. Then

$$\mathbb{B}^3 \setminus \{O\} = \mathbb{B}^3 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} \setminus \{O\}.$$

According to Step 1, $\mathbb{S}_{1/2} \setminus \{O\}$ is piecewise congruent to $\mathbb{S}_{1/2}$; hence, $\mathbb{B}^3 \setminus \{O\}$ is piecewise congruent to $\mathbb{B}^3 \setminus \mathbb{S}_{1/2} \sqcup \mathbb{S}_{1/2} = \mathbb{B}^3$.

This concludes the proof of Corollary 15.8, Parts 1 and 2, for $n \leqslant 3$. □

REMARK 15.12. Banach and Tarski's proof relies on the Hausdorff's paradox, discovered several years prior to their proof. Inspired by the Hausdorff's argument, R. M. Robinson, answering a question of von Neumann, proved in **[Rob47]** that five is the minimal number of pieces in a paradoxical decomposition of the unit 3-dimensional ball. See Proposition 16.90 for a proof of this statement, and Section 16.7 for a discussion on the minimal number of pieces in a paradoxical decomposition.

REMARK 15.13.
(1) The free group $F_2$ of rank 2 contains a free subgroup of countably infinite rank, see Proposition 4.47. This and a proof similar to the one of Theorem 15.7 yields that the unit sphere $\mathbb{S}^{n-1}$ is countably piecewise congruent to countably many copies of $\mathbb{S}^{n-1}$.
(2) It can be proved that the unit sphere $\mathbb{S}^{n-1}$ can be partitioned into $2^{\aleph_0}$ pieces, so that each piece is piecewise congruent to $\mathbb{S}^{n-1}$ (see **[Wag85]**).

CHAPTER 16

# Amenability and paradoxical decomposition.

In this chapter we discuss in detail two important concepts behind the Banach-Tarski paradox: Amenability and paradoxical decompositions. Although both properties were first introduced for groups (of isometries), it turns out that amenability can be defined in purely metric terms, in the context of graphs of bounded geometry. We shall begin by discussing the graph version of amenability, then we will turn to the case of groups, and then to the opposite property of being *paradoxical*.

## 16.1. Amenable graphs

DEFINITION 16.1. A graph $\mathcal{G}$ is called *amenable* if its Cheeger constant, as described in Definition 12.20, is zero. Equivalently, there exists a sequence $F_n$ of finite subsets of $V$ such that

$$\lim_{n\to\infty} \frac{|E(F_n, F_n^c)|}{|F_n|} = 0\,.$$

Such sequence $F_n$ is called a *Følner sequence for the graph* $\mathcal{G}$ .

A graph $\mathcal{G}$ is *non-amenable* if its Cheeger constant is strictly positive.

It is immediate from the definition that every finite graph is amenable (take $F_n = V$).

We describe in what follows various metric properties equivalent to non-amenability. Our arguments are adapted from **[dlHGCS99]**. The only tool that will be needed is Hall–Rado Marriage Theorem from graph theory, stated below.

Let $Bip(Y,Z;E)$ denote the bipartite graph with vertex set $V$ split as $V = Y \sqcup Z$, and the edge-set $E$. Given two integers $k,l \geqslant 1$, a *perfect* $(k,l)$*–matching* of $Bip(Y,Z;E)$ is a subset $M \subset E$ such that each vertex in $Y$ is the endpoint of exactly $k$ edges in $M$, while each vertex in $Z$ is the endpoint of exactly $l$ edges in $M$.

THEOREM 16.2 (Hall-Rado **[Bol79]**, §$III.2$). *Let $Bip(Y,Z;E)$ be a locally finite bipartite graph and let $k \geqslant 1$ be an integer such that:*

- *For every finite subset $A \subset Y$, its edge-boundary $E(A,A^c)$ contains at least $k|A|$ elements.*
- *For every finite subset $B$ in $Z$, its edge-boundary $E(B,B^c)$ contains at least $|B|$ elements.*

*Then $Bip(Y,Z;E)$ has a perfect $(k,1)$–matching.*

Given a discrete metric space $(X,\text{dist})$, two (not necessarily disjoint) subsets $Y,Z$ in $X$, and a real number $C \geqslant 0$, one defines a bipartite graph $Bip_C(Y,Z)$, with the vertex set $Y \sqcup Z$, where two vertices $y \in Y$ and $z \in Z$ are connected by an edge in $Bip_C(Y,Z)$ if and only if $\mathrm{dist}(y,z) \leqslant C$. (The reader will recognize here a version of the Rips complex of a metric space.) We will use this construction in the case when $Y = Z = X$, then the vertex set of $Bip(X,X)$ will consist of two copies of the set $X$.

In what follows, given a graph with the vertex-set $V$ we will use the notation $\overline{\mathcal{N}}_C(F)$ and $\mathcal{N}_C(F)$ to denote the "closed" and "open" $C$-neighborhood of $F$ in $V$:

$$\overline{\mathcal{N}}_C(F) = \{v \in V : \mathrm{dist}(v,F) \leqslant C\}, \quad \mathcal{N}_C(F) = \{v \in V : \mathrm{dist}(v,F) < C\}.$$

THEOREM 16.3. *Let $\mathcal{G}$ be a connected graph of bounded geometry, with vertex set $V$ and edge set $E$, endowed, as usual, with the standard metric. The following conditions are equivalent:*

(a) *$\mathcal{G}$ is non-amenable.*

(b) *$\mathcal{G}$ satisfies the following* expansion condition*: There exists a constant $C > 0$ such that for every finite non-empty subset $F \subset V$, the set $\overline{\mathcal{N}}_C(F) \subset V$ contains at least twice as many vertices as $F$.*

(c) *There exists a constant $C > 0$ such that the graph $Bip_C(V,V)$ has a perfect $(2,1)$–matching.*

(d) *There exists a map $f \in \mathcal{B}(V)$ (see Definition 5.10) such that for every $v \in V$ the preimage $f^{-1}(v)$ contains exactly two elements.*

(e) (Gromov's condition) *there exists a map $f \in \mathcal{B}(V)$ such that for every $v \in V$ the pre-image $f^{-1}(v)$ contains at least two elements.*

REMARK 16.4. Property (b) can be replaced by the property (b') that for some (equivalently, every) $\beta > 1$ there exists $C > 0$ such that $\overline{\mathcal{N}}_C(F) \cap V$ has cardinality at least $\beta$ times the cardinality of $F$. Indeed, it suffices to observe that for every $\alpha > 1$, $C > 0$,

$$\forall F, |\overline{\mathcal{N}}_C(F)| \geqslant \alpha |F| \Rightarrow \forall k \in \mathbb{N}, \quad |\overline{\mathcal{N}}_{kC}(F)| \geqslant \alpha^k |F|.$$

PROOF. We will now prove Theorem 16.3. Let $m \geq 1$ denote the valence of $\mathcal{G}$. (a) $\Rightarrow$ (b). The graph $\mathcal{G}$ is non-amenable if and only if its Cheeger constant is positive. In other words, there exists $\eta > 0$ such that for every finite set of vertices $F$, $|E(F,F^c)| \geqslant \eta |F|$. This implies that $\overline{\mathcal{N}}_1(F)$ contains at least $(1+\frac{\eta}{m})|F|$ vertices, which, according to Remark 16.4, implies property (b).

(b) $\Rightarrow$ (c). Let $C$ be the constant as in the expansion property. We form the bipartite graph $Bip_C(Y,Z)$, where $Y, Z$ are two copies of $V$. Clearly, the graph $Bip_C(Y,Z)$ is locally finite. For any finite subset $A$ in $V$, since $|\overline{\mathcal{N}}_C(A) \cap V| \geqslant 2|A|$, it follows that the edge–boundary of $A$ in $Bip_C(Y,Z)$ has at least $2|A|$ elements, where we embed $A$ in either one of the copies of $V$ in $Bip_C(Y,Z)$. It follows by Theorem 16.2 that $Bip_C(Y,Z)$ has a perfect $(2,1)$–matching.

(c) $\Rightarrow$ (d). The matching in (c) defines a map $f : Z = V \to Y = V$, so that $\mathrm{dist}_{\mathcal{G}}(z, f(z)) \leqslant C$. Hence, $f \in \mathcal{B}(V)$ and $|f^{-1}(y)| = 2$ for every $y \in V$.

The implication (d) $\Rightarrow$ (e) is obvious. We show that (e) $\Rightarrow$ (b). According to (e), there exists a constant $M > 0$ and a map $f : V \to V$ such that for every $x \in V$, $\mathrm{dist}(x, f(x)) \leqslant M$, and $|f^{-1}(y)| \geq 2$ for every $y \in V$. For every finite nonempty set $F \subset V$, $f^{-1}(F)$ is contained in $\mathcal{N}_M(F)$ and it has at least twice as many elements. Thus, (b) is satisfied.

Thus, we proved that the properties (b) through (e) are equivalent.

It remains to be shown that (b) $\Rightarrow$ (a). By hypothesis, there exists a constant $C$ such that for every finite non-empty subset $F \subset V$, $|\overline{\mathcal{N}}_C(F) \cap V| \geqslant 2|F|$. Without loss of generality, we may assume that $C$ is a positive integer. Recall that $\partial_V F$ is the vertex–boundary of the subset $F \subset V$. Since $\overline{\mathcal{N}}_C(F) = F \cup \mathcal{N}_C(\partial_V F)$, it follows that $|\mathcal{N}_C(\partial_V F) \setminus F| \geqslant |F|$.

Recall that the graph $\mathcal{G}$ has finite valence $m \geqslant 1$. Therefore,

$$|\overline{\mathcal{N}}_C(\partial_V F)| \leqslant m^C |\partial_V F| .$$

We have, thus, obtained that for every finite nonempty set $F \subset V$,

$$|E(F, F^c)| \geqslant |\partial_V F| \geqslant \frac{1}{m^C} |\mathcal{N}_C(\partial_V F)| \geqslant \frac{1}{m^C} |F|.$$

Therefore, the Cheeger constant of $\mathcal{G}$ is at least $\frac{1}{m^C} > 0$, and the graph is non-amenable. □

EXERCISE 16.5. Show that a sequence $F_n \subset V$ is Følner if and only if for every $C \in \mathbb{R}_+$

$$\lim_{n \to \infty} \frac{|\mathcal{N}_C(F_n)|}{|F_n|} = 1.$$

Some graphs with bounded geometry admit Følner sequences which consist of metric balls. A proof of the following property (in the context of Cayley graphs) first appeared in [**AVS57**].

PROPOSITION 16.6. *A graph $\mathcal{G}$ of bounded geometry and sub-exponential growth (in the sense of Definition 12.8) is amenable and has the property that for every basepoint $v_0 \in V$ (where $V$ is the vertex set of $\mathcal{G}$) there exists a Følner sequence consisting of metric balls with center $v_0$.*

PROOF. Let $v_0$ be an arbitrary vertex in $\mathcal{G}$. We equip the vertex set $V$ of $\mathcal{G}$ with the restriction of the standard metric on $\mathcal{G}$ and set

$$\mathfrak{G}_{v_0,V}(n) = |\bar{B}(v_0, n)|,$$

here and in what follows $\bar{B}(x, n)$ is the ball of center $x$ and radius $x$ in $V$. Our goal is to show that for every $\varepsilon > 0$ there exists a radius $R_\varepsilon$ such that $\partial_V \bar{B}(v_0, R_\varepsilon)$ has cardinality at most $\varepsilon \, |\bar{B}(v_0, R_\varepsilon)|$.

We argue by contradiction and assume that there exists $\varepsilon > 0$ such that for every integer $R > 0$,

$$|\partial_V \bar{B}(v_0, R)| \geqslant \varepsilon \, |\bar{B}(v_0, R)| .$$

(Since $\mathcal{G}$ has bounded geometry, considering vertex–boundary is equivalent to considering the edge-boundary.) This inequality implies that

$$|\bar{B}(v_0, R+1)| \geqslant (1+\varepsilon) |\bar{B}(v_0, R)| .$$

Applying the latter inequality inductively we obtain

$$\forall n \in \mathbb{N}, \quad |\bar{B}(v_0, n)| \geqslant (1+\varepsilon)^n ,$$

whence

$$\limsup_{n \to \infty} \frac{\ln \mathfrak{G}_{v_0,V}}{n} \geqslant \ln(1+\varepsilon) > 0 .$$

This contradicts the assumption that $\mathcal{G}$ has sub-exponential growth. □

For the sake of completeness we mention without proof two more properties equivalent to those in Theorem 16.3.

The first will turn out to be relevant to a discussion later on between non-amenability and existence of free sub-groups (the von Neumann-Day Question 16.77).

THEOREM 16.7 (Theorem 1.3 in [**Why99**]). *Let $\mathcal{G}$ be an infinite connected graph of bounded geometry. The graph $\mathcal{G}$ is non-amenable if and only if there exists a free action of a free group of rank two on $\mathcal{G}$ by bi-Lipschitz maps which are at finite distance from the identity.*

The second property is related to probability on graphs.

**An amenability criterion with random walks.** Let $\mathcal{G}$ be an infinite locally finite connected graph with set of vertices $V$ and set of edges $E$. For every vertex $x$ of $\mathcal{G}$ we denote by $\mathrm{val}(x)$ the valency at the vertex $X$. We refer the reader to [**Bre92, DS84, Woe00**] for the definition of Markov chains and detailed treatment of random walks on graphs and groups.

A *simple random walk* on $\mathcal{G}$ is a Markov chain with random variables

$$X_1, X_2, \ldots, X_n, \ldots$$

on $V$, with the transition probability $p(x,y) = \frac{1}{\mathrm{val}(x)}$ if $x$ and $y$ are two vertices joined by an edge, and $p(x,y) = 0$ if $x$ and $y$ are not joined by an edge.

We denote by $p_n(x,y)$ the probability that a random walk starting in $x$ will be at $y$ after $n$ steps. The *spectral radius* of the graph $\mathcal{G}$ is defined by

$$\rho(\mathcal{G}) = \limsup_{n\to\infty} \left[p_n(x,y)\right]^{\frac{1}{n}}.$$

It can be easily checked that the spectral radius does not depend on $x$ and $y$.

THEOREM 16.8 (J. Dodziuk, [**Dod84**]). *A graph of bounded geometry is non-amenable if and only if $\rho(\mathcal{G}) < 1$.*

Note that in the case of countable groups the corresponding theorem was proved by H. Kesten [**Kes59**].

COROLLARY 16.9. *In a non-amenable graph of bounded geometry, the simple random walk is transient, that is, for every $x, y \in V$,*

$$\sum_{n=1}^{\infty} p_n(x,y) < \infty .$$

## 16.2. Amenability and quasi-isometry

THEOREM 16.10 (Graph amenability is QI invariant). *Suppose that $\mathcal{G}$ and $\mathcal{G}'$ are quasi-isometric graphs of bounded geometry. Then $\mathcal{G}$ is amenable if and only if $\mathcal{G}'$ is.*

PROOF. We will show that non-amenability is a quasi-isometry invariant. We will assume that both $\mathcal{G}$ and $\mathcal{G}'$ are infinite, otherwise the assertion is clear. Note that according to Theorem 16.3, Part (b), nonamenability is equivalent to existence of a constant $C > 0$ such that for every finite non-empty set $F$ of vertices, its closed neighborhood $\bar{\mathcal{N}}_C(F)$ contains at least $2|F|$ vertices.

Let $V$ and $V'$ be the vertex sets of graphs $\mathcal{G}$ and $\mathcal{G}'$ respectively. We assume that $V, V'$ are endowed with the metrics obtained by restriction of the standard metrics on the respective graphs. Let $m < \infty$ be an upper bound on the valence of graphs $\mathcal{G}, \mathcal{G}'$. Let $f : V \to V'$ and $g : V' \to V$ be $L$–Lipschitz maps that are quasi-inverse to each other:

$$\mathrm{dist}(f \circ g, Id) \leqslant A, \quad \mathrm{dist}(g \circ f, Id) \leqslant A.$$

Assume that $\mathcal{G}'$ is amenable. Given a finite set $F$ in $V$, consider

$$F \xrightarrow{f} F' = f(F) \xrightarrow{g} F'' = g(F').$$

Since $F''$ is at Hausdorff distance $\leqslant A$ from $F$, it follows that $|F| \leqslant b|F''|$, where $b = m^L$. In particular,

$$|f(F)| \geqslant b^{-1}|F|.$$

Likewise, for every finite set $F'$ in $V'$ we obtain

$$|g(F')| \geqslant b^{-1}|F'| \ .$$

Remark 16.4 implies that for every number $\alpha > b^2$, there exists $C \geqslant 1$ such that for an arbitrary finite set $F' \subset V'$, its neighborhood $\bar{\mathcal{N}}_C(F')$ contains at least $\alpha|F'|$ vertices. Therefore, the set $g\left(\bar{\mathcal{N}}_C(F')\right)$ contains at least

$$\frac{1}{b}|\mathcal{N}_C(F')| \geqslant \frac{\alpha}{b}|F'|$$

elements.

Pick a finite nonempty subset $F \subset V$ and set $F' := f(F), F'' = gf(F)$. Then $|F'| \geq b^{-1}|F|$ and, therefore,

$$|g\left(\bar{\mathcal{N}}_C(F')\right)| \geqslant \frac{\alpha}{b^2}|F|.$$

Since $g$ is $L$–Lipschitz,

$$g\left(\bar{\mathcal{N}}_C(F')\right) \subset \bar{\mathcal{N}}_{LC}(F'') \subset \bar{\mathcal{N}}_{LC+A}(F).$$

We conclude that

$$|\bar{\mathcal{N}}_{LC+A}(F)| \geqslant \frac{\alpha}{b^2}|F|.$$

Setting $C' := LC + A$, and $\beta := \frac{\alpha}{b^2} > 1$, we conclude that $\mathcal{G}$ satisfies the expansion property (b') in Theorem 16.3. Hence, $\mathcal{G}$ is also non-amenable. □

We will see below that this theorem generalizes in the context connected Riemannian manifolds $M$ of bounded geometry and graphs $\mathcal{G}$ obtained by discretization of $M$, and, thus, quasi-isometric to $M$. More precisely, we will see that non-amenability of the graph is equivalent to positivity of the Cheeger constant of the manifold (see Definition 2.20). This may be seen as a version within the setting of amenability/isoperimetric problem of the Milnor–Efremovich–Schwartz Theorem 12.12 stating that the growth functions of $M$ and $\mathcal{G}$ are equivalent.

In what follows we use the terminology in Definitions 2.56 and 2.60 for the bounded geometry of a Riemannian manifold, respectively of a simplicial graph.

THEOREM 16.11. *Let $M$ be a complete connected $n$-dimensional Riemannian manifold and $\mathcal{G}$ a simplicial graph, both of bounded geometry. Assume that $M$ is quasi-isometric to $\mathcal{G}$. Then the Cheeger constant of $M$ is strictly positive if and only if the graph $\mathcal{G}$ is non-amenable.*

REMARKS 16.12. (1) Theorem 16.11 was proved by R. Brooks [**Bro82a**], [**Bro81a**] in the special case when $M$ is the universal cover of a compact Riemannian manifold and $\mathcal{G}$ is the a Cayley graph of the fundamental group of this compact manifold .

(2) A more general version of Theorem 16.11 requires a weaker condition of bounded geometry for the manifold than the one used in this book. See for instance [**Gro93**], Proposition $0.5.A_5$. A proof of that result can be obtained by combining the main theorem in [**Pan95**] and Proposition 11 in [**Pan07**].

PROOF. Since $M$ has bounded geometry it follows that its sectional curvature is at least $a$ and at most $b$, for some $b \geqslant a$. It also follows that the injectivity radius at every point of $M$ is at least $\rho$ , for some $\rho > 0$.

As in Theorem 2.24, we let $V_\kappa(r)$ denote the volume of ball of radius $r$ in the $n$-dimensional space of constant curvature $\kappa$.

Choose $\varepsilon$ so that $0 < \varepsilon < 2\rho$. Let $N$ be a maximal $\varepsilon$-separated set in $M$.

It follows that $\mathcal{U} = \{B(x, \varepsilon) \mid x \in N\}$ is a covering of $M$, and by Lemma 2.58, (2), its multiplicity is at most

$$m = \frac{V_a\left(\frac{3\varepsilon}{2}\right)}{V_b\left(\frac{\varepsilon}{2}\right)}.$$

We now consider the restriction of the Riemannian distance function on $M$ to the subset $N$. Define the Rips complex $Rips_{8\varepsilon}(N)$ (with respect to this metric on $N$), and the 1-dimensional skeleton of the Rips complex, the graph $\mathcal{G}_\varepsilon$. According to Theorem 5.41, the manifold $M$ is quasi-isometric to $\mathcal{G}_\varepsilon$. Furthermore, $\mathcal{G}_\varepsilon$ has bounded geometry as well. This and Theorem 16.10 imply that $\mathcal{G}_\varepsilon$ has strictly positive Cheeger constant if and only if $\mathcal{G}$ has. Thus, it suffices to prove the equivalence in Theorem 16.11 for the graph $\mathcal{G} = \mathcal{G}_\varepsilon$.

Assume that $M$ has positive Cheeger constant. This means that there exists $h > 0$ such that for every open submanifold $\Omega \subset M$ with compact closure and smooth boundary,

$$Area(\partial\Omega) \geqslant h\, Vol(\Omega)\,.$$

Our goal is to show that there exist uniform positive constants $B$ and $C$ such that for every finite subset $F \subset N$ there exists an open submanifold with compact closure and smooth boundary $\Omega$, such that (with the notation in Definition 1.11),

$$\text{card}\, E(F, F^c) \geqslant B\, Area(\partial\Omega) \quad \text{and} \quad C Vol(\Omega) \geqslant \text{card}\, F\,. \tag{16.1}$$

Then, it would follow that

$$|E(F, F^c)| \geqslant \frac{Bh}{C}|F|,$$

i.e., $\mathcal{G}$ would be non-amenable. Here, as usual, $F^c = N \setminus F$.

Since $M$ has bounded geometry, the open cover $\mathcal{U}$ admits a smooth partition of unity $\{\varphi_x\ ;\ x \in N\}$ in the sense of Definition 2.8, such that all the functions $\varphi_x$ are $L$–Lipschitz for some constant $L > 0$ independent of $x$, see Lemma 2.23. Let $F \subset N$ be a finite subset. Consider the smooth function $\Phi = \sum_{x\in F} \varphi_x$ . By hypothesis and since $\mathcal{U}$ has multiplicity at most $m$, the function $\Phi$ is $Lm$–Lipschitz. Furthermore, since the map $\Phi$ has compact support, the set $\Theta$ of singular values of $\Phi$ is compact and has Lebesgue measure zero.

For every $t \in (0,1)$, the preimage

$$\Omega_t = \Phi^{-1}((t,\infty)) \subset M$$

is an open submanifold in $M$ with compact closure. If we choose $t$ to be a regular value of $\Phi$, that is $t \notin \Theta$, then the hypersurface $\Phi^{-1}(t)$, which is the boundary of $\Omega_t$, is smooth (Theorem 2.4).

Since $N$ is $\epsilon$-separated, the balls $B\left(x, \frac{\varepsilon}{2}\right), x \in N$, are pairwise disjoint. Therefore, for every $x \in N$ the function $\varphi_x$ restricted to $B\left(x, \frac{\varepsilon}{2}\right)$ is identically equal to 1. Hence, the union

$$\bigsqcup_{x\in F} B\left(x, \frac{\varepsilon}{2}\right)$$

is contained in $\Omega_t$ for every $t \in (0,1)$, and in view of Part 2 of Theorem 2.24 we get

$$Vol(\Omega_t) \geqslant \sum_{x\in F} Vol\left(x, \frac{\varepsilon}{2}\right) \geqslant \operatorname{card} F \cdot V_b\left(\varepsilon/2\right) .$$

Therefore, for every $t \notin \Theta$, the domain $\Omega_t$ satisfies the second inequality in (16.1) with $C^{-1} = V_b(\varepsilon/2)$. Our next goal is to find values of $t \notin \Theta$ so that the first inequality in (16.1) holds.

Fix a constant $\eta$ in the open interval $(0,1)$, and consider the open set $U = \Phi^{-1}((0,\eta))$.

Let $F'$ be the set of points $x$ in $F$ such that $U \cap \overline{B(x,\varepsilon)} \neq \emptyset$. Since for every $y \in U$ there exists $x \in F$ such that $\varphi_x(y) > 0$, it follows that the set of closed balls centered in points of $F'$ and of radius $\varepsilon$ cover $U$.

Since $\{\varphi_x : x \in N\}$ is a partition of unity for the cover $\mathcal{U}$ of $M$, it follows that for every $y \in U$ there exists $z \in N \setminus F$ such that $\varphi_z(y) > 0$, whence $y \in \overline{B(z,\varepsilon)}$. Thus,

$$U \subset \left(\bigcup_{x\in F'} \overline{B(x,\varepsilon)}\right) \cap \left(\bigcup_{z\in N\setminus F} \overline{B(z,\varepsilon)}\right). \tag{16.2}$$

In particular, for every $x \in F'$ there exists $z \in N\setminus F$ such that $\overline{B(x,\varepsilon)} \cap \overline{B(z,\varepsilon)} \neq \emptyset$, whence $x$ and $z$ are connected by an edge in the graph $\mathcal{G}$.

Thus, every point $x \in F'$ belongs to the vertex-boundary $\partial_V F$ of the subset $F$ of the vertex set of the graph $\mathcal{G}$. We conclude that $\operatorname{card} F' \leqslant \operatorname{card} E(F, F^c)$.

Since $|\nabla\Phi| \leqslant mL$, by the Coarea Theorem 2.16, with $g \equiv 1$, $f = \Phi$ and $U = \Phi^{-1}(0,\eta)$, we obtain:

$$\int_0^\eta Area(\partial\Omega_t)\mathrm{dt} = \int_U |\nabla\Phi|\mathrm{d}V \leqslant mLVol(U) \leqslant mL \sum_{x\in F'} Vol(B(x,\varepsilon)).$$

The last inequality follows from the inclusion (16.2). At the same time, by applying Theorem 2.24, we obtain that for every $x \in M$

$$V_a(\varepsilon) \geqslant Vol(B(x,\varepsilon)).$$

By combining these inequalities, we obtain

$$\int_0^\eta Area(\partial\Omega_t)\mathrm{dt} \leqslant mLV_a(\varepsilon)\,|F'| \leqslant mLV_a(\varepsilon)\,|E\left(F,F^c\right)|.$$

Since $\Theta$ has measure zero, it follows that for some $t \in (0, \eta) \setminus \Theta$,

$$Area(\partial\Omega_t) \leqslant 2\frac{m}{\eta} L V_a(\varepsilon) \, |E\left(F, F^c\right)| = B|E\left(F, F^c\right)|.$$

This establishes the first inequality in (16.1) and, hence, shows that nonamenability of $M$ implies nonamenability of the graph $\mathcal{G}$.

We now prove the converse implication. To that end, we assume that for some $\delta$ satisfying $2\rho > \delta > 0$, some maximal $\delta$-separated set $N$ and the corresponding graph (of bounded geometry) $\mathcal{G} = \mathcal{G}_\delta$ are constructed as above, so that $\mathcal{G}$ has a positive Cheeger constant. Thus, there exists $h > 0$ such that for every finite subset $F$ in $N$

$$\operatorname{card} E(F, F^c) \geqslant h \operatorname{card} F\,.$$

Let $\Omega$ be an arbitrary open bounded subset of $M$ with smooth boundary. Our goal is to find a finite subset $F_k$ in $N$ such that for two constants $P$ and $Q$ independent of $\Omega$, we have

$$Area(\partial\Omega) \geqslant P\,|E(F_k, F_k^c)| \ \text{ and } |F_k| \geqslant QVol(\Omega)\,. \tag{16.3}$$

This would imply positivity of Cheeger constant of $M$. Note that, since the graph $\mathcal{G}$ has finite valence, in the first inequality of (16.3) we may replace the edge boundary $E(F_k, F_k^c)$ by the vertex boundary $\partial_V F_k$ (see Definition 1.11).

Consider the finite subset $F$ of points $x \in N$ such that $\Omega \cap B(x, \delta) \neq \emptyset$. It follows that $\Omega \subseteq \bigcup_{x \in F} B(x, \delta)$. We split the set $F$ into two parts:

$$F_1 = \left\{ x \in F \ : \ Vol[\Omega \cap B(x, \delta)] > \frac{1}{2} Vol[B(x, \delta)] \right\} \tag{16.4}$$

and

$$F_2 = \left\{ x \in F \ : \ Vol[\Omega \cap B(x, \delta)] \leqslant \frac{1}{2} Vol[B(x, \delta)] \right\}.$$

Set

$$v_k := Vol\left(\Omega \cap \bigcup_{x \in F_k} B(x, \delta)\right), k = 1, 2.$$

Thus,

$$\max\left(v_1, v_2\right) \geqslant \frac{1}{2} Vol(\Omega).$$

**Case 1:** $v_1 \geqslant \frac{1}{2} Vol(\Omega)$. In view of Theorem 2.24, this inequality implies that

$$\frac{1}{2} Vol(\Omega) \leqslant \sum_{x \in F_1} Vol\left(B(x, \delta)\right) \leqslant |F_1|\, V_a(\delta)\,. \tag{16.5}$$

This gives the second inequality in (16.3). A point $x$ in $\partial_V F_1$ is then a point in $N$ satisfying (16.4), such that within distance $8\delta$ of $x$ there exists a point $y \in N$ satisfying the inequality opposite to (16.4). The (unique) shortest geodesic $[x, y] \subset M$ will, therefore, intersect the set of points

$$\text{Half} = \left\{ x \in M \ ; \ Vol\left[B(x, \delta) \cap \Omega\right] = \frac{1}{2} Vol[B(x, \delta)] \right\}.$$

This implies that $\partial_V F_1$ is contained in the $8\delta$-neighborhood of the set Half $\subset M$. Given a maximal $\delta$–separated subset $H_\delta$ of Half (with respect to the restriction of the Riemannian distance on $M$), $\partial_V F_1$ will then be contained in the $9\delta$-neighborhood of $H_\delta$. In particular,

$$\bigsqcup_{x\in\partial_V F_1} B\left(x, \frac{\delta}{2}\right) \subseteq \bigcup_{y\in H_\delta} B(y, 10\delta)\,,$$

whence

$$V_b\left(\delta/2\right) |\partial_V F_1| \leqslant Vol\left[\bigsqcup_{x\in\partial_V F_1} B\left(x, \frac{\delta}{2}\right)\right] \leqslant$$

$$\sum_{y\in H_\delta} Vol\left[B(y, 10\delta)\right] \leqslant V_b(10\delta)\, |H_\delta|. \tag{16.6}$$

Since $H_\delta$ extends to a maximal $\delta$–separated subset $H'$ of $M$, Lemma 2.58, (2), implies that the multiplicity of the covering $\{B(x,\delta) \mid x \in H'\}$ is at most $\frac{V_a\left(\frac{3\delta}{2}\right)}{V_b\left(\frac{\delta}{2}\right)}$.

It follows that

$$m \cdot Area(\partial\Omega) \geqslant \sum_{y\in H_\delta} Area(\partial\Omega \cap B(y,\delta))\,.$$

We now apply Buser's Theorem 2.25 and deduce that there exists a constant $\lambda = \lambda(n, a, \delta)$ such that for all $y \in H_\delta$, we have,

$$\lambda Area(\partial\Omega \cap B(y,\delta)) \geqslant Vol\left[\Omega \cap B(y,\delta)\right] = \frac{1}{2} Vol[B(y,\delta)]\,.$$

It follows that

$$Area(\partial\Omega) \geqslant \frac{1}{2\lambda m} \sum_{y\in H_\delta} Vol[B(y,\delta)] \geqslant \frac{1}{2\lambda m} V_b(\rho)\, |H_\delta|\,.$$

Combining this estimate with the inequality (16.6), we conclude that

$$Area(\partial\Omega) \geqslant P|\partial_V F_1|\,,$$

for some constant $P$ independent of $\Omega$.

This establishes the first inequality in (16.3) and, hence, proves positivity of the Cheeger constant of $M$ in the Case 1.

**Case 2.** Assume now that $v_2$ is at least $\frac{1}{2}Vol(\Omega)$.

We obtain, using Buser's Theorem 2.25 for the second inequality below, that

$$mArea(\partial\Omega) \geqslant \sum_{y\in F_2} Area\left(\partial\Omega \cap B(y,\delta)\right) \geqslant \frac{1}{\lambda} \sum_{y\in F_2} Vol\left[\Omega \cap B(y,\delta)\right] \geqslant \frac{1}{2\lambda} Vol(\Omega)\,.$$

Thus, in the Case 2 we obtain the required lower bound on $Area(\partial\Omega)$ directly. □

COROLLARY 16.13. *Let $M$ and $M'$ be two complete connected Riemann manifolds of bounded geometry which are quasi-isometric to each other. Then $M$ has positive Cheeger constant if and only if $M'$ has positive Cheeger constant.*

PROOF. Consider graphs of bounded geometry $\mathcal{G}$ and $\mathcal{G}'$ that are quasi-isometric to $M$ and $M'$ respectively. Then $\mathcal{G}, \mathcal{G}'$ are also quasi-isometric to each other. The result now follows by combining Theorem 16.11 with Theorem 16.10. □

An interesting consequence of Corollary 16.13 is quasi-isometric invariance of a certain property of the Laplace-Beltrami operator for Riemannian manifolds of bounded geometry. Cheeger constant for Riemannian manifold $M$ is closely connected to the bottom of the spectrum of the Laplace-Beltrami operator $\Delta_M$ on $L^2(M) \cap C^\infty(M)$. Let $M$ be a complete connected Riemannian manifold of infinite volume, let $\lambda_0(M)$ denote the lowest eigenvalue of $\Delta_M$. Then $\lambda_0(M)$ can be computed as

$$\inf \left\{ \frac{\int_M |\nabla f|^2}{\int_M f^2} \mid f : M \to \mathbb{R} \text{ smooth with compact support } \right\}$$

(see [**CY75**] or [**SY94**], Chapter I). J. Cheeger proved in [**Che70**] that

$$\lambda_0(M) \geqslant \frac{1}{4} h^2(M)\,,$$

where $h(M)$ is the Cheeger constant of $M$. Even though Cheeger's original result was formulated for compact manifolds, his argument works for all complete manifolds, see [**SY94**]. Cheeger's inequality is complemented by the following inequality due to P. Buser (see [**Bus82**], or [**SY94**]) which holds for all complete Riemannian manifolds whose Ricci curvature is bounded below by some $a \in \mathbb{R}$:

$$\lambda_1(M) \leqslant \alpha h(M) + \beta h^2(M),$$

for some $\alpha = \alpha(a), \beta = \beta(a)$. Combined, Cheeger and Buser inequalities imply that $h(M) = 0 \iff \lambda_0(M) = 0$.

COROLLARY 16.14. *Let $M$ and $M'$ be two complete connected Riemann manifolds of bounded geometry which are quasi-isometric to each other. Then $\lambda_0(M) = 0 \iff \lambda_0(M') = 0$.*

We finish the section by noting a remarkable property of quasi-isometries between non-amenable graphs.

THEOREM 16.15 (K. Whyte [**Why99**]). *Let $\mathcal{G}_i, i = 1, 2,$ be two non-amenable graphs of bounded geometry. Then every quasi-isometry $\mathcal{G}_1 \to \mathcal{G}_2$ is at bounded distance from a bi-Lipschitz map.*

Note that this theorem was also implicit in [**DSS95**].

## 16.3. Amenability for groups

We now discuss the concept of amenability for groups. We introduce various versions of amenability and non-amenability, formulated in terms of actions and inspired by the Banach-Tarski paradox. We then show that in the case of finitely generated groups one of the notions of amenability is equivalent to the metric amenability for (arbitrarily chosen) Cayley graphs, as formulated in Definition 16.1.

Let $G$ be a group acting on a set $X$. We assume that the action is on the left (for an action on the right a similar discussion can be carried out). We denote the action by $\mu(g, x) = g(x) = g \cdot x$ .

We say that two subsets $A, B \subset X$ are *$G$–congruent* if there exists $g \in G$ such that $g \cdot A = B$.

We say subsets $A, B \subset X$ are *G–piecewise congruent* (or $A$ and $B$ are *G–equidecomposable*) if, for some $k \in \mathbb{N}$, there exist partitions $A = A_1 \sqcup \ldots \sqcup A_k$, $B = B_1 \sqcup \ldots \sqcup B_k$ such that $A_i$ and $B_i$ are $G$–congruent for every $i \in \{1, \ldots, k\}$.

The subsets $A, B$ are *G–countably piecewise congruent* (or *G–countably equidecomposable*) if they admit countable partitions $A = \bigsqcup_{n\in\mathbb{N}} A_n$, $B = \bigsqcup_{n\in\mathbb{N}} B_n$ such that $A_n$ and $B_n$ are $G$–congruent for every $n \in \mathbb{N}$.

Exercise 16.16. Verify that piecewise congruence and countably piecewise congruence are equivalence relations.

Definitions 16.17. (1) A *G–paradoxical subset of* $X$ is a subset $E$ that admits a *G-paradoxical decomposition*, i.e., a finite partition

$$E = X_1 \sqcup \ldots \sqcup X_k \sqcup Y_1 \sqcup \ldots \sqcup Y_m$$

such that for some elements $g_1, \ldots, g_k, h_1, \ldots, h_m$ of $G$,

$$g_1(X_1) \sqcup ... \sqcup g_k(X_k) = E \text{ and } h_1(Y_1) \sqcup ... \sqcup h_m(Y_m) = E\,.$$

(2) A *G–countably paradoxical subset of* $X$ is a subset $F$ admitting a countable partition

$$F = \bigsqcup_{n\in\mathbb{N}} X_n \sqcup \bigsqcup_{m\in\mathbb{N}} Y_m$$

such that for two sequences $(g_n)_{n\in\mathbb{N}}$ and $(h_m)_{m\in\mathbb{N}}$ in $G$,

$$\bigsqcup_{n\in\mathbb{N}} g_n(X_n) = F \text{ and } \bigsqcup_{m\in\mathbb{N}} \psi_m(Y_m) = F\,.$$

John von Neumann [**vN29**] studied properties of group actions that make paradoxical decompositions possible (like for the action of the group of isometries of $\mathbb{R}^n$ for $n \geqslant 3$) or, on the contrary forbid them (like for the action of the group of isometries of $\mathbb{R}^2$). He defined the notion of *amenable group*, based on the existence of a mean/finitely additive measure invariant under the action of the group, and equivalent to the *nonexistence* of paradoxical decompositions for any space on which the group acts. One can ask furthermore that no subset has a paradoxical decomposition, for any space endowed with an action of the group. This defines a strictly smaller class, that of *super-amenable groups*. In what follows we discuss all these variants of amenability and paradoxical behavior.

To clarify the setting, we recall the definition of a finitely additive (probability) measure.

Definition 16.18. An *algebra of subsets* of a set $X$ is a non-empty collection $\mathcal{A}$ of subsets of $X$ such that:

(1) $\emptyset$ and $X$ are in $\mathcal{A}$;
(2) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$;
(3) $A \in \mathcal{A} \Rightarrow A^c = X \setminus A \in \mathcal{A}$.

Definition 16.19. (1) A *finitely additive (f.a.) measure* $\mu$ on an algebra $\mathcal{A}$ of subsets of $X$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that $\mu(A \sqcup B) = \mu(A) + \mu(B)$ for all $A, B \in \mathcal{A}$.

(2) If moreover $\mu(X) = 1$ then $\mu$ is called a *finitely additive probability (f.a.p.) measure*.

(3) Let $G$ be a group acting on $X$ preserving $\mathcal{A}$, i.e., $gA \in \mathcal{A}$ for every $A \in \mathcal{A}$ and $g \in G$. If $\mu$ is a finitely additive measure on $\mathcal{A}$, so that $\mu(gA) = \mu(A)$ for all $g \in G$ and $A \in \mathcal{R}$, then $\mu$ is called *$G$–invariant*.

An immediate consequence of the f.a. property is that for any two sets $A, B \in \mathcal{A}$,

$$\mu(A\cup B) = \mu((A\setminus B)\sqcup(A\cap B)\sqcup(B\setminus A)) = \mu(A\setminus B)+\mu(A\cap B)+\mu(B\setminus A) \leqslant \mu(A)+\mu(B).$$

REMARK 16.20. In some texts the f.a. measures are called simply 'measures'. We prefer the terminology above, since in other texts a 'measure' is meant to be countably additive.

We recall without proof a strong result relating the existence of a finitely additive measure to the non-existence of paradoxical decompositions. It is due to Tarski ([**Tar38**], [**Tar86**, pp. 599–643]), see also [**Wag85**, Corollary 9.2].

THEOREM 16.21 (Tarski's alternative). *Let $G$ be a group acting on a space $X$ and let $E$ be a subset in $X$. Then $E$ is not $G$–paradoxical if and only if there exists a $G$–invariant finitely additive measure $\mu : \mathcal{P}(X) \to [0, \infty]$ such that $\mu(E) = 1$.*

## 16.4. Super-amenability, weakly paradoxical actions, elementary amenability

DEFINITION 16.22. (1) A group action $G \curvearrowright X$ is *weakly paradoxical* if there exists a $G$-paradoxical subset in $X$. An action $G \curvearrowright X$ is *super-amenable* if it is not weakly paradoxical.

(2) An action $G \curvearrowright X$ is *paradoxical* if the entire set $X$ is $G$-paradoxical.

(3) A group $G$ is *(weakly) paradoxical* if the action $G \curvearrowright G$ by left multiplications is (weakly) paradoxical.

(4) Likewise, a group $G$ is called *super-amenable* if the action $G \curvearrowright G$ by left multiplications is super-amenable.

Note that, by using the inversion map $x \mapsto x^{-1}$, one easily sees that in Definition 16.22, (3) and (4), it does not matter if one considers left or right multiplication.

PROPOSITION 16.23. (1) *A group is super-amenable if and only if every action of it is super-amenable.*

(2) *A group is weakly paradoxical if and only if it has at least one weakly paradoxical action.*

PROOF. (1) and (2) are equivalent, therefore it suffices to prove (1). The 'if' part of the statement is obvious. We prove the 'only if' part.

Consider an arbitrary action $G \curvearrowright X$ and an arbitrary non-empty subset $E$ of $X$. Without loss of generality we may assume that the action is $G \curvearrowright X$ is to the left.

Let $x$ be a point in $E$ and let $G_E$ be the set of $g \in G$ such that $gx \in E$. By hypothesis, $G$ is super-amenable, therefore $G_E$ is not paradoxical with respect to the left-action $G \curvearrowright G$. Theorem 16.21 implies that there exists a $G$–left-invariant finitely additive measure $\mu_G : \mathcal{P}(G) \to [0, \infty]$ such that $\mu(G_E) = 1$.

We define a $G$–invariant finitely additive measure $\mu : \mathcal{P}(X) \to [0, \infty]$ by

$$\mu(A) = \mu_G(\{g \in G \mid gx \in A\}) .$$

This measure satisfies $\mu(E) = 1$, hence, $E$ cannot be $G$–paradoxical. $\square$

PROPOSITION 16.24. *A weakly paradoxical group has exponential growth.*

PROOF. Let $G$ be weakly paradoxical and let $E$ be a $G$-paradoxical subset of $G$. Then

$$E = X_1 \sqcup \ldots \sqcup X_k \sqcup Y_1 \sqcup \ldots \sqcup Y_m$$

and there exist elements $g_1, \ldots, g_k, h_1, \ldots, h_m$ in $G$ such that

$$g_1 X_1 \sqcup \ldots \sqcup g_k X_k = E \text{ and } h_1 Y_1 \sqcup \ldots \sqcup h_m Y_m = E .$$

We define two piecewise left translations $\bar{g} : E \to E$ and $\bar{h} : E \to E$ as follows: The restriction of $\bar{g}$ to $g_i X_i$ coincides with the left translation by $g_i^{-1}$, for every $i \in \{1, \ldots, k\}$; the restriction of $\bar{h}$ to $h_j Y_j$ coincides with the left translation by $h_j^{-1}$, for every $j \in \{1, \ldots, m\}$. Both maps are injective. Indeed if $a, b$ are two distinct elements of $E$, either they are in the same subset $g_i X_i$ in which case the injectivity follows from the injectivity of left translations, or $a \in g_i X_i$ and $b \in g_j X_j$, for some $i \neq j$. In the latter case, $\bar{g}(a) \in X_i$ and $\bar{g}(b) \in X_j$ ad since $X_i \cap X_j = \emptyset$, the two images are distinct. A similar argument shows the injectivity of $\bar{h}$.

Given an alphabet of two letters $\{x, y\}$ we denote by $W_n$ the set of words of length $n$. For $w \in W_n$ we denote by $w(\bar{g}, \bar{h})$ the map $E \to E$ obtained by replacing $x$ with $\bar{g}$, $y$ with $\bar{h}$ and considering the composition of the finite sequence of maps thus obtained.

We prove by induction on $n \geqslant 1$ that the subsets $w(\bar{g}, \bar{h})(E)$, $w \in W_n$, are pairwise disjoint. For $n = 1$ this means that $\bar{g}(E)$ and $\bar{h}(E)$ are disjoint, which is obvious.

Assume that the statement is true for $n$. Let $u$ and $v$ be two distinct words of length $n + 1$. Assume that they both begin with the same letter, say $u = xu'$ and $v = xv'$, where $u'$ and $v'$ are distinct words of length $n$ (the case when the letter is $y$ is similar).

Then $u(\bar{g}, \bar{h})(E) = \bar{g}u'(\bar{g}, \bar{h})(E)$ and $v(\bar{g}, \bar{h})(E) = \bar{g}v'(\bar{g}, \bar{h})(E)$ . The induction hypothesis implies that the sets $u'(\bar{g}, \bar{h})(E)$ and $v'(\bar{g}, \bar{h})(E)$ are disjoint, and since $\bar{g}$ is injective, the same is true for the two initial sets.

If $u = xu'$ and $v = yv'$ then

$$u(\bar{g}, \bar{h})(E) \subset \bar{g}(E) \subset X_1 \sqcup \ldots \sqcup X_k$$

while

$$v(\bar{g}, \bar{h})(E) \subset \bar{h}(E) \subset Y_1 \sqcup \ldots \sqcup Y_m .$$

Thus, $u(\bar{g}, \bar{h})(E)$ and $v(\bar{g}, \bar{h})(E)$ are disjoint in this case too, which concludes the induction step, and the proof.

It follows from the statement just proved, that for every $n \geqslant 1$, given an arbitrary $a \in E$, the set $w(\bar{g}, \bar{h})(a)$, $w \in W_n$ , contains as many elements as $W_n$, that is $2^n$. By the definition of $\bar{g}$ and $\bar{h}$, for every $w \in W_n$, $w(\bar{g}, \bar{h})(a) = g_w a$, where $g_w$ is an element in $G$ obtained by replacing in $w$ every occurrence of the letter $x$ by one of the elements $g_1, \ldots, g_k$, every occurrence of the letter $y$ by one of the elements $h_1, \ldots, h_m$, and taking the product in $G$. Since $g_w a$, $w \in W_n$ , are pairwise distinct, the elements $g_w$, $w \in W_n$ , are pairwise distinct. With respect to a generating set $S$ containing $g_1, \dots, g_k, h_1, \dots, h_m$ we have $|g_w|_S \leqslant n$, whence, $\mathfrak{G}_S(n) \geqslant 2^n$. □

COROLLARY 16.25. *Every group with sub-exponential growth is super-amenable.*

Corollary 16.25 is a strengthening of Proposition 16.6 in the group-theoretic setting, in view of the discussion in Section 16.3.

COROLLARY 16.26. *Virtually nilpotent groups and finite extensions of Grigorchuk groups are super-amenable.*

EXERCISE 16.27. Given a finite group $G$ and a non-empty subset $E \subset G$, construct a $G$–left-invariant finitely additive measure $\mu : \mathcal{P}(G) \to [0, \infty]$ such that $\mu(E) = 1$.

It is not known if the converse of Proposition 16.24 is true or if on the contrary there exist super-amenable groups with exponential growth.

A weaker version of the converse of Proposition 16.24 is known though, and it runs as follows.

PROPOSITION 16.28. *A free two-generated sub-semigroup $S$ of a group $G$ is always $G$–paradoxical, where the action $G \curvearrowright G$ is either by left of by right multiplication.*

PROOF. Let $a, b$ be the two elements in $G$ generating the free sub-semigroup $S$, let $S_a$ and $S_b$ be the subsets of elements in $S$ represented by words beginning in $a$, respectively by words beginning in $b$. Then $S = S_a \sqcup S_b$, with $a^{-1}S_a = S$ and $b^{-1}S_b = S$. □

REMARK 16.29. The converse of Proposition 16.28, on the other hand is not true: a weakly paradoxical group does not necessarily contain a nonabelian free subsemigroup. There exist torsion groups that are paradoxical (see the discussion following Remark 16.81).

PROPOSITION 16.30. (1) *A subgroup of a super-amenable group is super-amenable.*

(2) *A finite extension of a super-amenable group is super-amenable.*

(3) *A quotient of a super-amenable group is super-amenable.*

(4) *A direct limit of a directed system of super-amenable groups is super-amenable.*

REMARKS 16.31. The list of group constructions under which super-amenability is stable cannot be completed with:

- if a normal subgroup $N$ in a group $G$ is super-amenable and $G/N$ is super-amenable then $G$ is super-amenable;
- a direct product of super-amenable groups is super-amenable.

It is simply not known if the second property is true, while the first property is known to be false. Otherwise, this property and Corollary 16.25 would imply that all solvable groups are super-amenable. On the other hand, solvable groups that are not virtually nilpotent contain a nonabelian free subsemigroup [**Ros74**].

Proof. (1) Let $H \leqslant G$ with $G$ super-amenable and let $E$ be a non-empty subset of $H$. By Theorem 16.21, there exists a $G$–left-invariant finitely additive measure $\mu : \mathcal{P}(G) \to [0, \infty]$ such that $\mu(E) = 1$. Theorem 16.21 applied to $\mu$ restricted to $\mathcal{P}(H)$ imply that $E$ cannot be $H$-paradoxical either.

(2) Let $H \leqslant G$ with $H$ super-amenable and $G = \bigsqcup_{i=1}^m Hx_i$. Let $E$ be a non-empty subset of $G$.

The group $H$ acts on $G$, whence by Proposition 16.23, (1), and Theorem 16.21, there exists an $H$–left-invariant finitely additive measure $\mu : \mathcal{P}(G) \to [0, \infty]$ such that $\mu\left(\cup_{i=1}^m x_i E\right) = 1$.

Define the measure $\nu : \mathcal{P}(G) \to [0, \infty]$ by

$$\nu(A) = \frac{\sum_{i=1}^m \mu(x_i A)}{\sum_{i=1}^m \mu(x_i E)}\,.$$

It is clearly finitely additive and satisfies $\nu(E) = 1$.

Let $A$ be an arbitrary non-empty subset of $G$ and $g$ an arbitrary element in $G$. We have that $G = \bigsqcup_{i=1}^m Hx_i = \bigsqcup_{i=1}^m Hx_i g$, whence there exists a bijection $\varphi : \{1, \dots, m\} \to \{1, \dots, m\}$ dependent on $g$ such that $Hx_i g = Hx_{\varphi(i)}$.

We may then rewrite the denominator in the expression of $\nu(gA)$ as

$$\sum_{i=1}^m \mu(x_i gA) = \sum_{i=1}^m \mu(h_i x_{\varphi(i)} A) = \sum_{i=1}^m \mu(x_{\varphi(i)} A) = \sum_{j=1}^m \mu(x_j A)\,.$$

For the second equality above we have used the $H$–invariance of $\mu$. We conclude that $\nu$ is $G$–left-invariant.

(3) Let $E$ be a non-empty subset of $G/N$. Theorem 16.21 applied to the action of $G$ on $G/N$ gives a $G$–left-invariant finitely additive measure $\mu : \mathcal{P}(G/N) \to [0, \infty]$ such that $\mu\,(E) = 1$. The same measure is also $G/N$–left-invariant.

(4) Let $h_{ij} : H_i \to H_j, i \leqslant j$, be the homomorphisms defining the direct system of groups $(H_i)$ and let $G$ be the direct limit. Let $h_i : H_i \to G$ be the homomorphisms to the direct limit, as defined in Section 1.1.

Consider a non-empty subset $E$ of $G$. Without loss of generality we may assume that all $h_i(H_i)$ intersect $E$: there exists $i_0$ such that for every $i \geqslant i_0$, $h_i(H_i)$ intersects $E$, and we can restrict to the set of indices $i \geqslant i_0$.

The set of functions

$$\{f : \mathcal{P}(G) \to [0, \infty]\} = \prod_{\mathcal{P}(G)} [0, \infty]$$

is compact according to Tychonoff's theorem (see Remark 7.2, 5).

Note that each group $H_i$ acts naturally on $G$ by left multiplication *via* the homomorphism $h_i : H_i \to G$. For each $i \in I$ let $\mathcal{M}_i$ be the set of $H_i$–left-invariant f.a. measures $\mu$ on $\mathcal{P}(G)$ such that $\mu(E) = 1$. Since $H_i$ is super-amenable, Proposition 16.23, (1), and Theorem 16.21 imply that the set $\mathcal{M}_i$ is non-empty.

Let us prove that $\mathcal{M}_i$ is closed in $\prod_{\mathcal{P}(G)}[0, \infty]$. Let $f : \mathcal{P}(G) \to [0, \infty]$ be an element of $\prod_{\mathcal{P}(G)}[0, \infty]$ in the closure of $\mathcal{M}_i$. Then, for every finite collection $A_1, \dots, A_n$ of subsets of $X$ and every $\epsilon > 0$ there exists $\mu$ in $\mathcal{M}_i$ such that $|f(A_j) - \mu(A_j)| \leqslant \epsilon$ for every $j \in \{1, 2, ..., n\}$. This implies that for every $\epsilon > 0$, $|f(E) - 1| \leqslant \epsilon$,

$$|f(A \sqcup B) - f(A) - f(B)| \leqslant 3\epsilon$$

and

$$|f(gA) - f(A)| \leqslant 2\epsilon,$$

$\forall A, B \in \mathcal{P}(X)$ and $g \in H_i$. By letting $\epsilon \to 0$ we obtain that $f \in \mathcal{M}_i$. Thus, the subset $\mathcal{M}_i$ is indeed closed.

By the definition of compactness, if $\{V_i : i \in I\}$ is a family of closed subsets of a compact space $X$ such that $\bigcap_{j \in J} V_j \neq \emptyset$ for every finite subset $J \subseteq I$, then $\bigcap_{i \in I} V_i \neq \emptyset$.

Consider a finite subset $J$ of $I$. Since $I$ is a directed set, there exists $k \in I$ such that $j \leqslant k, \forall j \in J$. Hence, we have homomorphisms $h_{jk} : H_j \to H_k, \forall j \in J$ , and all homomorphisms $h_j : H_j \to G$ factor through $h_k : H_k \to G$. Thus, $\bigcap_{j \in J} \mathcal{M}_j$ contains $\mathcal{M}_k$, in particular, this intersection is non-empty. It follows from the above that $\bigcap_{i \in I} \mathcal{M}_i$ is non-empty. Every element $\mu$ of this intersection is clearly a f.a. measure such that $\mu(E) = 1$, and $\mu$ is also $G$–left-invariant because

$$G = \bigcup_{i \in I} h_i(H_i).$$

□

In view of Corollary 16.26, Proposition 16.30 and Remarks 16.31 it is natural to consider the class of groups that contains all finite and abelian groups, that is stable with respect to the operations described in Proposition 16.30, plus the one of extension:

Definition 16.32. The class of *elementary amenable groups* $\mathcal{EA}$ is the smallest class of groups containing all finite groups, all abelian groups and closed under direct sums, finite-index extensions, direct limits, subgroups, quotients and extensions

$$1 \to G_1 \to G_2 \to G_3 \to 1,$$

where both $G_1, G_3$ are elementary amenable.

Neither of the two classes of super-amenable and of elementary amenable groups contains the other:

- solvable groups are all elementary amenable, while they are super-amenable only if they are virtually nilpotent;
- there exist Grigorchuk groups of intermediate growth that are not elementary amenable

The following result due to C. Chou (and proved previously for the smaller class of solvable groups by Rosenblatt [**Ros74**]) describes, within the setting of finitely generated groups, the intersection between the two classes, and brings information on the set of elementary amenable groups that are not super-amenable.

Theorem 16.33 ([**Cho80**]). *A finitely generated elementary amenable group either is virtually nilpotent or it contains a free non-abelian subsemigroup.*

## 16.5. Amenability and paradoxical actions

In this section we define amenable actions and amenable groups, and prove that paradoxical behavior is equivalent to non-amenability.

Definition 16.34. (1) A group action $G \curvearrowright X$ is *amenable* if there exists a $G$–invariant f.a.p. measure $\mu$ on $\mathcal{P}(X)$, the set of all subsets of $X$.

(2) A group is *amenable* if the action of $G$ on itself *via* left multiplication is amenable.

Lemma 16.35. *A paradoxical action $G \curvearrowright X$ cannot be amenable.*

Proof. Suppose to the contrary that $X$ admits a $G$-invariant f.a.p. measure $\mu$ and

$$X = X_1 \sqcup \ldots \sqcup X_k \sqcup Y_1 \sqcup \ldots \sqcup Y_m$$

is a $G$-paradoxical decomposition, i.e., for some $g_1, \ldots, g_k, h_1, \ldots, h_m \in G$,

$$g_1(X_1) \sqcup \ldots \sqcup g_k(X_k) = X \text{ and } h_1(Y_1) \sqcup \ldots \sqcup h_m(Y_m) = X\,.$$

Then

$$\mu(X_1 \sqcup \ldots \sqcup X_k) = \mu(Y_1 \sqcup \ldots \sqcup Y_k) = \mu(X),$$

which implies that $2\mu(X) = \mu(X)$, contradicting the assumption that $\mu(X) = 1$. □

Remark 16.36. We will prove in Corollary 16.63 that a finitely-generated group is amenable if and only if it is non-paradoxical.

Example 16.37. If $X$ is a finite set, then every group action $G \curvearrowright X$ is amenable. In particular, every finite group is amenable. Indeed, for a finite set $G$ define $\mu : \mathcal{P}(X) \to [0,1]$ by $\mu(A) = \frac{|A|}{|X|}$, where $|\cdot|$ denotes cardinality of a subset.

Example 16.38. The free group of rank two $F_2$ is non-amenable since $F_2$ is paradoxical, as explained in Chapter 15, Section 15.4.

Yet another equivalent definition for amenability can be formulated using the concept of an *invariant mean*, which is responsible for the terminology 'amenable':

Definition 16.39. (1) A *mean* on a set $X$ is a linear functional $m : \ell^\infty(X) \to \mathbb{C}$ defined on the set $\ell^\infty(X)$ of bounded functions on $X$, with the following properties:

(M1) if $f$ takes values in $[0,\infty)$ then $m(f) \geqslant 0$;

(M2) $m(\mathbf{1}_X) = 1$.

Assume, moreover, that $X$ is endowed with the action of a group $G$, $G \times X \to X, (g,x) \mapsto g\cdot x$. This induces an action of $G$ on the set $\ell^\infty(X)$ of bounded complex-valued functions on $X$ defined by $g \cdot f(x) = f(g^{-1} \cdot x)$.

A mean is called *left-invariant* if $m(g\cdot f) = m(f)$ for every $f \in \ell^\infty(X)$ and $g \in G$.

A special case of the above is when $G = X$ and $G$ acts on itself by left translations.

Proposition 16.40. *A group action $G \curvearrowright X$ is amenable (in the sense of Definition 16.34) if and only if it admits a left-invariant mean.*

PROOF. Given a f.a.p. measure $\mu$ on $X$ one can apply the standard construction of integrals (see [**Rud87**, Chapter 1] or [**Roy68**, Chapter 11]) and define, for any function $f : X \to \mathbb{C}$, $m(f) = \int f\, d\mu$. Since $\mu(X) = 1$, for every bounded function $f$, $m(f) \in \mathbb{C}$. Thus, we obtain a linear functional $m : \ell^\infty(X) \to \mathbb{C}$. If the measure $\mu$ is $G$-invariant then $m$ is also $G$-invariant.

Conversely, given a $G$-invariant mean $m$ on $X$, one defines an invariant f.a.p. measure $\mu$ on $X$ by $\mu(A) = m(\mathbf{1}_A)$.

EXERCISE 16.41. Prove that $\mu$ thus defined is a f.a.p. measure and that $G$-invariance of $m$ implies $G$-invariance of $\mu$.

□

REMARK 16.42. Suppose that in Proposition 16.40, $X = G$ and $G \curvearrowright X$ is the action by left multiplication. Then:

(a) In the above proposition, left-invariance can be replaced by right-invariance.

(b) Moreover, both can be replaced by bi-invariance.

PROOF. (a) It suffices to define $\mu_r(A) = \mu(A^{-1})$ and $m_r(f) = m(f_1)$, where $f_1(x) = f(x^{-1})$.

(b) Let $\mu$ be a left-invariant f.a.p. measure and $\mu_r$ the right-invariant measure in (a). Then for every $A \subseteq X$ define

$$\nu(A) = \int \mu(Ag^{-1}) d\mu_r(g) \,.$$

□

QUESTION 16.43. Suppose that $G$ is a group which admits a mean $m : \ell^\infty(G) \to \mathbb{R}$ that is quasi-invariant, i.e., there exists a constant $\kappa$ such that

$$|m(f \circ g) - m(f)| \leqslant \kappa$$

for all functions $f \in \ell^\infty(G)$ and all group elements $g$. Is it true that $G$ is amenable?

LEMMA 16.44. *Every action $G \curvearrowright X$ of an amenable group $G$ is also amenable.*

PROOF. Choose a point $x \in X$ and define $\nu : \mathcal{P}(X) \to [0, 1]$ by

$$\nu(A) = \mu(\{g \in G \; ; \; gx \in A\}).$$

We leave it to the reader to verify that $\nu$ is again a $G$-invariant f.a.p. measure. □

COROLLARY 16.45. *If $G$ is a group which admits a paradoxical action, then $G$ is non-amenable. In other words, if an amenable group $G$ acts on a space $X$, then $X$ cannot be $G$–paradoxical.*

This corollary and the fact that the sphere $\mathbb{S}^2$ is $O(3)$–paradoxical imply that the group $O(3)$ is not amenable (as an abstract group). More generally, in view of Tits' Alternative, if $G$ is a connected Lie group then either $G$ is solvable or non-amenable (since every non-solvable connected Lie group contains a free nonabelian subgroup).

The converse to Lemma 16.44 is false: The action of any group on a one-point set is clearly amenable, see, however, Proposition 16.58. On the other hand, Glasner and Monod [**GM07**] proved that every countable group admits an amenable faithful action on a set $X$.

A natural question to ask is whether on an amenable group there exists only one invariant finitely additive probability measure. It turns out that this is far from being true:

THEOREM 16.46 (J. Rosenblatt [**Ros76**]). *Let $G$ be a non-discrete $\sigma$-compact locally compact metric group. If $G$ is amenable as a discrete group, then there are $2^{\aleph_0}$ mutually singular $G$-invariant means on $L^\infty(G)$.*

REMARK 16.47. Theorem 16.21 and the Banach-Tarski paradox prove that there exists no $\mathrm{Isom}(\mathbb{R}^3)$–left-invariant finitely additive measure $\mu : \mathcal{P}(\mathbb{R}^3) \to [0,\infty]$ such that the unit ball has positive measure.

PROPOSITION 16.48. (1) *A subgroup of an amenable group is amenable.*

(2) *Let $N$ be a normal subgroup of a group $G$. The group $G$ is amenable if and only if both $N$ and $G/N$ are amenable.*

(3) *The direct limit $G$ (see Section 1.1) of a directed system $(H_i)_{i\in I}$ of amenable groups $H_i$, is amenable.*

PROOF. (1) Let $\mu$ be a f.a.p. measure on an amenable group $G$, and let $H$ be a subgroup. By Axiom of Choice, there exists a subset $D$ of $G$ intersecting each right coset $Hg$ in exactly one point. Then $\nu(A) = \mu(AD)$ defines a left-invariant f.a.p. measure on $H$.

(2) "⇒" Assume that $G$ is amenable and let $\mu$ be a f.a.p. measure on $G$. The subgroup $N$ is amenable according to (1). Amenability of $G/N$ follows from Lemma 16.44, since $G$ acts on $G/N$ by left multiplication.

(2) "⇐" Let $\nu$ be a left-invariant f.a.p. measure on $G/N$, and $\lambda$ a left-invariant f.a.p. measure on $N$. On every left coset $gN$ one defines a f.a.p. measure by $\lambda_g(A) = \lambda(g^{-1}A)$. The $H$–left-invariance of $\lambda$ implies that $\lambda_g$ is independent of the representative $g$, i.e. $gN = g'N \Rightarrow \lambda_g = \lambda_{g'}$.

For every subset $B$ in $G$ define

$$\mu(B) = \int_{G/N} \lambda_g(B \cap gN) d\nu(gN) .$$

Then $\mu$ is a $G$–left-invariant probability measure.

(3) The proof is along the same lines as that of Proposition 16.30, (4). The only difference is that the compact $\prod_{\mathcal{P}(G)}[0,\infty]$ is replaced in this argument by

$$\{f : \mathcal{P}(G) \to [0,1]\} = \prod_{\mathcal{P}(G)} [0,1] .$$

□

COROLLARY 16.49. *Let $G_1$ and $G_2$ be two groups that are co-embeddable in the sense of Definition 3.40. Then $G_1$ is amenable if and only if $G_2$ is amenable.*

COROLLARY 16.50. *Any group containing a free nonabelian subgroup is non-amenable.*

PROOF. Note that every non-abelian free group contains a subgroup isomorphic to $F_2$, free group of rank 2. Now, the statement follows from Proposition 16.48, Part (1), and Example 16.38. □

Corollary 16.51. *A semidirect product $N \rtimes H$ is amenable if and only if both $N$ and $H$ are amenable.*

Proof. The statement follows immediately from Part (2) of the above proposition. □

Corollary 16.52. *1. If $G_i$, $i = 1, \ldots, n$, are amenable groups, then the Cartesian product $G = G_1 \times \ldots \times G_n$ is also amenable.*

*2. Direct sum $G = \oplus_{i \in I} G_i$ of amenable groups is again amenable.*

Proof. 1. The statement follows from inductive application of Corollary 16.51.
2. This is a combination of Part 1 and the fact that $G$ is a direct limit of finite direct products of the groups $G_i$. □

Corollary 16.53. *A group $G$ is amenable if and only if all finitely generated subgroups of $G$ are amenable.*

Proof. The direct part follows from (1). The converse part follows from (3), where, given the group $G$, we let $I$ be the set of all finite subsets in $G$, and for any $i \in I$, $H_i$ is the subgroup of $G$ generated by the elements in $i$. We define the directed system of groups $(H_i)$ by letting $h_{ij} : H_i \to H_j$ be the natural inclusion map whenever $i \subset j$. Then $G$ is the direct limit of the system $(H_i)$ and the assertion follows from Proposition 16.48. □

Corollary 16.54. *Every abelian group $G$ is amenable.*

Proof. Since every abelian group is a direct limit of finitely-generated abelian subgroups, by Part (3) of the above proposition, it suffices to prove the corollary for finitely-generated abelian groups. Amenability of such groups will be proven in Proposition 16.69 as an application of the Følner criterion for amenability. □

Remark 16.55. Even for the infinite cyclic group $\mathbb{Z}$, amenability is nontrivial, it depends on a form of Axiom of Choice, e.g., ultrafilter lemma: One can show that Zermelo–Fraenkel axioms are insufficient for proving amenability of $\mathbb{Z}$.

Corollary 16.56. *Every solvable group is amenable.*

Proof. We argue by induction on the derived length. If $k = 1$ then $G$ is abelian and, hence, are amenable by Corollary 16.54.

Assume that the assertion holds for $k$ and take a group $G$ such that $G^{(k+1)} = \{1\}$ and $G^{(i)} \neq \{1\}$ for any $i \leqslant k$. Then $G^{(k)}$ is abelian and $\bar{G} = G/G^{(k)}$ is solvable with derived length equal to $k$. Whence, by the induction hypothesis, $\bar{G}$ is amenable. This and Proposition 16.48, (2), imply that $G$ is amenable. □

In view of the above results, every elementary amenable group is amenable. On the other hand, all finitely generated groups of intermediate growth are amenable but not elementary amenable.

Example 16.57 (Infinite direct products of amenable groups need not be amenable). Let $F = F_2$ be free group of rank 2. Recall, Corollary 3.49, that $F$ is residually finite, hence, for every $g \in F \setminus \{1\}$ there exists a homomorphism $\varphi_g : F \to \Phi_g$ so that $\varphi_g(g) \neq 1$ and $\Phi_g$ is a finite group. Each $\Phi_g$ is, of course, amenable. Consider the direct product of these finite groups:

$$G = \prod_{g \in F} \Phi_g.$$

Then the product of homomorphisms $\varphi_g : F \to \Phi_g$, defines a homomorphism $\varphi : F \to G$. This homomorphism is injective since for every $g \neq 1$, $\varphi_g(g) \neq 1$. Thus, $G$ cannot be amenable.

The following is a generalization of Proposition 16.48, (2); this proposition also completes the result in Lemma 16.44.

PROPOSITION 16.58. *Let $G$ be a group acting on a set $X$. The group $G$ is amenable if and only if $G \curvearrowright X$ is amenable and for every $p \in X$ the stabilizer* $\mathrm{Stab}(p)$ *of the point $p$ is amenable.*

PROOF. The direct implication follows from Lemma 16.44 and from Proposition 16.48, (1).

Assume now that for every $p \in X$ its $G$-stabilizer $S_p$ is amenable and that $m_X : \ell^\infty(X) \to \mathbb{C}$ is a $G$-invariant mean. By proposition 16.40, for every $p \in X$ there exists a left-invariant mean $m_p : \ell^\infty(S_p) \to \mathbb{C}$.

We define a left-invariant mean on $\ell^\infty(G)$ using a construction in the spirit of Fubini's Theorem.

Let $F \in \ell^\infty(G)$. We split $X$ into $G$–orbits $X = \bigsqcup_{p \in \Re} Gp$.

For every $p \in \Re$ we define a function $F_p$ on the orbit $Gp$ by $F_p(gp) = m_p\left(F|_{gS_p}\right)$. Then we define a function $F_X$ on $X$ which coincides with $F_p$ on each orbit $Gp$.

The fact that $F$ is bounded implies that $F_X$ is bounded. We define

$$m(F) = m_X\left(F_X\right).$$

The linearity of $m$ follows from the linearity of every $m_p$ and of $m_X$. The two properties (M1) and (M2) in Definition 16.39 are easily checked for the mean $m$. We now check that $m$ is left-invariant. Let $h$ be an arbitrary element of $G$, and let $h \cdot F$ be defined by $h \cdot F(x) = F(h^{-1} \cdot x)$, for every $x \in G$.

Then

$$(h \cdot F)_p(gp) = m_p\left((h \cdot F)|_{gS_p}\right) = m_p\left(F|_{h^{-1}gS_p}\right) = F_p(h^{-1}gp).$$

We deduce from this that $(h \cdot F)_X = F_X \circ h^{-1} = h \cdot F_X$, whence

$$m(h \cdot F) = m_X\left((h \cdot F)_X\right) = m_X\left(h \cdot F_X\right) = m_X\left(F_X\right) = m(F).$$

□

COROLLARY 16.59. *Amenability is preserved by virtual isomorphisms of groups.*

PROOF. The only nontrivial part of this statement is: If $H$ is an amenable subgroup of finite index in a group $G$, then $G$ is also amenable. Consider the action of $G$ on $X = G/H$ by left multiplications. Stabilizers of points under this action are conjugates of the group $H$ in $G$, hence, they are amenable. The set $X$ is finite and, hence, the action $G \curvearrowright X$ is amenable. Thus, $G$ is amenable by Proposition 16.58. □

For topological groups and topological group actions one can refine the notion of amenability as follows:

DEFINITION 16.60 (Amenability for topological group actions). 1. Let $G \times X \to X$ be a topological action of a topological group $G$. Then this action is *topologically amenable* if there exists a continuous $G$-invariant linear functional $m$ defined on the space of all Borel measurable bounded functions $X \to \mathbb{C}$, such that:

- $m(f) \geqslant 0$ when $f \geqslant 0$;

- $m(\mathbf{1}_X) = 1$;

Such a linear functional is called an *invariant mean.*

2. A topological group $G$ is said to be *amenable* if the action of $G$ on itself *via* left multiplication is amenable. The corresponding linear functional $m$ is called a *left-invariant mean.*

REMARK 16.61. With this notion, for instance, every compact group is topologically amenable (we can take $m$ to be the integral with respect to a left Haar measure). In particular, the group $SO(3)$ is topologically amenable. On the other hand, as we saw, $SO(3)$ is not amenable as an abstract group. More generally, if $\mathcal{H}$ is a separable Hilbert space and $G = U(\mathcal{H})$ is the group of unitary operators on $\mathcal{H}$ endowed with the weak operator topology, then $G$ is topologically amenable, see [**dlH73**]. We refer to [**BdlHV08**] for further details on topological amenability.

### 16.6. Equivalent definitions of amenability for finitely generated groups

In view of Corollary 16.53, amenability in the case of finitely generated groups is particularly significant. In this case, one can relate the group amenability to the metric amenability for Cayley graphs.

THEOREM 16.62. *Let $G$ be a finitely-generated group. The following are equivalent:*

(1) *$G$ is amenable in the sense of Definition 16.34;*

(2) *one (every) Cayley graph of $G$ is amenable in the sense of Definition 16.1.*

PROOF. According to Theorem 16.10, if one Cayley graph of $G$ is amenable then all the other Cayley graphs are. Thus, in what follows we fix a finite generating set $S$ of $G$, the corresponding Cayley graph $\mathcal{G} = \text{Cayley}(G, S)$, and word metric, and we assume that the statement (2) refers to $\mathcal{G}$.

(2)⇒ (1). We first illustrate the proof in the case $G = \mathbb{Z}$ and the Følner sequence

$$\Omega_n = [-n, n] \subset \mathbb{Z},$$

since the proof is more transparent in this case and illustrates the general argument. Puck a non-principal ultrafilter $\omega$ on $\mathbb{N}$. For a subset $A \subset \mathbb{Z}$ we define a f.a.p. measure $\mu$ by

$$\mu(A) := \omega\text{-lim} \frac{|A \cap \Omega_n|}{2n+1}.$$

Let us show that $\mu$ is invariant under the unit translation $g : z \mapsto z + 1$. Note that

$$||A \cap \Omega_n| - |gA \cap \Omega_n|| \leqslant 1.$$

Thus,

$$|\mu(A) - \mu(gA)| \leqslant \omega\text{-lim} \frac{1}{2n+1} = 0.$$

This implies that $\mu$ is $\mathbb{Z}$-invariant.

We now consider the general case. Since $\mathcal{G}$ is amenable, there exists a Følner sequence of subsets $(\Omega_n) \subset G$ (since $G$ is the vertex set of $\mathcal{G}$). We use the sets $\Omega_i$ to construct a $G$-invariant f.a.p. measure on $\mathcal{P}(G)$. Following Remark 16.42, we can and will use the action to the right of $G$ on itself in this discussion.

Let $\omega$ be a non-principal ultrafilter on $\mathbb{N}$. For every $A \subset G$ define

$$\mu(A) = \omega\text{-lim} \frac{|A \cap \Omega_n|}{|\Omega_n|}.$$

We leave it to the reader to check that $\mu$ is a f.a.p. measure on $G$. Now consider an arbitrary generator $g \in S$. We have

$$|\mu(Ag) - \mu(A)| = \omega\text{-lim} \frac{||Ag \cap \Omega_n| - |A \cap \Omega_n||}{|\Omega_n|} = \omega\text{-lim} \frac{\big||A \cap \Omega_n g^{-1}| - |A \cap \Omega_n|\big|}{|\Omega_n|}.$$

Now $A \cap \Omega_n g^{-1} = \big(A \cap \Omega_n g^{-1} \cap \Omega_n\big) \sqcup \big(A \cap \Omega_n g^{-1} \setminus \Omega_n\big)$. Likewise,

$$A \cap \Omega_n = \big(A \cap \Omega_n \cap \Omega_n g^{-1}\big) \sqcup \big(A \cap \Omega_n \setminus \Omega_n g^{-1}\big).$$

Therefore, the ultralimit above can be rewritten as

$$\omega\text{-lim} \frac{\big||A \cap (\Omega_n g^{-1} \setminus \Omega_n)| - |A \cap (\Omega_n \setminus \Omega_n g^{-1})|\big|}{|\Omega_n|} \leqslant$$
$$\omega\text{-lim} \frac{|A \cap (\Omega_n g^{-1} \setminus \Omega_n)| + |A \cap (\Omega_n \setminus \Omega_n g^{-1})|}{|\Omega_n|}$$
$$= \omega\text{-lim} \frac{|A \cap (\Omega_n g^{-1} \setminus \Omega_n)| + |Ag \cap (\Omega_n g \setminus \Omega_n)|}{|\Omega_n|} \leqslant \omega\text{-lim} \frac{2|E(\Omega_n, \Omega_n^c)|}{|\Omega_n|} = 0.$$

The last equality follows from amenability of the graph $\mathcal{G}$. Therefore, $\mu(Ag) = \mu(A)$ for every $g \in S$. Since $S$ is a generating set of $G$, we obtain the equality $\mu(Ag) = \mu(A)$ for every $g \in G$.

(1)$\Rightarrow$ (2). We prove this implication by proving the contrapositive, that is $\neg(2) \Rightarrow \neg(1)$. We shall, in fact, prove that $\neg(2)$ implies that $G$ is paradoxical.

Assume that $\mathcal{G}$ is non-amenable. According to Theorem 16.3, this implies that there exists a map $f : G \to G$ which is at finite distance from the identity map, such that $|f^{-1}(y)| = 2$ for every $y \in G$. Lemma 5.27 implies that there exists a finite set $\{h_1, ..., h_n\}$ and a decomposition $G = T_1 \sqcup ... \sqcup T_n$ such that $f$ restricted to $T_i$ coincides with the multiplication on the right $R_{h_i}$.

For every $y \in G$ we have that $f^{-1}(y)$ consists of two elements, which we label as $\{y_1, y_2\}$. This gives a decomposition of $G$ into $Y_1 \sqcup Y_2$. Now we decompose $Y_1 = A_1 \sqcup ... \sqcup A_n$, where $A_i = Y_1 \cap T_i$, and likewise $Y_2 = B_1 \sqcup ... \sqcup B_n$, where $B_i = Y_2 \cap T_i$. Clearly $A_1 h_1 \sqcup ... \sqcup A_n h_n = G$ and $B_1 h_1 \sqcup ... \sqcup B_n h_n = G$. We have, thus, proved that $G$ is paradoxical. □

The equivalence in Theorem 16.62 allows to give another proof that the free group on two generators $F_2$ is paradoxical: Consider the map $f : F_2 \to F_2$ which consists in deleting the last letter in every reduced word. This map satisfies Gromov's condition in Theorem 16.3. Hence, the Cayley graph of $F_2$ is non-amenable; thus, $F_2$ is non-amenable as well.

Another consequence of the proof of Theorem 16.62 is the following weaker version of the Tarski's Alternative Theorem 16.21:

Corollary 16.63. *A finitely generated group is either paradoxical or amenable.*

Proof. Indeed, in the proof of Theorem 16.62, we proved that Cayley graph $\mathcal{G}$ of $G$ is amenable if and only if the group $G$ is, and that if $\mathcal{G}$ is non-amenable then $G$ is paradoxical. Thus, we have that group amenability is equivalent not only to the Cayley graph amenability but also to non-paradoxical behavior. □

Note that the above proof uses existence of ultrafilters on $\mathbb{N}$. One can show that ZF axioms of the set theory are insufficient to conclude that $\mathbb{Z}$ has an invariant mean. In particular, for any group $G$ containing an element of infinite order, ZF are not enough to conclude that $G$ admits an invariant mean.

QUESTION 16.64. Is there a finitely-generated infinite group which admits an invariant mean under the ZF axioms in set theory?

COROLLARY 16.65. *Every super-amenable group is amenable.*

LEMMA 16.66. *Let $(\Omega_n)$ be a sequence of subsets of a finitely-generated group $G$. The following are equivalent:*

(1) *$(\Omega_n)$ is a Følner sequence for one of the Cayley graphs of $G$.*

(2) *For every $g \in G$*

$$\lim_{n\to\infty} \frac{|\Omega_n g \,\Delta\, \Omega_n|}{|\Omega_n|} = 0\,. \tag{16.7}$$

(3) *For every element $g$ of a generating set $S$ of $G$,*

$$\lim_{n\to\infty} \frac{|\Omega_n g \,\Delta\, \Omega_n|}{|\Omega_n|} = 0\,. \tag{16.8}$$

PROOF. Let $S$ be a finite generating set that determines a Cayley graph $\mathcal{G}$ of $G$, we will assume that $1 \notin S$. Let $\Omega \subset G$, i.e., $\Omega$ is a subset of the vertex set of $\mathcal{G}$. Then the vertex boundary of $\Omega$ in $\mathcal{G}$ is

$$\partial_V \Omega = \bigcup_{s\in S} \Omega \setminus \Omega s^{-1}\,.$$

Thus, for a sequence $(\Omega_n)$ the equality

$$\lim_{n\to\infty} \frac{|E(\Omega_n, \Omega_n^c)|}{|\Omega_n|} = 0\,.$$

is equivalent to the set of equalities

$$\lim_{n\to\infty} \frac{|\Omega_n \setminus \Omega_n s^{-1}|}{|\Omega_n|} = 0 \text{ for every } s \in S\,,$$

which in its turn is equivalent to (16.8) for every $g \in S^{-1}$. Thus, (1) is equivalent to (3).

It remains to show that (1) implies that (16.7) holds for an arbitrary $g \in G$. In view of Exercise 16.5, if $\Omega_n$ is a Følner sequence for one finite generating set of $G$, the sequence $\Omega_n$ is also Følner for every finite generating set of $G$. By taking a finite generating set of $G$ which contains given $g \in G$, we obtain the desired conclusion. □

DEFINITION 16.67. If $G$ is a group, then a sequence of subsets $\Omega_n \subset G$ satisfying property (16.7) in Lemma 16.66, is called a *Følner sequence for the group $G$*. Note that the form of this definition makes sense even if $G$ is not finitely-generated.

EXERCISE 16.68. Prove that the subsets $\Omega_n = \mathbb{Z}^k \cap [-n, n]^k$ form a Følner sequence for $\mathbb{Z}^k$.

PROPOSITION 16.69. (1) *If* $(\Omega_n)$ *is a Følner sequence in a countable group* $G$ *and* $\omega$ *is a non-principal ultrafilter on* $\mathbb{N}$ *then a left-invariant mean* $m : \ell^\infty(G) \to \mathbb{C}$ *may be defined by*

$$m(f) = \omega\text{-}lim \frac{1}{|\Omega_n|} \sum_{x \in \Omega_n} f(x)$$

.

(2) *For any* $k \in \mathbb{N}$ *the group* $\mathbb{Z}^k$ *has an invariant mean* $m : \ell^\infty\left(\mathbb{Z}^k\right) \to \mathbb{C}$ *is defined by*

$$m(f) = \omega\text{-}lim \frac{1}{(2n+1)^k} \sum_{x \in \mathbb{Z}^k \cap [-n,n]^k} f(x) .$$

PROOF. (1) It suffices to note that $\mu(A) = m(\mathbf{1}_A)$ is the left invariant f.a.p. measure defined in the proof of (2)$\Rightarrow$ (1) above.

(2) is a consequence of (1) and Exercise 16.68. □

We are now able to relate amenable groups to the Banach–Tarski paradox.

PROPOSITION 16.70. (1) *The group of isometries* $\mathrm{Isom}(\mathbb{R}^n)$ *with* $n = 1, 2$ *is amenable.*

(2) *The group of isometries* $\mathrm{Isom}(\mathbb{R}^n)$ *with* $n \geqslant 3$ *is non-amenable.*

PROOF. (1) The group $\mathrm{Isom}(\mathbb{R}^n)$ is the semidirect product of $O(n)$ and $\mathbb{R}^n$. The group $\mathbb{R}^n$ is abelian and, hence, amenable, by Corollary 16.54. Furthermore, $O(1) \cong \mathbb{Z}_2$ is finite and, hence, amenable. The group $O(2)$ contains the abelian subgroup $SO(2)$ of index 2. Hence, $O(2)$ is also amenable. Thus, $\mathrm{Isom}(\mathbb{R}^n)$ $(n \leqslant 2)$ is amenable as a semidirect product of two amenable groups, see Corollary 16.51.

(2) This follows from Corollary 16.45 and Banach-Tarski paradox. □

In many textbooks one finds the following property (first introduced by Følner in [**Fø55**]) as an alternative characterization of amenability. Though it is close to the one provided by Lemma 16.66, we briefly discuss it here, for the sake of completeness.

DEFINITION 16.71. A group $G$ is said to have *the Følner Property* if for every finite subset $K$ of $G$ and every $\epsilon > 0$ there exists a finite non-empty subset $F$ such that for all $g \in K$

$$\frac{|Fg \,\triangle\, F|}{|F|} \leqslant \epsilon . \tag{16.9}$$

REMARK 16.72. The relation (16.9) can be rewritten as

$$\frac{|FK \,\triangle\, F|}{|F|} \leqslant \epsilon , \tag{16.10}$$

where $FK = \{fk \; : \; f \in F \, , \, k \in K\}$.

EXERCISE 16.73. Verify that a group has Følner property if and only if it contains a Følner sequence in the sense of Definition 16.67.

LEMMA 16.74. (1) *In both Definition 16.71 and in the characterization of the Følner Property provided by Lemma 16.66, one can take the action of* $G$ *on the left, i.e.* $\frac{|gF \,\triangle\, F|}{|F|} \leqslant \epsilon$ *in (16.9) etc.*

(2) *When $G$ is finitely generated, it suffices to check Definition 16.71 for a finite generating set.*

Proof. (1) One formulation is equivalent to the other *via* the anti-automorphism $G \to G$ given by the inversion $g \mapsto g^{-1}$.

In Definition 16.71, for every finite subset $K$ and every $\epsilon > 0$ it suffices to apply the property with multiplication on the left to the set $K^{-1} = \{g^{-1} \,;\, g \in K\}$, obtain the set $F$, then take $F' = F^{-1}$. This set will verify $\frac{|F'K \,\Delta\, F'|}{|F'|} \leqslant \epsilon$. The proof for Lemma 16.66 is similar.

(2) Let $S$ be an arbitrary finite generating set of $G$. The general statement implies the one for $K = S$. Conversely, assume that the condition holds for $K = S$. In other words, there exists a sequence $F_n$ of finite subsets of $G$, so that for every $s \in S$,

$$\lim_n \frac{|F_n s \,\Delta\, F_n|}{|F|} = 0.$$

In view of Lemma 16.66, for every $g \in G$

$$\lim_n \frac{|F_n g \,\Delta\, F_n|}{|F|} = 0.$$

Thus, there exists a sequence of finite subsets $F_n$ so that for every $g \in G$ there exists $N = N_g$ so that

$$\forall n \geq N, \quad \frac{|F_n g \,\Delta\, F_n|}{|F|} < \epsilon.$$

Taking $N = \max\{N_g : g \in K\}$, we obtain the required statement with $F = F_N$. □

Corollary 16.75. *A finitely-generated group is amenable if and only if it has Følner property.*

We already know that subgroups of amenable groups are again amenable, below we show how to construct Følner sequences for subgroups directly.

Proposition 16.76. *Let $H$ be a subgroup of an amenable group $G$, and let $(\Omega_n)_{n\in\mathbb{N}}$ be a Følner sequence for $G$. For every $n \in \mathbb{N}$ there exists $g_n \in G$ such that the intersection $g_n^{-1}\Omega_n \cap H = F_n$ form a Følner sequence for $H$.*

Proof. Consider a finite subset $K \subset H$, let $s$ denote the cardinality of $K$. Since $(\Omega_n)_{n\in\mathbb{N}}$ is a Følner sequence for $G$, the ratios

$$\alpha_n = \frac{|\Omega_n K \,\Delta\, \Omega_n|}{|\Omega_n|} \tag{16.11}$$

converge to 0. We partition each subset $\Omega_n$ into intersections with left cosets of $H$:

$$\Omega_n = \Omega_n^{(1)} \sqcup \ldots \sqcup \Omega_n^{(k_n)},$$

where

$$\Omega_n^{(i)} = \Omega_n \cap g_i H, i = 1, \ldots, k_n, \quad g_i H \neq g_j H, \forall i \neq j.$$

Then $\Omega_n K \cap g_i H = \Omega_n^{(i)} K$. We have that

$$\Omega_n K \,\Delta\, \Omega_n = \left(\Omega_n^{(1)} K \,\Delta\, \Omega_n^{(1)}\right) \sqcup \cdots \sqcup \left(\Omega_n^{(k_m)} K \,\Delta\, \Omega_n^{(k_n)}\right).$$

The inequality

$$\frac{|\Omega_n K \,\Delta\, \Omega_n|}{|\Omega_n|} \leqslant \alpha_n$$

implies that there exists $i \in \{1, 2, \ldots, k_n\}$ such that

$$\frac{|\Omega_n^{(i)} K \,\triangle\, \Omega_n^{(i)}|}{|\Omega_n^{(i)}|} \leqslant \alpha_n.$$

In particular, $g_i^{-1}\Omega_n^{(i)} = F_n$, with $F_n \subseteq H$, and we obtain that

$$\frac{|F_n K \,\triangle\, F_n|}{|F_n|} \leqslant \alpha_n.$$

□

Since many examples and counterexamples display a connection between amenability and the algebraic structure of a group, it is natural to ask whether there exists a purely algebraic criterion of amenability. J. von Neumann made the observation that the existence of a free subgroup excludes amenability in the very paper where he introduced the notion of amenable groups, under the name of *measurable groups* [**vN28**]. It is this observation that has led to the following question:

QUESTION 16.77 (the von Neumann-Day problem). Does every non-amenable group contain a free non-abelian subgroup?

The question is implicit in [**vN29**], and it was formulated explicitly by Day [**Day57**, §4].

When restricted to the class of subgroup of Lie groups with finitely many components (in particular, subgroups of $GL(n, \mathbb{R})$), Question 16.77 has an affirmative answer, since, in view of the Tits' alternative (see Chapter 13, Theorem 13.1), every such group without either contains a free non-abelian subgroup or is virtually solvable. Since all virtually solvable groups are amenable, the claim follows.

Note that more can be said about finitely generated amenable subgroups $\Gamma$ of a Lie group $L$ with finitely many connected components than just the fact that $\Gamma$ is virtually solvable. To begin with, according to Theorem 13.78, $\Gamma$ contains a solvable subgroup $\Sigma$ of derived length $\leqslant \delta(L)$ so that $|\Gamma : \Sigma| \leqslant \nu(L)$.

THEOREM 16.78 (Mostow–Tits). *A discrete amenable subgroup* $\Gamma$ *of a Lie group* $L$ *with finitely many components, contains a polycyclic group of index at most* $\eta(L)$.

PROOF. We will prove this theorem for subgroups of $GL(n, \mathbb{C})$ as the general case is obtained *via* the adjoint representation of $L$. Let $G$ denote the Zariski closure of $\Gamma$ in $GL(n, \mathbb{C})$. Then, by Part 1 of Theorem 13.78, $G$ contains a connected solvable subgroup $S$ of derived length at most $\delta = \delta(n)$ and $|G : S| \leqslant \nu = \nu(n)$. Note that, up to conjugation, $S$ is a subgroup of the group $B$ of upper-triangular matrices in $GL(n, \mathbb{C})$, see Proposition 13.36. The intersection $\Lambda := \Gamma \cap B$ is a discrete subgroup of a connected solvable Lie group. Mostow proved in [**Mos57**] that such a group is necessarily polycyclic. Furthermore, he established an upper bound on ranks of quotients $\Lambda^{(k)}/\Lambda^{(k+1)}$. □

When the subgroup $\Gamma < GL(n, \mathbb{C})$ is not discrete, not much is known. We provide below a few examples to illustrate that when one removes the hypothesis of discreteness, the variety of subgroups that may occur is much larger. Since this already occurs in $SL(2, \mathbb{R})$, it is natural to ask the following.

Question 16.79. 1. What are the possible solvable subgroups of $SL(2,\mathbb{R})$? Equivalently, what are the possible subgroups of the group of affine transformations of the real line?

2. What are the possible solvable subgroups of $SL(2,\mathbb{C})$ ?

Examples 16.80. 1. We first note that for all integers $m,n \geqslant 1$, the wreath product $\mathbb{Z}^m \wr \mathbb{Z}^n$ is a subgroup of $SL(2,\mathbb{R})$. Indeed, consider $\mathcal{O}_K$, the ring of integers of a totally-real algebraic extension $K$ of $\mathbb{Q}$ of degree $m$. This ring is a free $\mathbb{Z}$-module with a basis $\omega_1,...,\omega_m$. Let $t_1,...,t_n$ be transcendental numbers that are independent over $\mathbb{Q}$, i.e. for every $i \in \{1,...,n\}$, $t_i$ is transcendental over $\mathbb{Q}(t_1,\dots,t_{i-1},t_{i+1},\dots,t_n)$.

Then the subgroup $G$ of $SL(2,\mathbb{R})$ generated by the following matrices

$$s_1 = \begin{pmatrix} t_1 & 0 \\ 0 & 1 \end{pmatrix}, \dots, s_n = \begin{pmatrix} t_n & 0 \\ 0 & 1 \end{pmatrix},$$

$$u_1 = \begin{pmatrix} 1 & \omega_1 \\ 0 & 1 \end{pmatrix}, \dots, u_m = \begin{pmatrix} 1 & \omega_m \\ 0 & 1 \end{pmatrix}$$

is isomorphic to $\mathbb{Z}^m \wr \mathbb{Z}^n$.

Indeed, $G$ is a semidirect product of its unipotent subgroup consisting of matrices

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \text{ with } x \in O_K(t_1,...,t_n),$$

isomorphic to the direct sum $\bigoplus_{z\in\mathbb{Z}^n} O_K$, and of its abelian subgroup consisting of matrices

$$\begin{pmatrix} t_1^{k_1}\cdots t_n^{k_n} & 0 \\ 0 & 1 \end{pmatrix} \text{ with } (k_1,...,k_n) \in \mathbb{Z}^n .$$

2. Every free metabelian group (see Definition 11.22) is a subgroup of $SL(2,\mathbb{R})$. This follows from the fact that a free metabelian group with $m$ generators appears as a subgroup of $\mathbb{Z}^m \wr \mathbb{Z}^m$, using the Magnus embedding (Theorem 11.26).

3. All the examples above can be covered by the following general statements. Given an arbitrary free solvable group $S$ with derived length $k \geqslant 1$, we have:

- $S$ is a subgroup of $SL\left(2^{k-1},\mathbb{R}\right)$;
- for every $m \in \mathbb{N}$ the wreath product $\mathbb{Z}^m \wr S$ is a subgroup of $SL\left(2^k,\mathbb{R}\right)$.

Indeed, one can construct by induction on $k$ the necessary injective homomorphisms. The initial step for both statements above is represented by the examples 1 and 2. We assume that the second statement is true for $k$ and we deduce that the first statement is true for $k+1$. This implication and the Magnus embedding described in Theorem 11.26 suffice to finish the inductive argument.

Consider the free solvable group $S_{n,k}$ of derived length $k$ with $n$ generators $s_1,...,s_n$. According to the hypothesis, $S_{n,k}$ embeds as a subgroup of $SL\left(2^{k-1},\mathbb{R}\right)$; thus, we will regard $s_1,...,s_n$ as $2^{k-1}\times 2^{k-1}$ real matrices. Let $\mathcal{O}_K$ be the ring of integers of a totally-real algebraic extension of degree $m$, and let $\{\omega_1,...,\omega_m\}$ be a basis of $\mathcal{O}_K$ as a free $\mathbb{Z}$-module.

We consider the subgroup $G$ of $SL\left(2^k, \mathbb{R}\right)$ generated by the following matrices (described by square $2^{k-1} \times 2^{k-1}$ blocks; in particular the notations $I$ and $0$ below signify the identity respectively the zero square $2^{k-1} \times 2^{k-1}$ matrices):

$$\sigma_1 = \begin{pmatrix} s_1 & 0 \\ 0 & I \end{pmatrix}, \dots, \sigma_n = \begin{pmatrix} s_n & 0 \\ 0 & I \end{pmatrix},$$

$$u_1 = \begin{pmatrix} I & \omega_1 I \\ 0 & I \end{pmatrix}, \dots, u_m = \begin{pmatrix} I & \omega_m I \\ 0 & I \end{pmatrix}.$$

The group $G$ is isomorphic to $\mathbb{Z}^m \wr S_{n,k}$: It is a semidirect product of the unipotent subgroup consisting of matrices

$$\begin{pmatrix} I & x \\ 0 & I \end{pmatrix}, \quad \text{with } x \text{ in the group ring } O_K S_{n,k} \simeq \bigoplus_{S_{n,k}} \mathbb{Z}^m,$$

and the subgroup isomorphic to $S_{n,k}$ consisting of matrices

$$\begin{pmatrix} g & 0 \\ 0 & I \end{pmatrix} \quad \text{with } g \in S_{n,k}.$$

REMARK 16.81. Other classes of groups satisfying the Tits' alternative are:

(1) finitely generated subgroups of $GL(n, \mathbb{K})$ for some integer $n \geqslant 1$ and some field $\mathbb{K}$ of finite characteristic [**Tit72**];
(2) subgroups of Gromov hyperbolic groups ([**Gro87**, §8.2.$F$], [**GdlH90**, Chapter 8]);
(3) subgroups of the mapping class group, see [**Iva92**];
(4) subgroups of $Out(F_n)$, see [**BFH00, BFH05, BFH04**];
(5) fundamental groups of compact manifolds of nonpositive curvature, see [**Bal95**].

Hence, for all such groups Question 16.77 has positive answer.

The first examples of finitely-generated non-amenable groups with no (non-abelian) free subgroups were given in [**Ol'80**]. In [**Ady82**] it was shown that the free Burnside groups $B(n, m)$ with $n \geqslant 2$ and $m \geqslant 665$, $m$ odd, are also non-amenable. The first finitely presented examples of non-amenable groups with no (non-abelian) free subgroups were given in [**OS02**].

Still, metric versions of the von Neumann-Day Question 16.77 have positive answers. One of these versions is Whyte's Theorem 16.7 (a graph of bounded geometry is non-amenable if and only if it admits a free action of a free non-Abelian group by bi-Lipschitz maps at finite distance from the identity).

Another metric version of the von Neumann-Day Question was established by Benjamini and Schramm in [**BS97b**]. They proved that:

- *An infinite locally finite simplicial graph $\mathcal{G}$ with positive Cheeger constant contains a tree with positive Cheeger constant.*

  Note that in the result above uniform bound on the valency is not assumed. The definition of the Cheeger constant is considered with the edge boundary.

- *If, moreover, the Cheeger constant of $\mathcal{G}$ is at least an integer $n \geqslant 0$, then $\mathcal{G}$ contains a spanning subgraph, where each connected component is a rooted tree with all vertices of valency $n$, except the root, which is of valency $n+1$.*
- *If $X$ is either a graph or a Riemannian manifold with infinite diameter, bounded geometry and positive Cheeger constant (in particular, if $X$ is the Cayley graph of a paradoxical group) then $X$ contains a bi-Lipschitz embedding of the binary rooted tree.*

Related to the above, the following is asked in [**BS97b**]:

Open question 16.82. Is it true every Cayley graph of every finitely generated group with exponential growth contains a tree with positive Cheeger constant?

Note that the open case is that of amenable non-linear groups with exponential growth.

### 16.7. Quantitative approaches to non-amenability

One can measure "how paradoxical" a group or a group action is *via* the *Tarski numbers*. In what follows, groups are not required to be finitely generated.

Definition 16.83. (1) Given an action of a group $G$ on a set $X$, and a subset $E \subset X$, which admits a $G$–paradoxical decomposition in the sense of Definition 16.17, the *Tarski number of the paradoxical decomposition* is the number $k+m$ of elements of that decomposition.

(2) The *Tarski number* $\mathrm{Tar}_G(X, E)$ is the infimum of the Tarski numbers taken over all $G$-paradoxical decompositions of $E$. We set $\mathrm{Tar}_G(X, E) = \infty$ in the case when there exists no $G$-paradoxical decomposition of the subset $E \subset X$.
   We use the notation $\mathrm{Tar}_G(X)$ for $\mathrm{Tar}_G(X, X)$.

(3) We define the *lower Tarski number* $\mathrm{tar}(G)$ *of a group* $G$ to be the infimum of the numbers $\mathrm{Tar}_G(X, E)$ for all the actions $G \curvearrowright X$ and all the non-empty subsets $E$ of $X$.

(4) When $X = G$ and the action is by left multiplication, we denote $\mathrm{Tar}_G(X)$ simply by $\mathrm{Tar}(G)$ and we call it the *Tarski number of* $G$.

Note that $G$–invariance of the subset $E$ is not required in Definition 16.83.

It is easily seen that $\mathrm{tar}(G) \leqslant \mathrm{Tar}(G)$ for every group $G$.

Of course, in view of the notion of countably paradoxical sets, one could refine the discussion further and use other cardinal numbers besides the finite ones. We do not follow this direction here.

Proposition 16.84. *Let $G$ be a group, $G \curvearrowright X$ be an action and $E \subset X$ be a nonempty subset.*

(1) *If $H$ is a subgroup of $G$ then* $\mathrm{Tar}_G(X, E) \leqslant \mathrm{Tar}_H(X, E)$.

(2) *The lower Tarski number* $\mathrm{tar}(G)$ *of a group is at least two.*
   *Moreover,* $\mathrm{tar}(G) = 2$ *if and only if $G$ contains a free two-generated sub-semigroup $S$.*

Proof. (1) If the subset $E$ does not admit a paradoxical decomposition with respect to the action of $H$ on $X$ then there is nothing to prove. Consider an $H$-paradoxical decomposition

$$E = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

such that

$$E = h_1X_1 \sqcup ... \sqcup h_kX_k = h_1'Y_1 \sqcup ... \sqcup h_m'Y_m,$$

and $k+m = \mathrm{Tar}_H(X,E)$. The above decomposition is paradoxical for the action of $G$ on $X$ as well, hence $\mathrm{Tar}_G(X,E) \leqslant \mathrm{Tar}_H(X,E)$.

(2) The fact that every $\mathrm{Tar}_G(X,E)$ is at least two is immediate.

We prove the direct part of the equivalence.

Assume that $\mathrm{tar}(G) = 2$. Then there exists an action $G \curvearrowright X$, a subset $E$ of $X$ with a decomposition $E = A \sqcup B$ and two elements $g, h \in G$ such that $gA = E$ and $hB = E$. Set $g' := g^{-1}, h' := h^{-1}$. We claim that $g'$ and $h'$ generate a free subsemigroup in $G$. Indeed every non-trivial word $w$ in $g', h'$ cannot equal the identity because, depending on whether its first letter is $g'$ or $h'$, it will have the property that $wE \subseteq A$ or $wE \subseteq B$.

Two non-trivial words $w$ and $u$ in $g', h'$ cannot be equal either. Indeed, without loss of generality we may assume that the first letter in $w$ is $g'$, while the first letter in $u$ is $h'$. Then $wE \subseteq A$ and $uE \subseteq B$, whence $w \neq u$.

We now prove the converse part of the equivalence. Let $x, y$ be two elements in $G$ generating the free sub-semigroup $S$, let $S_x$ be the set of words beginning in $x$ and $S_y$ be the set of words beginning in $y$. Then $S = S_x \sqcup S_y$, with $x^{-1}S_x = S$ and $y^{-1}S_y = S$. □

R. Grigorchuk constructed in [**Gri87**] examples of finitely-generated amenable torsion groups $G$ which are weakly paradoxical, thus answering Rosenblatt's conjecture [**Wag85**, Question 12.9.b]. Thus, every such amenable group $G$ satisfies

$$3 \leqslant \mathrm{tar}(G) < \infty.$$

Question 16.85. What are the possible values of $\mathrm{tar}(G)$ for an amenable group $G$? How different can it be from $\mathrm{Tar}(G)$ ?

We now move on to study values of Tarski numbers $\mathrm{Tar}_G(X)$ and $\mathrm{Tar}(G)$, that is for $G$–paradoxical sets that are moreover $G$–invariant.

Proposition 16.86. *Let $G$ be a group, and let $G \curvearrowright X$ be an action.*

(1) $\mathrm{Tar}_G(X) \geq 4$.

(2) *If $G$ acts freely on $X$ and $G$ contains a free subgroup of rank two, then* $\mathrm{Tar}_G(X) = 4$.

Proof. (1) Since in every paradoxical decomposition of $X$ one must have $k \geqslant 2$ and $m \geqslant 2$, the Tarski number is always at least 4.

(2) The proof of this statement is identical to the one appearing in Chapter 15, Section 15.4, Step 3, for $E = \mathbb{S}^2 \setminus C$. □

Proposition 16.86, (2), has a strong converse, appearing as a first statement in the following proposition.

Proposition 16.87. *1. If* $\mathrm{Tar}_G(X) = 4$*, then* $G$ *contains a non-abelian free subgroup.*

*2. If* $X$ *admits a* $G$*-paradoxical decomposition*

$$X = X_1 \sqcup X_2 \sqcup Y_1 \sqcup \ldots \sqcup Y_m,$$

*then* $G$ *contains an element of infinite order. In particular, if* $G$ *is a torsion group then for every* $G$*-action on a set* $X$*,* $\mathrm{Tar}_G(X) \geq 6$*.*

Proof. 1. By hypothesis, there exists a decomposition

$$X = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2$$

and elements $g_1, g_2, h_1, h_2 \in G$, such that

$$g_1X_1 \sqcup g_2X_2 = h_1Y_1 \sqcup h_2Y_2 = X\,.$$

Set $g := g_1^{-1}g_2$ and $h := h_1^{-1}h_2$; then

$$X_1 \sqcup gX_2 = X, Y_1 \sqcup hY_2 = X. \tag{16.12}$$

This implies that

$$gX_1 \sqcup gY_1 \sqcup gY_2 = X \setminus g(X_2) = X_1$$

and, similarly,

$$hX_1 \sqcup hX_2 \sqcup hY_1 = Y_1.$$

In particular, $gX_1 \subset X_1$, $hY_1 \subset Y_1$. It follows that for every $n \in \mathbb{N}$,

$$g^nX_1 \subseteq X_1, \quad \text{ and } h^nY_1 \subseteq Y_1.$$

It also follows that for every $n \in \mathbb{N}$,

$$g^n(Y_1 \sqcup Y_2) \subseteq g^{n-1}(X_1) \subseteq X_1$$

and that

$$h^n(X_1 \sqcup X_2) \subseteq h^{n-1}(Y_1) \subseteq Y_1.$$

Equations (16.12) also imply that

$$X = g^{-1}X_1 \sqcup X_2 = h^{-1}Y_1 \sqcup Y_2.$$

Furthermore, for every $n \in \mathbb{N}$,

$$g^{-n}(X_2) \subseteq X_2 \quad \text{and} \quad h^{-n}(Y_2) \subseteq Y_2$$

and

$$g^{-n}(Y_1 \sqcup Y_2) \subseteq X_2 \quad \text{and} \quad h^{-n}(X_1 \sqcup X_2) \subseteq Y_2.$$

Now we can apply Lemma 4.37 with $A := Y_1 \sqcup Y_2$ and $B := X_1 \sqcup X_2$; it follows that bijections $g$ and $h$ of $X$ generate a free subgroup $F_2$.

2. Let $g_1, g_2 \in G$ be such that

$$g_1X_1 \sqcup g_2X_2 = X\,.$$

Again, set $g := g_1^{-1}g_2$. The same arguments as in the proof of Part 1 show that for every $n > 0$,

$$g^n(Y_1 \sqcup \ldots \sqcup Y_m) \subseteq X_1.$$

Therefore, $g^n \neq 1$ for all $n > 0$. □

S. Wagon (Theorems 4.5 and 4.8 in [**Wag85**]) proved a stronger form of Proposition 16.87 and Proposition 16.86, part (2):

Theorem 16.88 (S. Wagon). *Let $G$ be a group acting on a set $X$. The Tarski number $\mathrm{Tar}_G(X)$ is four if and only if $G$ contains a free non-abelian subgroup $F$ such that the stabilizer in $F$ of each point in $X$ is abelian.*

As an immediate consequence of Proposition 16.86 is the following

Corollary 16.89. *The Tarski number for the action of $SO(n)$ on the $(n-1)$–dimensional sphere $\mathbb{S}^{n-1}$ is 4, for every $n \geqslant 3$.*

The result on the paradoxical decomposition of Euclidean balls can also be refined, and the Tarski number computed. We begin by noting that the Euclidean unit ball $\mathbb{B}$ in $\mathbb{R}^n$ centered in the origin 0 is never paradoxical with respect to the action of the orthogonal group $O(n)$. Indeed, assume that there exists a decomposition

$$\mathbb{B} = X_1 \sqcup \cdots \sqcup X_n \sqcup Y_1 \sqcup \cdots \sqcup Y_m$$

such that

$$\mathbb{B} = g_1 X_1 \sqcup \cdots \sqcup g_n X_n = h_1 Y_1 \sqcup \cdots \sqcup h_m Y_m$$

with

$$g_1, \ldots, g_n, h_1, \ldots, h_m \in O(n).$$

Then the origin 0 is contained in only one of the sets of the initial partition, say, in $X_1$. It follows that none of the sets $Y_j$ contains 0; hence, neither does the union

$$h_1 Y_1 \sqcup \cdots \sqcup h_m Y_m$$

which contradicts the fact that this union equals $\mathbb{B}$.

The following result was first proved by R. M. Robinson in [**Rob47**].

Proposition 16.90. *The Tarski number for the unit ball $\mathbb{B}$ in $\mathbb{R}^n$ with respect to the action of the group of isometries $G$ of $\mathbb{R}^n$ is 5.*

Proof. We first prove that the Tarski number cannot be 4. Assume to the contrary that there exists a decomposition

$$\mathbb{B} = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2$$

and $g_1, g_2, h_1, h_2 \in G = \mathrm{Isom}(\mathbb{R}^n)$, such that

$$\mathbb{B} = g_1 X_1 \sqcup g_2 X_2 = h_1 Y_1 \sqcup h_2 Y_2.$$

By Proposition 3.62, the elements $g_i$ and $h_j$ are compositions of linear isometries and translations. Since, as we observed above, elements $g_i, h_j$ cannot all belong to $O(n)$, it follows that, say, $g_1$ has a non-trivial translation component:

$$g_1(x) = U_1 x + T_1, \quad U_1 \in O(n), T_1 \neq 0.$$

We claim that $g_2 \in O(n)$ and that $X_2$ contains a closed hemisphere of the unit sphere $\mathbb{S} = \partial\mathbb{B}$.

Indeed, $g_1 X_1 \subset T_1\mathbb{B}$. As $T_1$ is non-trivial, $T_1\mathbb{B} \neq \mathbb{B}$; hence, $T_1\mathbb{S}$ contains no subsets of the form $\{x, -x\}$, where $x$ is a unit vector. Therefore, $T_1\mathbb{B} \cap \mathbb{S}$ is contained in an open hemisphere of the unit sphere $\mathbb{S}$. Since the union $g_1(X_1) \cup g_2(X_2)$ contains the sphere $\mathbb{S}$, it follows that $g_2 X_2$ contains a closed hemisphere in $\mathbb{S}$, and, hence, so does $g_2\mathbb{B}$. Since $g_2\mathbb{B} \subset \mathbb{B}$, it follows that $g_2\mathbb{B} = \mathbb{B}$, hence, $g_2(0) = 0$ and, thus, $X_2$ contains a closed hemisphere of $\mathbb{S}$.

This claim implies that $(Y_1 \sqcup Y_2) \cap \mathbb{S}$ is contained in an open hemisphere of $\mathbb{S}$. By applying the above arguments to the isometries $h_1, h_2$, we see that both $h_1, h_2$ belong to $O(n)$. We then have that

$$\mathbb{S} = h_1(Y_1 \cap \mathbb{S}) \sqcup h_2(Y_2 \cap \mathbb{S}).$$

On the other hand, both $Y_1, Y_2$ and, hence, $h_1(Y_1), h_2(Y_2)$ are contained in open hemispheres of $\mathbb{S}$. Union of two open hemispheres in $\mathbb{S}$ cannot be the entire $\mathbb{S}$. Contradiction. Thus, $\mathrm{Tar}_G(\mathbb{B}) \geq 5$.

We now show that there exists a paradoxical decomposition of $\mathbb{B}$ with five elements. Corollary 16.89 implies that there exist $g_1, g_2, h_1, h_2$ in $SO(n)$ such that

$$\mathbb{S} = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2 = g_1 X_1 \sqcup g_2 X_2 = h_1 Y_1 \sqcup h_2 Y_2$$

As in the proof of Proposition 16.87, we take $g := g_1^{-1} g_2, h := h_1^{-1} h_2$ and obtain

$$\mathbb{S} = X_1 \sqcup X_2 \sqcup Y_1 \sqcup Y_2 = X_1 \sqcup g X_2 = Y_1 \sqcup h Y_2.$$

It follows that for every $\lambda > 0$ the sphere $\lambda\mathbb{S}$ (of radius $\lambda$) has the paradoxical decomposition

$$\lambda\mathbb{S} = \lambda X_1 \sqcup \lambda X_2 \sqcup \lambda Y_1 \sqcup \lambda Y_2 = \lambda X_1 \sqcup g\, \lambda X_2 = \lambda Y_1 \sqcup h\, \lambda Y_2$$

The group $\Gamma := \langle g, h \rangle$ generated by $g$ and $h$ contains countably many nontrivial orthogonal transformations; the fixed-point set of every such transformation is a proper linear subspace in $\mathbb{R}^n$. Therefore, there exists a point $P \in \mathbb{S}$ not fixed by any nontrivial element of $\Gamma$. Let $\Omega$ denote the $\Gamma$-orbit of $P$. Since the action of $\Gamma$ on $\Omega$ is free, the map

$$\gamma \mapsto \gamma P$$

is a bijection $\Gamma \to \Omega$. The group $\Gamma$ is a free group of rank two with free generators $g, h$, hence as in equation (15.1) of Section 15.4, we have the following paradoxical decomposition of the group $\Gamma$:

$$\langle g, h \rangle = \{1\} \sqcup \mathcal{W}_g \sqcup \mathcal{W}_{g^{-1}} \sqcup \mathcal{W}_h \sqcup \mathcal{W}_{h^{-1}}\,,$$

where

$$\Gamma = \mathcal{W}_g \sqcup g\mathcal{W}_{g^{-1}}, \quad \Gamma = \mathcal{W}_h \sqcup h\mathcal{W}_{h^{-1}}.$$

We now replace the original paradoxical decomposition of $\mathbb{S}$ by

$$\mathbb{S} = X_1' \sqcup X_2' \sqcup Y_1' \sqcup Y_2' \sqcup \{P\}$$

where

$$\begin{aligned} X_1' &= (X_1 \setminus \Omega) \sqcup \mathcal{W}_g P, \\ X_2' &= (X_2 \setminus \Omega) \sqcup \mathcal{W}_{g^{-1}} P, \\ Y_1' &= (X_1 \setminus \Omega) \sqcup \mathcal{W}_h P, \\ Y_2' &= (Y_2 \setminus \Omega) \sqcup \mathcal{W}_{h^{-1}} P\,. \end{aligned}$$

Clearly, $X_1' \sqcup g X_2' = Y_1' \sqcup h Y_2' = \mathbb{S}$.

We now consider the decomposition

$$\mathbb{B} = U_1 \sqcup U_2 \sqcup V_1 \sqcup V_2 \sqcup \{P\}\,,$$

where

$$U_1 = \{O\} \sqcup \bigsqcup_{0<\lambda<1} \lambda X_1 \sqcup X_1',$$

$$U_2 = \bigsqcup_{0<\lambda<1} \lambda X_2 \sqcup X_2',$$

$$V_1 = \bigsqcup_{0<\lambda<1} \lambda Y_1 \sqcup Y_1',$$

and

$$V_2 = \bigsqcup_{0<\lambda<1} \lambda Y_2 \sqcup Y_2'.$$

Then $U_1 \sqcup gU_2 = \mathbb{B}$, while $V_1 \sqcup hV_2 \sqcup \{T(P)\} = \mathbb{B}$, where $T$ is the translation sending the point $P$ to the origin $O$. $\square$

Below we describe the behavior of the Tarski number of groups with respect to certain group operations.

PROPOSITION 16.91. (1) *If $H$ is a subgroup of $G$ then* $\mathrm{Tar}(G) \leqslant \mathrm{Tar}(H)$.

(2) *Every paradoxical group $G$ contains a finitely generated subgroup $H$ such that* $\mathrm{Tar}(G) = \mathrm{Tar}(H)$.

(3) *If $N$ is a normal subgroup of $G$ then* $\mathrm{Tar}(G) \leqslant \mathrm{Tar}(G/N)$.

PROOF. (1) If $H$ is amenable then there is nothing to prove. Consider a decomposition

$$H = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

such that

$$H = h_1 X_1 \sqcup ... \sqcup h_k X_k = h_1' Y_1 \sqcup ... \sqcup h_m' Y_m$$

and $k + m = \mathrm{Tar}(H)$.

Let $\mathcal{R}$ be the set of representatives of right $H$–cosets inside $G$. Then $\widetilde{X}_i = X_i\mathcal{R}$, $i \in \{1, 2, ..., k\}$ and $\widetilde{Y}_j = Y_j\mathcal{R}$, $j \in \{1, 2, ..., m\}$ form a paradoxical decomposition for $G$.

(2) Given a decomposition

$$G = X_1 \sqcup ... \sqcup X_k \sqcup Y_1 \sqcup ... \sqcup Y_m$$

such that

$$G = g_1 X_1 \sqcup ... \sqcup g_k X_k = h_1 Y_1 \sqcup ... \sqcup h_m Y_m$$

and $k + m = \mathrm{Tar}(G)$, consider the subgroup $H$ generated by $g_1, ..., g_k, h_1, ..., h_m$. Thus $\mathrm{Tar}(H) \leqslant \mathrm{Tar}(G)$; since the converse inequality is also true, the equality holds.

(3) Set $\overline{Q} := G/N$. As before, we may assume, without loss of generality, that $\overline{Q}$ is paradoxical. Let

$$\overline{Q} = \overline{X}_1 \sqcup ... \sqcup \overline{X}_k \sqcup \overline{Y}_1 \sqcup ... \sqcup \overline{Y}_m$$

be a decomposition such that

$$\overline{Q} = g_1 \overline{X}_1 \sqcup ... \sqcup g_k \overline{X}_k = h_1 \overline{Y}_1 \sqcup ... \sqcup h_m \overline{Y}_m$$

and $k + m = \mathrm{Tar}\left(\overline{Q}\right)$.

Consider an (injective) section $\sigma : \overline{Q} \to G$, for the projection $G \to \overline{Q}$; set $Q := s(\overline{Q})$. Then $G = QN$ and the sets $X_i = \sigma\left(\overline{X}_i\right) N$, $i \in \{1, 2, ..., k\}$ and $Y_j = \sigma\left(\overline{Y}_j\right) N$, $j \in \{1, 2, ..., m\}$ form a paradoxical decomposition for $G$. $\square$

Proposition 16.91, (1), allows to formulate the following quantitative version of Corollary 16.49.

Corollary 16.92. *If two groups are co-embeddable then they have the same Tarski number.*

It is proven in [**Šir76**], [**Ady79**, Theorem VI.3.7] that, for every odd $m \geqslant 665$, two free Burnside groups $B(n;m)$ and $B(k;m)$ of exponent $m$ and with $n \geqslant 2$ and $k \geqslant 2$, are co-embeddable. Thus:

Corollary 16.93. *For every odd $m \geqslant 665$, and $n \geqslant 2$, the Tarski number of a free Burnside groups $B(n;m)$ of exponent $m$ is independent of the number of generators $n$.*

Corollary 16.94. *A group has the Tarski number 4 if and only if it contains a non-abelian free subgroup.*

Proof. If a group $G$ contains a non-abelian free subgroup then the result follows by Proposition 16.86, (1), (2), and Proposition 16.91, (1). If a group $G$ has $\mathrm{Tar}(G) = 4$ then the claim follows from Proposition 16.87. □

Thus, the Tarski number helps to classify the groups that are non-amenable and do not contain a copy of $F_2$. This class of groups is not very well understood and, as noted in the end of Section 16.6, its only known members are "infinite monsters". For torsion groups $G$ as we proved above $\mathrm{Tar}(G) \geqslant 6$. On the other hand, Ceccherini, Grigorchuk and de la Harpe proved:

Theorem 16.95 (Theorem 2, [**CSGdlH98**]). *The Tarski number of every free Burnside group $B(n;m)$ with $n \geqslant 2$ and $m \geqslant 665$, $m$ odd, is at most 14.*

Natural questions, in view of Corollary 16.93, are the following:

Question 16.96. How does the Tarski number of a free Burnside group $B(n;m)$ depend on the exponent $m$? What are its possible values?

Question 16.97 (Question 22 [**dlHGCS99**], [**CSGdlH98**]). What are the possible values for the Tarski numbers of groups? Do they include 5 or 6? Are there groups which have arbitrarily large Tarski numbers?

It would also be interesting to understand how much of the Tarski number is encoded in the large scale geometry of a group. In particular:

Question 16.98. 1. Is the Tarski number of a group $G$ equal to that of its direct product $G \times F$ with an arbitrary finite group $F$?
2. Is the same true when $F$ is an arbitrary amenable group?
3. Is the Tarski number invariant under virtual isomorphisms?

Note that the answers to Questions 16.98 are positive for the Tarski number equal to $\infty$ or 4.

Question 16.99. 1. Is the Tarski number of groups a quasi-isometry invariant?
2. Is it at least true that the existence of an $(L,C)$-quasi-isometry between groups implies that their Tarski number differ at most by a constant $K = K(L,A)$?

The answer to Question 16.98 (Part 1) is, of course, positive for $\mathrm{Tar}(G) = \infty$, but, already, for $\mathrm{Tar}(G) = 4$ this question is equivalent to a well-known open problem below. A group $G$ is called *small* if it contains no free nonabelian subgroups. Thus, $G$ is small iff $\mathrm{Tar}(G) > 4$.

Question 16.100. Is smallness invariant under quasi-isometries of finitely generated groups?

### 16.8. Uniform amenability and ultrapowers

In this section we discuss a *uniform version of amenability* and its relation to ultrapowers of groups.

Recall (Definition 16.71) that a (discrete) group $G$ is *amenable* (has *the Følner Property*) iff for every finite subset $K$ of $G$ and every $\epsilon \in (0,1)$ there exists a finite non-empty subset $F \subset G$ satisfying:

$$|KF \bigtriangleup F| < \epsilon |F|.$$

Definition 16.101. A group $G$ has the *uniform Følner Property* if, in addition, one can bound the size of $F$ in terms of $\epsilon$ and $|K|$, i.e. there exists a function $\phi : (0,1) \times \mathbb{N} \to \mathbb{N}$ such that

$$|F| \leqslant \phi(\epsilon, |K|) \, .$$

Examples 16.102.

1. Nilpotent groups have the uniform Følner property, [**Boż80**].
2. A subgroup of a group with the uniform Følner Property also has this property, [**Boż80**].
3. Let $N$ be a normal subgroup of $G$. The group $G$ has the uniform Følner Property if and only if $N$ and $G/N$ have this property, [**Boż80**].
4. There is an example of a countable (but infinitely generated) group that is amenable but does not satisfy the uniform Følner Property, see [**Wys88**, §*IV*].

Theorem 16.103 (G. Keller [**Kel72**], [**Wys88**]).

1. *If for some non-principal ultrafilter $\omega$ the ultrapower $G^\omega$ has the Følner Property, then $G$ also has the uniform Følner Property.*
2. *If $G$ has the uniform Følner property, then for every non-principal ultrafilter $\omega$, the ultrapower $G^\omega$ also has the uniform Følner property.*

Proof. (1) The group $G$ can be identified with the "diagonal" subgroup $\widehat{G}$ of $G^\omega$, represented by constant sequences in $G$. It follows by Proposition 16.76 that $G$ has the Følner property. Assume that it does not have the uniform Følner property. Then there exists $\varepsilon > 0$ and a sequence of subsets $K_n$ in $G$ of fixed cardinality $k$ such that for every sequence of subsets $\Omega_n \subset G$

$$|K_n \Omega_n \bigtriangleup \Omega_n| < \epsilon |\Omega_n| \Rightarrow \lim_{n \to \infty} |\Omega_n| = \infty.$$

Let $K_\omega = (K_n)^\omega$. According to Lemma 7.32, $K$ has cardinality $k$. Since $G^\omega$ is amenable it follows that there exists a finite subset $U \in G^\omega$ such that $|KU \bigtriangleup U| < \epsilon |U|$. Let $c$ be the cardinality of $U$. According to Lemma 7.32, (3), $U = (U_n)^\omega$, where each $U_n \subset G$ has cardinality $c$. Moreover, $\omega$-almost surely $|KU_n \bigtriangleup U_n| < \epsilon |U_n|$. Contradiction. We, therefore, conclude that $G$ has the uniform Følner Property.

(2) Let $k \in \mathbb{N}$ and $\epsilon > 0$; define $m := \phi(\epsilon, k)$ where $\phi$ is the function in the uniform Følner property of $G$. Let $K$ be a subset of cardinality $k$ in $G^\omega$. Lemma 7.32 implies that $K = (K_n)^\omega$, for some sequence of subsets $K_n \subset G$ of cardinality

$k$. The uniform Følner Property of $G$ implies that there exists $\Omega_n$ of cardinality at most $m$ such that

$$|K_n\Omega_n \bigtriangleup \Omega_n| < \epsilon|\Omega_n|.$$

Let $F := (\Omega_n)^\omega$. The description of $K$ and $F$ given by Lemma 7.32, (1), implies that

$$KF \bigtriangleup F = (K_n\Omega_n \bigtriangleup \Omega_n)^\omega,$$

whence $|KF \bigtriangleup F| < \epsilon|F|$. Since $|F| \leqslant m$ according to Lemma 7.32, (1), the claim follows. □

PROPOSITION 16.104 (G. Keller, [**Kel72**], Corollary 5.9). *Every group with the uniform Følner property satisfies a law.*

PROOF. Indeed, by Theorem 16.103, (2), if $G$ has the uniform Følner Property then any ultrapower $G^\omega$ has the uniform Følner Property. Assume that $G$ does not satisfy any law, i.e., in view of Lemma 7.39, the group $G^\omega$ contains a subgroup isomorphic to the free group $F_2$. By Proposition 16.76 it would then follow that $F_2$ has the Følner Property, a contradiction. □

## 16.9. Quantitative approaches to amenability

One quantitative approach to amenability is due to A.M. Vershik, who introduced in [**Ver82**] the *Følner function*. Given an amenable graph $\mathcal{G}$ of bounded geometry, its *Følner function* $\mathrm{F}_o^{\mathcal{G}} : (0,\infty) \to \mathbb{N}$ is defined by the condition that $\mathrm{F}_o^{\mathcal{G}}(x)$ is the minimal cardinality of a finite non-empty set $F$ of vertices satisfying the inequality

$$|E(F,F^c)| \leqslant \frac{1}{x}|F|\,.$$

According to the inequality (1.1) relating the cardinalities of the vertex and edge boundary, if one replaces in the above $E(F,F^c)$ by the vertex boundary $\partial_V F$ of $F$, one obtains a Følner function asymptotically equal to the first, in the sense of Definition 1.7.

The following is a quantitative version of Theorem 16.10.

PROPOSITION 16.105. *If two graphs of bounded geometry are quasi-isometric then they are either both non-amenable or both amenable and their Følner functions are asymptotically equal.*

PROOF. Let $\mathcal{G}$ and $\mathcal{G}'$ be two graphs of bounded geometry, and let $f : \mathcal{G} \to \mathcal{G}'$ and $g : \mathcal{G}' \to \mathcal{G}$ be two $(L,C)$–quasi-isometries such that $f \circ g$ and $g \circ f$ are at distance at most $C$ from the respective identity maps (in the sense of the inequalities (5.3)). Without loss of generality we may assume that both $f$ and $g$ send vertices to vertices. Let $m$ be the maximal valency of a vertex in either $\mathcal{G}$ or $\mathcal{G}'$.

We begin by some general considerations. We denote by $\alpha$ the maximal cardinality of $B(x,C) \cap V$, where $B(x,C)$ is an arbitrary ball of radius $C$ in either $\mathcal{G}$ or $\mathcal{G}'$. Since both graphs have bounded geometry, it follows that $\alpha$ is finite.

Let $A$ be a finite subset in $V(\mathcal{G})$, let $A' = f(A)$ and $A'' = g(A')$. It is obvious that $|A''| \leqslant |A'| \leqslant |A|$. By hypothesis, the Hausdorff distance between $A''$ and $A$ is $C$, therefore $|A| \leqslant \alpha|A''|$. Thus we have the inequalities

$$\frac{1}{\alpha}|A| \leqslant |f(A)| \leqslant |A|\,, \tag{16.13}$$

and similar inequalities for finite subsets in $V(\mathcal{G}')$ and their images by $g$.

The first part of the statement follows from Theorem 16.10.

Assume now that both $\mathcal{G}$ and $\mathcal{G}'$ are amenable, and let $\mathrm{F}_o^{\mathcal{G}}$ and $\mathrm{F}_o^{\mathcal{G}'}$ be their respective Følner functions. Without loss of generality we assume that both Følner functions are defined using the vertex boundary.

Fix $x \in (0, \infty)$, and let $A$ be a finite subset in $V(\mathcal{G})$ such that $|A| = \mathrm{F}_o^{\mathcal{G}}(x)$ and

$$|\partial_V(A)| \leqslant \frac{1}{x}|A|\,.$$

Let $A' = f(A)$ and $A'' = g(A')$. We fix the constant $R = L(2C+1)$, and consider the set $B = \mathcal{N}_R(A')$. The vertex-boundary $\partial_V(B)$ is composed of vertices $u$ such that $R \leqslant \mathrm{dist}(u, A') < R+1$.

It follows that

$$\mathrm{dist}(g(u), A) \geqslant \mathrm{dist}(g(u), A'') - C \geqslant \frac{1}{L}R - 2C = 1$$

and that

$$\mathrm{dist}(g(u), A) \leqslant L(R+1) + C\,.$$

In particular every vertex $g(u)$ is at distance at most $L(R+1)+C-1$ from $\partial_V(A)$ and it is not contained in $A$. We have thus proved that

$$g\left(\partial_V(B)\right) \subseteq \mathcal{N}_{L(R+1)+C-1}\left(\partial_V(A)\right) \setminus A\,.$$

It follows that if we denote $m^{L(R+1)+C-1}$ by $\lambda$, then we can write, using (16.13),

$$|\partial_V(B)| \leqslant \alpha\left|g\left(\partial_V(B)\right)\right| \leqslant \alpha\lambda\left|\partial_V(A)\right| \leqslant \alpha\lambda\frac{1}{x}|A| \leqslant$$

$$\alpha^2\lambda\frac{1}{x}|A'| \leqslant \alpha^2\lambda\frac{1}{x}|B|\,.$$

We have thus obtained that, for $\kappa = \alpha^2\lambda$ and every $x > 0$, the value $\mathrm{F}_o^{\mathcal{G}'}\left(\frac{x}{\kappa}\right)$ is at most $|B| \leqslant m^R|A'| \leqslant m^R|A| = m^R\,\mathrm{F}_o^{\mathcal{G}}(x)$. We conclude that $\mathrm{F}_o^{\mathcal{G}'} \preceq \mathrm{F}_o^{\mathcal{G}}$.

The opposite inequality $\mathrm{F}_o^{\mathcal{G}} \preceq \mathrm{F}_o^{\mathcal{G}'}$ is obtained similarly. □

Proposition 16.105 implies that, given a finitely generated amenable group $G$, any two of its Cayley graphs have asymptotically equal Følner functions. We will, therefore, write $\mathrm{F}_o^{G}$, for the equivalence class of all these functions.

Definitions 16.106. (1) We say that a sequence $(F_n)$ of finite subsets in a group *realizes the Følner function* of that group if for some generating set $S$, $\mathrm{card}\, F_n = \mathrm{F}_o^{\mathcal{G}}(n)$, where $\mathcal{G}$ is the Cayley graph of $G$ with respect to $S$, and

$$|E(F_n, F_n^c)| \leqslant \frac{1}{n}\,|F_n|\,.$$

(2) We say that a sequence $(A_n)$ of finite subsets in a group *quasi-realizes the Følner function* of that group if $\mathrm{card}\, A_n \asymp \mathrm{F}_o^{G}(n)$ and there exists a constant $a > 0$ and a finite generating set $S$ such that for every $n$,

$$|E(A_n, A_n^c)| \leqslant \frac{a}{n}\,|A_n|\,,$$

where $|E(A_n, A_n^c)|$ is the edge boundary of $A_n$ in the Cayley graph of $G$ with respect to $S$.

LEMMA 16.107. *Let $H$ be a finitely generated subgroup of a finitely generated amenable group $G$. Then $\mathrm{F}_o^H \preceq \mathrm{F}_o^G$.*

PROOF. Consider a generating set $S$ of $G$ containing a generating set $X$ of $H$. We shall prove that for the Følner functions defined with respect to these generating sets, we can write $\mathrm{F}_o^H(x) \leqslant \mathrm{F}_o^G(x)$ for every $x>0$. Let $F$ be a finite subset in $G$ such that $|F| = \mathrm{F}_o^G(x)$ and $|\partial_V F| \leqslant \frac{1}{x}|F|$.

The set $F$ intersects finitely many cosets of $H$, $g_1 H, \dots, g_k H$. In particular $F = \bigsqcup_{i=1}^k F_i$, where $F_i = F \cap g_i H$. We denote by $\partial_V^i F_i$ the set of vertices in $\partial_V F_i$ joined to vertices in $F_i$ by edges with labels in $X$. The sets $\partial_V^i F_i$ are contained in $g_i H$ for every $i \in \{1,2,\dots,k\}$, hence they are pairwise disjoint subsets of $\partial_V F$. We can thus write

$$\sum_{i=1}^k \left|\partial_V^i F_i\right| \leqslant |\partial_V F| \leqslant \frac{1}{x}|F| = \frac{1}{x}\sum_{i=1}^k |F_i|\,.$$

It follows that there exists $i \in \{1,2,\dots,k\}$ such that $\left|\partial_V^i F_i\right| \leqslant \frac{1}{x}|F_i|$. By construction, $F_i = g_i K_i$ with $K_i$ a subset of $H$, and the previous inequality is equivalent to $|\partial_V K_i| \leqslant \frac{1}{x}|K_i|$, where the vertex-boundary $\partial_V K_i$ is considered in the Cayley graph of $H$ with respect to the generating set $X$. We then have that $\mathrm{F}_o^H(x) \leqslant |K_i| \leqslant |F| = \mathrm{F}_o^G(x)$. □

One may ask how do the Følner functions relate to the growth functions, and when do the sequences of balls of fixed centre quasi-realize the Følner function, especially under the extra hypothesis of subexponential growth, see Proposition 16.6.

THEOREM 16.108. *Let $G$ be an infinite finitely generated group.*

(1) $\mathrm{F}_o^G(n) \succeq \mathfrak{G}_G(n)$.

(2) *If the sequence of balls $B(1,n)$ quasi-realizes the Følner function of $G$ then $G$ is virtually nilpotent.*

PROOF. (1) Consider a sequence $(F_n)$ of finite subsets in $G$ that realizes the Følner function of that group, for some generating set $S$. In particular

$$|E(F_n, F_n^c)| \leqslant \frac{1}{n}\,|F_n|\,.$$

We let $\mathfrak{G}$ denote the growth function of $G$ with respect to the generating set $S$.

Inequality (12.2) in Proposition 12.23 implies that

$$\frac{|F_n|}{2dk_n} \leqslant \frac{1}{n}\,|F_n|\,,$$

where $d=|S|$ and $k_n$ is such that $\mathfrak{G}(k_n-1) \leqslant 2|F_n| < \mathfrak{G}(k_n-1)$.

This implies that

$$k_n - 1 \geqslant \frac{n}{2d} - 1 \geqslant \frac{n}{4d}\,,$$

whence,

$$2\,\mathrm{F}_o^G(n) \geqslant \mathfrak{G}\left(k_n-1\right) \geqslant \mathfrak{G}\left(\frac{n}{4d}\right)\,.$$

(2) The inequality in (2) implies that for some $a>0$,

$$|S(1,n+1)| \leqslant \frac{a}{n}\,|B(1,n)|\,.$$

In terms of the growth function, this inequality can be re-written as

$$\frac{\mathfrak{G}(n+1)-\mathfrak{G}(n)}{\mathfrak{G}(n)} \leqslant \frac{a}{n}\,. \tag{16.14}$$

Let $f(x)$ be the piecewise-linear function on $\mathbb{R}_+$ whose restriction to $\mathbb{N}$ equals $\mathfrak{G}$ and which is linear on every interval $[n, n+1], n \in \mathbb{N}$. Then the inequality (16.14) means that for all $x \notin \mathbb{N}$,

$$\frac{f'(x)}{f(x)} \leqslant \frac{a}{x}\,.$$

which, by integration, implies that $\ln|f(x)| \leqslant a\ln|x| + b$. In particular, it follows that $\mathfrak{G}(n)$ is bounded by a polynomial in $n$, whence, $G$ is virtually nilpotent. □

In view of Theorem 16.108, (1), one may ask if there is a general upper bound for the Følner functions of a group, same as the exponential function is a general upper bound for the growth functions; related to this, one may ask how much can the Følner function and the growth function of a group differ. The particular case of the wreath products already shows that there is no upper bound for the Følner functions, and that consequently they can differ a lot from the growth function.

Theorem 16.109 (A. Erschler, [**Ers03**]). *Let $G$ and $H$ be two amenable groups and assume that some representative $F$ of $\mathrm{F}_o^H$ has the property that for every $a > 0$ there exists $b > 0$ so that $aF(x) < F(bx)$ for every $x > 0$.*

*Then the Følner function of the wreath product $A \wr B$ is asymptotically equal to $[\mathrm{F}_o^B(x)]^{\mathrm{F}_o^A(x)}$.*

A. Erschler proved in [**Ers06**] that for every function $f : \mathbb{N} \to \mathbb{N}$, there exists a finitely generated group $G$, which is a subgroup of a group of intermediate growth (in particular, $G$ is amenable) whose Følner function satisfies $\mathrm{F}_o^G(n) \geqslant f(n)$ for all sufficiently large $n$.

## 16.10. Amenable hierarchy

We conclude this chapter with the following diagram illustrating hierarchy of amenable groups:

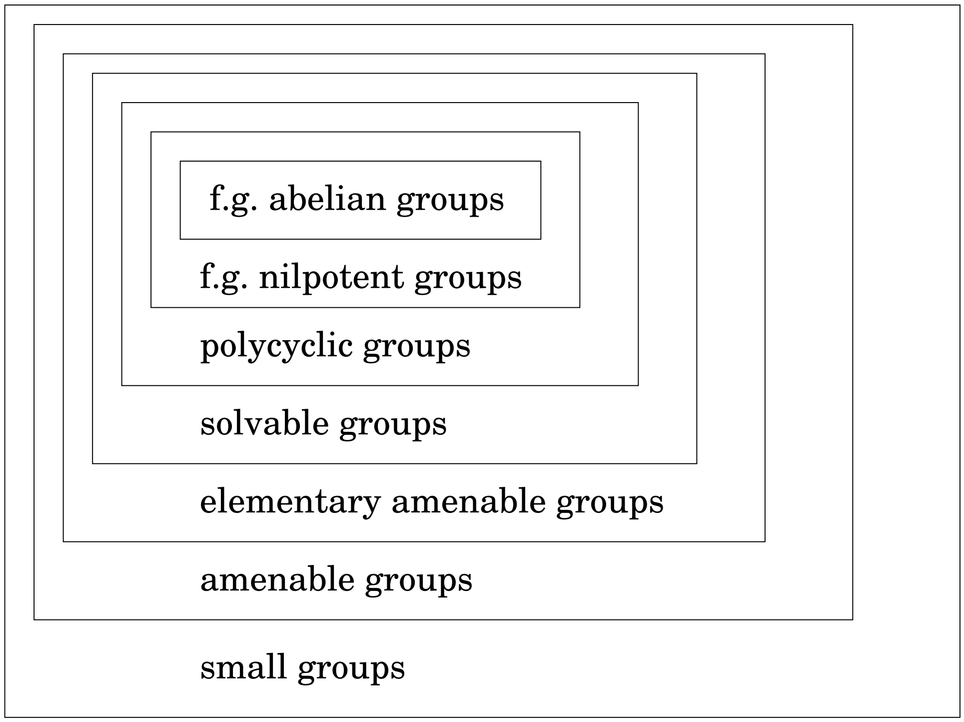



FIGURE 16.1. Hierarchy of amenable groups

CHAPTER 17

# Ultralimits, embeddings and fixed point properties

## 17.1. Classes of spaces stable with respect to (rescaled) ultralimits

DEFINITION 17.1. *Consider a class $\mathcal{C}$ of metric spaces. We say that $\mathcal{C}$ is* stable with respect to ultralimits *if for every set of indices $I$, every nonprincipal ultrafilter $\omega$ on $I$, every collection $(X_i, \mathrm{dist}_i)_{i\in I}$ of metric spaces in $\mathcal{C}$ and every sets of base-points $(e_i)_{i\in I}$ with $e_i \in X_i$, the ultralimit $\omega$-lim$(X_i, e_i, \mathrm{dist}_i)$ is isometric to a metric space in $\mathcal{C}$ .*

*We say that $\mathcal{C}$ is* stable with respect to rescaled ultralimits *if for an arbitrary choice of $I$, $\omega$, $(X_i, \mathrm{dist}_i)_{i\in I}$ and $(e_i)_{i\in I}$ as above, and, moreover, an arbitrary indexed set of positive real numbers $(\lambda_i)_{i\in I}$, the ultralimit of rescaled spaces $\omega$-lim$(X_i, e_i, \lambda_i \mathrm{dist}_i)$ is isometric to a metric space in $\mathcal{C}$ .*

EXAMPLE 17.2. The class of $CAT(0)$ spaces is stable with respect to rescaled ultralimits.

Since in a normed space $V$ the scaling $x \mapsto \lambda x$, $\lambda \in \mathbb{R}_+$, scales the metric by $\lambda$, the space $(V, \lambda\mathrm{dist})$ is isometric to $(V, \mathrm{dist})$, where $\mathrm{dist}(u, v) = \|u - v\|$. Therefore, taking rescaled ultralimits of normed spaces is the same as taking their ultralimits.

In this section we show that certain classes of Banach spaces are stable with respect to ultralimits. It is easy to see that ultralimits of Banach spaces are Banach spaces. Below, we will see that in the class of Banach spaces, Hilbert spaces and $L^p$–spaces can be distinguished by properties that are preserved under ultralimits. The main references for this section are **[LT79]**, **[Kak41]** and **[BDCK66]**.

CONVENTION 17.3. Unless otherwise stated, for every ultralimit of Banach spaces, the base-points are the zero vectors. This assumption is harmless since translations of Banach spaces are isometries.

THEOREM 17.4 (Jordan–von Neumann **[JvN35]**.). *A (real or complex) Banach space $(X, \| \ \|)$ is Hilbert (i.e. the norm $\| \ \|$ comes from an inner product) if and only if for every pair of vectors $x, y \in X$ satisfies the parallelogram identity:*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 .$$

*Note that we do not assume Hilbert spaces to be separable.*

PROOF. We claim that the inner/hermitian product on $X$ is given by the formula:

$$(x, y) := \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2\right) = \frac{1}{4}\sum_{k=0}^{1}(-1)^k\|x + (-1)^k y\|^2, \quad \text{if } X \text{ is real}$$

and

$$(x, y) := \frac{1}{4}\sum_{k=0}^{3} i^k\|x + i^k y\|^2, \quad \text{if } X \text{ is complex,}$$

where $i = \sqrt{-1}$.

Note that it is clear that $(x, x) = \|x\|^2$ (real case), $(x, \bar{x}) = \|x\|^2$ (complex case). We will verify that $(\cdot, \cdot)$ is a hermitian inner product in the complex case; the real case is similar and is left to the reader. We leave it to the reader to show that

$$(ix, y) = (x, -iy) = i(x, y), \quad (x, y) = \overline{(y, x)}$$

and that the parallelogral identity implies that

$$\|u + v\|^2 = -\|u\|^2 + \frac{\|v\|^2}{2} + 2\|u + \frac{1}{2}v\|^2. \tag{17.1}$$

By the definition of $(\cdot, \cdot)$, we have:

$$4(x/2, z) = \sum_{k=0}^{3} i^k \|\frac{x}{2} + i^k z\|^2 =$$

(by applying the equation (17.1) to each term of this sum)

$$\sum_{k=0}^{3} i^k (2\|\frac{x}{2} + i^k \frac{z}{2}\|^2 + \|i^k \frac{z}{2}\|^2) - \|x/2\|^2) =$$

(again, by the definition of $(\cdot, \cdot)$)

$$\sum_{k=0}^{3} i^k (2\|\frac{x}{2} + i^k \frac{z}{2}\|^2 + \|i^k \frac{z}{2}\|^2) = 2(x, z).$$

Thus, $(x/2, z) = (x, z)$ and, clearly,

$$(2x, z) = 2(x, z) \tag{17.2}$$

By the symmetry of $(\cdot, \cdot)$ it follows that

$$(x, 2z) = 2(x, z). \tag{17.3}$$

Instead of proving the multiplicative property for $(\cdot, \cdot)$ for all scalars, we now prove the additivity property of $(\cdot, \cdot)$.

By the definition of $(\cdot, \cdot)$ , we have

$$4(x + y, z) = \sum_{k=0}^{3} \|(x + y) + i^k z\|^2 =$$

(by applying the parallelogram to each term of this sum)

$$\sum_{k=0}^{3} i^k (2\|x + i^k (z/2)\|^2 + \|y + i^k (z/2)\|^2) - \|x - y\|^2) =$$

$$\sum_{k=0}^{3} i^k (2\|x + i^k (z/2)\|^2 + \|y + i^k (z/2)\|^2) = 8(x, z/2) + 8(y, z/2) =$$

(by applying (17.3))

$$4(x, z) + 4(y, z).$$

Thus, $(x + y, z) = (x, z) + (y, z)$.

By applying the additivity property of $(\cdot, \cdot)$ inductively, we obtain

$$(nx, y) = n(x, y), \forall n \in \mathbb{N}.$$

For every $n \in \mathbb{N}$ we also have

$$(x,y) = (n\frac{1}{n}x, y) = n(\frac{1}{n}x, y) \Rightarrow (\frac{1}{n}x, y) = \frac{1}{n}(x,y).$$

Combined with the additivity property, this implies that $(rx, y) = r(x, y)$ holds for all $r \in \mathbb{Q}, r \geqslant 0$. Since $(-x, y) = -(x, y)$, we have the same multiplicative identity for all $r \in \mathbb{Q}$. Note that so far we did not use the triangle inequality in $X$. Observe that the triangle inequality in $X$ implies that for all $x, y \in X$ the function

$$t \mapsto (tx, y) = \frac{1}{4}(\|tx + y\|^2 - \|tx - y\|^2)$$

is continuous. Continuity implies that the identity $(tx, y) = t(x, y)$ holds for all $t \in \mathbb{Q}$. Hence, by the symmetry of $(\cdot,\cdot)$, it follows that $(x, y)$ is indeed an inner product on $X$. ☐

Corollary 17.5. *Every ultralimit of a sequence of Hilbert spaces is a Hilbert space.*

Exercise 17.6. Every closed linear subspace of a Hilbert space is a Hilbert space.

A key feature of Banach function spaces is the existence of an order relation satisfying certain properties with respect to the algebraic operations and the norm.

Definition 17.7. A *Banach lattice* is a real Banach space $(X, \|\ \ \|)$ endowed with a partial order $\leqslant$ such that:

(1) if $x \leqslant y$ then $x + z \leqslant y + z$ for every $x, y, z \in X$;
(2) if $x \geqslant 0$ and $\lambda$ is a non-negative real number then $\lambda x \geqslant 0$;
(3) for every $x, y$ in $X$ there exists a *least upper bound (l.u.b)*, denoted by $x \vee y$, and a *greatest lower bound (g.l.b)*, denoted by $x \wedge y$; this allows to define *the absolute value* of a vector $|x| = x \vee (-x)$;
(4) if $|x| \leqslant |y|$ then $\|x\| \leqslant \|y\|$.

Remarks 17.8. (a) It suffices to require the existence of one of the two bounds in Definition 17.7, (3). Either the relation $x \vee y + x \wedge y = x + y$ or the relation $x \wedge y = -[(-x) \vee (-y)]$ allows to deduce the existence of one bound from the existence of the other.

(b) The conditions (1), (2) and (3) in Definition 17.7 imply that

$$|x - y| = |x \vee z - y \vee z| + |x \wedge z - y \wedge z|. \tag{17.4}$$

This and condition (4) imply that both operations $\vee$ and $\wedge$ on $X$ are continuous.

(c) Condition (4) applied to $x = u$ and $y = |u|$, and to $x = |u|$ and $y = u$ imply that $\|u\| = \|\,|u|\,\|$.

Definition 17.9. A *sublattice* in a Banach lattice $(X, \|\ \ \|, \leqslant)$ is a linear subspace $Y$ of $X$ such that if $y, y'$ are elements of $Y$ then $y \vee y'$ is in $Y$ (hence $y \wedge y' = y + y' - y \vee y'$ is also in $Y$).

Definition 17.10. Two elements $x, y \in X$ of a Banach lattice are called *disjoint* if $x \wedge y = 0$.

Exercise 17.11. Prove that:

(1) For every $p \in [1, \infty)$ and every measure space $(X, \mu)$, the space $L^p(X, \mu)$ with the order defined by

$$f \leqslant g \Leftrightarrow f(x) \leqslant g(x),\ \mu - \text{almost surely},$$

is a Banach lattice.

(2) For every compact Hausdorff topological space $X$, the space $C(K)$ of continuous functions on $X$ with the pointwise partial order and the sup-norm is a Banach lattice.

(3) For both (1) and (2) prove that two functions are disjoint in the sense of Definition 17.10 if and only if both are non-negative functions with disjoint supports (up to a set of measure zero in the first case).

DEFINITION 17.12. Two Banach lattices $X, Y$ are *order isometric* if there exists a linear isometry $T : X \to Y$ which is also an order isomorphism.

PROPOSITION 17.13 (Ultralimits of Banach lattices). *An ultralimit of Banach lattices is a Banach lattice.*

PROOF. Let $(X_i, \| \|_i), i \in I$, be a family of Banach lattices and let $\omega$ be a nonprincipal ultrafilter on $I$. Consider the ultralimit $X_\omega$ endowed with the limit norm $\| \|_\omega$. We will need:

LEMMA 17.14. *Suppose that $a_i, b_i \in X_i$ are such that $u = \omega\text{-}lim\, a_i = \omega\text{-}lim\, b_i$. Then $u = \omega\text{-}lim(a_i \vee b_i) = \omega\text{-}lim(a_i \wedge b_i)$.*

PROOF. Equation (17.4) and Definition 17.7, (4), imply that

$$|x - y| \geqslant |x \vee z - y \vee z| \text{ and } |x - y| \geqslant |x \wedge z - y \wedge z|.$$

These inequalities applied to $x = a_i$ and $y = z = b_i$ imply that $|a_i \vee b_i - b_i| \leqslant |a_i - b_i|$ and $|a_i \wedge b_i - b_i| \leqslant |a_i - b_i|$. Part (4) of Definition 17.7 concludes the proof. □

We define on $X_\omega$ a relation $\leqslant$ as follows:

Points $u, v \in X_\omega$ satisfy $u \leqslant v$ if and only if there exist representatives $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ of $u$ and $v$ (i.e., $u = \omega\text{-lim}\, x_i$ and $v = \omega\text{-lim}\, y_i$) such that $x_i \leqslant y_i$ $\omega$-almost surely.

We now verify that $\leqslant$ is an order. Reflexivity of $\leqslant$ is obvious. Let us check anti-symmetry. If $u \leqslant v$ and $v \leqslant u$ then $u = \omega\text{-lim}\, x_i = \omega\text{-lim}\, x_i'$ and $v = \omega\text{-lim}\, y_i = \omega\text{-lim}\, y_i'$ such that $\omega$-almost surely $x_i \leqslant y_i$ and $y_i' \leqslant x_i'$. The vectors $z_i = x_i - y_i'$ satisfy $z_i \leqslant y_i - y_i'$ and $-z_i \leqslant x_i' - x_i$. This implies that

$$|z_i| \leqslant (y_i - y_i') \vee (x_i' - x_i) \leqslant |y_i - y_i'| \vee |x_i' - x_i| \leqslant |y_i - y_i'| + |x_i' - x_i|.$$

Property (4) in Definition 17.7, the triangle inequality and Remark 17.8, (c), imply that $\|z_i\| \leqslant \|y_i - y_i'\| + \|x_i' - x_i\|$. It follows that $\omega\text{-lim}\, z_i = 0$, hence $u = v$.

We now check transitivity. Let $u = \omega\text{-lim}\, x_i$, $v = \omega\text{-lim}\, y_i = \omega\text{-lim}\, y_i'$, $w = \omega\text{-lim}\, z_i$ be such that $\omega$-almost surely $x_i \leqslant y_i$ and $y_i' \leqslant z_i$. Then $x_i \leqslant z_i + y_i - y_i'$. Since $w = \omega\text{-lim}(z_i + y_i - y_i')$, it follows that $u \leqslant w$.

We will now verify that $X_\omega$ is a Banach lattice with respect to the order $\leqslant$. Properties (1) and (2) in Definition 17.7 are immediate.

Given $u = \omega\text{-lim}\, x_i$ and $v = \omega\text{-lim}\, y_i$ define $u \vee v$ as $\omega\text{-lim}(x_i \vee y_i)$. We claim that $u \vee v$ is well–defined, i.e. does not depend on the choice of representatives for $u$ and $v$. Indeed, assume that $u = \omega\text{-lim}\, x_i'$ and $v = \omega\text{-lim}\, y_i'$ and take $w = \omega\text{-lim}(x_i \vee y_i)$ and $w' = \omega\text{-lim}(x_i' \vee y_i')$. Let $a_i = x_i \wedge x_i'$ and $A_i = x_i \vee x_i'$; likewise, $b_i = y_i \wedge y_i'$

and $B_i = y_i \vee y_i'$. Clearly, $\omega$-$\lim(a_i \vee b_i) \leqslant w \leqslant \omega$-$\lim(A_i \vee B_i)$, and the same for $w'$. The inequalities $a_i \vee b_i \leqslant A_i \vee B_i \leqslant a_i \vee b_i + A_i - a_i + B_i - b_i$ imply that $\omega$-$\lim(a_i \vee b_i) = \omega$-$\lim(A_i \vee B_i)$, hence $w = w'$. We conclude that the vector $u \vee v = \omega$-$\lim(x_i \vee y_i)$ is well–defined. Clearly, $u \vee v \geqslant u$ and $u \vee v \geqslant v$. We need to verify that $u \vee v$ is the l.u.b. for the vectors $u, v$.

Suppose that $z \geqslant u$, $z \geqslant v$, where $u = \omega$-$\lim x_i$, $v = \omega$-$\lim y_i$ and $z = \omega$-$\lim z_i = \omega$-$\lim z_i'$ such that $\omega$-almost surely $z_i \geqslant x_i$ and $z_i' \geqslant y_i$. Lemma 17.14 implies that $z = \omega$-$\lim(z_i \vee z_i')$ and $z_i \vee z_i' \geqslant x_i \vee y_i$, whence, $z \geqslant (u \vee v)$.

Consider now $u, v \in X_\omega$ such that $|u| \leqslant |v|$. It follows that $u = \omega$-$\lim x_i = \omega$-$\lim x_i'$ and $v = \omega$-$\lim y_i = \omega$-$\lim y_i'$, where $x_i \vee (-x_i') \leqslant y_i \vee (-y_i')$. Then

$$|x_i| = x_i \vee (-x_i) \leqslant x_i \vee (-x_i') + |x_i - x_i'| \leqslant |y_i| + |y_i - y_i'| + |x_i - x_i'|.$$

This inequality, part (4) in Definition 17.7 and the triangle inequality imply that

$$\|x_i\| \leqslant \|y_i\| + \|y_i - y_i'\| + \|x_i - x_i'\|.$$

In particular, $\|u\| \leqslant \|v\|$. □

It is a remarkable fact that $L^p$–spaces can be identified, up to an order isometry, in the class of Banach lattices by a simple criterion that we will state below.

Definition 17.15. Let $p \in [1, \infty)$. An *abstract $L^p$–space* is a Banach lattice such that for every pair of disjoint vectors $x, y \in X$,

$$\|x + y\|^p = \|x\|^p + \|y\|^p.$$

Clearly, every space $L^p(X, \mu)$, with $(X, \mu)$ a measure space, is an abstract $L^p$–space. S. Kakutani proved that the converse is also true:

Theorem 17.16 (Kakutani representation theorem [**Kak41**], see also Theorem 3 in [**BDCK66**] and Theorem 1.b.2 in [**LT79**]). *For every $p \in [1, \infty)$ every abstract $L^p$–space is order isometric to a space $L^p(X, \mu)$ for some measure space $(X, \mu)$.*

Corollary 17.17. *For every $p \in [1, \infty)$ any closed sublattice of a space $L^p(X, \mu)$ is order isometric to a space $L^p(Y, \nu)$.*

Corollary 17.18. *Consider an indexed family of spaces $L^{p_i}(X_i, \mu_i)$, $i \in I$, such that $p_i \in [1, \infty)$. If $\omega$ is an ultrafilter on $I$ such that $\omega$-$\lim p_i = p$ then the ultralimit $\omega$-$\lim L^{p_i}(X_i, \mu_i)$ is order isometric to a space $L^p(Y, \nu)$.*

Corollary 17.19. *For fixed $p$, the family of spaces $L^p(X, \mu)$, where $(X, \mu)$ are measure spaces, is stable with respect to (rescaled) ultralimits.*

Remark 17.20. The measure space $(Y, \nu)$ in Corollary 17.18 can be identified with the ultralimit of measure spaces $(X_i, \mu_i)$. We refer to [**Cut01**] and [**War12**] for details of the construction of the *Loeb measure*, which is the ultralimit of the measure spaces $(X_i, \mu_i)$. The same can be done in the context of von Neumann algebras.

## 17.2. Assouad–type theorems

In this section we prove Assouad's Theorem (and some of its generalizations) for a collection of metric spaces stable with respect to ultralimits. The arguments in this section were inspired by arguments in [**BDCK66**, Troisième partie, §2, pp. 252].

In what follows, we suppose that $\rho_\pm : \mathbb{R}_+ \to \mathbb{R}_+$ are two continuous functions such that $\rho_-(x) \leqslant \rho_+(x)$ for every $x \in \mathbb{R}_+$. The following definition generalizes the concept of quasi-isometry between two metric spaces:

DEFINITION 17.21. A $(\rho_-, \rho_+)$–*embedding* of a metric space $(X, \mathrm{dist}_X)$ into a metric space $(Y, \mathrm{dist}_Y)$ is an embedding $\varphi : X \to Y$ such that

$$\rho_-(\mathrm{dist}_X(x,y)) \leqslant \mathrm{dist}_Y(\varphi(x)\,,\,\varphi(y)) \leqslant \rho_+(\mathrm{dist}_X(x,y))\,. \tag{17.5}$$

Assume that $\rho_-$ is the inverse of $\rho_+$ (in particular, both are bijections). A $(\rho_-, \rho_+)$–*transformation* of a metric space $(X, \mathrm{dist}_X)$ is a bijection $\varphi : X \to X$ such that both $\varphi$ and its inverse $\varphi^{-1}$ satisfy the inequalities in (17.5).

If $\rho_+(x) = Lx$ for some $L \geqslant 1$, the corresponding transformation is an *L-bi-Lipschitz transformation.*

THEOREM 17.22. *Let $\mathcal{C}$ be a collection of metric spaces stable with respect to ultralimits.*

*A metric space $(X, \mathrm{dist})$ has a $(\rho_-, \rho_+)$–embedding into a space $Y$ in $\mathcal{C}$ if and only if every finite subset of $X$ has a $(\rho_-, \rho_+)$–embedding into a space in $\mathcal{C}$.*

PROOF. The direct implication is obvious, we will prove the converse. Let $(X, \mathrm{dist})$ be a metric space such that for every finite subset $F$ in $X$ endowed with the induced metric, there exists a $(\rho_-, \rho_+)$–embedding $\varphi_F : F \to Y_F$, where $(Y_F\,, \mathrm{dist}_F)$ is a metric space in $\mathcal{C}$. We fix a base-point $e$ in $X$. In every finite subset of $X$ we fix a base-point $e_F$, such that $e_F = e$ whenever $e \in F$, and we denote $\varphi_F(e_F)$ by $y_F$.

Let $I$ be the collection of all finite subsets of $X$. Let $\mathcal{B}$ be the collection of subsets of $I$ of the form $I_F = \{F' \in I \mid F \subseteq F'\}$, where $F$ is a fixed element of $I$. Then $\mathcal{B}$ is the base of a filter. Indeed:

1. $I_{F_1} \cap I_{F_2} = I_{F_1 \cup F_2}$.
2. For every $F$, $I_F$ contains $F$ and, hence, is non-empty.
3. $I = I_\emptyset \in \mathcal{B}$.

Therefore, it follows from Exercise 7.9 and the Ultrafilter Lemma 7.16 that there exists an ultrafilter $\omega$ on $I$ such that for every finite subset of $F \subset X$, $\omega(I_F) = 1$.

Consider the ultralimits $X_\omega = \omega\text{-lim}(X,\, e,\, \mathrm{dist})$ and $Y_\omega = \omega\text{-lim}(Y_F,\, y_F,\, \mathrm{dist}_F)$. By hypothesis, the space $(Y_\omega, y_\omega, \mathrm{dist}_\omega)$ belongs to the class $\mathcal{C}$.

We have the diagonal isometric embedding $\iota : X \to X_\omega$, $\iota(x) = x_\omega$. Set $X^0_\omega := \iota(X)$. We define a map

$$\varphi_\omega : X^0_\omega \to Y_\omega$$

by $\varphi_\omega(x_\omega) := \omega\text{-lim}\, z_F$, where $z_F = \varphi_F(x)$ whenever $x \in F$, and $z_F = y_F$ when $x \notin F$.

Let us check that $\varphi_\omega$ is a $(\rho_-, \rho_+)$–embedding. Consider two points $x_\omega, x'_\omega$ in $X^0_\omega$. Recall that $\omega(I_{\{x,x'\}}) = 1$ by the definition of $\omega$. Therefore, if $\varphi_\omega(x_\omega) = \omega\text{-lim}\, z_F$ and $\varphi_\omega(x'_\omega) = \omega\text{-lim}\, z'_F$, then $\omega$-almost surely $z_F = \varphi_F(x)$ and $z'_F = \varphi_F(x')$. Hence, $\omega$-almost surely

$$\rho_-(\mathrm{dist}(x,x')) \leqslant \mathrm{dist}_F\,(z_F\,,\, z'_F) \leqslant \rho_+(\mathrm{dist}(x,x'))\,.$$

By passing to the ultralimit we obtain

$$\rho_-(\mathrm{dist}(x,x')) \leqslant \mathrm{dist}_\omega\,(\varphi_\omega(x_\omega)\,,\, \varphi_\omega(x'_\omega)) \leqslant \rho_+(\mathrm{dist}(x,x'))\,.$$

□

The following result first appeared in [**BDCK66**, Troisième partie, §2, pp. 252].

Corollary 17.23. *Let $p$ be an real number in $[1,\infty)$. A metric space $(X,\mathrm{dist})$ has a $(\rho_-,\rho_+)$–embedding into an $L^p$–space if and only if every finite subset of $X$ has such a $(\rho_-,\rho_+)$–embedding.*

Corollary 17.24 (Assouad's Theorem [**WW75**], Corollary 5.6). *Let $p$ be a real number in $[1,\infty)$. A metric space $(X,\mathrm{dist})$ has an isometric embedding into an $L^p$–space if and only if every finite subset of $X$ has such an isometric embedding.*

Note that the same statement holds if one replaces "isometry" by "$(L,C)$–quasi-isometry", with fixed $L \geqslant 1$ and $C \geqslant 0$.

### 17.3. Limit actions, fixed point properties

First, we recall some topological notions that shall be used in what follows. All the topological spaces that we consider are assumed to be Hausdorff. A topological space $X$ is called *$\sigma$-compact* if there exists a sequence of compact subsets $(K_n)_{n\in\mathbb{N}}$ in $X$ such that $X=\bigcup_{n\in\mathbb{N}} K_n$. A special case is that of a topological group which is *compactly generated*, that is, for which there exists a compact subset generating the whole group. Note that a locally compact group with the topological Property (T) (see Definition 17.40 below) is compactly generated, as proved by Y. Shalom in [**Sha00**].

A *second countable topological space* is a topological space which admits a countable base of topology (this is sometimes called *the second axiom of countability*). A second countable space is *separable* (i.e. contains a countable dense subset) and *Lindelöf* (i.e. every open cover has a countable sub-cover). The converse implications are not true in general, but they hold for metric spaces.

A locally compact second countable space is $\sigma$-compact. Moreover a locally compact second countable group has a proper left-invariant metric, see [**Str74**].

The converse is not true in general: For locally compact $\sigma$-compact spaces (even compactly generated groups) second countability may not hold. Nevertheless, for every locally compact $\sigma$-compact group $G$ there exists a compact normal subgroup $N$ such that $G/N$ is second countable [**Com84**, Theorem 3.7].

For compactly generated groups a *limit* of a family of actions may naturally occur in various settings, as noted in [**Gro03**] (see also [**BFGM07**], §3, $c$).

Let $G$ be a compactly generated topological group, and let $K$ be a compact generating subset of $G$. Let $I$ be an ordered set and let $(N_i)_{i\in I}$ be a collection of normal subgroups in $G$ such that if $i\leqslant j$ then $N_i \leqslant N_j$. Each quotient group $G_i=G/N_i$ is compactly generated by $K_i = KN_i/N_i$. The direct system formed by the quotients $G_i$ and the natural projections $G_i \to G_j$ for $i \leqslant j$ defines a direct limit which we denote by $\overline{G}$, this is a group generated by a compact $\overline{K}$, the image of all compact sets $K_i$ under the corresponding homomorphisms.

Let $(X_i,\mathrm{dist}_i), i\in I$, be a family of complete metric spaces on which the group $G_i$ acts by $L$–bi-Lipschitz transformations, $L\geqslant 1$.

Proposition 17.25 (Point-selection theorem). *For each $i$ we define $F_i\subset X_i$, the set of points fixed by $G_i$. Let $x_i \in X_i \setminus F_i$ be a family of base-points.*

*Assume that for some ultrafilter $\omega$ on $I$ either the limit $\omega$-lim $\frac{\mathrm{dist}(x_i,F_i)}{\mathrm{diam}(K_i x_i)}$ is infinite or $F_i = \emptyset$ $\omega$-almost surely. Let $\delta_i$ be either equal to $\frac{\mathrm{dist}(x_i,F_i)}{2\,\mathrm{diam}(K_i x_i)}$ in the first case, or be positive real numbers such that $\omega$-lim $\delta_i = +\infty$ in the second case.*

*Then the group $\overline{G}$ acts by $L$–bi-Lipschitz transformations on some rescaled ultralimit of the form*

$$X_\omega = \omega\text{-}lim(X_i, y_i, \lambda_i \mathrm{dist}_i)\,,\ \ with\ \lambda_i \geqslant \frac{2}{(1+\delta_i)\,\mathrm{diam}(K_i x_i)}\,, \tag{17.6}$$

*such that for every point $z_\omega$ in $X_\omega$ the diameter of $\overline{K} z_\omega$ is at least $1$.*

Proof. *Step 1.* We claim that $\omega$-almost surely there exists a point $y_i$ in $B(x_i, 2\delta_i\,\mathrm{diam}(K_i x_i))$ such that for every point $z$ in the ball $B\left(y_i, \frac{\delta_i\,\mathrm{diam}(K_i y_i)}{2}\right)$ the diameter of $K_i z$ is at least $\frac{\mathrm{diam}(K_i y_i)}{2}$.

Indeed, assume to the contrary that $\omega$-almost surely for every point $y_i$ in $B(x_i, 2\delta_i\,\mathrm{diam}(K_i x_i))$ there exists $z_i$ in the ball $B\left(y_i, \frac{\delta_i\,\mathrm{diam}(K y_i)}{2}\right)$ such that the diameter of $K_i z$ is strictly less than $\frac{\mathrm{diam}(K_i y_i)}{2}$. Let $J \subset I$ be the set of indices such that the above holds, $\omega(J) = 1$, and let $i$ be a fixed index in $J$. In what follows the argument is only for the index $i$ and for simplicity we suppress the index $i$ in our notation.

Set

$$D := 2\delta\,\mathrm{diam}(Kx).$$

Then for every point $y$ in the ball $B(x, D)$, there exists

$$z \in B\left(y, \frac{\delta\,\mathrm{diam}(Ky)}{2}\right)$$

such that $\mathrm{diam}(Kz) < \frac{\mathrm{diam}(Ky)}{2}$. Applied to $y = x$, this gives that there exists

$$x_1 \in B\left(x, \frac{R}{2}\right),$$

with $R = \frac{D}{2}$ such that $\mathrm{diam}(Kx_1) < \frac{\mathrm{diam}(Kx)}{2}$. Applied to $x_1$, the same statement implies that there exists

$$x_2 \in B\left(x_1, \frac{\delta\,\mathrm{diam}(Kx_1)}{2}\right) \subset B\left(x, \frac{R}{2} + \frac{R}{4}\right)$$

such that

$$\mathrm{diam}(Kx_2) < \frac{\mathrm{diam}(Kx_1)}{2} < \frac{\mathrm{diam}(Kx)}{2^2}\,.$$

Assume that we thus found $x_1, x_2, \ldots, x_n$ such that for every $j \in \{1, 2, \ldots, n\}$,

$$x_j \in B\left(x_{j-1}, \frac{\delta\,\mathrm{diam}(Kx_{j-1})}{2}\right) \subset B\left(x, \frac{R}{2} + \frac{R}{4} + \ldots + \frac{R}{2^j}\right)$$

and $\mathrm{diam}(Kx_j) < \frac{\mathrm{diam}(Kx)}{2^j}$. Then, by taking $y = x_n$, we conclude that there exists

$$x_{n+1} \in B\left(x_n, \frac{\delta\,\mathrm{diam}(Kx_n)}{2}\right) \subset B\left(x, \frac{R}{2} + \frac{R}{4} + \ldots + \frac{R}{2^n} + \frac{R}{2^{n+1}}\right)$$

such that

$$\mathrm{diam}(Kx_{n+1}) < \frac{\mathrm{diam}(Kx_n)}{2} < \frac{\mathrm{diam}(Kx)}{2^{n+1}}\,.$$

We thus obtain a Cauchy sequence $(x_n)$ in a complete metric space; therefore, $(x_n)$ converges to a point $u$ in $X$. By construction,

$$\operatorname{diam}(Ku) = 0,$$

and, hence, $u$ is fixed by $K$, hence by the entire group $G$ (since $K$ generates $G$). On the other hand, $\operatorname{dist}(x, u) \leqslant R = \frac{\operatorname{dist}(x,F)}{2}$, where $F$ is the set of points fixed by $G$, a contradiction.

*Step 2.* Thus, $\omega$-almost surely there exists $y_i$ in $B(x_i, 2\delta_i \operatorname{diam}(K_i x_i))$ such that for every point $z$ in the ball $B\left(y_i, \frac{\delta_i \operatorname{diam}(K_i y_i)}{2}\right)$ the diameter of $K_i z$ is at least $\frac{\operatorname{diam}(K_i y_i)}{2}$. Consider the ultralimit

$$X_\omega = \omega\text{-lim}\left(X_i, y_i, \frac{2}{\operatorname{diam}(K_i y_i)} \operatorname{dist}_i\right).$$

Note that (by our choice of $\delta_i$) $X_\omega$ coincides with the ultralimit of the family of sets $B\left(y_i, \frac{\delta_i \operatorname{diam}(K_i y_i)}{2}\right)$ with the same base-points $y_i$ and same scaling constants. The direct limit $\overline{G}$ acts on $X_\omega$ by $L$–bi-Lipschitz transformations. Moreover for every point $z_\omega$ in $X_\omega$ the diameter of $\overline{K} z_\omega$ is at least $1$. □

Proposition 17.25 allows one to prove certain fixed point properties for actions of groups using ultralimits. Let $\mathcal{C}$ be a collection of metric spaces, let $L \geqslant 1$ and let $G$ be a group.

DEFINITION 17.26. We say that a group $G$ has *the fixed point property F$\mathcal{C}$ for L-actions* if every action of $G$ by $L$–bi-Lipschitz transformations on a space $X$ in $\mathcal{C}$ has a global fixed point. Here and in what follows we adopt the convention that for a topological group we consider only topological actions.

Several special cases of this property are important in group theory. We list them below.

Given a (real or complex) Hilbert space $(\mathcal{H}, \langle\ ,\ \rangle)$, its *unitary group* $U(\mathcal{H})$ is the group of unitary linear invertible operators $U : \mathcal{H} \to \mathcal{H}$, i.e. $\langle Ux, Uy\rangle = \langle x, y\rangle$. By Mazur-Ulam theorem for real Hilbert spaces (see e.g. [**FJ03**, p. 6]), every isometry of $\mathcal{H}$ is an affine transformation. Thus, in this case, $\operatorname{Isom}(\mathcal{H})$, the isometry group of $\mathcal{H}$, has the form $U(\mathcal{H}) \rtimes \mathcal{H}$, where the second factor $\mathcal{H}$ is identified to the group of translations on $\mathcal{H}$. In the case of complex Hilbert spaces one also has to add *conjugate-linear* isometries to $U(\mathcal{H})$ in order to get the entire group $\operatorname{Isom}(\mathcal{H})$. Recall that a real-linear map $L : \mathcal{H} \to \mathcal{H}$ of a complex Hilbert space $\mathcal{H}$ is *conjugate-linear* if

$$L(\alpha x) = \bar{\alpha} L(x), \forall x \in \mathcal{H}, \alpha \in \mathbb{C}.$$

An *isometric affine action* of a group $G$ on a Hilbert space $\mathcal{H}$ is a homomorphism $\alpha : G \to \operatorname{Isom}(\mathcal{H})$. For a topological group $G$, such an action is *continuous* if the map $G \times \mathcal{H} \to \mathcal{H}$

$$(g, x) \mapsto \alpha(g)x$$

is continuous.

REMARK 17.27. Instead of continuity of the map $G \times \mathcal{H} \to \mathcal{H}$, it suffices to assume that for every $x \in \mathcal{H}$, the map $g \mapsto \alpha(g)x$ is a continuous map $G \to \mathcal{H}$, see e.g., [**CM70**].

Definition 17.28. A topological group $G$ has *Property* FH if every affine isometric continuous action of $G$ on a Hilbert space has a global fixed point.

Exercise 17.29. Show that a discrete group $G$ has Property FH if and only if $H^1(G, \mathcal{H}_\pi) = 0$ for every unitary representation $\pi : G \to U(\mathcal{H})$. Hint: Use Lemma 3.86.

In view of the fixed-point theorem for isometric group actions on $CAT(0)$ spaces (Theorem 2.42), we obtain:

Corollary 17.30 (A. Guichardet). *A group $G$ has Property FH if and only if every affine isometric continuous action of $G$ on a Hilbert space has a bounded orbit.*

Recall (see Definition 9.22) that a group $G$ has Property FA if every isometric action of $G$ on a real tree has a fixed point.

Here is a link between the two fixed-point notions:

Theorem 17.31 (See e.g. [**BdlHV08**]). *$FH \Rightarrow FA$: Every (discrete) group with Property FH also has Property FA.*

Another interesting connection is between Property FH and isometric group actions on $\mathbb{H}^n$:

Theorem 17.32 (See e.g. [**BdlHV08**]). *If a topological group $G$ has Property FH, then every isometric continuous action of $G$ on $\mathbb{H}^n$ has a fixed point.*

Given these two examples, the reader might wonder if FH implies that every isometric action on a Rips-hyperbolic space has a bounded orbit. It turns out that the answer is negative, as there are infinite hyperbolic groups which have Property FH. The oldest example of infinite hyperbolic groups with Property FH comes from the theory of symmetric spaces. Let $X = \mathbf{H}\mathbb{H}^n, n \geqslant 2$ be the *quaternionic symmetric space* of dimension $\geqslant 2$; it is a rank 1 symmetric space, in particular, $X$ is negatively curved. Then the isometry group $G$ of $X$ is a simple Lie group, which contains uniform lattices $\Gamma$. Every such lattice is a hyperbolic group. On the other hand, the Lie group $G$ has Property FH, so does $\Gamma$. Other interesting examples of infinite hyperbolic groups with Property FH could be found, for instance, in [**BS97a**]. Furthermore, it turns out that in Gromov's model of randomness for groups, for a certain range of a parameter $d$ called *density*, "majority of groups" are infinite hyperbolic with Property FH, see [**Ż03**] (on the other hand, in for values of $d$ varying in other intervals, majority of groups are infinite, hyperbolic and without Property FH, see [**OW11**]).

Below is a generalization of the Property FH in the context of other Banach spaces.

Definition 17.33. A topological group $G$ has *Property $FL^p$*, where $p$ is a real number in $(0, +\infty)$, if every affine isometric continuous action of $G$ on a space $L^p(X, \mu)$ has bounded orbits (for $p > 1$, this is equivalent to the requirement that the action has a global fixed point.)

Note that for discrete groups Property FH is equivalent to the property in Definition 17.26 for $L = 1$ and $\mathcal{C}$ the class of Hilbert spaces; Property $FL^p$ is equivalent to the property in Definition 17.26 for $L = 1$ and $\mathcal{C}$ the class of $L^p$–spaces.

Corollary 17.34. *Let $\mathcal{C}$ be a class of metric spaces stable with respect to ultralimits, and let $G$, $K$, $G_i$, $K_i$ and $\overline{G}$, $\overline{K}$ be defined as in the paragraph preceding Proposition 17.25.*

*If $\overline{G}$ has the Property FC for L–actions then there exists $i_0$ and $\varepsilon > 0$ such that for every $i > i_0$ the group $G_i$ has the Property FC for L–actions. Furthermore, for every L–action of $G_i$ on a space $X_i \in \mathcal{C}$, if $F_i$ is the set of points in $X_i$ fixed by $G_i$, then for every point $x \in X_i$ the diameter of $K_i x$ at least $\varepsilon \mathrm{dist}_i(x, F_i)$ (and, obviously, at most $2\mathrm{dist}_i(x, F_i)$ ).*

Proof. Assume to the contrary that for every $\varepsilon > 0$ and $i_0$ there exists an $i > i_0$ such that $G_i$ has an action on a space $X_i \in \mathcal{C}$ either without fixed points or such that for a point $x_i \in X_i$ the diameter of $K_i x_i$ is $< \varepsilon \mathrm{dist}_i(x, F_i)$. Then one can find a strictly increasing sequence of indexes $i_n$ and a sequence of actions of $G_{i_n}$ on some $X_{i_n} \in \mathcal{C}$ either without fixed points or with points $x_n \in X_{i_n}$ with $\mathrm{diam}(K_{i_n} x_n) < \frac{1}{n}\mathrm{dist}_{i_n}(x_n, F_{i_n})$. Proposition 17.25 then yields a contradiction. □

Remark 17.35. Note that when $G$ is a topological group, one is usually interested in continuous actions.

One may ask in what case, given continuous actions by isometries of a group $G$ on complete metric spaces $(X_i, \mathrm{dist}_i)$, Proposition 17.25 gives a continuous action of $G$ on the ultralimit $X_\omega$. If $G$ is locally compact second countable and compactly generated, there are results (which involve imposing extra conditions on actions) which ensure continuity in the limit, see for instance [**Sha00**, Lemma 6.3] or [**CCS04**, Lemma 4.3]. We have the following partial result, proving that at least some orbit maps for the limit actions are continuous.

Proposition 17.36. *Let $G$ be a topological group locally compact second countable and compactly generated and let $K$ be a compact generating set which is the closure of an open neighborhood of the identity. Suppose that we have:*

1. *$(X_i, \| \,\|_i)$, a sequence of Banach spaces on which the group $G$ acts continuously and isometrically.*
2. *A sequence of subsets $F_i \subset X_i$, and points $x_i \in X_i \setminus F_i$, $y_i \in X_i$ and numbers $\delta_i$, as in Proposition 17.25. In particular, we have the corresponding action by isometries of $G$ on the rescaled ultralimit $X_\omega = \omega\text{-}lim(X_i, y_i, \lambda_i \mathrm{dist}_i)$.*
3. *$f : G \to \mathbb{R}$, a continuous nonnegative function with support in the interior of $K$, satisfying $\int f(g)\, d\mu(g) \leqslant 1$, where $\mu$ is a G-left-invariant Haar measure (see Definition 3.13).*
4. *Point $z_\omega = \omega\text{-}lim\, z_i$, where $z_i = \int f(g) g w_i d\,\mu(g)$ and $w_i$ are points such that $\omega\text{-}lim \frac{\|w_i - y_i\|_i}{\mathrm{diam}\, K y_i} < \infty$.*

*Then the map $G \to X_\omega$ defined by $g \mapsto g z_\omega$, is continuous.*

Proof. Since both the distance on $G$ and that on $X_\omega$ are $G$-invariant, it suffices to prove that the map $K \to X_\omega$ defined by $g \mapsto g z_\omega$ is continuous at 1.

By hypothesis, $\frac{\mathrm{diam}\, K w_i}{\mathrm{diam}\, K y_i} \leqslant M$ $\omega$-almost surely, for some constant $M$. Let $\varepsilon > 0$ be an arbitrary small number. There exists $\delta > 0$ such that if $\mathrm{dist}(1, h) \leqslant \delta$ then

$\|f(h^{-1}g) - f(g)\|_1 \leqslant \frac{\varepsilon}{M}$ . Then for every $i \in I$,

$$\|hz_i - z_i\|_i = \|\int f(g)hgw_i d\,\mu(g) - \int f(g)gw_i d\,\mu(g)\| =$$

$$\|\int f(h^{-1}g)gw_i d\,\mu(g) - \int f(g)gw_i d\,\mu(g)\| \leqslant \|f(h^{-1}g) - f(g)\|_1 \operatorname{diam}(Kw_i) \leqslant$$

$$\varepsilon \operatorname{diam} Ky_i\,.$$

It follows that $\|hz_\omega - z_\omega\|_\omega \leqslant 2\varepsilon$ . □

Below we will give another application of Proposition 17.25, which allows one to compare various (almost) fixed point properties.

Definition 17.37. Given an action of a group $G$ by $L$–bi-Lipschitz transformations on a metric space $X$, $K$ a compact subset in $G$ and $\varepsilon > 0$, a $(K, \varepsilon)$–*almost fixed point in* $X$ is a point $x \in X$ such that the diameter of $Kx$ in $X$ is at most $\varepsilon$ .

We say that $G$ has *the almost fixed point Property* $\alpha F\mathcal{C}$ *for* $L$*–actions* if for every action of $G$ by $L$–bi-Lipschitz transformations on a space $X \in \mathcal{C}$, for every compact subset $K$ in $G$ and every $\varepsilon > 0$, there exists a $(K, \varepsilon)$–almost fixed point in $X$.

A consequence of Proposition 17.25 is the following.

Corollary 17.38. *Let* $\mathcal{C}$ *be a collection of metric spaces stable with respect to rescaled ultralimits, and let* $L \geqslant 1$. *A finitely generated group* $G$ *has the fixed point Property* $F\mathcal{C}$ *for* $L$*–actions if and only if it has the almost fixed point Property* $\alpha F\mathcal{C}$ *for* $L$*–actions.*

Proof. The "only if" part is immediate. We prove the "if" part. Assume to the contrary that a finitely generated group $G$ has the almost fixed point Property $\alpha F\mathcal{C}$ for $L$–actions, but not the fixed point property. It follows that there exists an $L$–action of $G$ on a metric space $X \in \mathcal{C}$ with no fixed point but such that for every compact subset $K$ in $G$ and every $\varepsilon > 0$ there exists a $(K, \varepsilon)$–almost fixed point in $X$. Take $K$ to be a fixed finite set of generators of $G$ and $\varepsilon = \frac{1}{n}$ . Let $x_n \in X$ be a $(K, \frac{1}{n})$–almost fixed point. Proposition 17.25 implies that $G$ admits an $L$–action on a rescaled ultralimit $X_\omega = \omega\text{-lim}(X, x_n, \lambda_n \mathrm{dist})$ such that for every $z_\omega \in X_\omega$ the diameter of $Kz_\omega$ is at least 1. By hypothesis, $X_\omega$ is also in $\mathcal{C}$ , but for $K$ and $\epsilon = \frac{1}{2}$ there exists no $(K, \epsilon)$–almost fixed point in $X_\omega$, a contradiction. □

## 17.4. Property (T)

Our discussion of Property (T) is somewhat sketchy, we refer the reader to **[BdlHV08]** for the missing details. Recall that finitely generated groups, by default, are endowed with discrete topology.

A unitary representation of a topological group $G$ in a Hilbert space $\mathcal{H}$ is a homomorphism $\pi : G \to U(\mathcal{H})$ such that for every $x \in \mathcal{H}$, the map from $G$ to $\mathcal{H}$ defined by $g \mapsto gx$, is continuous.

Definition 17.39. Let $(\pi, \mathcal{H})$ be a unitary representation of a topological group $G$.

(1) Given a subset $S \subseteq G$ and a number $\varepsilon > 0$, a unit vector $x$ in $\mathcal{H}$ is $(S, \varepsilon)$-*invariant* if

$$\sup_{g \in S} \|\pi(g)x - x\| \leqslant \varepsilon \|x\| .$$

(2) The representation $(\pi, \mathcal{H})$ *almost has invariant vectors* if it has $(K, \varepsilon)$–invariants vectors for every compact subset $K$ of $G$ and every $\varepsilon > 0$.

(3) The representation $(\pi, \mathcal{H})$ has *invariant vectors* if there exists a unit vector $x$ in $\mathcal{H}$ such that $\pi(g)x = x$ for all $g \in G$.

Clearly, existence of invariant vectors implies existence of almost invariant vectors. It is a remarkable fact that there are many groups for which the converse holds as well.

Definition 17.40. A topological group $G$ has *Kazhdan's Property* (T) if for every unitary representation $\pi$ of $G$, if $\pi$ has an almost invariant vector, then it also has an invariant vector.

Theorem 17.41. *$G$ has Property (T) if and only if there exists a compact $K \subset G$ and $\epsilon > 0$, so that whenever a unitary representation $\pi$ has a $(K, \varepsilon)$–invariant vector, it has a non-zero invariant vector.*

In view of this theorem, a pair $(K, \varepsilon) \subset G \times \mathbb{R}$ satisfying this theorem is called a *Kazhdan pair* for $G$; the subset $K$ is called a *Kazhdan set*, and the number $\varepsilon$ is called a *Kazhdan constant.* Below we will prove the above theorem for a large class of groups, including all finitely generated groups.

Theorem 17.42. *Let $G$ be a $\sigma$-compact locally compact compactly group. The the following are equivalent:*

*1. $G$ has Property FH.*

*2. $G$ has Property (T).*

*3. There exists a compact $K \subset G$ and $\epsilon > 0$, so that whenever a unitary representation $\pi$ has a $(K, \varepsilon)$–invariant vector, it has a non-zero invariant vector.*

Proof. Clearly, (3) $\Rightarrow$ (2).

We will deduce the implication (1) $\Rightarrow$ (3) from Proposition 17.25. Our proof follows **[Sil]** and **[Gro03]**.

Let $G$ be a finitely generated group with Property FH and assume that it does not satisfy (3). Fix a compact generating set $S$ of $G$. Then, for every $n \in \mathbb{N}$ there exists a unitary representation $\pi_n : G \to U(\mathcal{H}_n)$ with an $(S, \frac{1}{n})$–invariant (unit) vector $x_n$ and no invariant vectors. Let $X_n$ be the unit sphere $\{u \in \mathcal{H}_n \; : \; \|u\| = 1\}$ with the induced path metric $\mathrm{dist}_n$. Proposition 17.25 applied to the sequence of isometric actions of $G$ on $X_n$ and a choice of $\delta_n$ such that $\omega\text{-lim}\, \delta_n = +\infty$ and $\omega\text{-lim}[\delta_n \,\mathrm{diam}(Sx_n)] = 0$, implies that $G$ acts by isometries on a rescaled ultralimit

$$X_\omega = \omega\text{-lim}(X_n, x_n, \lambda_n \mathrm{dist}_n)\,, \text{ with } \lambda_n \geqslant \frac{2}{(1 + 2\delta_n)\,\mathrm{diam}(Sx_n)}\,.$$

Note that $\omega\text{-lim}\, \lambda_n = +\infty$. Moreover, for every point $z_\omega \in X_\omega$ the diameter of $Sz_\omega$ is at least 1. Observe that such an ultralimit $X_\omega$ is a Hilbert space $\mathcal{H}$, where $x_\omega$ is zero. Indeed, completeness of $X_\omega$ follows from Proposition 7.44. In order to define structure of a Hilbert space on $X_\omega$, it suffices to consider ultralimits of 2-dimensional sub-spheres $Y_n$ in $X_n$. Now, suppose that $Y_n$ is a family of 2-dimensional spheres in $\mathbb{R}^3$ passing through a point $p$ with the common tangent space $P$ at $p$, so that the radii of $Y_n$ diverge to infinity. Then, clearly, Gromov–Hausdorff limit of the sequence of spheres $(Y_n, p)$ is the Euclidean plane $P$. It follows that the ultralimit of the sequence $(Y_n, p)$ is isometric to $P$.

We, thus, obtain an action of $G$ by isometries on a Hilbert space $\mathcal{H}$ without a global fixed point, contradicting Property FH.

The implication (2) $\Rightarrow$ (1) relies on the notion of *kernel* which we introduced in section 1.9.

The main source of examples of conditionally negative semidefinite kernels comes from norms in $L^p$–spaces (the case $p = 2$ is covered by Theorem 1.83):

PROPOSITION 17.43 ([**WW75**], Theorem 4.10). *Let $(Y, \mu)$ be a measure space. Let $0 < p \leqslant 2$, and let $E = L^p(Y, \mu)$ be endowed with the norm $\| \cdot \|_p$. Then $\psi : E \times E \to \mathbb{R}$, $\psi(x, y) = \|x - y\|_p^p$ is a conditionally negative semidefinite kernel.*

On the other hand, according to Schoenberg's theorem 1.83, every conditionally negative semidefinite kernels comes from maps to Hilbert spaces.

The connection between square powers of Hilbert norms and $p$-powers of $L^p$ norms is given by the following.

THEOREM 17.44 (Theorems 1 and 7 in [**BDCK66**]). *Let $1 \leqslant p \leqslant q \leqslant 2$.*

(1) *The normed space $(L^q(X, \mu)\, ,\, \| \cdot \|_q)$ can be embedded linearly and isometrically into*
$$(L^p(X', \mu')\, ,\, \| \cdot \|_p)$$
*for some measure space $(X', \mu')$.*

(2) *If $L^p(X, \mu)$ has infinite dimension, then $(L^p(X, \mu)\, ,\, \| \cdot \|_p^\alpha)$ can be embedded isometrically into $(L^q(X', \mu')\, ,\, \| \cdot \|_q)$ for some measured space $(X', \mu')$, if and only if $0 < \alpha \leqslant \frac{p}{q}$.*

Work of Delorme and Guichardet relates Kazhdan's Property (T) and conditionally negative definite kernels:

THEOREM 17.45 ([**Del77**], [**Gui77**], [**dlHV89**], [**CCJ$^+$01**]). *A second countable, locally compact group $G$ has Property (T) if and only if every continuous left-invariant conditionally negative definite kernel on $G$ is bounded (as a function).*

This and the above results on kernels, allow to relate Property (T) to actions on $L^p$–spaces. As in the case of the Hilbert spaces, an action by affine isometries of a topological group $G$ on a Banach space $B$ (or a subset of it) is called *continuous* if for every vector $v \in B$ the orbit map $g \mapsto gv$ from $G$ to $B$ is continuous.

COROLLARY 17.46 ([**Del77**], [**AW81**], [**WW75**]). *Let $G$ be a second countable, locally compact group. If $G$ has Property (T), then for every $p \in (0, 2]$, every continuous action by isometries of $G$ on a subset of a space $L^p(X, \mu)$ has bounded orbits.*

PROOF. If $G$ acts by isometries on a subset $A$ of a space $L^p(X, \mu)$ then for any $a \in A$ the map $\psi(g, h) = \|g \cdot a - h \cdot a\|_p^p$ is a continuous left invariant conditionally negative definite kernel on $G$. $\square$

In view of this corollary, every continuous isometric action of a group $G$ with Property (T) on a Hilbert space has bounded orbits. By Corollary 17.30, every such action has a fixed point. Thus, $G$ has the Property FH. Theorem 17.42 follows. $\square$

REMARK 17.47. Yves de Cornulier constructed in **[dC06]** examples of uncountable discrete groups with Property FH that do not satisfy Property (T).

Note that (topological) amenability and (topological) Property (T) are incompatible in the class of noncompact groups according to the following result:

THEOREM 17.48 (See Theorem 1.1.6, **[BdlHV08]**). *Let $G$ be a locally compact group. The following properties are equivalent:*

(1) *$G$ is topologically amenable and has topological Property (T);*

(2) *$G$ is compact.*

PROOF. We will use yet another characterization of topologically amenable groups:

PROPOSITION 17.49. *A locally compact group $G$ is amenable if and only if the action of $G$ on $L^2(G,\mu)$ via left multiplication has almost invariant vectors. Here $\mu$ is a left Haar measure on $G$.*

PROOF. We will prove only the direct implication (needed for the proof of Theorem 17.48) and only in the case of groups with discrete topology. Let $F_i \subset G$ be a Følner sequence. Let $f_i = \frac{1}{N_i}\mathbf{1}_{F_i}$, where $N_i := |F_i|^{1/2}$ and $\mathbf{1}_{F_i}$ denotes the characteristic function of $F_i$. Then for every $g \in G$

$$\|g(f_i) - f_i\|^2 \leqslant \frac{|gF_i \,\Delta\, F_i|}{|F_i|},$$

which converges to zero by the definition of a Følner sequence. □

Suppose that $G$ satisfies the topological Property (T). Thus, there exists a nonzero $G$-invariant vector $f \in L^2(G,\mu)$; the function $f$, hence, is constant. Since $f \in L^2(G,\mu)$, it follows that $G$ has finite total measure. However, $G$-invariance of $\mu$ then implies that $G$ is compact. □

**Further properties of groups with Property (T).**

In view of equivalence of Property (T) and Property FH, it is clear that every quotient group of a group with Property (T) also has Property (T). Since a (discrete) amenable group has property (T) if and only if such group is finite, it follows that every amenable quotient of a group with Property (T) has to be finite. In particular, every discrete group with Property (T) has finite abelianization. For instance, free groups and surface groups never have Property (T). On the other hand, we will see below that, unlike amenability, Property (T) is not inherited by subgroups.

LEMMA 17.50. *Property (T) is a VI-invariant.*

PROOF. 1. Suppose that a group $H$ has Property (T) and $G$ is a group containing $H$ as a finite index subgroup. Suppose that $G \curvearrowright \mathcal{H}$ is an isometric affine action of $G$ on a Hilbert space. Since $H$ has Property (T), there exists $x \in hh$ fixed by $H$. Therefore, the $G$-orbit of $x$ is finite. Therefore, by Theorem 2.42, $G$ fixes a point in $\mathcal{H}$ as well.

2. Suppose that $H \leqslant G$ is a finite index subgroup and $G$ has Property (T). Let $H \curvearrowright \mathcal{H}$ be an isometric affine action. Define the *induced* action $Ind_H^G$ of $G$ on the space $V$:

$$V = \{\phi : G \to \mathcal{H} : \phi(gh^{-1}) = h\phi(g), \forall h \in H, g \in G\}.$$

Every such function is, of course, determined by its values on $\{g_1, \ldots, g_n\}$, coset representatives for $G/H$. The group $G$ acts on $V$ by the left multiplication $g : \phi(x) \mapsto \phi(gx)$. Therefore, as a vector space, $V$ is naturally isomorphic to the $n$-fold sum of $\mathcal{H}$. We equip $V$ with the inner product

$$\langle \phi, \psi \rangle := \sum_{i=1}^{n} \langle \phi(g_i), \psi(g_i) \rangle ,$$

making it a Hilbert space. We leave it to the reader to verify that the action of $G$ on $V$ is affine and isometric. The initial Hilbert space $\mathcal{H}$ embeds diagonally in $V$; this embedding is a $H$-equivariant, linear and isometric. Since $G$ has Property (T), it has a fixed vector $\psi \in V$. Therefore, the orthogonal projection of $\psi$ to the diagonal in $V$ is fixed by $H$. Hence, $H$ also has Property (T).

3. Consider a short exact sequence

$$1 \to F \to G \to H \to 1.$$

If $G$ has property (T), then so does $H$ (as a quotient of $G$).

Conversely, suppose that $H$ and $F$ both have Property (T) (we will use it in the case where $F$ is a finite group). Consider an affine isometric action $G \curvearrowright \mathcal{H}$ on a Hilbert space. Since $F$ has Property (T), it has nonempty fixed-point set $V \subset \mathcal{H}$. Then $V$ is a closed affine subspace in $\mathcal{H}$, which implies that $V$ (with the restriction of the metric from $\mathcal{H}$) is isometric to a Hilbert space. The group $G$ preserves $V$ and the affine isometric action $G \curvearrowright V$ factors through the group $H$. Since $H$ has Property (T), it has a fixed point $v \in V$. Thus, $v$ is fixed by the entire group $G$. In particular, every co-extension of a group with Property (T) with finite kernel, also has Property (T).

Putting all these facts together, we conclude that Property (T) is invariant under virtual isomorphisms. ☐

Moreover (see e.g. [**BdlHV08**]):

THEOREM 17.51. *Let $G$ be a locally compact group and $\Gamma < G$ is a lattice. Then $G$ has Property (T) if and only if $\Gamma$ does.*

Examples and non-examples of groups with Property (T):

| **Groups with Property (T)** | **Groups without Property (T)** |
|---|---|
| All simple Lie groups of rank $\geqslant 2$ | $O(n,1)$ and $U(n,1)$ |
| Lattices in simple Lie groups of rank $\geqslant 2$ | Unbounded subgroups of $O(n,1)$ and $U(n,1)$ |
| $SL(n,\mathbb{Z}), n \geqslant 3$ | $SL(2,\mathbb{Z})$ |
| Lattices in the isometry group of $\mathbb{HH}^n, n \geqslant 2$ | Lattices in $O(n,1)$ and $U(n,1)$ |
| $SL(n,\mathbb{Z}[t]), n \geqslant 3$ | All Mapping class groups |
| | Thompson group |
| | All finitely generated infinite Coxeter groups |
| | Infinite 3-manifold groups |
| Some hyperbolic groups | Some hyperbolic groups |
| | Groups which admit nontrivial splittings as amalgams |
| | Infinite amenable groups |
| | Infinite fundamental groups of closed conformally-flat manifolds |

Property (T) is unclear for the following groups:

- $Out(F_n)$, $n \geqslant 3$.
- Infinite Burnside groups.
- Diffeomorphism groups of various smooth manifolds.
- Symplectomorphism groups.
- Groups of volume-preserving diffeomorphisms of smooth manifolds.
- $SL(\infty, \mathbb{R})$.
- Infinite permutation group $S_\infty$.
- Shephard groups. (Property (T) fails at least for some of these groups.)
- Generalized van Dyck groups. (Property (T) fails at least for some of these groups.)
- Automorphism group of a rooted tree of constant valence.
- Hyperbolic Kähler groups. (Property (T) fails at least for some of these groups.)
- Infinite fundamental groups of closed manifolds which admit flat (real or complex) projective structure. (Property (T) fails at least for some of these groups.)

## 17.5. Failure of quasi-isometric invariance of Property (T)

THEOREM 17.52. *The Property (T) is not a QI invariant.*

PROOF. This theorem should be probably attributed to S. Gersten and M. Ramachandran; the example below is a variation on the *Raghunathan's example* discussed in [**Ger92**].

Let $\Gamma$ be a hyperbolic group which satisfies Property (T) and such that $H^2(\Gamma, \mathbb{Z})$ is nontrivial. To construct such a group, start for instance with an infinite hyperbolic group $F$ satisfying Property (T) which has an aspherical presentation complex (see for instance [**BS97a**] for the existence of such groups). Then $H^1(F, \mathbb{Z}) = 0$ (since $F$ satisfies (T)), if $H^2(F, \mathbb{Z}) = 0$, we add more random relations to $F$, keeping the resulting groups $F'$ hyperbolic, infinite, 2-dimensional. Then $H^1(F', \mathbb{Z}) = 0$ since $F'$ also satisfies (T). For large number of relators we get a group $\Gamma = F'$ such that $\chi(\Gamma) > 0$ (the number of relators is larger than the number of generators), hence $H^2(\Gamma, \mathbb{Z}) \neq 0$. Now, pick a nontrivial element $\omega \in H^2(\Gamma, \mathbb{Z})$ and consider a central extension

$$1 \to \mathbb{Z} \to G \to \Gamma \to 1$$

with the extension class $\omega$. Since the group $\Gamma$ is hyperbolic, theorem Theorem 9.113 implies that the groups $G$ and $G' := \mathbb{Z} \times \Gamma$ are quasi-isometric (see also [**Ger92**] for a more general version of this argument in the case of central co-extensions defined by *bounded cohomology classes*). The group $G'$ does not satisfy (T), since it surjects to $\mathbb{Z}$. On the other hand, the group $G$ satisfies (T), see [**dlHV89**, 2.c, Theorem 12]. □

## 17.6. Properties $FL^p$

Throughout the section $p$ is a real number in $(0, +\infty)$ and $G$ is a topological group.

In view of Corollary 17.46, for every locally compact second countable group, Property $FH$ implies Property $FL^p$ for every $p$ in $(0, 2]$. Moreover:

Theorem 17.53 (See [?]). *For every* $p \geqslant 2$,

$$FL^p \Rightarrow FH.$$

On the other hand:

Theorem 17.54 ([**BFGM07**], [**CDH10**]). *Let* $G$ *be a second countable locally compact group. If* $G$ *has Property* $FL^p$ *for some* $p \in (0,2]$, *then* $G$ *has Property (T).*

It follows that Property (T) is equivalent to any of the properties $FL^p$ with $p \in (0,2]$.

A natural question to ask is whether this equivalence can be extended beyond 2.

Theorem 17.55 (D. Fisher and G. Margulis, see [**BFGM07**], §3.$c$). *For every group* $G$ *with Property (T) there exists* $\varepsilon = \varepsilon(G)$ *such that* $G$ *has Property* $FL^p$ *for every* $p \in [1, 2+\varepsilon)$.

This result generalizes as follows:

Theorem 17.56. *Let* $G$ *be a finitely generated group. The set* $\mathcal{FP}_G$ *of* $p \in (0,\infty)$ *such that* $G$ *has Property* $FL^p$, *is open.*

Proof. 1. Consider $p \in \mathcal{FP}_G$. If $p < 2$, then by Theorem 17.54, $(0,2) \subset \mathcal{FP}_G$, so $p$ belongs to the interior of $\mathcal{FP}_G$.

2. We shall prove that the set of $p \in [2,\infty)$ such that $G$ does not have Property $FL^p$ is closed. Indeed, let $(p_n)$ be a sequence in $[2,\infty)$ converging to $p < \infty$, such that for every $n$, $G$ has an action on a space $L^{p_n}(X_n, \mu_n)$ without a fixed point. Proposition 17.25 and Corollary 17.18 imply that $G$ also acts on a space $L^p(Y,\nu)$ without a fixed point. □

For $p$ much larger than 2, the converse of the implication $FL^p \Rightarrow FH$ is not true in general. P. Pansu proved in [**Pan95**] that the group $G = Sp(n,1)$ ($n \geqslant 2$) of isometries of the quaternionic-hyperbolic space $\mathbb{HH}^n$, while satisfying Property (T), does not have Property $FL^p$ for $p > 4n+2$. Y. de Cornulier, R. Tessera and A. Valette later proved in [**dCTV08**] that any simple algebraic group of rank one over a local field has a proper action on a space $L^p(X,\mu)$ for $p$ large enough. For the group $G = Sp(n,1)$ such a proper action exists for any $p > 4n+2$.

Every infinite hyperbolic group $G$ (in particular, any hyperbolic group with Property (T)) has a fixed-point-free isometric action on $\ell^p(G)$ for sufficiently large $p$ (depending on $G$), hence do not have Property $FL^p$, due to results of M. Bourdon and H. Pajot [**BP03**]. Such groups even have proper actions on $L^p$–spaces for $p$ large enough according to the b later work of G. Yu [**Yu05**].

A consequence of this is that the family of properties $FL^p$ achieves separation between the semisimple Lie groups of rank one and the semisimple Lie groups with all factors of rank at least 2 (and their respective lattices). By the above results, all cocompact rank one lattices do not have Property $FL^p$ for $p$ large enough. On the other hand, lattices in simple Lie groups of higher rank have Property $FL^p$ for all $p \in (1,\infty)$, by results of Bader, Furman, Gelander and Monod [**BFGM07**].

Another natural variation on the Property (T) is given by replacing the unitary representations in Definition 17.40 with linear isometric actions on $L^p$–spaces. It turns out that for every $1 < p < \infty$ this $L^p$ version of Property (T) (using almost invariant vectors) is equivalent to the original Property (T), see [**BFGM07**, Theorem A]. Hence, this definition is no longer equivalent to the Property $FL^p$ for certain $p$ and groups $G$.

In view of Theorem 17.56, one can ask the following natural questions:

QUESTION 17.57. 1. Given a finitely generated group $G$ with Property (T), is the open set $\mathcal{FP}_G$ connected?

2. In case $\mathcal{FP}_G$ is bounded, does its supremum have any geometric significance ?

Note that the results of Pansu and Cornulier-Tessera-Valette quoted above, suggest that it does.

Below is yet another application of limits of actions:

THEOREM 17.58. *Let $G$ be a finitely generated group with Property $FL^p$ , for some $p > 1$. Then $G$ can be written as $H/N$, where $H$ is a finitely presented group with Property $FL^p$ and $N$ is a normal subgroup in $H$.*

REMARK 17.59. As noted before, for $p \in (0, 2]$ Property $FL^p$ is equivalent to Property (T). In this case the theorem was proved by Y. Shalom [**Sha00**, Theorem p. 5], answering a question of R. Grigorchuk and A. Zuk.

PROOF. Consider an infinite presentation of $G$, $G = \langle S \mid r_1, \ldots, r_n, \ldots \rangle$ , where $S$ is a finite set generating $G$ and $(r_i)$ is a sequence of relators in $S$. Let $F(S)$ be the free group in the alphabet $S$ and $N_i$ the normal closure in $F(S)$ of the finite set $\{r_1, \ldots, r_i\}$ . The groups $G_i = F(S)/N_i$ are all finitely presented, and form a direct system whose direct limit is $G$. Assume that none of these groups has Property $FL^p$ . It follows that for each $i$ there exists a space $L^p(Y_i, \mu_i)$ and an affine isometric action of $G_i$ on $L^p(Y_i, \mu_i)$ without a fixed point. Proposition 17.25 and Corollary 17.18 imply that $G$ acts by affine isometries and without a global fixed point on a space $L^p(Z, \nu)$ , contradicting the hypothesis. □

**Other generalizations.**

**a-T-menability.** Let $G$ be a locally compact topological group. The group $G$ is said to be *a-T-menable* or *have Haagerup property* if there exists a proper isometric continuous action $G \curvearrowright \mathcal{H}$ of $G$ on a Hilbert space $\mathcal{H}$. In this context, *proper action* means that for every bounded subset $B \subset \mathcal{H}$ the set

$$\{g \in G : gB \cap B \neq \emptyset\}$$

is relatively compact. Clearly, for such action, point-stabilizers $G_x \leqslant G, x \in \mathcal{H}$ have to be compact. In particular: *If a group if a-T-menable and has Property (T) then $G$ is compact.* It is known that every amenable group is also a-T-menable [**CCJ$^+$01**]. Other examples of a-T-menable groups include:

1. Closed subgroups of $SO(n, 1)$ and $SU(n, 1)$, see [**CCJ$^+$01**].
2. Various small cancelation groups, see [**OW11**].
3. Groups which admit properly discontinuous actions on CAT(0) cubical complexes, see [**CCJ$^+$01**].

**a-T-$L^p$**-property: Groups having such property admit proper continuous isometric action $G \curvearrowright L^p$ for some $L^p$-space. In particular, if a group $G$ has properties a-T-$L^p$ and $FL^p$, then $G$ is compact. According to [**BP03**], every hyperbolic group has a-T-$L^p$ property for all sufficiently large $p$.

As for the fixed point (bounded orbits) property, one can associate to a group $G$ the set $\mathcal{P}_G$ of numbers $p \in (0, \infty)$ such that $G$ has a proper action on an $L^p$–space. Not much is known about the topological features of this set either.

QUESTION 17.60. 1. Given a finitely generated group $G$, is the set $\mathcal{P}_G$ connected? Is it closed?

2. For what groups $G$ are the sets $\mathcal{FP}_G$ and $\mathcal{P}_G$ complementary to each other ?

**Uniformly Lipschitz actions.** In properties (T) and a-T-menability, one can weaken the assumption that the action $G \curvearrowright \mathcal{H}$ is isometric as follows. An affine action $\alpha : G \curvearrowright V$ of $G$ on a Banach space $V$ is said to be *uniformly Lipschitz* if the linear parts of the transformations $\alpha(g), g \in G$ are linear operators in $V$ with uniformly bounded norms. In particular, every isometric action satisfies this property. We say that a group $G$ has *Property LT* if for every uniformly Lipschitz affine continuous action $G \curvearrowright \mathcal{H}$ on a Hilbert space, $G$ fixes a point in $\mathcal{H}$. Similarly, we say that a group $G$ has *Property* a-LT it admits a proper uniformly Lipschitz affine continuous action $G \curvearrowright \mathcal{H}$.

Note that if $V$ were a finite-dimensional Euclidean space, then every uniformly Lipschitz linear action $G \curvearrowright V$ admits an invariant positive-definite bilinear form. Hence, such an action *unitarizable*, i.e., is conjugate to an isometric action. The same holds for (continuous) actions of amenable groups on infinite-dimensional Hilbert spaces, see [**Dix50**]. On the other hand, Monod and Ozawa in [**MO10**] construct a large class of non-unitarizable uniformly Lipschitz linear actions of (non-amenable) groups $G$ .

One can further modify the definitions to the properties $LFL^p$ and a-$LTL^p$ by considering uniformly Lipschitz affine actions on $L^p$-spaces and asking for existence of a fixed point or existence of a proper action.

CONJECTURE 17.61. *1. If $\Gamma$ is a lattice in an irreducible semisimple Lie group of rank $\geqslant 2$, then $\Gamma$ has Property LT.*

*2. If $\Gamma$ is Gromov-hyperbolic then it has Property a-LT.*

*3. If $\Gamma_1, \Gamma_2$ are quasi-isometric groups then $\Gamma_1$ has Property LT if and only if $\Gamma_2$ does.*

*4. Let $\Gamma$ be a finitely generated discrete subgroup of $QI(\mathbb{H}^n)$. Suppose that $\Gamma$ acts on $\mathbb{H}^n$* uniformly quasi-isometrically*. Then $\Gamma$ has Property a-LT.*

## 17.7. Map of the world of infinite finitely generated groups

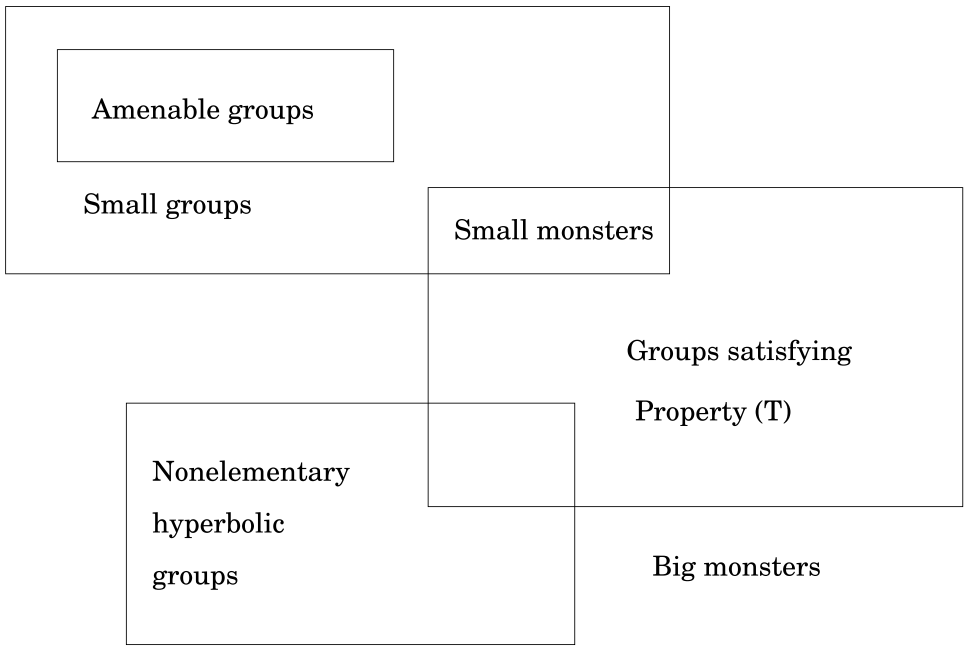



Figure 17.1. World of infinite finitely generated groups.

CHAPTER 18

# Stallings Theorem and accessibility

The goal of this chapter is to prove Stallings Theorem (Theorem 6.10) on ends of groups in the class of (almost) finitely-presented groups and Dunwoody's Accessibility Theorem for finitely-presented groups. As a corollary we obtain QI rigidity of the class of virtually free groups. Our proof is a combination of arguments due to Dunwoody [**Dun85**], Swarup [**Swa93**] and Jaco and Rubinstein [**JR88**].

## 18.1. Maps to trees and hyperbolic metrics on 2-dimensional simplicial complexes

**Collapsing maps.** Let $\Delta$ be a 2-dimensional simplex with the vertices $x_i, i = 1, 2, 3$. Our goal is to define a class of maps $\Delta \to Y$, where $Y$ is a simplicial tree with the standard metric (the same could be done when targets are arbitrary real trees but we will not need it). The construction of $f$ is, as usual, by induction on skeleta, it is analogous to the construction of collapsing maps $\kappa$ in 9.7. (The difference with the maps $\kappa$ is that the maps $f$ will not be isometric on edges, only linear.) Let $f : \Delta^{(0)} = \{x_1, x_2, x_3\} \to Y$ be given. If the image of this map is contained in a geodesic segment $\alpha$ in $Y$, then we extend $f$ to be a linear map $f : \Delta \to \alpha$. Otherwise, the points of $f(\Delta^{(0)})$ span a tripod $T$ in $Y$ with the centroid $o$ and extreme vertices $y_i := f(x_i)$. We extend $f$ to the map $f : \Delta^{(1)} \to Y$ by sending edges $[x_i, x_{i+1}]$ of $\Delta$ to the geodesics $[y_i, y_{i+1}] \subset T$ by linear maps. The preimage $f^{-1}(o)$ consists of three interior points $x_{ij}$ of the edges $[x_i, x_j]$ of $\Delta$, called *center points* of $\Delta$ (with respect to $f$). The 1-dimensional triangle $T(x_{12}, x_{23}, x_{31})$ (called *middle triangle*) splits $\Delta$ in four solid sub-triangles $\blacktriangle_i, i = 0, 1, 2, 3$ ($\blacktriangle_0$ is spanned by the center points while each $\blacktriangle_i$ contains $x_i$ as a vertex). Then $f$ sends the vertices of each $\blacktriangle_i$ to points in one of the legs of $T$. We then extend $f$ to a linear map on each of these four sub-triangles; clearly, $f(\blacktriangle_0) = \{o\}$.

Definition 18.1. The resulting map $f : \Delta \to \tau \subset Y$ is called a *canonical collapsing* map.

It is clear that if $X$ is a simplicial complex and $f : X^{(0)} \to Y$ is a map, then $f$ admits a unique extension to $f : X \to Y$ which is linear on every edge of $X$ and is a canonical collapsing map on each 2-simplex. We refer to the map $f : X \to Y$ as a *canonical map* $X \to Y$ (it depends, of course, on the initial map $f : X^{(0)} \to Y$). Suppose that $G$ is a group acting simplicially on $X$ and isometrically on $Y$. By uniqueness of the extension $f$ from $X^{(0)}$ to $X$, if $f : X^{(0)} \to Y$ is a $G$-equivariant map, then its extension $f : X \to Y$ is also $G$-equivariant. Such equivariant map $f : X \to Y$ is called a *canonical resolution* of the $G$-tree $Y$.

**Existence of resolutions of simplicial $G$-trees.** Recall that every finite group acting isometrically on a real tree $T$ has a fixed point (Corollary 9.21). If $T$

is a simplicial tree with the standard metric and the action is without inversions, then $G$ has to fix a vertex of $T$ (since a fixed point in the interior of an edge implies that the edge is fixed pointwise).

Let $T$ be a simplicial tree and $G \curvearrowright T$ be a cocompact simplicial action (without inversions). Let $X$ be a connected simplicial 2-dimensional complex on which $G$ acts properly discontinuously and cocompactly (possibly non-freely). We construct a resolution $f : X \to T$ as follows. Let $v \in X^{(0)}$ be a vertex. This vertex has finite stabilizer $G_v$ in $G$, therefore, this stabilizer fixes a vertex $w$ in $T$. We then set $f(v) := w$. (Fixed vertex may be non-unique, then we choose it arbitrarily). We then extend this map to the orbit $G \cdot v$ by equivariance. Repeating this for each vertex-orbit we obtain an equivariant map $f : X^{(0)} \to T^{(0)}$. Note that without loss of generality, by subdividing $X$ barycentrically if necessary, we may assume that $f : X^{(0)} \to T^{(0)}$ is onto (all what we need for this is that $X/G$ has more vertices than $T/G$). We then extend $f$ to the rest of $X$ by the canonical collapsing map, therefore obtaining the resolution.

**Piecewise-canonical maps.** In the proof of Theorem 18.30 we will need a mild generalization of the canonical maps and resolutions. Suppose that in the 2-simplex $\Delta$ we are given a subdivision into the solid triangles $\blacktriangle_i, i = 0, ..., 3$ with vertices $x_i, x_{jk}$. Suppose we are also given structure of a polygonal cell complex $P$ on $\Delta$ so that:

(1) Every vertex belongs to the boundary of $\Delta$.
(2) Every edge is geodesic.
(3) Every segment $[x_{ij}, x_{jk}]$ is an edge.
(4) Every vertex has valence 3 except for $x_{jk}, x_i, i, j, k = 1, 2, 3$.

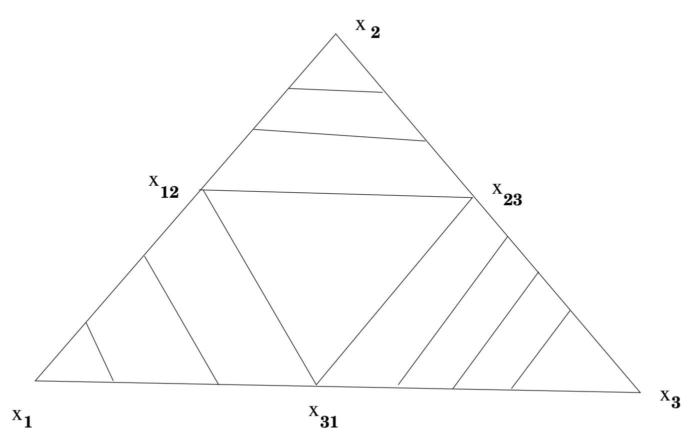



FIGURE 18.1. *P*olygonal subdivision of a simplex.

Edges of $P$ not contained in the boundary of $\Delta$ are called *interior edges*.

DEFINITION 18.2. A map $f : \Delta \to Y$ is called *piecewise-canonical* (PC) if it is constant on every interior edge and linear on each 2-cell. Note that the map $f$ could be constant on some 2-faces of $P$ (for instance, it is always constant on the solid middle triangle).

Clearly, a map $f$ of the 1-skeleton of $P$ which is constant on every interior edge, admits a unique PC extension to $\Delta$. A map $f : X \to Y$ from a simplicial complex to a tree is *piecewise-canonical* (PC) if it is PC on every 2-simplex of $X$ and piecewise-linear on every edge not contained in a 2-face. Every canonical map $f : \Delta \to Y$ is also almost canonical: The vertices of $P$ are the points $x_i, x_{jk}$.

Let $X$ be a simplicial complex, $Y$ a simplicial tree and $f : X \to Y$ be an PC map. We say that a point $y \in Y$ is a *regular value* of $f$ if for every 2-simplex $\Delta$ in $X$ we have:

a. $f^{-1}(y)$ is disjoint from the vertex set of $\Delta$.

b. $f^{-1}(y)$ is either empty or is a single topological arc (which necessarily connects distinct edges of $\Delta$).

A point $y \in Y$ which is not a regular value of $f$ is called a *critical value* of $f$. The following is an analogue of Sard's Theorem in the context of PC maps.

LEMMA 18.3. *Let $X$ be a countable simplicial complex, $Y$ a simplicial tree and $f : X \to Y$ be an PC map Then almost every point $y \in Y$ is a regular point of $f$.*

PROOF. Let $\Delta \subset X$ be a 2-simplex and $P$ be its polygonal cell complex structure. Then there are only finitely many critical values of $f$, namely the images of the vertices of $\Delta$ and of all the 2-faces of $P$ where $f$ is constant. Since $X$ is countable, this means that the set of critical values of $f$ is at most countable. □

**Complete hyperbolic metrics on punctured 2-dimensional simplicial complexes.** Our next goal is to introduce a path metric on $X' := X \setminus X^{(0)}$, so that each 2-simplex (minus vertices) is isometric to a solid ideal hyperbolic triangle.

PROPOSITION 18.4. *Let $X$ be a locally finite 2-dimensional simplicial complex. Then there exists a proper path-metric on $X' := X \setminus X^{(0)}$ so that each 2-simplex in $X$ is isometric to the ideal hyperbolic triangle. Moreover, this metric is invariant under all automorphisms of $X$.*

PROOF. We identify each 2-simplex $s$ in $X$ with the solid ideal hyperbolic triangle $\blacktriangle$ (so that vertices of $s$ correspond to the ideal vertices of the hyperbolic triangle). We now would like to glue edges of the solid triangles isometrically according to the combinatorics of the complex $X$. However, this identification is not unique since for each complete geodesic in $\mathbb{H}^2$ is invariant under a group of translations. Moreover, some of the identifications will yield incomplete hyperbolic metrics. (Even if we glue two ideal triangles along their boundaries!) Therefore, we have to choose gluing isometries appropriately.

The ideal triangle $\blacktriangle$ admits a unique inscribed circle; the points of tangency of this circle and the sides $\tau_{ij}$ of $\blacktriangle$ are the *central points* $x_{ij} \in \tau_{ij}$, see Section 9.7.

Now, given two solid ideal triangles $\blacktriangle_i, i = 1, 2$ and oriented sides $\tau_i, i = 1, 2$ of these triangles, there is a unique isometry $\tau_1 \to \tau_2$ which sends center-point to center-point and preserves orientation. We use these gluings to obtain a path-metric on $X'$. Clearly, this metric is invariant under all automorphisms of $X$ in the following sense:

If $g \in Aut(X)$ then the restriction of $g$ to $X^{(0)}$ admits a unique extension $\widehat{g} : X \to X$ which is an isometry of $X'$.

We claim that $X'$ is proper. The proof relies upon a certain collection of functions $b_\xi$ on $X'$ defined below, $\xi \in X^{(0)}$.

We first define three functions $b_1, b_2, b_3$ on the ideal triangle $\blacktriangle$. Let $\xi_i$, $i = 1, 2, 3$, be the ideal vertices of $\blacktriangle$, $\tau_k$ the ideal edge connecting $\xi_i$ to $\xi_j$; $\xi_{ij} \in \tau_k$ be the central point ($k = i + j \mod 3$).

Each pair of ideal vertices $\xi_{ij}, \xi_{jk}$ belongs to a unique horocycle $H_j$ in $\mathbb{H}^2$ with the ideal center $\xi_j$. To see this, consider the upper half-plane model of $\mathbb{H}^2$ so that $\xi_j = \infty$. Then $H_j$ is the horizontal line passing through $\xi_{ij}, \xi_{jk}$.

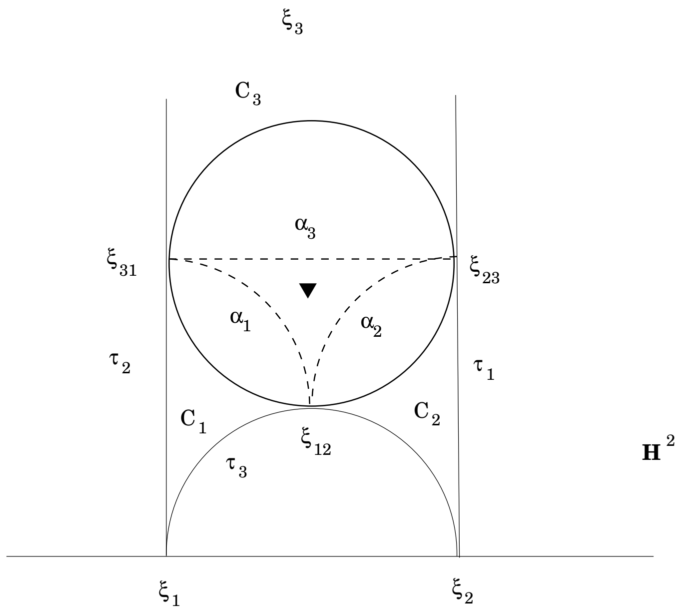



FIGURE 18.2. *G*eometry of the ideal hyperbolic triangle.

Consider circular arcs $\alpha_i := H_i \cap \blacktriangle$. The arcs $\alpha_1, \alpha_2, \alpha_3$ cut out a solid triangle $\blacktriangledown$ (with horocyclic arcs $\alpha_i$'s as it edges) from $\blacktriangle$. We refer to the complementary components $C_i$ of $\blacktriangle \setminus \tau$ as *corners* of $\blacktriangle$ with the ideal vertices $\xi_i$, $i = 1, 2, 3$. Their closures in $\Delta$ are the *closed corners* $\overline{C}_i$. We then define a 1-Lipschitz function $b_i : \overline{C}_i \to \mathbb{R}_+$ by

$$b_i(x) = \operatorname{dist}(x, \alpha_i), \quad i = 1, 2, 3.$$

The level sets of $b_i : C_i \to \mathbb{R}_+$ are arcs of horocycles in $C_i$. (The functions $b_i$ are the negatives of *Busemann functions*, see [**Bal95**].) We extend each $b_i$ by zero to $\blacktriangle \setminus C_i$.

For each vertex $\xi$ of $X$ we define $\overline{C}_\xi$ to be the union of closed corners $\overline{C}_i$ (with the vertex $\xi = \xi_i$) of 2–simplices $s \subset X$ which have $\xi$ as a vertex. Then the functions $b_i : \overline{C}_i \to \mathbb{R}_+$ match (since the central points do), thus, we obtain a collection of 1-Lipschitz functions $b_\xi : X' \to \mathbb{R}_+$. It is clear from the construction that each $b_\xi$ is proper on $\overline{C}_\xi$.

Set

$$\blacktriangle_r := \{x \in \blacktriangle : \forall i \quad b_i(x) \leqslant r\}, \quad X_r := \{x \in X' : \forall \xi \in X^{(0)}, b_\xi(x) \leqslant r\}.$$

Since each $b_\xi$ is 1-Lipschitz, for every path $\mathfrak{p}$ in $X'$ of length $\leqslant r$, if $Image(\mathfrak{p}) \cap X_0 \neq \emptyset$ then $p \subset X_r$.

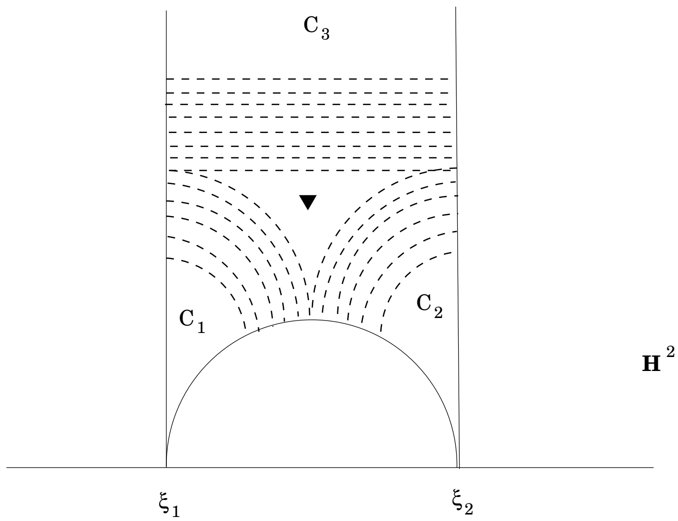



FIGURE 18.3. *P*artial foliation of corners of an ideal triangle by level sets of the functions $b_i$.

For $x, y \in X'$ we let $\rho(x,y)$ be the minimal number of edges that a path in $X'$ from $x$ to $y$ has to intersect. Since $X$ is locally finite, for every $x \in X$, $k \in \mathbb{N}$, the set $\{y \in X' : \rho(x,y) \leqslant k\}$ is a union of finitely many cells.

Every ideal side $\tau_i$ of $\blacktriangle$ intersects $\blacktriangle_r$ in a compact subset. Thus, there exists $D(r) > 0$ so that the minimal distance between the geodesics

$$\tau_i \cap \blacktriangle_r, \tau_j \cap \blacktriangle_r \quad (i \neq j)$$

is at least $D(r)$. Therefore, if $\mathfrak{p}$ is a path in $X_r$ connecting $x$ to $y$, then its length is at least

$$D(r)\rho(x,y).$$

Thus, for every $x \in X_0$ the metric ball $B_r(x)$ intersects only finitely many cells in $X$ and is contained in $X_r$. Since intersection of $X_r$ with any finite subcomplex in $X$ is compact, it is now immediate that $X'$ is a proper metric space. □

### 18.2. Transversal graphs and Dunwoody tracks

We continue with the notation of the previous section.

Our goal is to introduce for $X'$ notions analogous to the transversality in the theory of smooth manifolds. We define the *vertex-complexity* of a finite graph $\Gamma$, denoted $\nu(\Gamma)$, to be cardinality of the vertex set $V(\Gamma)$. We say that a properly embedded graph $\Gamma \subset X'$ is *transversal* if the following hold:

1. $\Gamma \cap X^{(1)} = V(\Gamma) = \Gamma^{(0)}$.
2. $\Gamma$ intersects every 2-face of $X$ along a (possibly empty) disjoint union of edges.

Transversal graphs generalize the concept of properly embedded *smooth codimension 1 submanifolds* in a smooth manifold. Note that every transversal graph satisfies the property:

For every edge $e \subset X^{(1)}$, every pair of distinct 2-faces $s_1, s_2 \subset X$ containing $e$ and every vertex $v \in \Gamma \cap e$, there are exactly two edges $\gamma_1 \subset \Gamma \cap s_1, \gamma_2 \subset \Gamma \cap s_2$ which have $v$ as a vertex.

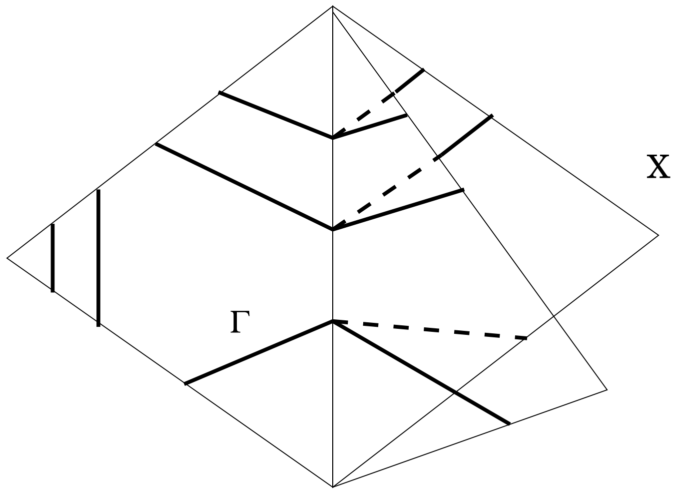



FIGURE 18.4. *D*unwoody track.

If a transversal graph $\Gamma$ satisfies the property:

3. For every edge $\gamma$ of $\Gamma$, the end-points of $\gamma$ belong to distinct edges of $X^{(1)}$,

then $\Gamma$ is called a *Dunwoody track*, or simply a *track*.

EXERCISE 18.5. Let $f : X \to Y$ be a PC map from a simplicial 2-complex $X$ to a simplicial tree $Y$ and $y \in Y$ be a regular value of $f$. Then $f^{-1}(y)$ is a Dunwoody track in $X$.

The following lemma is left as an exercise to the reader, it shows that every Dunwoody track in $X$ behaves like a codimension 1 smooth submanifold in a differentiable manifold.

LEMMA 18.6. *Let $\Gamma$ be a Dunwoody track. Then for every $x \in \Gamma$ there exists a neighborhood $U$ of $x$ which is naturally homeomorphic to the product*

$$\Gamma_U \times [-1, 1],$$

*where $\Gamma_U = \Gamma \cap U$ and the above homeomorphism sends $\Gamma_U$ to $\Gamma_U \times \{0\}$. We will refer to the neighborhoods $U$ as* product neighborhoods *of points of* $\Gamma$.

Note that the entire track need not have a product neighborhood. For instance, let $\Gamma$ be a non-separating loop in the Moebius band $X$. Triangulate $X$ so that $\Gamma$ is a track. Then every regular neighborhood of $\Gamma$ in $X$ is again a Moebius band. However, the neighborhoods $\Gamma_U$ combine in a neighborhood $N_\Gamma$ of $\Gamma$ in $X$ which is an interval bundle over $\Gamma$, where the product neighborhoods $U \cong \Gamma_U \times [-1, 1]$ above serve as coordinate neighborhoods in the fibration.

We say that the track $\Gamma$ is *1-sided* if the interval bundle $N_\Gamma \to \Gamma$ is non-trivial and *2-sided* otherwise.

EXERCISE 18.7. Suppose that $\Gamma$ is connected and 1-sided. Then $N_\Gamma \setminus \Gamma$ is connected.

For each edge transversal graph $\Gamma \subset X$ we define the *counting function* $m_\Gamma : Edges(X) \to \mathbb{Z}$:

$$m_\Gamma(e) := |\Gamma \cap e|.$$

**The $\mathbb{Z}_2$-cocycle of a transversal graph.** Recall that, by the Poincaré duality, every proper codimension $k$ embedding of smooth manifolds

$$N \hookrightarrow M$$

defines an element $[N]$ of $H^k(M, \mathbb{Z}_2)$. Our goal is to introduce a similar concept for transversal graphs $\Gamma$ in $X$. Observe that for every 2-face $s$ in $X$

$$\sum_{i=1}^{3} m_\Gamma(e_i) = 0, \quad \mod 2,$$

where $e_1, e_2, e_3$ are the edges of $s$ (since every edge $\gamma$ of $\Gamma \cap s$ contributes zero to this sum. Therefore, $m_\Gamma$ determines an element of $Z^1(X, \mathbb{Z}_2)$. If $\Gamma$ is finite, then $m_\Gamma \in Z^1_c(X, \mathbb{Z}_2)$ since the cocycle $m_\Gamma$ is supported only on the finitely many edges which cross $\Gamma$. We let $[\Gamma]$ denote the cohomology class in $H^1_c(X, \mathbb{Z}_2)$ determined by $m_\Gamma$. It is clear that $[\Gamma]$ depends only on the isotopy class of $\Gamma$.

LEMMA 18.8. *Suppose that $\Gamma$ is 1-sided. Then $m_\Gamma$ represents a nontrivial class in $H^1(X, \mathbb{Z}_2)$.*

PROOF. We subdivide $X$ so that $N_\Gamma$ is a subcomplex in $X$. We will compute $H^1(X, \mathbb{Z}_2)$ using this cell complex. Since $\Gamma$ is 1-sided, there are vertices $u, v \in N_\Gamma \cap X^{(1)}$ in the boundary of $N_\Gamma$ which are connected by an arc $\tau \subset X^{(1)}$ intersecting $\Gamma$ in exactly one vertex. Hence, $m(u) + m(v) = 1 \mod 2$. Since $N_\Gamma \setminus \Gamma$ is connected, there exists a path $\alpha \subset \partial N_\Gamma$ connecting $u$ to $v$. Suppose that $m_f = \delta\eta$ in $C^1(X, \mathbb{Z}_2)$. Then $\eta(u) + \eta(v) = 1$. Since $\alpha$ is disjoint from $\Gamma$, it follows that none of the edges in $\alpha$ cross $\alpha$, thus $\eta(u) = \eta(v)$. Contradiction. ☐

LEMMA 18.9. *Suppose that $H^1(X, \mathbb{Z}_2) = 0$. Then a connected finite transversal graph $\Gamma$ separates $X$ in at least two unbounded components if and only if$[\Gamma]$ is a nontrivial class in $H^1_c(X, \mathbb{Z}_2)$. Such a track is said to be* essential.

PROOF. The proof is similar to the argument of the previous lemma.

1. Suppose that $X \setminus \Gamma$ contains at least two unbounded complementary components $U$ and $V$, but $[\Gamma] = 0$ in $H^1_c(X, \mathbb{Z}_2)$. Then there exists a compactly-supported function $\sigma : X^{(0)} \to \mathbb{Z}_2$ so that $\delta(\sigma) = m_\Gamma$, mod 2. Since $\sigma$ is compactly supported, there exists a compact subset $C \subset X$ so that $\sigma = 0$ on $U \setminus C, V \setminus C$. Let $\alpha \subset X^{(1)}$ be a path connecting $u \in U \cap X^{(0)}$ to $v \in V \cap X^{(0)}$. We leave it to the reader to verify that if an edge $e = [x, y]$ of $X$ crosses $\Gamma$ in the even number of points then $x, y$ belong to the same connected component of $X \setminus \Gamma$ (this is the only place where we use the assumption that $\Gamma$ is connected). Therefore, the path $\alpha$ crosses $\Gamma$ in an odd number of points, which implies that $\langle m_\Gamma, \alpha \rangle = 1 \in \mathbb{Z}_2$. However, $\langle m_\Gamma, \alpha \rangle = \langle \sigma, \partial\alpha \rangle = \sigma(u) + \sigma(v) = 0$. Contradiction.

2. Suppose that $[\Gamma] \neq 0$ in $H^1_c(X, \mathbb{Z}_2)$. Since $H^1(X, \mathbb{Z}_2) = 0$, there exists a 0-cochain $\sigma : X^{(0)} \to \mathbb{Z}_2$ so that $\delta(C) = m_\Gamma \mod 2$. Then $[\Gamma]$ takes nonzero value on some edge $e = [u, v]$ of $X$, which means that, say, $\sigma(u) = 0, \sigma(v) = 1$.

If $\{u \in X^{(0)} : \sigma(u) = 1\}$ is finite, then $\sigma \in C^0_c(X, \mathbb{Z}_2)$ and, hence $[\sigma] = 0$, which is a contradiction. Therefore, the set of such vertices is unbounded. Consider another 0-cochain $\sigma + 1$ (which equals to $\sigma(w) + 1$ on every vertex $w \in X^{(0)}$). Then $\delta(\sigma + 1) = \delta(\sigma)$ and

$$\{u \in X^{(0)} : \sigma(u) = 0\} = \{u \in X^{(0)} : \sigma(u) + 1 = 1\}.$$

Therefore, by the above argument, the set $\{u \in X^{(0)} : \sigma(u) = 0\}$ is also unbounded. Thus, since $\Gamma$ is a finite graph, there are unbounded connected subsets $U, V \subset X \setminus \Gamma$ such that

$$\forall u \in U \cap X^{(0)}, \sigma(u) = 0, \quad \forall v \in V \cap X^{(0)}, \sigma(v) = 1.$$

These are the required unbounded complementary components of $\Gamma$. □

EXERCISE 18.10. If $H^1(X, \mathbb{Z}_2) = 0$, then every connected essential Dunwoody track $\Gamma \subset X$ has exactly two complementary components, both of which are unbounded.

The following key lemma due to Dunwoody is a direct generalization of the Kneser–Haken finiteness theorem for triangulated 3-dimensional manifolds, see e.g. [**Hem78**].

LEMMA 18.11 (M. Dunwoody). *Suppose that $X$ has $F$ faces and $H^1(X, \mathbb{Z}_2) \cong \mathbb{Z}_2^r$. Suppose that $\Gamma_1, ..., \Gamma_k$ are pairwise disjoint pairwise non-isotopic connected tracks in $X$. Then*

$$k \leqslant 6F + r.$$

PROOF. The union $\Gamma$ of tracks $\Gamma_i$ cuts each 2-simplex $s$ in $X$ in triangles, quadrilaterals and hexagons. Note that some of the complementary rectangles might contain vertices of $X$. In what follows, we regard such rectangles as *degenerate hexagons* (and not as rectangles). The boundary of each complementary rectangle has two disjoint edges contained in $X^{(1)}$, we call these edges *vertical*. Consider an edge of $\Gamma$ which is contained in the boundary of a complementary triangle or a (possibly degenerate) hexagon. The number of such edges is at most $6F$. Thus, the number of tracks $\Gamma_i$ containing such edges is at most $6F$ as well. We now remove from $X$ the union of closures of all components of $X \setminus \Gamma$ which contain complementary triangles and (possibly degenerate) hexagons.

The remainder $R$ is a union of rectangles $Q_j$ glued together along their vertical edges. Therefore, $R$ is homeomorphic to an open interval bundle over a track $\Lambda \subset X$: The edges of $\Lambda$ are geodesics connecting midpoints of vertical edges of $Q_j$'s. If a component $R_i$ of $R$ is a trivial interval bundle then the boundary of $R_i$ is the union of tracks $\Gamma_j, \Gamma_k$ which are therefore isotopic. This contradicts our assumption on the tracks $\Gamma_i$. Therefore, each component of $R$ is a nontrivial interval bundle. For each $R_i$ we define the cohomology class $[\Lambda_i] \in H^1(X, \mathbb{Z}_2) = H^1_c(X, \mathbb{Z}_2)$ (using the counting function $m_\Lambda$). We claim that these classes are linearly independent. Suppose to the contrary that

$$\sum_{i=1}^{\ell} [\Lambda_i] = 0.$$

This means that the track $\Lambda' := \Lambda_1 \cup ... \cup \Lambda_\ell$ determines a trivial cohomology class $[\Lambda'] = 0$. Since $\Lambda'$ is 1-sided, we obtain a contradiction with Lemma 18.8. Therefore, the number of components of $R$ is at most $r$, the dimension of $H^1(X, \mathbb{Z}_2)$. Each component of $R$ is bounded by a track $\Gamma_i$. Therefore, the total number of tracks $\Gamma_i$ is at most $6F + r$. □

### 18.3. Existence of minimal Dunwoody tracks

Our next goal is to deform finite transversal graphs to Dunwoody tracks, so that the cohomology class is preserved and so that the counting function $m_\Gamma : Edges(X) \to \mathbb{Z}$ decreases as the result of the deformation. To this end, we define the operation *pull* on transversal graphs $\Gamma \subset X$.

**Pull.** Suppose that $v_1, v_2$ are distinct vertices of $\Gamma$ which belong to a common edge $e$ of $X$ and which are not separated by any vertex of $\Gamma \cap e$ on $e$. We call such vertex pair $\{v_1, v_2\}$ *innermost.* Then for every 2-face $s$ of $X$ containing $e$ and every pair of distinct edges $\gamma_i = [u_i, v_i], i = 1, 2$ of $\Gamma$ we perform the following operation. We replace $\gamma_1 \cup \gamma_2 \subset \Gamma$ by a single edge $\gamma = [u_1, u_2] \subset s$, keeping the rest of $\Gamma' = \Gamma \setminus \gamma_1 \cup \gamma_2$ unchanged, so that $\gamma$ intersects $\Gamma'$ only at the end-points $u_1, u_2$. In case $\gamma_1 = \gamma_2$ we simply eliminate this edge from $\Gamma$. Let $pull(\Gamma)$ denote the resulting graph. It is clear that $\nu(pull(\Gamma)) < \nu(\Gamma)$ and $pull(\Gamma)$ is again a transversal graph. Note that *a priori*, $pull(\Gamma)$ need not be connected even if $\Gamma$ is.

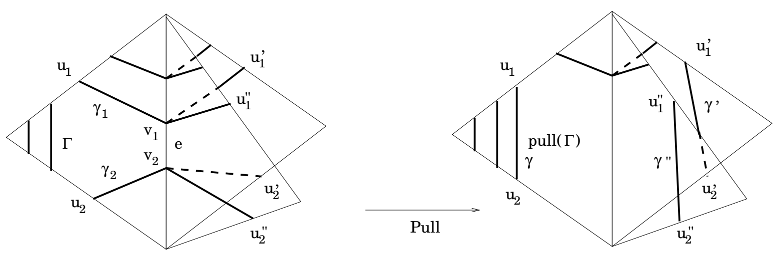



FIGURE 18.5. *P*ull.

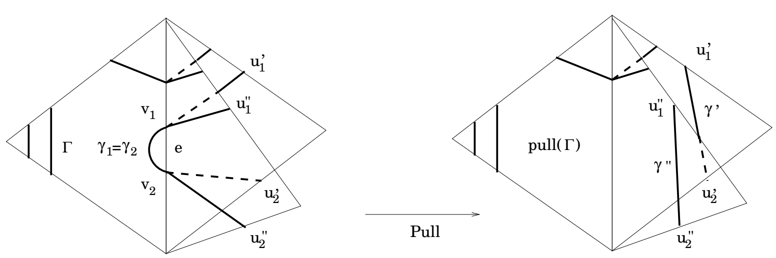



FIGURE 18.6. *E*liminating edge $\gamma_1 = \gamma_2$. In this example, $u_1 = v_2, u_2 = v_1$

EXERCISE 18.12. Verify that $[pull(\Gamma)] = [\Gamma]$; actually, the functions $m_\Gamma$ and $m_{pull(\Gamma)}$ are equal as $\mathbb{Z}_2$-cochains.

LEMMA 18.13. *Given a finite transversal graph $\Gamma \subset X$, there exists a finite sequence of pull-operations which transforms $\Gamma$ to a new graph $\Gamma'$; the graph $\Gamma'$ is a track so that for every edge $e$, $m_{\Gamma'}(e) \in \{0, 1\}$. Moreover, $[\Gamma] = [\Gamma']$.*

Proof. We apply the pull-operation to $\Gamma$ as long as possible; since the vertex-complexity under *pull* is decreasing, this process terminates at some transversal graph $\Gamma'$. If $m_{\Gamma'}(e) \geqslant 2$ for some edge $e$ of $X$, we can again perform *pull* using an innermost pair of vertices of $\Gamma'$ on $e$, which is a contradiction. Since *pull* preserves the cohomology class of a transversal graph, $[\Gamma] = [\Gamma']$. □

Lemma 18.14. *Assume that $H^1(X, \mathbb{Z}_2) = 0$. If $|Ends(X)| > 1$ then there exists a connected essential transversal graph $\Gamma \subset X$.*

Proof. Let $\epsilon_+, \epsilon_-$ be distinct ends of $X$. We claim that there exists a proper 1-Lipschitz function $h : X \to \mathbb{R}$ so that

$$\lim_{x \to \epsilon_\pm} = \pm\infty.$$

Indeed, let $K$ be a compact which separates the ends $\epsilon_+, \epsilon_-$. We define $h$ to be constant on $K$. We temporarily re-metrize $X$ by equipping it with the standard metric (every simplex is isometric to the standard Euclidean simplex with unit edges). Let $U_\pm$ be the unbounded components of $X \setminus K$ which are neighborhoods of the ends $\epsilon_\pm$. We then set

$$h|_{U_\pm} := \pm \text{dist}(\cdot, K).$$

For every other component $V$ of $X \setminus K$ we set

$$h|_V := \text{dist}(\cdot, K).$$

It is immediate that this function satisfies the required properties. We give $\mathbb{R}$ the structure of a simplicial tree $T$, where integers serve as vertices. We next approximate $h$ by a proper canonical map $f : X \to T$. Namely, for every vertex $v$ of $X$ we let $f(v)$ be a vertex of $T$ within distance $\leqslant 1$ from $f(v)$. We extend $f : X^{(0)} \to T$ to a canonical map $f : X \to T$. Then $\text{dist}(f, h) \leqslant 3$ and, hence, $f$ is again proper. By Lemma 18.3, $\Gamma := f^{-1}(y)$ is a finite transversal graph separating $\epsilon_+$ from $\epsilon_-$ for almost every $y$. Hence, $[\Gamma] \neq 0$ in $H^1_c(X, \mathbb{Z}_2)$. The graph $\Gamma$ need not be connected, let $\Gamma_1, ..., \Gamma_n$ be its connected components: They are still transversal graphs. Thus,

$$[\Gamma] = \sum_{i=1}^{n} [\Gamma_i],$$

which implies that at least one of the graphs $\Gamma_i$ is essential. □

Note, that the graph $\Gamma$ constructed in the above proof need not be a Dunwoody track. However, Lemma 18.13 implies that we can replace $\Gamma$ with a essential Dunwoody track $\Gamma'$ which intersects every edge in at most one point. The graph $\Gamma'$ need not be connected, but it has an essential connected component. Therefore:

Corollary 18.15. *Assume that $H^1(X, \mathbb{Z}_2) = 0$ and $|Ends(X)| > 1$. Then there exists a connected essential Dunwoody track $\Gamma \subset X$. Moreover,*

$$m_\Gamma : Edges(X) \to \{0, 1\}.$$

We define the *complexity* of a transversal graph $\Gamma \subset X$, denoted $c(\Gamma)$, to be the pair $(\nu(\Gamma), \ell(\Gamma))$, where $\nu(\Gamma)$ is the number of vertices in $\Gamma$ and $\ell(\Gamma)$ is the total length of $\Gamma$. We give the set of complexities the lexicographic order. It is clear that $c(\Gamma)$ is preserved by isometric actions $G \curvearrowright X'$.

An essential connected Dunwoody track $\Gamma \subset X$ is said to be *minimal* if it has minimal complexity among all connected essential Dunwoody tracks in $X$.

Definition 18.16. A vertex $v$ of $X$ is said to be a *cut-vertex* if $X \setminus \{v\}$ contains at least two unbounded components. (Note that our definition is slightly stronger than the usual definition of a cut-vertex, where it is only assumed that $X \setminus \{v\}$ is not connected.)

Lemma 18.17. *Suppose that $X$ admits a cocompact simplicial action $G \curvearrowright X$, $H^1(X, \mathbb{Z}_2) = 0$, $|Ends(X)| > 1$ and $X$ has no cut-vertices. Then there exists an essential minimal track $\Gamma_{min}$.*

Proof. By Corollary 18.15, the set of connected essential tracks in $X$ is nonempty. Let $\Gamma_i$ be a sequence of such graphs whose complexity converges to the infimum. Without loss of generality, we can assume that each $\Gamma_i$ has minimal vertex-complexity $\nu = \nu(\Gamma)$ among all connected essential tracks in $X$. Since $X$ is a simplicial complex, it is easy to see that each $\Gamma_i$ is also a simplicial complex. Therefore, the number of edges of the graphs $\Gamma_i$ is also uniformly bounded (by $\frac{\nu(\nu-1)}{2}$). In particular, there are only finitely many combinatorial types of these graphs; therefore, after passing to a subsequence, we can assume that the graphs $\Gamma_i$ are combinatorially isomorphic to a fixed graph $\Gamma$.

Replace each edge of $\Gamma_i$ with the hyperbolic geodesic (in the appropriate 2-simplex of $X$). This does not increase the complexity of $\Gamma_i$, keeps the graph embedded and preserves all the properties of Dunwoody tracks. Therefore, we will assume that each edge of $\Gamma_i$ is geodesic. We let $h_i : \Gamma \to \Gamma_i$ be graph isomorphisms. Since $\ell(\Gamma_i)$ are uniformly bounded from above, there exists a path-metric on $\Gamma$ so that all the maps $h_i$ are 1-Lipschitz. We let $\bar{h}_i$ denote the composition of $h_i$ with the quotient map $X \to Y = X/G$. If there exists a compact set $C \subset Y' := X'/G$ such that $\bar{h}_i(\Gamma) \cap C \neq \emptyset$, then the Arzela–Ascoli Theorem implies that the sequence $(\bar{h}_i)$ subconverges to a 1-Lipschitz map $\Gamma \to Y'$. On the other hand, if such compact does not exist, then, since edges of $Y'$ have infinite length and $\Gamma$ is connected, the sequence of maps $\bar{h}_i$ subconverges to a constant map sending $\Gamma$ to one of the vertices of $Y$. Hence, in this case, by post-composing the maps $h_i$ with $g_i \in G$, we conclude that the sequence $g_i \circ h_i$ subconverges to a constant map whose image is one of the vertices of $X$. Recall that, by our assumption, $X$ has no cut-vertices. Therefore, every sufficiently small neighborhood of a vertex $v$ of $X$ does not separate $X$ into several unbounded components. This contradicts the assumption that each $\Gamma_i$ is essential.

Therefore, by replacing $h_i$ with $g_i \circ h_i$ (and preserving the notation $h_i$ for the resulting maps), we conclude that the maps $h_i$ subconverge to a 1-Lipschitz map $h : \Gamma \to X'$. In view of Lemma 18.13 (and the fact that $\Gamma_i$'s have minimal vertex-complexity), for every face $s$ and edge $e \subset s$ of $X$ there exists at most one edge of $\Gamma_i$ contained in $s$ and intersecting $e$. Therefore, the map $h$ is injective and $\Gamma_{min} = h(\Gamma)$ is a track in $X$. Moreover, for each sufficiently large $i$, the graph $\Gamma_{min}$ is isotopic to $\Gamma_i$ as they have the same counting function $m_\Gamma = m_{\Gamma_i}$. Thus, $\Gamma_{min}$ is essential. Therefore, it is the required minimal track. □

## 18.4. Properties of minimal tracks

**18.4.1. Stationarity.** The following discussion is local and does not require any assumptions on $H^1(X, \mathbb{Z}_2)$.

We say that a transversal graph $\Gamma$ is *stationary* if for every small smooth isotopy $\Gamma_t$ of $\Gamma$ (through transversal graphs), with $\Gamma_0 = \Gamma$, we have

$$\frac{d}{dt}\ell(\Gamma_t)|_{t=0} = 0.$$

In particular, every edge of $\Gamma$ is geodesic.

EXAMPLE 18.18. Every minimal essential Dunwoody track is stationary.

Let $\Gamma$ be a Dunwoody track with geodesic edges. Let $v$ be a vertex of $\Gamma$ which belongs to an edge $e$ of $X$ and $\gamma$ be an edge of $\Gamma$ incident to $v$. We assume that $\gamma = \gamma(t)$ is parameterized by its arc-length so that $\gamma(0) = v$. We define $\pi_e(\gamma')$ to be the orthogonal projection of the vector $\gamma'(0) \in T_e\mathbb{H}^2$ to the tangent line of $e$ at $v$.

LEMMA 18.19. *If $\Gamma$ is stationary then for every vertex $v$ as above we have*

$$\sum_{\gamma} \pi_e(\gamma') = 0 \tag{18.1}$$

*where the sum is taken over all edges $\gamma_1, ..., \gamma_k$ of $\Gamma$ incident to $v$.*

PROOF. We construct a small isotopy $\Gamma_t$ of $\Gamma$ by fixing all the vertices and edges of $\Gamma$ except for the vertex $v$ which is moved along $e$, so that $v(t), t \in [0, 1]$, is a smooth function. We assume that all edges of $\Gamma_t$ are geodesic. This variation of $v$ uniquely determines $\Gamma_t$. It is clear that

$$0 = \frac{d}{dt}\ell(\Gamma_t)|_{t=0} = \sum_{i=1}^{k} \frac{d}{dt}\ell(\gamma_i(t))|_{t=0}.$$

By the first variation formula (8.10), we conclude that

$$0 = \sum_{i=1}^{k} \frac{d}{dt}\ell(\gamma_i(t))|_{t=0} = \sum_{i=1}^{k} \pi_e(\gamma'). \quad \square$$

REMARK 18.20. The proof of the above lemma also shows the following. Suppose that $\Gamma$ fails the stationarity condition (18.1) at a vertex $v$. Orient the edge $e$ and assume that the vector

$$\sum_{\gamma} \pi_e(\gamma')$$

points to the "right" of zero. Construct a small isotopy $\Gamma_t$, $\Gamma_0 = \Gamma, t \in [0, 1)$, so that all edges of $\Gamma_t$ are geodesic, vertices of $\Gamma_t$ except for $v$ stay fixed, while the vertex $v(t)$ moves to the "right" of $v = v(0)$. Then

$$\ell(\Gamma_t) < \ell(\Gamma)$$

for all small $t > 0$.

LEMMA 18.21 (The Maximum Principle). *Let $\Lambda_1, \Lambda_2$ be stationary Dunwoody tracks. Then in a small product neighborhood $U$ of every common vertex $u$ of these graphs, either the graphs $\Lambda_1, \Lambda_2$ coincide, or one "crosses" the other. The latter means that*

$$\Lambda_1 \cap U_+ \neq \emptyset, \Lambda_2 \cap U_- \neq \emptyset.$$

*Here $U_\pm = \Lambda_{1,U} \times (0, \pm 1]$ where we identify the product neighborhood $U$ with $\Lambda_{1,U} \times [-1, 1]$, $\Lambda_{1,U} = \Lambda_1 \cap U$. In other words, if $h : U = \Lambda_{1,U} \times [-1, 1] \to [-1, 1]$ is the* *projection to the second factor, then $h|\Lambda_2$ cannot have maximum or minimum at $u$, unless $h|\Lambda_2$ is identically zero.*

Proof. Let $e$ be the edge of $X$ containing $u$. Since $\Lambda_1, \Lambda_2$ are tracks, every 2-simplex $s$ of $X$ adjacent to $e$ contains (unique) edges $\gamma_{i,s} \subset \Lambda_i, i = 1, 2$, which are incident to $u$. Suppose that $\Lambda_2$ does not cross $\Lambda_1$. Then either for every $s, \gamma_i = \gamma_{i,s}, i = 1, 2$ as above,

$$\pi_e(\gamma_1') \geqslant \pi_e(\gamma_2')$$

or for every $s, \gamma_1, \gamma_2$

$$\pi_e(\gamma_1') \geqslant \pi_e(\gamma_2').$$

Since, by the previous lemma,

$$\sum_s \pi_e(\gamma_{i,s}') = 0, \quad i = 1, 2,$$

we conclude that $\pi_e(\gamma_{1,s}') = \pi_e(\gamma_{2,s}')$. Therefore, since geodesic is uniquely determined by its derivative at a point, it follows that $\gamma_{1,s} = \gamma_{2,s}$ for every 2-simplex $s$ containing $e$. Thus, $\Lambda_1 \cap U = \Lambda_2 \cap U$. □

**18.4.2. Disjointness of minimal tracks.** The following proposition is the key for the proof of Stallings Theorem presented in the next section:

Proposition 18.22. *If $H^1(X, \mathbb{Z}_2) = 0$ then any two minimal tracks in $X$ are either equal or disjoint.*

Proof. Our proof follows **[JR89]**. The three central ingredients in the proof are: Exchange and round-off arguments as well as the Meeks–Yau trick, all coming from theory of least area surfaces in 3-dimensional Riemannian manifolds.

Suppose that $\Lambda, \Lambda'$ are distinct minimal tracks which have nonempty intersection.

**Step 1: Transversal case.** We first present an argument that this is impossible under the assumption that the graphs $\Lambda$ and $\Lambda'$ are *transverse*, i.e., $\Lambda \cap \Lambda'$ is disjoint from the 1-skeleton of $X$. We consider the four sets

$$\Lambda_+ \cap \Lambda'_+, \Lambda_+ \cap \Lambda'_-, \Lambda_- \cap \Lambda'_+, \Lambda_- \cap \Lambda'_-.$$

Since both $\Lambda, \Lambda'$ separate ends of $X$, at least two of the above sets are unbounded. After relabeling, we can assume that $\Lambda_+ \cap \Lambda'_+, \Lambda_- \cap \Lambda'_-$ are unbounded. Observe that

$$\Lambda \cup \Lambda' = \partial(\Lambda_+ \cap \Lambda'_+) \cup \partial(\Lambda_- \cap \Lambda'_-).$$

Set

$$\Gamma_+ := \partial(\Lambda_+ \cap \Lambda'_+), \Gamma_- := \partial(\Lambda_- \cap \Lambda'_-).$$

It is immediate that both graphs are transversal (here and below we disregard valency 2 vertices of $\Gamma_\pm$ contained in the interiors of 2-simplices of $X$). We now compare complexity of $\Lambda$ (which is the same as the complexity of $\Lambda'$) and complexities of the graphs $\Gamma_+, \Gamma_-$. After relabeling, $\ell(\Gamma_+) \leqslant \ell(\Gamma_-)$. We leave it to the reader to verify that the numbers of edges of the graphs $\Gamma_\pm$ are the same as the number of edges of $\Lambda$. Clearly, the total length of $\Gamma_+ \cup \Gamma_-$ is the same as $2\ell(\Lambda) = 2\ell(\Lambda')$. Therefore,

$$\ell(\Gamma_+) \leqslant \ell(\Lambda).$$

Hence, $c(\Gamma_+) \leqslant c(\Lambda)$. The transition from $\Lambda$ to the graph $\Gamma_+$ is called the *exchange argument*: We replaced parts of $\Lambda$ with parts of $\Lambda'$ in order to get $\Gamma_+$.

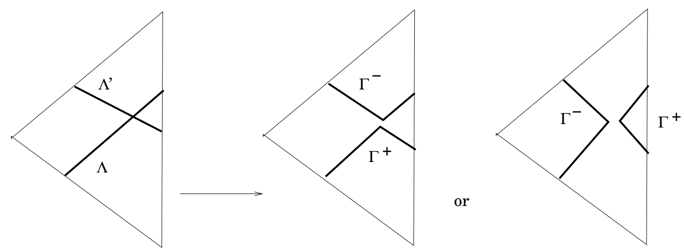



FIGURE 18.7. *E*xchange argument.

Note that both $\Lambda_+ \cap \Lambda'_+$ and its complement are unbounded. Therefore, the graph $\Gamma_+$ separates ends of $X$. Thus, $[\Gamma_+] \neq 0$ in $H^1_c(X, \mathbb{Z}_2)$. Hence, at least one connected component of $\Gamma_+$ represents a nontrivial element of $H^1_c(X, \mathbb{Z}_2)$. Since $H^1(X, \mathbb{Z}_2) = 0$, this component is essential. By minimality of the graph $\Lambda$, it follows that $\Gamma_+$ is connected (otherwise we replace it with the above essential component thereby decreasing the vertex-complexity). Since $\Lambda, \Lambda'$ crossed at a point $x \notin X^{(1)}$, there exists an edge $\gamma$ of $\Gamma_+$ which is a broken geodesic containing $x$ in its interior. Replacing the broken edge $\gamma$ with a shorter path (and keeping the end-points) we get a new graph $\widehat{\Lambda}$ whose total length is strictly smaller than the one of $\Gamma_+$. (This part of the proof is called the "round-off" argument.) We obtain a contradiction with minimality of $\Lambda$. This finishes the proof in the case of transversal intersections of $\Lambda$ and $\Lambda'$.

**Step 2: Weakly transversal case.** We assume now that $\Lambda \cap \Lambda'$ contains at least one point $p$ of transversal intersection which is not in the 1-skeleton of $X$. We say that in this situation $\Lambda, \Lambda'$ are *weakly transversal* to each other. Note that doing "exchange and round-off" at $p$ we have some definite reduction in the complexity of the tracks, which depended only on the intersection angle $\alpha$ between $\Lambda, \Lambda'$ at $p$. Then, the weakly transversal case is handled *via* the "original Meeks-Yau trick" [**MY81**], which reduces the proof to the transversal case. This trick was introduced in the work by Meeks and Yau in the context of minimal surfaces in 3-dimensional manifolds and generalized by Jaco and Rubinstein in the context of PL minimal surfaces in 3-manifolds, see [**JR88**]. The idea is to isotope $\Lambda$ to a (non-minimal) geodesic graph $\Lambda_t$, whose total length is slightly larger than $\Lambda$ but which is transversal to $\Lambda'$: $\ell(\Lambda_t) = \ell(\Lambda) + o(t)$.

The the intersection angle $\alpha_t$ between $\lambda_t$ and $\lambda$ near $p$ can be made arbitrarily close to the original angle $\alpha$. Therefore, by taking $t$ small, one can make the complexity loss $\epsilon$ to be higher than the length gain $\ell(\Lambda_t) - \ell(\Lambda)$. Then, as in Case 1, we obtain a contradiction with minimality of $\lambda, \Lambda'$.

**Step 3: Non-weakly transversal case.** We, thus assume that $\Lambda \cap \Lambda'$ contains no points of transversal intersection. (This case does not happen in the context of minimal surfaces in 3-dimensional Riemannian manifolds.) The idea is again to isotope $\Lambda$ to $\Lambda_t$, so that $\ell(\Lambda_t) = \ell(\Lambda) + o(t)$. One then repeats the arguments from Step 1 (exchange and round off) verifies that the new graph $\widehat{\Lambda}_t$ satisfies

$$\ell(\Lambda_t) - \ell(\widehat{\Lambda}_t) \geqslant O(t).$$

It will then follow that $\ell(\widehat{\Lambda}_t) < \ell(\Lambda)$ when $t$ is sufficiently small, contradicting minimality of $\Lambda$.

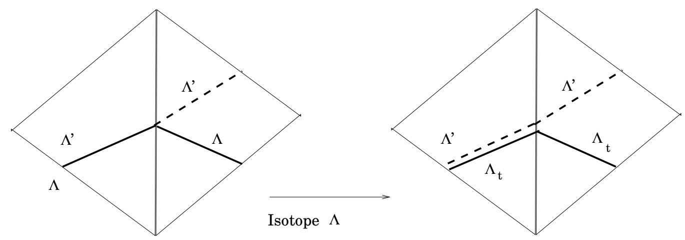



FIGURE 18.8. *M*eeks-Yau trick: Initially, graphs $\Lambda, \Lambda'$ had a common edge. After isotopy of $\Lambda$, this edge is no longer common. The isotopy $\Lambda_t$ is through geodesic graphs, which no longer satisfy the balancing condition.

We now provide the details of the Meeks–Yau trick in this setting. We first push the graph $\Lambda$ in the direction of $\Lambda^+$, so that the result is a smooth family of isotopic Dunwoody tracks $\Lambda_t, t \in [0, t_0], \Lambda_0 = \Lambda$, where each $\Lambda_t$ has geodesic edges and so that each vertex of $\Lambda_t$ is within distance $t$ from the corresponding vertex of $\Lambda$. Since $\Lambda$ was stationary, we have

$$\ell(L_t) = \ell(\Lambda) + ct^2 + o(t^2).$$

It follows from the Maximum Principle (Lemma 18.21) that the graphs $\Lambda_t$ and $\Lambda'$ have to intersect. For sufficiently small values of $t \neq 0$, the intersection is necessarily disjoint from $X^{(1)}$. We now apply the exchange argument and obtain a graph $\Gamma_{t+}$, so that

$$\ell(\Gamma_{t+}) \leqslant \ell(L) + ct^2 + o(t^2).$$

Let $\widehat{\Lambda}_t$ be obtained from $\Gamma_{t+}$ by the round-off argument (straightening the broken edges). Lastly, we need to estimate from below the difference

$$\ell(\widehat{\Lambda}_t) - \ell(\Gamma_{t+}).$$

It suffices to analyze what happens within a single 2-simplex $s$ of $X$ where the graphs $\Lambda_t$ and $\Lambda'$ intersect. We will consider the most difficult case when edges $\lambda \subset \Lambda \cap s, \lambda' \subset \Lambda' \cap s$ share only a vertex $A$ (which, thus, belongs to an edge $e$ of $s$) and $\lambda_t \subset \Lambda_t \cap s$ crosses $\lambda'$ for small $t > 0$. In the case when $\lambda = \lambda'$, the edges $\lambda_t$ and $\lambda'$ will be disjoint for small $t > 0$ and nothing interesting happens during the exchange and round-off argument.

Thus, $\lambda = [A, B], \lambda' = [A, C]$, $\lambda_t = [A_t, B_t], t \in [0, t_0]$ is the variation of $\lambda$ with $\lambda = \lambda_0$. By construction, $\mathrm{dist}(A, A_t) = t, \mathrm{dist}(B, B_t) = t$. Set $D_t := \lambda' \cap \lambda_t$. There are several possibilities for the intersection $\Gamma_t^+ \cap s$. If this intersection contains the broken geodesics $AD_tA_t$ or $B_tP_tC$, then the round-off of $\Gamma_t^+$ will result in reduction of the number of edges, contradicting minimality of $\Lambda$. We, therefore, consider the case when $\Gamma_{t+} \cap s$ contains the broken geodesic $A_tD_tC$, as the case of the path $AD_tB_t$ is similar.

Consider the triangle $\Delta(A, D_t, A_t)$. We note that the angles of this triangle are bounded away from zero and $\pi$ if $t_0$ is sufficiently small. Therefore, the Sine Law for hyperbolic triangles (8.3) implies that $\mathrm{dist}(A_t, D_t) \sim t$ as $t \to 0$. Consider then the triangle $\Delta A_tD_tC$. Again, the angles of this triangle are bounded away from zero and $\pi$ if $t_0$ is sufficiently small. Therefore, Lemma 8.24 implies that

$$\mathrm{dist}(A_t, D_t) + \mathrm{dist}(D_tC) - \mathrm{dist}(A_t, C) \geqslant c_1 \mathrm{dist}(A_t, D_t) \geqslant c_2 t = O(t)$$

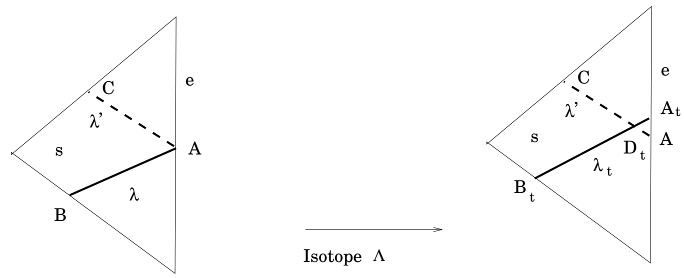



FIGURE 18.9. *M*eeks-Yau trick: Isotopying edge $\lambda$ so that $D_t = \lambda_t \cap \lambda'$ no longer in the edge $e$.

if $t_0$ is sufficiently small. Here $c_1, c_2$ are positive constants. Observe that when we replace $\Gamma_{t+}$ with $\widehat{\Lambda}_t$ (the round-off), the path $[A_t, D_t] \cup [D_t, C]$ is replaced with the geodesic $[A_t, C]$. Therefore,

$$\ell(\Gamma_{t+}) - \ell(\widehat{\Lambda}_t) = O(t).$$

Since

$$\ell(\Gamma_{t+}) - \ell(\Lambda) = o(t),$$

we conclude that

$$\ell(\widehat{\Lambda}_t) - \ell(\Lambda) < 0$$

if $t$ is sufficiently small. This contradicts minimality of $\Lambda$. □

## 18.5. Stallings Theorem for almost finitely-presented groups

DEFINITION 18.23. A group $G$ is said to be *almost finitely-presented* (afp) if it admits a properly discontinuous cocompact simplicial action on a 2-dimensional simplicial complex $X$ such that $H^1(X, \mathbb{Z}_2) = 0$.

For instance, every finitely-presented group is also afp: Take $Y$ to be a finite presentation complex of $G$, subdivide it to obtain a triangulated complex $W$, then let $X$ be the universal cover of $W$. Note that, in view of Lemma 3.25, in the definition of an afp group one can replace a complex $X$ with a new simplicial complex $\widehat{X}$ which is 2-dimensional, has $H^1(X, \mathbb{Z}_2) = 0$, and the action $G \curvearrowright X$ is free properly discontinuous cocompact action $G \curvearrowright X$.

We are now ready to prove

THEOREM 18.24. *Let $G$ be an almost finitely-presented group with at least 2 ends. Then $G$ splits as the fundamental group of a finite graph of finitely-generated groups with finite edge-groups.*

PROOF. Since $G$ is afp, it admits a properly discontinuous cocompact simplicial action on a (locally finite) 2-dimensional simplicial complex $X$ with $H^1(X, \mathbb{Z}_2) = 0$. We give $X' := X \setminus X^{(0)}$ the piecewise-hyperbolic path metric as in Section 18.1.

DEFINITION 18.25. A subset $Z \subset X$ is called *precisely-invariant* (under its $G$-stabilizer) if for every $g \in G$ either $gZ = Z$ or $gZ \cap Z = \emptyset$.

PROPOSITION 18.26. *There exists an finite subgraph $\Lambda \subset X$ which separates ends of $X$ and is precisely-invariant.*

Proof. If $X$ has a cut-vertex, then we take $\Lambda$ to be this vertex. Suppose, therefore, that $X$ contains no such vertices. Then, by Lemma 18.17, $X$ contains a minimal (essential) Dunwoody track $\Lambda \subset X$. By Proposition 18.22, for every $g \in G$, the track $\Lambda' := g\Lambda$ (which is also minimal) is either disjoint from $\Lambda$ or equal to $\Lambda$. □

The proof of Theorem 18.24 then reduces to:

Proposition 18.27. *Every finite subgraph $\Lambda \subset X$ as in Proposition 18.26 gives rise to a nontrivial action of $G$ on a simplicial tree $T$ with finite edge-stabilizers and finitely-generated vertex groups.*

Proof. Let $\Lambda$ be either a cut-vertex of $X$ or a finite connected essential precisely-invariant track $\Gamma \subset X$ (see Definition 18.25). We first consider the more interesting case of when $\Gamma$ is a Dunwoody track.

We partition of $X$ in components of $G \cdot \Gamma$ and the complementary regions. Each complementary region $C_v$ is declared to be a *vertex* $v$ of the partition and each $\Gamma_e := g \cdot \Gamma$ is declared to be an *edge* $e$. Since $\Gamma$ is a Dunwoody track and $H_1(X, \mathbb{Z}_2) = 0$, the complement $X \setminus \Gamma_e$consists of exactly two components $\Gamma_e^{\pm}$; therefore, each edge of the partition is incident to exactly two (distinct) complementary regions. These regions represent vertices incident to $e$. Thus, we obtain a graph $T$. Since the action of $G$ preserves the above partition of $X$, the group $G$ acts on the graph $T$.

Lemma 18.28. *The group $G$ does not fix any vertices of $T$ and does not stabilize any edges.*

Proof. Suppose that $G$ fixes a vertex $v$ of $T$. Let $E_v$ denote the set of edges of $T$ incident to $v$. By relabeling, we can assume that the corresponding component $C_v$ of $X \setminus G \cdot \Gamma$ equals

$$\bigcap_{e \in E_v} \Gamma_e^+.$$

Therefore, for every $x \in C_v$ and every $g \in G$, $g(x) \notin \Gamma_e^-$, $e \in E_v$. Recall that the action $\Gamma \curvearrowright X$ is cocompact. Therefore, there exists a finite subcomplex $K \subset X$ whose $G$-orbit is the entire $X$. Clearly, $x \in K$ for some $x \in C_v$. On the other hand, by the above observation, the intersection

$$G \cdot K \cap \Gamma_e^-$$

is a finite subcomplex. This contradicts the fact that $\Gamma_e^-$ is unbounded. Thus, $G$ does not fix any vertex in $T$. Similarly, we see that $G$ does not preserve any edge of $T$. □

Lemma 18.29. *The graph $T$ is a tree.*

Proof. Connectedness of $T$ immediately follows from connectedness of $X$. If $T$ were to contain a circuit, it would follow that some $\Gamma_e$ did not separate $X$, which is a contradiction. □

Lastly, we observe that compactness of $\Gamma_e$'s and proper discontinuity of the action $G \curvearrowright X$ imply that the stabilizer $G_e$ of every edge $e$ in $G$ is finite. Note that, a prori, $G$ acts on $T$ with inversions since $g \in G$ can preserve $\Gamma_e$ and interchange $\Gamma_e^+, \Gamma_e^-$.

Since the closure $\overline{C}_v$ each vertex-space $C_v$ is connected and $\overline{C}_v/G_v$ is compact, it follows that the stabilizer $G_v$ of each vertex $v \in T$ is finitely-generated (Milnor-Schwartz Lemma).

Suppose now that $\Lambda$ is a single vertex $v$. If $\Lambda$ were to separate $X$ in exactly two components, we would be done by repeating the arguments above. Otherwise, we modify $X$ by replacing $v$ with an edge $e$ whose mid-point $m$ separates $X$ in exactly two components both of which are unbounded. We repeat this for every point in $G \cdot v$ in $G$-equivariant fashion. The result is a new complex $\widehat{X}$ with a cocompact action $G \curvearrowright \widehat{X}$. Clearly, $\Lambda := \{m\}$ is a precisely-invariant, so we are done as above. Proposition 18.27 follows. □

In both cases, the quotient graph $\Gamma = T/G$ is finite since the action $G \curvearrowright X$ is cocompact.

We can now finish the proof of Theorem 18.24. In view of Proposition 18.27 Bass–Serre correspondence (Section 4.7.5), implies that the group $G$ is the fundamental group of a nontrivial finite graph of groups $\mathcal{G}$ with finite edge groups and finitely-generated vertex groups. □

### 18.6. Accessibility

Let $G$ be a finitely-generated group which splits nontrivially as an amalgam $G_1 \star_H G_2$ or $G_1 \star_H$ with finite edge-group $H$. Sometimes, this decomposition process can be iterated, by decomposing the groups $G_i$ as amalgams with finite edge groups, etc. The key issue that we will be addressing in this section is: Does the decomposition process terminates after finitely many steps. Groups for which termination does occur are called *accessible*. In the case of trivial edge groups (e.g., when $G$ is torsion-free) Grushko's Theorem (see e.g. [**SW79**] for a topological proof), the decomposition process does terminate, so torsion-free finitely-generated groups are accessible. M. Dunwoody constructed an example of a finitely-generated group which is not accessible [**Dun93**]. The main result of this section is

THEOREM 18.30 (M. Dunwoody, [**Dun85**]). *Every almost finitely-presented group is accessible.*

Before proving this theorem, we will establish several technical facts.

**Refinements of graphs of groups.** Let $\mathcal{G}$ be a graph of groups with the underlying graph $\Gamma$, let $H = G_v$ be one of its vertex groups. Let $\mathcal{H}$ be a graph-of groups decomposition of $H$ with the underlying graph $\Lambda$. Suppose that:

ASSUMPTION 18.31. For every edge $e \subset \Gamma$ incident to $v$, the subgroup $G_e \subset H$ is conjugate in $H$ to a subgroup of one of the vertex groups $H_w$ of $\mathcal{H}$, $w = w(e)$ (this vertex need not be unique). For instance, if every $G_e$ is finite, then, in view of Property FA for finite groups, $G_e$ will fix a vertex in the tree corresponding to $\mathcal{H}$. Thus, our assumption will hold in this case.

Under this assumption, we can construct a new graph of groups decomposition $\mathcal{F}$ of $G$ as follows. Cut $\Gamma$ open at $v$, i.e. remove $v$ from $\Gamma$ and then replace each open or half-open edge of the resulting space with a closed edge. The resulting graph $\Gamma'$ could be disconnected. We have the natural map $r : \Gamma' \to \Gamma$. Let $\Phi$ denote the graph obtained from the union $\Gamma' \sqcup \Lambda$ by identifying each vertex $v_i' \in r^{-1}(v) \in e_i' \subset \Gamma'$ with the vertex $w(e) \in \Lambda$ as in the above assumption. Then $\Phi$ is connected. We retain for $\Phi$ the vertex and edge groups and the group homomorphisms coming from $\Gamma$ and $\Phi$. The only group homomorphisms which need to be defined are for the edges $e = [e^-, e^+]$, where $e^- = w(e) = w$. In this case, the embedding $G_e \to G_w$ is the one given by the conjugation of $G_e$ to the corresponding subgroup of $G_w$.

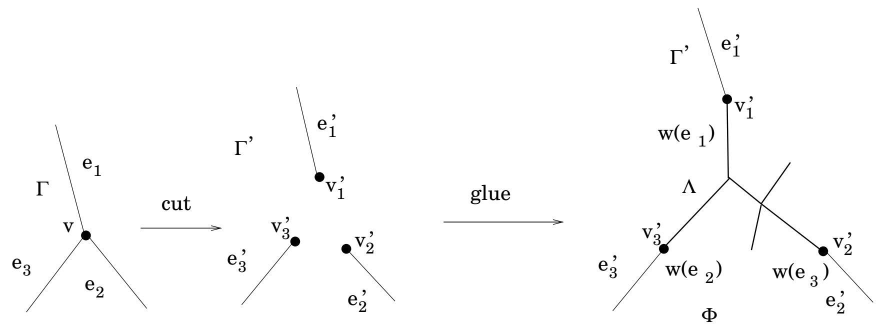



FIGURE 18.10. *C*ut $\Gamma$ open and glue $\Phi$ from $\Gamma'$ and $\Lambda$.

We leave it to the reader to verify (using Seifert – Van Kampen theorem) that $\pi_1(\Phi) = G$.

DEFINITION 18.32. The new graph of groups decomposition $\mathcal{F}$ of the group $G$ is called the *refinement* of $\mathcal{G}$ *via* $\mathcal{H}$. A refinement is said to be *trivial* if $\mathcal{H}$ is a trivial graph of groups. We use the notation $\mathcal{G} \prec \mathcal{F}$ for a refinement.

DEFINITION 18.33. A group $G$ is said to be *accessible* if it admits a (finite) graph of groups decomposition with finite edge groups and 1-ended vertex groups.

In [**TW93**] it was proven that a finitely-generated group $G$ is accessible if and only if one (equivalently, every) Cayley graph $\Gamma$ of $G$ satisfies the following property:

There exists a number $D$ so that every two ends of $\Gamma_G$ can be separated by a bounded subset of $\Gamma$ of diameter $\leqslant D$.

As an immediate corollary of their result we obtain

THEOREM 18.34. *Accessibility is QI invariant.*

Our goal below is to give a proof (mostly following papers by Dunwoody [**Dun85**] and Swarup [**Swa93**]) of Dunwoody's theorem that every almost finitely-presented group is accessible.

PROPOSITION 18.35. *Let $\mathcal{G}$ be a finite graph of finitely-generated groups with finite edge-subgroups (with the underlying graph $\Lambda$). Then:*

*1. Every vertex subgroup $G_v$ is QI embedded in $G$. (Note that finite generation of the vertex groups implies that $G$ is itself finitely-generated.)*

*2. If, in addition, the group $G$ is finitely-presented, then every vertex group of $\mathcal{G}$ is also finitely-presented.*

PROOF. The proofs of both statements are very similar. We first construct (as in Section 4.7.4) a tree of graphs $\mathcal{Z}$, corresponding to $\mathcal{G}$.

Namely, let $G \curvearrowright T$ be the action of $G$ on a tree corresponding to the graph of groups $\mathcal{G}$, see Section 4.7.4). For every edge-group $G_e$ in $\mathcal{G}$ we take $S_e := G_e \setminus \{1\}$ as the generating set of $G_e$. For every vertex-group $G_v$ in $\mathcal{G}$ we pick a finite generating set $S_v$ of $G_v$, so that for every edge $e = [v, w]$, the sets $S_v, S_w$ contain the images of $S_e$ under the embeddings $G_e \to G_v, G_e \to G_w$.

Then, using the projection $p : T \to \Lambda = T/G$, we define generating sets $S_{\tilde{v}}, S_{\tilde{e}}$ of $G_{\tilde{v}}$, $G_{\tilde{e}}$ using isomorphisms $G_{\tilde{v}} \to G_v, G_{\tilde{e}} \to G_e$, where $\tilde{v}, \tilde{e}$ project to $v, e$ under the map $T \to \Lambda$. Let $Z_{\tilde{v}}, Z_{\tilde{e}}$ denote the Cayley graphs of the groups $G_{\tilde{v}}, G_{\tilde{e}}$ ($\tilde{v} \in V(T), \tilde{e} \in E(T)$) with respect to the generating sets $S_{\tilde{v}}, S_{\tilde{e}}$. (provided that $\tilde{v}$ projects to

Now, to simplify the notation, we will use the notation $v$ and $e$ for vertices and edges of $T$. Let $Z_v, Z_e$ denote the Cayley graphs of the groups $G_v, G_e$ ($v \in V(T), e \in E(T)$) with respect to the generating sets $S_v, S_e$ defined above. We have natural injective maps of graphs $f_{ev} : Z_e \hookrightarrow Z_v$, whenever $v$ is incident to $e$. The resulting collection of graphs $Z_v, Z_e$ and embeddings $Z_e \hookrightarrow Z_v$, defines a *tree of graphs* $\mathcal{Z}$ with the underlying tree $T$. For each $Z_e$ we consider the product $Z_e \times [-1, 1]$ with the standard triangulation of the product of simplicial complexes. Let $\tilde{Z}_e$ be the 1-skeleton of this product. Lastly, let $Z$ denote the graph obtained by identifying vertices and edges of each $\tilde{Z}_e$ with their images in $Z_v$ under the maps $f_{ev} \times \{\pm 1\}$. We endow $Z$ with the standard metric.

Clearly, the group $G$ acts on $Z$ freely, preserving the labels; the quotient graph $Z/G$ has only finitely many vertices, the graph $Z/G$ is finite if each $G_v$ is finitely-generated.

For every $v \in V(T)$ define the map $\rho_v : Z^0 \to Z_v^0 = G_v$ to be the $G_v$-equivariant nearest-point projection. Since for every $e = [v, w]$, $f_v(Z_e) \subset Z_v$ separates $Z$, and every two distinct vertices in $f_v(Z_e)$ are connected by an edge in this graph, it follows that for $x, y \in Z^0$ within unit distance from each other,

$$\text{dist}(\rho_v(x), \rho_v(y)) \leqslant 1.$$

Hence, the map $\rho_v$ is 1-Lipschitz. Hence, we extend each $\rho_v$ ($G_v$-equivariantly) to the entire graph $Z$.

We now can prove the assertions of the proposition.

1. Since each $G_v$ is finitely-generated, the action $G \curvearrowright Z_v$ is geometric and, hence, the action $G \curvearrowright Z$ is geometric as well. Thus, the space $Z$ is QI to the group $G$ and $Z_v$ is QI to $G_v$ for every vertex $v$. Let $x, y \in Z_v$ be two vertices and $\alpha \subset Z$ be the shortest path connecting them. Then $\rho(\alpha) \subset Z$ still connects $x$ to $y$ and has length which is at most the length of $\alpha$. It follows that $Z_v$ is isometrically embedded in $Z$. Hence, each $G_v$ is QI embedded in $G$. This proves (1).

2. Since $G$ is finitely-presented and $G \curvearrowright Z$ is geometric, the space $Z$ is coarsely simply-connected by Corollaries 6.19 and 6.40. Our goal is to show that each vertex space $Z_v$ of $Z$ is also coarsely simply-connected. Let $\alpha$ be a loop in a vertex space $Z_v$. Since $Z$ is coarsely simply-connected, there exists a constant $C$ (independent of $\alpha$) and a collection of (oriented) loops $\alpha_i$ in the 1-skeleton of $Z$ so that

$$\alpha = \prod_i \alpha_i$$

and each $\alpha_i$ has length $\leqslant C$. We then apply the retraction $\rho$ to the loops $\alpha_i$. Then

$$\alpha = \prod_i \rho(\alpha_i)$$

and $\text{length}(\rho(\alpha_i)) \leqslant \text{length}(\alpha_i) \leqslant C$ for each $i$. Thus, $Z_v$ is coarsely simply-connected and, therefore, $G_v$ is finitely-presented. □

We are now ready to prove Dunwoody's accessibility theorem.

*Proof of Theorem 18.30.* We will construct inductively a finite chain of refinements of

$$\mathcal{G}_1 \prec \mathcal{G}_2 \prec \mathcal{G}_3 \ldots$$

of graph-of-groups decompositions with finite edge groups, so that the terminal graph of groups has 1-ended vertex groups.

Let $X$ be a 2-dimensional simplicial complex with $H^1(X, \mathbb{Z}_2) = 0$, so that $X$ admits a simplicial properly discontinuous, cocompact action $G \curvearrowright X$. We let $\sigma(G, X)$ denote the total number of simplices in $X/G$ and let $\sigma(G)$ denote the minimum of the numbers $\sigma(G, X)$ where the minimum is taken over all complexes $X$ and group actions $G \curvearrowright X$ as above. Suppose $X$ minimizes $\sigma(G)$ and contain an cut-vertex $v$ (see Definition 18.16). Then, as in the proof of Theorem 18.24, we split $G$ as a graph of groups (with the edge-groups stabilizing $v$) so that vertex-group $G_i$ acts on a subcomplex $X_i \subset X$, where the frontier of $X_i$ in $X$ is contained in $G \cdot v$. It follows from the Mayer-Vietoris sequence that $H^1(X_i, \mathbb{Z}_2) = 0$ for each $i$. Thus, $\sigma(G_i) < \sigma(G)$. Therefore, this decomposition process eventually terminates.

Hence, without loss of generality, we may assume that $X$ has no cut-vertices. If the group $G$ is 1-ended we are done. Suppose that $G$ has at least 2 ends. Then, by Propositions 18.26, 18.27, there exists a (connected) finite precisely-invariant track $\tilde{\tau}_1 \subset X_1 := X$ which determines a nontrivial graph of groups decomposition $\mathcal{G}_1$ of $G_1 := G$ with finite edge groups. Our assumption that $H^1(X, \mathbb{Z}_2) = 0$ implies that $\tau_1$ is 2-sided in $X_1 := X$. Let $X_2$ be the closure of a connected component of $X \setminus G \cdot \tilde{\tau}_1$. By compactness of $X/G$ and Milnor-Schwartz Lemma, stabilizer $G_2$ of $X_2$ in $G$ is finitely-generated. Since $H^1(X, \mathbb{Z}_2) = 0$, it follows by excision and Mayer-Vietoris sequence that

$$H^1(X_2, \partial X_2; \mathbb{Z}_2) = 0,$$

where $\partial X_2$ is the frontier of $X_2$ in $X_1$.

Therefore, if define $W_2$ by pinching each boundary component of $X_2$ to a point, then $H^1(W_2, \mathbb{Z}_2) = 0$. The stabilizer $G_2$ of $X_2$ in $G$ acts on $W_2$ properly discontinuously cocompactly. Therefore, each vertex group of $\mathcal{G}_1$ is again afp.

If each vertex group of $\mathcal{G}_1$ is 1-ended, we are again done. Suppose therefore that the closure $X_2$ of some component $X_1 \setminus G \cdot \tilde{\tau}_1$ as above has stabilizer $G_2 < G_1$ which has at least two ends. According to Theorem 6.10, $G_2$ splits (nontrivially) as a graph of groups with finite edge groups. Let $G_2 \curvearrowright T_2$ be a nontrivial action of $G_2$ on a simplicial tree (without inversions) which corresponds to this decomposition. Since each edge-group of $\mathcal{G}_1$ is finite, if such group is contained in $G_2$, it has to fix a vertex in $T_2$, see Corollary 9.21. Recall that the edge-groups of $\mathcal{G}_1$ are conjugate to the stabilizers of components of $G \cdot \tilde{\tau}_1$ in $G$. Therefore, every such stabilizer fixes a vertex in $T_2$. We let $X_2^+$ denote the union $X_2$ with all simplices in $X$ which have nonempty intersection $\partial X_2$. Clearly, $G_2$ still acts properly discontinuously cocompactly on $X_2^+$. The stabilizer of each component of $X_2^+ \setminus X_2$ is an edge group of $\mathcal{G}_1$.

We then construct a ($G_1$-equivariant) PC map (see Definition 18.2) $f_2 : X_2^+ \to T_2$ which sends components of $X_2^+ \setminus int(X_2)$ to the corresponding fixed vertices in $T_2$, sends vertices to vertices of $T$ and is linear on each edge of the cell-complex $X_2$. Since $G_2 \curvearrowright T_2$ is nontrivial, the image of the map $f_2$ is unbounded, otherwise, this image will contain a bounded $G_2$-orbit contradicting Cartan's Theorem (Theorem

2.42). Therefore, the image of $f_2$ contains an edge $e \subset T_2$ which separates $T_2$ in at least two unbounded subsets.

Then, by Lemma 18.3, there exists a point $t \in e$ which is a regular value of $f_2$. Thus, by Exercise 18.5, $f^{-1}(t)$ is a track. It is immediate that $f^{-1}(t)$ is precisely-invariant in $X_2$ with finite $G_2$-stabilizer. By the choice of $e$, the graph $f^{-1}(t)$ separates $X$ in at least two unbounded component. Let $\tilde{\tau}_2$ be an essential component of $f^{-1}(t)$.

Thus, by Proposition 18.27, the track $\tau_2$ determines a decomposition of $G_2$ as a graph of groups $\mathcal{G}_3$ with finite edge groups. We continue this decomposition inductively. We obtain a collection of pairwise disjoint connected tracks $\tau_1, \tau_2, ... \subset Y = X/G$ which are projections of the tracks $\tilde{\tau}_i \subset X$.

Suppose that the number of tracks $\tau_i$ is $> 6F + r$, where $F$ is the number of faces in $X$ and $r$ is the dimension of $H^1(X, \mathbb{Z}_2)$. Then, by Lemma 18.11, every $\tau_k$, $k > 6F + r$ is isotopic to some graph $\tau_{i(k)}, i = i(k) \leqslant 6F + r$. Let $R$ be the product region in $Y$ bounded by such tracks. Lifting this region in $X$ we again obtain a product region $\tilde{R}$ bounded by tracks $g_i\tilde{\tau}_i, g_k\tilde{\tau}_k$, $g_i, g_k \in G$. Therefore, the stabilizers of $g_i\tilde{\tau}_i, g_k\tilde{\tau}_k$ and $R$ in $G$ have to be the same. It follows that every $X_k$, $k > 6F + r$ is a product region whose stabilizer fixes its boundary components. The corresponding tree $T_k$ is just a union of two edges which are fixed by the entire group $G_k$. This contradicts the fact that each graph of groups $\mathcal{G}_k$ is nontrivial. Therefore, the decomposition process of $G$ terminates after $6F + r$ steps and $G$ is accessible. □

### 18.7. QI rigidity of virtually free groups

Proposition 18.36. *Let $G$ be the fundamental group of a (finite) graph of finite groups. Then $G$ is virtually free.*

Proof. Arguing inductively (since every graph of groups converts to an amalgam, see Section 4.7.3), it suffices to prove the following:

Suppose that $G = G_v \star_{G_e} G_w$ or $G = G_v \star_{G_e}$ where $G_v, G_w$ are virtually free and $G_e$ is finite. Then $G$ is again virtually free.

We consider the case of an amalgamated free product as the other case is similar. The graph of groups $\mathcal{G}$ corresponding to this amalgam consists of a single edge $[v, w]$. We let $u$ be the midpoint of this edge. Let $G'_v \lhd G_v, G'_w \lhd G_w$ denote normal free subgroups of finite indices $m, n$ respectively.

We now construct a new graph of groups $\mathcal{G}'$ as follows. Define graphs $\Lambda_v, \Lambda_w$ which are stars with the center $v$ (respectively, $w$) and $m$ (respectively $n$) edges $[v, u_i], [w, u_j]$ emanating from these centers. The groups $G_v/G'_v, G_w/G'_w$ act naturally on these graphs fixing the centers, so that the quotients are the edges $[v, u], [w, u]$ respectively.

We label the vertex $v$ of $\Lambda_v$ by the group $G'_v$ and the vertex $w$ of $\Lambda_w$ by the group $G'_w$.

Now, take $m$ copies $\Lambda_{v,i}, i = 1, \ldots, n$ of $\Lambda_v$ and $n$ copies $\Lambda_{w,i}, j = 1, \ldots, m$ of $\Lambda_w$, these are stars with centers $v_1, \ldots, v_n, w_1, \ldots, w_m$ respectively. Note that the disjoint union $U_v$ of the graphs $\Lambda_{v,i}$ has $nm$ legs while the disjoint union $U_w$ of the graphs $\Lambda_{v,j}$ also has $mn$ legs. Therefore, we can bijectively match the valence 1 vertices of $U_v$ to the valence 1 vertices of $U_w$. Let $\Gamma'$ denote the graph resulting from this gluing, where we erase the valence two vertices (corresponding to the vertices

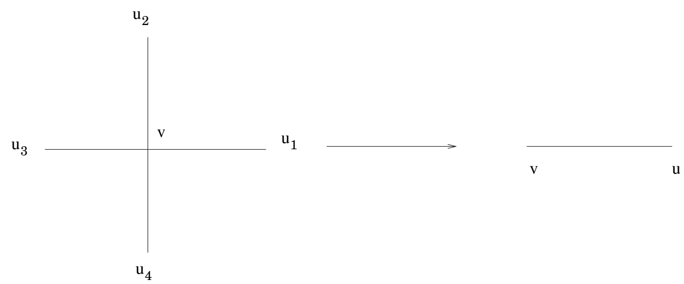



FIGURE 18.11. *S*tar $\Lambda_v$.

$u_i, u_j$ in $\Lambda_v, \Lambda_w$). We convert $\Gamma$ to a graph of groups $\mathcal{G}'$ with trivial edge-groups and the vertex groups labeled by $G'_v, G'_w$ as before.

We have a natural projection (a *covering* of graphs of groups) $p : \Gamma' \to \Gamma$ of degree $d = nm$, so that $p(v_i) = v, p(w_j) = w$ and which induces the embeddings $G'_{v_i} \to G'_v < G_v, G_{w_j} \to G'_w < G$. We claim that this projection induces a finite index embedding $G' := \pi_1(\mathcal{G}') \hookrightarrow G = \pi_1(\mathcal{G})$. This follows easily by constructing the corresponding finite covering (of degree $d$) $q : Y' \to Y$ between graphs of spaces corresponding to the graphs of groups $\mathcal{G}, \mathcal{G}'$. (One constructs these graphs of spaces using the Milnor's construction of the classifying spaces for $G_v, G_w, G_e$ and taking appropriate finite coverings over these classifying spaces in order to build $Y'$.) Since finite covering corresponds to an embedding $G' \to G$ as a finite index subgroup, the claim follows.

Now, the group $G'$ is the fundamental group of a graph of groups $\mathcal{G}'$ with trivial edge groups and free vertex groups (of finite rank).

LEMMA 18.37. *The group $G' = \pi_1(\mathcal{G}')$ is free.*

PROOF. We construct a new graph of spaces $Z$ realizing $\mathcal{G}'$ by using finite roses $Z_{v,i}, Z_{w,j}$ as the classifying spaces for the (free) vertex groups of $\mathcal{G}'$ and points as the classifying spaces for the trivial edge groups. The result is a graph whose fundamental group is $G'$. Therefore, $G'$ is free. □

This concludes the proof of Proposition as well. Thus, $G' < G$ is a free subgroup of finite index. □

THEOREM 18.38. *If $G$ is virtually free, then every group $G'$ which is QI to $G$ is also virtually free.*

PROOF. If the group $G$ is finite, the assertion is clear. If $G$ is virtually cyclic, then $G'$ and $G$ are 2-ended, which, by Part 3 of Theorem 6.8, implies that $G'$ is also virtually cyclic.

Suppose now that $G$ has infinitely many ends. Since $G$ is finitely-presented, by Corollary 6.40, the group $G'$ is finitely-presented as well. The group $G$ acts geometrically on a locally finite simplicial tree $T$ with infinitely many ends, therefore, $G$ and $G'$ are QI to $T$. Since $G'$ is finitely-presented, by Theorem 18.30, the group $G'$ splits as a graph of groups where every edge group is finite, every vertex group is finitely-generated and each vertex group has 0 or 1 ends.

By Proposition 18.35, every vertex group $G'_v$ of this decomposition is QI embedded in $G'$. In particular, every $G'_v$ is quasi-isometrically embedded in the tree.

The image of such an embedding is coarsely-connected (with respect to the restriction of the metric on $T$) and, therefore, is within finite distance from a subtree $T'_v \subset T$. Thus, each $G'_v$ is QI to a locally-finite simplicial tree (embedded in $T$).

LEMMA 18.39. *$T'_v$ cannot have one end.*

PROOF. Suppose that $T'_v$ has one end. The group $G'_v$ is infinite and finitely-generated. Therefore, its Cayley graph contains a bi-infinite geodesic (see Exercise 4.74). Such geodesic $\gamma$ cannot embed quasi-isometrically in a 1-ended tree (since both ends of $\gamma$ would have to map to the same end of $T'_v$). □

Thus, every vertex group $G'_v$ has 0 ends and, hence, is finite. By Proposition 18.36, the group $G$ is virtually free. □

CHAPTER 19

# Proof of Stallings Theorem using harmonic functions

In his essay [**Gro87**, Pages 228–230], Gromov gave a proof of the Stallings theorem on ends of groups using harmonic functions. The goal of this chapter is to provide details for Gromov's arguments. One advantage is that this proof works for finitely generated groups without finite presentability assumption. However, the geometry behind the proof is less transparent as in Chapter 18. The proof that we present in this chapter is a simplified form of the arguments which appear in [**Kap07**].

Suppose that $G$ is a finitely-generated group with infinitely many ends. Then $G$ admits an isometric free properly discontinuous cocompact action $G \curvearrowright M$ on a Riemannian manifold $M$, which, then, necessarily has bounded geometry (since it covers a compact Riemannian manifold). For instance, if $G$ is $k$–generated, and $F$ is a Riemann surface of genus $k$, we have an epimorphism

$$\phi : \pi_1(F) \to G.$$

Then $G$ acts isometrically and cocompactly on the covering space $M$ of $F$ so that $\pi_1(M) = \mathrm{Ker}(\phi)$. Then the space $\epsilon(M)$ of ends of $M$ is naturally homeomorphic to the space of ends of $G$, see Proposition 6.6. Let $\overline{M} := M \cup \epsilon(M)$ denote the compactification of $M$ by its space of ends, which is necessarily $G$-equivariant.

We will see in Section 19.2 that given a continuous function $\chi : \epsilon(M) \to \{0, 1\}$, there exists a function

$$h = h_\chi : \overline{M} \to [0, 1],$$

a continuous extension of $\chi$, so that the function $h|M$ is harmonic.

Let $H(M)$ denote the space of harmonic functions

$$\{h = h_\chi, \chi : \epsilon(M) \to \{0, 1\} \text{ is nonconstant}\}.$$

We give $H(M)$ the topology of uniform convergence on compacts in $M$. Let $E : H(M) \to \mathbb{R}_+ = [0, \infty)$ denote the energy functional.

Definition 19.1. Given the manifold $M$, define its *energy gap* $e(M)$ as

$$e(M) := \inf\{E(h) : h \in H(M)\}.$$

Clearly, the isometric group action $G \curvearrowright M$ yields an action $G \curvearrowright H(M)$ preserving the functional $E$. Therefore $E$ projects to a lower semi-continuous (see Theorem 2.26) functional $E : H(M)/G \to \mathbb{R}_+$, where we give $H(M)/G$ the quotient topology. The main technical result needed for the proof is

Theorem 19.2. *1. $e(M) \geqslant \mu > 0$, where $\mu$ depends only on $R$, $\lambda_1(M)$.*
*2. The functional $E : H(M)/G \to \mathbb{R}_+$ is proper in the sense that the preimage*

$$E^{-1}([0, T])$$

*is compact for every $T \in \mathbb{R}_+$. In particular, $e(M)$ is attained.*

We now sketch our proof of the Stallings' theorem. Let $h \in H(M)$ be an energy-minimizing harmonic function guaranteed by Theorem 19.2. We then verify that the set $\{h(x) = \frac{1}{2}\}$ is *precisely-invariant* with respect to the action of $G$. By choosing $t$ sufficiently close to $\frac{1}{2}$ we obtain a smooth hypersurface $S = \{h(x) = t\}$ which is precisely-invariant under $G$ and separates the ends of $M$. If this hypersurface were connected, we could use the standard construction of a dual simplicial tree $T$ whose edges are the "walls", i.e., the images of $S$ under the elements of $G$ and the vertices are the components of $M \setminus G \cdot S$. In the general case, a "wall" can be adjacent to more than two connected component of $M \setminus G \cdot S$. We show however that each wall is adjacent to exactly two "indecomposable" subset of $M \setminus G \cdot S$, i.e., a subset which cannot be separated by one wall. These indecomposable sets are the vertices of $T$. We then verify that the graph $T$ is actually a tree.

It was observed by W. Woess that the arguments in this paper generalize directly to harmonic functions on graphs. In particular, smoothness of harmonic functions (emphasized by Gromov in [**Gro87**, Pages 228–230]) becomes irrelevant. One advantage of this approach is to avoid the discussion of nodal sets of harmonic functions. We observe that many of our arguments simplify if we take $M$ to be a Riemann surface (or a graph), which suffices for the proof of Stallings theorem. We wrote the proofs in greater generality because the compactness theorem for harmonic functions appears to be of independent interest.

### 19.1. Proof of Stallings' theorem

The goal of this section is to prove Stallings theorem assuming Theorem 19.2. Our argument is a slightly expanded version of Gromov's proof in [**Gro87**, Pages 228–230].

Let $H(M)$ denote the space of harmonic functions $h : M \to (0,1)$ as in the Introduction. According to Theorem 19.2, there exists a function $h \in H(M)$ with minimal energy $E(h) = e(M) > 0$. Then, for every $g \in G$, the function

$$g^*h := h \circ g$$

has the same energy as $h$ and equals

$$h_{g^*(\chi)}.$$

For $g \in G$, define

$$g_+(h) := \max(h, g^*(h)), \quad g_-(h) := \min(h, g^*(h)).$$

We will see (Lemma 19.8) that

$$E(g_+(h)) + E(g_-(h)) = 2E(h).$$

Note that the functions $g_+(h), g_-(h)$ have continuous extension to $\overline{M}$ (since $h$ does and $G$ acts on $\overline{M}$ by homeomorphisms). By construction, the restrictions

$$\chi_+ := g_+(h)|\epsilon(M), \quad \chi_- := g_-(h)|\epsilon(M)$$

take the values 0 and 1 on $\epsilon(M)$. Let

$$h_\pm := h_{\chi_\pm}$$

denote the corresponding harmonic functions on $M$. Then

$$E(h_\pm) \leqslant E(g_\pm(h)),$$

$$E(h_+) + E(h_-) \leqslant 2E(h) = 2e(M).$$

Note that it is, *a priori*, possible that $\chi_-$ or $\chi_+$ is constant. Set

$$G_c := \{g \in G : \chi_- \text{ or } \chi_+ \text{ is constant}\}.$$

We first analyze the set $G \setminus G_c$. For $g \notin G_c$, both $h_-$ and $h_+$ belong to $H(M)$ and, hence,

$$E(h_+) = E(h_-) = E(h) = e(M).$$

Therefore,

$$E(g_+(h)) = E(h_+), \quad E(g_-(h)) = E(h_-).$$

It follows that $g_\pm(h)$ are both harmonic. Since

$$g_-(h) \leqslant g_+(h),$$

the maximum principle implies that either $g_-(h) = g_+(h)$ or $g_-(h) < g_+(h)$. Hence, the set $\Lambda_g$ is either empty or equals the entire $M$, in which case $g^*(h) = h$. Therefore, for every $g \in G \setminus G_c$ one of the following holds:

1. $g^*h = h$.
2. $g^*h(x) < h(x), \ \forall x \in M$.
3. $g^*h(x) > h(x), \ \forall x \in M$.

Thus, the set

$$L := h^{-1}\left(\frac{1}{2}\right)$$

is *precisely–invariant* under the elements of $G \setminus G_c$: for every $g \in G \setminus G_c$, either

$$g(L) = L$$

or

$$g(L) \cap L = \emptyset.$$

We now consider the elements of $G_c$. Suppose that $g$ is such that $\chi_- = 0$. Then

$$g^*(\chi) \leqslant 1 - \chi$$

and, hence,

$$g^*(h) \leqslant 1 - h.$$

Since these functions are harmonic, in the case of the equality at some $x \in M$, by the maximum principle we obtain $g^*(h) = 1 - h$. The latter implies that

$$g(L) = L.$$

If

$$g^*(h) < 1 - h$$

then $g(L) \cap L = \emptyset$. The same argument applies in the case when $\chi_+$ is constant.

To summarize, for every $g \in G$ one of the following holds:

$$g^*h = h, \quad g^*h < h, \quad g^*h > h, \quad g^*h = 1 - h, \quad g^*h < 1 - h, \quad g^*h > 1 - h. \tag{19.1}$$

We conclude that $L$ is precisely–invariant under the action of the entire group $G$. Moreover, if $g(L) = L$ then either $g^*h = h$ or $g^*h = 1 - h$. Since $L$ is compact, its stabilizer $G_L$ in $G$ is finite.

By construction, the hypersurface $L$ separates $M$ in at least two unbounded components.

Since $L$ is compact, by Sard's Theorem, there exists $t \in (0, 1) \setminus \frac{1}{2}$ sufficiently close to $\frac{1}{2}$, which is a regular value of $h$, so that the hypersurface $S := h^{-1}(t)$ is still precisely–invariant under $G$. Let $G_S \subset G_L$ denote the stabilizer of $S$ in $G$.

We now show that $G$ splits nontrivially over a subgroup of $G_S$. (The proof is straightforward under the assumption that $S$ is connected, but requires extra work in general.) We proceed by constructing a simplicial $G$–tree $T$ on which $T$ acts without inversions, with finite edge–stabilizers and without a global fixed vertex.

**Construction of $T$.** Consider the family of functions $\mathcal{H} = \{f = g^* h : g \in G\}$. Each function $f \in \mathcal{H}$ defines the *wall* $W_f = \{x : f(x) = t\}$ and the *half-spaces* $W_f^+ := \{x : f(x) > t\}$, $W_f^- := \{x : f(x) < t\}$ (these spaces are not necessarily connected).

Let $\mathcal{E}$ denote the set of walls. We say that a wall $W_f$ *separates* $x, y \in M$ if

$$x \in W_f^+, \quad y \in W_f^-.$$

Maximal subsets $V$ of

$$M^o := M \setminus \bigcup_{f \in \mathcal{H}} W_f$$

consisting of points which cannot be separated from each other by a wall, are called *indecomposable* subsets of $M^o$. Note that such sets need not be connected. Set

$$\mathcal{V} := \{\text{indecomposable subsets of } M^o\}.$$

We say that a wall $W$ is *adjacent* to $V \in \mathcal{V}$ if $W \cap cl(V) \neq \emptyset$.

The next lemma follows immediately from the inequalities (19.1), provided that $t$ is sufficiently close to $\frac{1}{2}$:

LEMMA 19.3. *No wall $W_{f_1}$ separates points of another wall $W_{f_2}$.*

LEMMA 19.4. *1. Let $V \in \mathcal{V}$ and $W \in \mathcal{E}$ be adjacent to $V$. Then, for each component $C$ of $V$, we have $C \cap W \neq \emptyset$.*

*2. $W \in \mathcal{E}$ is adjacent to $V \in \mathcal{V}$ if and only if $W \subset cl(V)$.*

PROOF. 1. Suppose that $V \subset W^+$. A generic point $x \in C$ is connected to $W = W_f$ by a gradient curve $p : [0, 1] \to M$ of the function $f$. The curve $p$ crosses each wall at most once. Since $V$ is indecomposable and for sufficiently small $\epsilon > 0$, $p(1 - \epsilon) \in V$, it follows that $p$ does not cross any walls. Therefore the image of $p$ is contained in the closure of $C$ and $p(1) \in W \cap cl(C)$

2. Lemma 19.3 implies that for $x, y \in W^+$ (resp. $x, y \in W^-$) which are sufficiently close to $W$, there is no wall which separates $x$ from $y$. Therefore, such points $x, y$ belong to the same indecomposable set $V^+$ (resp. $V^-$) which is adjacent to $W$ and $W \subset cl(V^\pm)$. Clearly, $V^+, V^-$ are the only indecomposable sets which are adjacent to $W$. □

Hence, each wall $W$ is adjacent to exactly two elements of $\mathcal{V}$ (contained in $W^+, W^-$ respectively). We obtain a graph $T$ with the vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, where a vertex $V$ is incident to an edge $W$ if and only if the wall $W$ is adjacent to the indecomposable set $V$.

From now on, we abbreviate $W_{f_i}$ to $W_i$.

LEMMA 19.5. *$T$ is a tree.*

PROOF. By construction, every point of $M$ belongs to a wall or to an indecomposable set. Hence, connectedness of $T$ follows from connectedness of $M$.

Let

$$W_1 - V_1 - W_2 - .... - W_k - V_k - W_1$$

be an embedded cycle in $T$. This cycle corresponds to a collection of paths $p_j : [0,1] \to cl(V_j)$, so that

$$p_j(0) \in W_j, \quad p_j(1) \in W_{j+1}, j = 1, ..., k$$

and points of $p_j([0,1])$ are not separated by any wall, $j = 1, ..., k$. By Lemma 19.3, the points $p_j(1), p_{j+1}(0)$ are not separated by any wall either. Therefore, the points of

$$\bigcup_{j=1}^{k} p_j([0,1])$$

are not separated by $W_1$. However,

$$p_1((0,1]) \subset W_1^+, \quad p_k([0,1)) \subset W_1^-$$

or vice–versa. Contradiction. □

We next note that $G$ acts naturally on $T$ since the sets $\mathcal{H}$, $\mathcal{E}$ and $\mathcal{V}$ are $G$–invariant and $G$ preserves adjacency. If $g(W_f) = W_f$, then $g^*f = f$, which implies that $g$ preserves $W_f^+, W_f^-$. Hence, $g$ fixes the end-points of the edge corresponding to $W$, which means that $G$ acts on $T$ without inversions. The stabilizer of an edge in $T$ corresponding to a wall $W$ is finite, since $W$ is compact and $G$ acts on $M$ properly discontinuously.

Suppose that $G \curvearrowright T$ has a fixed vertex. This means that the corresponding indecomposable subset $V \subset M$ is $G$–invariant. Since $G$ acts cocompactly on $M$, it follows that $M = B_r(V)$ for some $r \in \mathbb{R}_+$. The indecomposable subset $V$ is contained in the half-space $W^+$ for some wall $W$. Since $W$ is compact and $W^-$ is not, the subset $W^-$ is not contained in $B_r(W)$. Thus $W^- \setminus B_r(V) \neq \emptyset$. Contradiction.

Therefore $T$ is a nontrivial $G$–tree and we obtain a nontrivial graph of groups decomposition of $G$ where the edge groups are conjugate to subgroups of the finite group $G_S$. □

### 19.2. An existence theorem for harmonic functions

Theorem 19.6 below was originally proven by Kaimanovich and Woess in Theorem 5 of [**KW92**] using probabilistic methods (they also proved it for functions with values in $[0,1]$). At the same time, an analytical proof of this result was given by Li and Tam [**LT92**], see also [**Li04**, Theorem 4.1] for a detailed and more general treatment.

We owe the following proof to Mohan Ramachandran:

Theorem 19.6. *Let $\chi : \epsilon(M) \to \{0,1\}$ be a continuous function. Then $\chi$ admits a continuous harmonic extension to $M$.*

Proof. Let $\varphi$ denote a smooth extension of $\chi$ to $M$ so that $d\varphi$ is compactly supported.

We let $W_o^{1,2}(M)$ denote the closure of $C_c^\infty(M)$ with respect to the norm

$$\|u\| := \|u\|_{L_2} + \sqrt{E(u)}.$$

Consider the affine subspace of functions

$$\mathcal{G} := \varphi + W_o^{1,2}(M) \subset L^2_{loc}(M).$$

Then the energy is well-defined on $\mathcal{G}$ and we set $E := \inf_{f\in\mathcal{G}} E(f)$.

Note that, since $\mathcal{G}$ is affine, for $u, v \in \mathcal{G}$ we also have

$$\frac{u+v}{2} \in \mathcal{G},$$

in particular,

$$E(\frac{u+v}{2}) \geqslant E$$

and we set

$$E(u,v) := 2E(\frac{u+v}{2}) - \frac{E(u)+E(v)}{2}.$$

The latter equals

$$E(u,v) := \int_M \langle \nabla u, \nabla v \rangle$$

in the case when $u, v$ are smooth. We thus obtain

$$E(u,v) \geqslant 2E - \frac{E(u)+E(v)}{2}$$

for all $u, v \in \mathcal{G}$. Hence,

$$E(u-v) = E(u) + E(v) - 2E(u,v) \leqslant 2E(u) + 2E(v) - 4E. \tag{19.2}$$

Pick a sequence $u_n \in \mathcal{G}$ such that

$$\lim_{n\to\infty} E(u_n) = E.$$

Then, according to (19.2),

$$E(u_m - u_m) \leqslant 2E(u_n) + 2E(u_m) - 4E = 2(E(u_n) - E) + 2(E(u_m) - E).$$

Since $\lambda := \lambda_1(M) > 0$, we obtain

$$\lambda \int_M f^2 \leqslant E(f) \tag{19.3}$$

for all $f \in W_o^{1,2}(M)$. Therefore, the functions $v_n := u_n - \varphi \in W_o^{1,2}(M)$ satisfy

$$\|v_n - v_m\| \leqslant (2 + \lambda^{-1})(E(u_n) - E + E(u_m) - E).$$

Hence, the sequence $(v_n)$ is Cauchy in $W_o^{1,2}(M)$. Set

$$v := \lim_n v_n, u := \varphi + v \in \mathcal{F}.$$

By semicontinuity of energy, $E(u) = E$. Therefore, $u$ is harmonic and, hence, $u$ is smooth (see Section 2.1.7). Since $d\varphi$ is compactly supported, the function $v$ is also harmonic away from a compact subset $K \subset M$. By the inequality (19.3), we have

$$\int_M v^2 \leqslant \lambda^{-1} E(v) < \infty. \tag{19.4}$$

Let $r > 0$ denote the injectivity radius of $M$. Pick a base-point $o \in M$. Then (19.4) implies that there exists a function $\rho : M \to \mathbb{R}_+$ which converges to 0 as $d(x,o) \to \infty$, so that

$$\int_{B_r(x)} v^2(x) \leqslant \rho(x)$$

for all $x \in M$. By the gradient estimate (Theorem 2.1.7), there exists $C_1 < \infty$ so that

$$\sup_{B_r(x)} v^2 \leqslant C_1 \inf_{B_r(x)} v^2$$

provided that $d(x, K) \geqslant r$. Therefore,

$$v^2(x) \leqslant \frac{C_1}{Vol(B_r(x))} \int_{B_r(x)} v^2 \leqslant C_2 \rho(x).$$

Thus

$$\lim_{d(x,o)\to\infty} v(x) = 0.$$

Therefore the harmonic function $u$ extends to the function $\chi$ on $\epsilon(M)$. □

### 19.3. Energy of minimum and maximum of two smooth functions

The arguments here are again due to Mohan Ramachandran.

Let $M$ be a smooth manifold and $f$ be a $C^1$-smooth function on $M$. Define the function $f^+ := max(f, 0)$ and the closed set

$$\Gamma := \{x \in M : f(x) = 0, df(x) = 0\}.$$

Set

$$\Omega := \{x \in M : f(x) = 0, df(x) \neq 0\} = f^{-1}(0) \setminus \Gamma.$$

By the implicit function theorem, $\Omega$ is a smooth submanifold in $M$ and, hence, has measure zero. Clearly, $f^+$ is smooth on $M \setminus \Omega$.

Lemma 19.7. *Under the above conditions, a.e. on $M$ we have: $df^+(x) = df(x)$ if $f(x) > 0$ and $df^+(x) = 0$ if $f(x) \leqslant 0$.*

Proof. Since $\Omega$ has measure zero, it suffices to prove the assertion for points $x_0 \in \Gamma$. Choose local coordinates on $M$ at a point $x_0 \in \Gamma$, so that $x_0 = 0$. Since $f$ has zero derivative at 0, we have:

$$\lim_{v\to 0} \frac{|f(v)|}{\|v\|} = 0.$$

Since $0 \leqslant |f^+| \leqslant |f|$, it follows that

$$\lim_{v\to 0} \frac{|f^+(v)|}{\|v\|} = 0.$$

Therefore, $f^+$ is differentiable at $x_0$ and $df^+(x_0) = 0$. □

Consider now two $C^1$-smooth functions $f_1, f_2$ on $M$. Define

$$f_{max} := \max(f_1, f_2), \quad f_{min} := \min(f_1, f_2), \quad f := f_1 - f_2.$$

Lemma 19.8. $E(f_1) + E(f_2) = E(f_{max}) + E(f_{min})$.

Proof. Set

$$M_1 := \{f_1 > f_2\}, M_2 := \{f_2 > f_1\}, M_0 := \{f_1 = f_2\}.$$

Since

$$f_{max} = f_2 - f^+, \quad f_{min} = f_1 - f^+,$$

by the above lemma we have:

$$\nabla f_{max} = \nabla f_2, \quad \nabla f_{min} = \nabla f_1 \quad \text{a.e. on} \quad M_0.$$

Clearly,

$$\nabla f_{max} = \nabla f_i|_{M_i}, \nabla f_{min} = \nabla f_{i+1}|_{M_{i+1}}, i = 1, 2.$$

Hence,

$$\int_{M_i} |\nabla f_{max}|^2 + |\nabla f_{min}|^2 = \int_{M_i} |\nabla f_1|^2 + |\nabla f_2|^2, i = 0, 1, 2.$$

Therefore,

$$E(f_1) + E(f_2) = E(f_{max}) + E(f_{min}). \quad \square$$

## 19.4. A compactness theorem for harmonic functions

**19.4.1. The main results.** Let $M$ be a bounded geometry Riemannian manifold, and $\bar{M} = M \cup \text{Ends}(M)$ be the end compactification of $M$. Let $\mathbb{F}$ denote the collection of continuous functions $u$ on $\bar{M}$, whose restriction to Ends$(M)$ is nonconstant, and takes values in $\{0,1\}$, while $u$ is differentiable almost everywhere on $M$.

Given a function $f$, we set $Var(f) := \sup(f) - \inf(f)$. Given an $m$-dimensional Riemannian manifold $N$ (possibly with boundary), we let $|N|$ denote the $m$-dimensional volume of $N$.

The main result of the appendix is

Theorem 19.9. *Suppose that $M$ has positive Cheeger constant, $\eta(M) \geq c > 0$. Then there is a $\mu > 0$ such that any $u \in \mathbb{F}$ has energy at least $\mu$.*

Let $U \subset M$ be a smooth codimension 0 submanifold with compact boundary $C$. Recall that the *capacitance* of the pair $(U, C)$ is the infimal energy of compactly supported functions $u : U \to [0,1]$ which are equal to 1 on $C$.

Corollary 19.10. *For each $U$ and $C$ as above, the capacitance is at least $\mu$.*

Proof. Given a function $u : U \to [0,1]$ which equals 1 on $C$, we extend $u$ by 1 to the rest of $M$. Then, clearly, the extension $\tilde{u}$ has the same energy as $u$ and $u \in \mathbb{F}$. Therefore, $E(u) = E(\tilde{u}) \geq \mu$. $\square$

Corollary 19.11. *Assume that every point in $M$ belongs to an $R$-neck. Then for all $0 < a < b < 1$, $E \in [0, \infty)$, there is an $r \in (0, \infty)$ with the following property. If $u : M \to (0,1)$ is a proper map, and $p \in M$, then either*

(1) *$u(B_r(p))$ is not contained in the interval $[a, b]$, or*
(2) *The energy of $u$ is at least $E$.*

Proof. Given $0 < R < \infty$ and $p \in M$, we let $\mathcal{C}$ denote the collection of unbounded components of $M \setminus B_R(p)$. Let $u : M \to (0,1)$ be a proper map so that $u(B_r(p) \subset [a,b]$. For each $U \in \mathcal{C}$, the function $u$ takes the values in $[a,b]$ on $C = \partial U$. Consider the two functions $u^+ = \max\{b, u\}$ and $u^- = \min\{a, u\}$ on $U$. Then

$$E(u^{\pm}) \leq E(u|_U)$$

and $u^+|_C = b, u^-|_C = a$. let $\tilde{u}^{\pm}$ denote the extension of $u^{\pm}$ to the rest of $M$ so that

$$\tilde{u}^{\pm}|_{U^c} \equiv u^{\pm}|_C.$$

Then

$$E(\tilde{u}^{\pm}) = E(u^{\pm}) \leq E(u|_U).$$

Consider the function $\tilde{u}^-$: Its values on Ends$(M)$ belong to $\{0, a\}$. If it does not take zero values on Ends$(M)$ then $u|_{End(U)}$ takes only the value 1. Assuming that this does not happen, we see that $\frac{1}{a}\tilde{u}^-$ belongs to $\mathbb{F}$ and, hence,

$$E(u|U) \geq E(\tilde{u}^-) \geq a^2 \mu$$

by Theorem 19.9. If $u\big|_{End(U)}$ takes only the value 1, then $\tilde{u}^-$ is constant (equal to $a$) on Ends($M$) and we obtain no contradiction. In this case, we use the function $\tilde{u}^+$: It takes the values $b$ and 1 on Ends($M$). We then consider the function

$$\tilde{v} := 1 - \tilde{u}^+$$

and, similarly, obtain

$$E(u\big|_U) \geq E(\tilde{v}) \geq b^2\mu.$$

In either case, we conclude that

$$E(u\big|_U) \geq a^2\mu > 0.$$

Since the number of elements of $\mathcal{C}$ grows exponentially with $R$, the statement follows. □

COROLLARY 19.12. *Suppose $M$ is as above, and $E \in (0,\infty)$. If $u \in H$ has energy at most $E$, and $u$ is nearly constant on a large ball $B$, then it is nearly equal to 0 or 1 on $B$. (I.e., the supremum-norm of $u\big|_B$ or of $(u-1)\big|_B$ converges to zero as $Var(u\big|_B) \to 0$.)*

We can now prove Theorem 19.2. Recall that $H = H(M)$ is the space of functions $f \in \mathbb{F}$ which are harmonic on $M$. Consider a sequence $f_n \in H$ and $x_n \in f_n^{-1}(1/2)$. By applying elements of the isometry group $G$, we can assume that the points $x_n$ belong to a fixed compact $K \subset M$. By passing to a subsequence, we may assume that $\lim x_n = x \in K$. The functions $f_n$ then subconvere uniformly on compacts to a harmonic function $f$ which attains values $1/2$ at $x \in K$. We have to show that $f \in \mathbb{F}$. Suppose first that $f$ is constant on $M$. Then for each $\epsilon > 0$ and $r > 0$ there exists $n$ so that

$$Var(f_n\big|_{B_r(x)}) < \epsilon.$$

By taking $r$ sufficiently large, we conclude that $f_n$ is approximately equal to 0 or 1 on $B_r(x)$, which contradicts the assumption that $f_n(x_n) = 1/2$. Therefore, $f$ cannot be constant.

Suppose now that $f$ takes a value $y \notin \{0,1\}$ at a point $\xi \in$ Ends($M$). Then for every $r$, there exists $z \in M$ sufficiently close to $\xi$ in the topology of $\bar{M}$, so that $f\big|_{B_r(z)}$ is approximately equal to $y$. Therefore, the same will be true for the functions $f_n$ if $n$ is sufficiently large. By Corollary 19.12, it then follows that $y = 0$ or $y = 1$. Contradiction. □

REMARK 19.13. One could remove the cocompactness assumption by saying that any sequence $u_i \in H$ has a pointed limit living in a pointed Gromov-Hausdorff limit of a seqeunce $(M, x_n)$ (which will be another bounded geometry manifold with a linear isoperimetric inequality and ubiquitous $R$-necks).

The proof of Theorem 19.9 occupies the rest of the appendix.

**19.4.2. Some coarea estimates.** Recall that if $u : M \to \mathbb{R}$ is a smooth function on a Riemannian manifold $M$, then for a.e. $t \in \mathbb{R}$, the level set $u^{-1}(\{t\}) = \{u = t\}$ is a smooth hypersurface, and for any measurable function $\phi : M \to \mathbb{R}$ such that $\phi|\nabla u|$ is integrable, we have the *coarea formula*

$$\int_M \phi|\nabla u| = \int_{\mathbb{R}} \left(\int_{\{u=t\}} \phi\right) dt\,, \tag{19.5}$$

where the integration $\int_{\{u=t\}} \phi$ is w.r.t. Riemannian measure on the hypersurface. See e.g. **[Fed69]**.

The two applications of this appearing below are:

$$\int_{\{t_1 \le u \le t_2\}} |\nabla u|^2 = \int_{t_1}^{t_2} \left( \int_{\{u=t\}} |\nabla u| \right) dt\,, \tag{19.6}$$

where we take $\phi = |\nabla u|^2$ on $\{t_1 \le u \le t_2\}$ and zero otherwise, and

$$|\{t_1 \le u \le t_2\}| = \int_{\{t_1 \le u \le t_2\}} 1 = \int_{t_1}^{t_2} \left( \int_{\{u=t\}} \frac{1}{|\nabla u|} \right) dt\,, \tag{19.7}$$

where we take $\phi = \frac{1}{|\nabla u|}$ under the assumption that $\nabla u \neq 0$ a.e. on $M$.

We first combine these in the following general inequality:

Lemma 19.14. *If $u : M \to [t_1, t_2]$ is a smooth proper function, such that $\nabla u \neq 0$ on a full measure subset, and so that $A(t) = |\{u = t\}| \ge A_0 > 0$ for all $t$. Then*

$$\int_{\{t_1 \le u \le t_2\}} |\nabla u|^2 \ge \frac{A_0^2 (t_2 - t_1)^2}{V}\,, \tag{19.8}$$

*where $V$ is the volume of the "slab" $\{t_1 \le u \le t_2\}$.*

Proof. The argument combines (19.6), (19.7), and Jensen's inequality.

We have

$$\int_{\{u=t\}} |\nabla u| = A(t) ⨍_{\{u=t\}} |\nabla u|$$

$$\ge A(t) \frac{1}{⨍_{\{u=t\}} \frac{1}{|\nabla u|}} \quad \text{by Jensen's inequality}$$

$$= \frac{A^2(t)}{\int_{\{u=t\}} \frac{1}{|\nabla u|}}\,, \tag{19.9}$$

with the equality in the case when $|\nabla u|$ is constant a.e. on $M$.

Now

$$\int_{\{t_1 \le u \le t_2\}} |\nabla u|^2 = \int_{t_1}^{t_2} \left( \int_{\{u=t\}} |\nabla u| \right) dt \qquad \text{by (19.6)}$$

$$\ge \int_{t_1}^{t_2} \frac{A(t)}{\left( ⨍_{\{u=t\}} \frac{1}{|\nabla u|} \right)} dt \qquad \text{by (19.9)}$$

$$\ge A_0 (t_2 - t_1) ⨍_{t_1}^{t_2} \frac{dt}{\left( ⨍_{\{u=t\}} \frac{1}{|\nabla u|} \right)}$$

$$\ge A_0 (t_2 - t_1) \frac{1}{⨍_{t_1}^{t_2} \left( ⨍_{\{u=t\}} \frac{1}{|\nabla u|} \right) dt} \qquad \text{by Jensen's inequality}$$

$$\ge \frac{A_0^2 (t_2 - t_1)^2}{\int_{t_1}^{t_2} \left( \int_{\{u=t\}} \frac{1}{|\nabla u|} \right) dt} = \frac{A_0^2 (t_2 - t_1)^2}{|\{t_1 \le u \le t_2\}|} \tag{19.10}$$

by (19.7). □ □

**19.4.3. Energy comparison in the case of a linear isoperimetric inequality.** Now suppose $M$ is a manifold with compact boundary $\partial M$, which satisfies a linear isoperimetric inequality

$$|\partial D| \geq c|D|\,,$$

where $D \subset M$ is any compact domain with smooth boundary.

Suppose $u : M \to [0,1]$ is a compactly supported smooth function, such that $u^{-1}(\{1\}) = \partial M$. Let $\hat{M}$ be a scaled copy of $\mathbb{H}^2/\mathbb{Z}$, where $\mathbb{Z}$ is a parabolic isometry, and let $\hat{u} : \hat{M} \to [0,1]$ be the radial function (i.e., $\hat{u}$ is constant along projections of $\mathbb{Z}$-invariant horocycles), such that the superlevel sets of $\hat{u}$ have the same volume as the corresponding superlevel sets of $u$:

$$|\{\hat{u} \geq t\}| = |\{u \geq t\}| \quad \text{for all } t \in (0,1).$$

For $t \in [0,1]$, let $A(t) = |\{u = t\}|$ and $V(t) = |\{t \leq u \leq 1\}|$, and let $\hat{A}$ and $\hat{V}$ denote the corresponding quantities for $\hat{u}$. We recall that if horocircles in $\hat{M}$ have geodesic curvature $k$, then $\hat{M}$ satisfies a $k$-linear isoperimetric inequality, and $|\partial\hat{D}| = k|\hat{D}|$ for every horoball $\hat{D} \subset \hat{M}$.

We compare the energy of $u$ with the energy of $\hat{u}$.

LEMMA 19.15. *Suppose that for some $T \in (0,1]$, we have $V(T) \geq \frac{2}{c}\,A(1) = \frac{2}{c}\,|\partial M|$, and $\hat{A}(t) \leq \frac{c}{2}\hat{V}(t)$ for all $t \in (0,T]$. Then*

$$\int_{\{0 \leq u \leq T\}} |\nabla u|^2 \geq \int_{\{0 \leq \hat{u} \leq T\}} |\nabla \hat{u}|^2\,.$$

PROOF. Since $V(t) = \hat{V}(t)$, differentiating the version

$$V(t) = \int_t^1 \int_{u=\tau} \frac{1}{|\nabla u|} d\tau$$

of (19.7) with respect to $t$, we get that

$$\int_{\{u=t\}} \frac{1}{|\nabla u|} \;=\; \int_{\{\hat{u}=t\}} \frac{1}{|\nabla \hat{u}|} \tag{19.11}$$

for a.e. $t \in [0,1]$.

For all $t \leq T$, we have

$$|\partial\{u \geq t\}| = |\partial M| + A(t) \geq cV(t)$$

so

$$A(t) \geq cV(t) - |\partial M| \geq cV(t) - \frac{c}{2}V(t) \geq \frac{c}{2}V(t) = \frac{c}{2}\hat{V}(t) \geq \hat{A}(t)\,. \tag{19.12}$$

Now

$$\begin{aligned}
\int_{\{u=t\}} |\nabla u| &\geq \frac{A^2(t)}{\int_{\{u=t\}} \frac{1}{|\nabla u|}} && \text{see (19.9)}\\
&\geq \frac{\hat{A}^2(t)}{\int_{\{\hat{u}=t\}} \frac{1}{|\nabla \hat{u}|}} && \text{by (19.11) and (19.12)}\\
&= \int_{\{\hat{u}=t\}} |\nabla \hat{u}| && (19.13)
\end{aligned}$$

because $|\nabla\hat{u}|$ is constant on $\{u = t\}$, so the equality case of (19.9) applies.

The lemma now follows from (19.6) and (19.13). □

□

**19.4.4. The proof of Theorem 19.9.** Suppose $v \in \mathbb{F}$. Every level set of $v$ defines a nontrivial homology class in $M$, hence, by the Federer-Fleming deformation lemma [**Fed69**],

$$\inf_{\tau\in(0,1)} |\{v=\tau\}| \geq A_0 > 0,$$

where $A_0$ depends only on the bounds on the geometry of $M$ (i.e., a lower bound on the injectivity radius and curvature bounds). Choose a regular value $t_1 \in (0,1)$ of the function $v$ where $A(t)$ almost attains its infimum, i.e.,

$$A(t_1) \geq \inf_{\tau\in(0,1)} |\{v=\tau\}| \geq A(t_1)/2.$$

We may assume that $t_1 \geq \frac{1}{2}$ and we focus attention on the codimension 0 submanifold $N \subset M$ given by the sublevel set $\{v \leq t_1\}$. Replacing $v$ with $\frac{1}{t_1}v$, we get a function $u : N \to [0,1]$ which is 1 on $\partial N$, tends to zero at infinity, and all the level sets $\{u = t\}$ have area at least $\frac{1}{2}\,\mathrm{Area}(\partial N)$. By an approximation argument, we may assume that $u$ is compactly supported, and that $\nabla u \neq 0$ on a full measure subset of $\{u > 0\}$.

By continuity, there exists a superlevel set $\{u \geq t_0\} \subset N$ whose volume is $\frac{2}{c}|\partial N|$, where $c$ is the constant in the linear isoperimetric inequality for $M$.

Applying Lemma 19.14, we get

$$\int_{\{t_0\leq u\leq 1\}} |\nabla u|^2 \geq \frac{(1-t_0)^2 A_0^2}{\frac{2}{c}|\partial N|},$$

$$\geq \frac{\left((1-t_0)\frac{|\partial N|}{2}\right)^2}{\frac{2}{c}|\partial N|} = \frac{(1-t_0)^2}{8}c|\partial N| \geq \frac{(1-t_0)^2}{8}cA_0\,.$$

Therefore we get a lower bound on the energy of $u$ if $t_0 \leq \frac{1}{2}$. So we may assume that $t_0 \geq \frac{1}{2}$, in which case we may apply Lemma 19.15 to see that the energy of $u$ is at least a big as that of the comparison function $\hat{u}$ on a hyperbolic cusp, which is scaled to have the $\frac{c}{2}$-linear isoperimetric inequality. Since the $t_0$-superlevel set $\{\hat{u} \geq t_0\} \subset \hat{M}$ has volume $\frac{2}{c}|\partial N| \geq \frac{2}{c}A_0$, the fact that $\lambda_1(\hat{M}) > 0$ gives a lower bound on the energy of $\hat{u}$ which depends only on $c$ and $A_0$.

We give a self-contained proof of the lower energy bound for $\hat{u}$. We identify $\hat{M}$ with the punctured unit disk $D^* = \{z \in \mathbb{C} : 0 < |z| < 1\}$. Let $r : \hat{M} \to \mathbb{R}$ be the horospherical coordinate normalized so that the level set $\{r = h\}$ has length $e^{kh}$, for $k = \frac{c}{2}$; in particular, $r(z)$ converges to $-\infty$ as $z \to 0$ and converges to $\infty$ as $|z| \to 1$. (The coordinate $r$ is constant on the circles $|z| = R$ in $D^*$, which are the horocycles in $\hat{M}$ with the center $z = 0$.) We will regard $\hat{u}$ as a function of $r$ so that

$$\lim_{r\to-\infty} \hat{u}(r) = 1, \quad \lim_{r\to\infty} \hat{u}(r) = 0.$$

Then, for $s \in (0, t_0)$,

$$|t_0 - s| \leq \int_{\hat{u}(r)=s}^{\hat{u}(r)=t_0} \left|\frac{\partial \hat{u}}{\partial r}\right| dr = \int_{\hat{u}(r)=s}^{\hat{u}(r)=t_0} \left(\left|\frac{\partial \hat{u}}{\partial r}\right| e^{\frac{k}{2}r}\right) e^{-\frac{k}{2}r} dr$$

$$\leq \left(\int_{\hat{u}(r)=s}^{\hat{u}(r)=t_0} \left|\frac{\partial \hat{u}}{\partial r}\right|^2 e^{kr}\, dr\right)^{\frac{1}{2}} \left(\int_{\hat{u}(r)=s}^{\hat{u}(r)=t_0} e^{-kr}\, dr\right)^{\frac{1}{2}}. \tag{19.14}$$

Since the volume element $d\hat{V}$ in $\hat{M}$ is $e^{kr}dr\,d\theta$, we get

$$\int_{\{s\le\hat{u}\le t_0\}}\left|\frac{\partial\hat{u}}{\partial r}\right|^2 d\hat{V} \;=\; \int_{\hat{u}(r)=s}^{\hat{u}(r)=t_0}\left|\frac{\partial\hat{u}}{\partial r}\right|^2 e^{kr}\,dr \ge \frac{(t_0-s)^2}{\int_{\hat{u}(r)=s}^{\hat{u}(r)=t_0} e^{-kr}\,dr} \qquad \text{by (19.14)}$$

$$\ge \frac{(t_0-s)^2}{\frac{1}{k}e^{-kr}\big|_{\hat{u}(r)=t_0}} \ge k^2(t_0-s)^2\hat{V}(r)\big|_{\hat{u}(r)=t_0} \ge k^2(t_0-s)^2\left(\frac{2}{c}A_0\right) = \frac{c}{2}(t_0-s)^2\,A_0.$$

Here $\hat{V}(r)$ denotes the volume of the punctured disk $\{z\in\hat{M} : r(z)\le r\}$. Since $t_0\ge\frac{1}{2}$, letting $s\to 0$ we obtain

$$E(\hat{u}\big|_{\hat{u}\le t_0}) \ge \frac{c}{2}t_0^2\,A_0 \ge \frac{c}{8}A_0.$$

By Lemma 19.15,

$$E(u\big|_N) \ge E(\hat{u}\big|_{\hat{u}\le t_0}) \ge \frac{c}{8}A_0.$$

This completes the proof of Theorem 19.9. □

CHAPTER 20

# Quasiconformal mappings

We refer the reader to the books [**Res89**], [**Vuo88**], [**Väi71**], [**IM01**] for the detailed discussion of quasiconformal maps.

### 20.1. Linear algebra and eccentricity of ellipsoids

Suppose that $M \in GL(n, \mathbb{R})$ is an invertible linear transformation of $\mathbb{R}^n$. We would like to measure deviation of $M$ from being a conformal linear transformation, i.e., from being an element of $\mathbb{R}_+ \cdot O(n)$. Geometrically speaking we are interested in measuring deviation of the ellipsoid $E = M(B) \subset \mathbb{R}^n$ from a round ball, where $B$ is the unit ball in $\mathbb{R}^n$.

In case $n = 2$, there is essentially only one way for such measurement, namely, eccentricity of the ellipsoid: The ratio of major to minor axes of $E$. In higher dimensions, there are several invariants which are useful in different situations. This reflects the simple fact that the matrix $A$ has $n$ singular values, while an invariant we are looking for is a single real number.

Recall that every invertible $n \times n$ matrix $M$ has *singular value decomposition* (see Theorem 1.70)

$$M = UDV = UDiag(\lambda_1, ..., \lambda_n)V,$$

where the (positive) diagonal entries $\lambda_1 \leqslant ... \leqslant \lambda_n$ of the diagonal matrix $D = Diag(\lambda_1, ..., \lambda_n)$ are the *singular values* of $A$. Here $U, V$ are orthogonal matrices. Equivalently, if we symmetrize $M$: $A = MM^T$, then the numbers $\lambda_i$ are square roots of the eigenvalues of $A$. Geometrically speaking, the singular values $\lambda_i$ are the half-lengths of the axes of the ellipsoid $E = M(B)$.

We define the following *distortion quantities* for the matrix $M$:

- Linear dilatation:

$$H(M) := \frac{\lambda_n}{\lambda_1} = \|M\| \cdot \|M^{-1}\|,$$

  where $\|M\|$ is the operator norm of a matrix $M$:

$$\max_{v \in \mathbb{R}^n \setminus 0} \frac{|Mv|}{|v|}.$$

  Thus, $H(M) = \epsilon(E)$ is the *eccentricity* of the ellipsoid $E$, the ratio of lengths of major and minor axes of $E$. This is the invariant that we will be using most of the time.
- Inner dilatation:

$$H_I(M) := \frac{\lambda_1 .... \lambda_n}{\lambda_1^n} = |\det(M)| \cdot \|M^{-1}\|^n.$$

- Outer dilatation:

$$H_O(M) := \frac{\lambda_n^n}{\lambda_1 .... \lambda_n} = \|M\|^n |\det(M)|^{-1}.$$

- Maximal dilatation:

$$K(M) := \max(H_I(M), H_O(M)).$$

Thus, geometrically speaking, inner and outer dilatations compute volume ratios of $E$ and inscribed/circumscribed balls, while linear dilatation compares the radii of inscribed/circumscribed balls. Note that all four dilatations agree if $n = 2$.

EXERCISE 20.1. $M$ is conformal $\iff H(M) = 1 \iff H_I(M) = 1 \iff H_O(M) = 1 \iff K(M) = 1$.

EXERCISE 20.2. Logarithms of linear and maximal dilatations are comparable:

$$(H(M))^{n/2} \leqslant K(M) \leqslant (H(M))^{n-1}$$

Hint: It suffices to consider the case when $M = Diag(\lambda_1, ...\lambda_n)$ is a diagonal matrix.

EXERCISE 20.3. 1. $H(M) = H(M^{-1})$ and $H(M_1 \cdot M_2) \leqslant H(M_1) \cdot H(M_2)$.
2. $K(M) = K(M^{-1})$ and $K(M_1 \cdot M_2) \leqslant K(M_1) \cdot K(M_2)$.
Hint: Use geometric interpretation of the four dilatations.

## 20.2. Quasi-symmetric maps

Our next goal is to generalize the dilatation constants for linear maps in the context of non-linear maps. The linear dilatation is easiest to generalize since it deals only with distances. Recall the geometric meaning of the linear dilatation $H(M)$: If $B$ is a round ball, its image is an ellipsoid $E$ and $H(M)$ is the ratio of the "outer radius" of $E$ by its "inner radius." Such ratio makes sense not only for ellipsoids but also for arbitrary topological balls $D \subset \mathbb{R}^n$ where we have a chosen "center" , a point $x'$ in the interior of $D$: Then we have two real numbers $r$ and $R$, so that $B(x', r) \subset D$ is the largest metric ball (centered at $x'$) contained in $D$ and $D \subset B(x', R)$ is the smallest metric ball containing $D$. Then $r$ and $R$ are the inner and outer radii of $D$. In other words,

$$\frac{R}{r} = \max \frac{|y' - x'|}{|z' - x'|}$$

where maximum is taken over all $y', z' \in \partial D$. This ratio is "eccentricity" of the topological ball $D \subset \mathbb{R}^n$. The idea then is to consider homeomorphisms $f$ which send round balls $B(x, \rho)$ to topological balls of uniformly bounded eccentricity with respect to the "center" $x' = f(x)$.

This leads to

DEFINITION 20.4. A homeomorphism $f : U \to U'$ between two domains in $\mathbb{R}^n$ is *c-weakly quasi-symmetric* if

$$\frac{|f(x) - f(y)|}{|f(x) - f(z)|} \leqslant c \tag{20.1}$$

for all $x, y, z \in U$ so that $|x - y| = |y - z| > 0$. Note that we do not assume that $f$ preserves orientation. We will be mostly interested in the case $U = U' = \mathbb{R}^n$.

The name *quasi-symmetric* comes from the case $n = 1$ (and quasi-symmetric maps were originally introduced only for $n = 1$ by Ahlfors and Beurling [**AB56**]). Namely, a homeomorphism $f : \mathbb{R} \to \mathbb{R}, f(0) = 0$ is *symmetric* at the origin if it sends any pair of points symmetric about 0 to points symmetric about 0, i.e., these are *odd* functions: $f(-y) = -f(y)$. In the case of $c$-weakly quasi–symmetric maps, the exact symmetry is lost, but is replaced by a uniform bound on the ratio of absolute values.

Exercise 20.5. Show that 1-weakly quasi-symmetric homeomorphisms $f : \mathbb{R} \to \mathbb{R}$ are compositions of dilations and isometries of $\mathbb{R}$.

It turns out that there is a slightly stronger condition, which is a bit easier to work with and which generalizes naturally to metric spaces other than $\mathbb{R}^n$:

Definition 20.6. Let $\eta : [1, \infty) \to [1, \infty)$ be a continuous surjective monotonic function. A homeomorphism $f : U \subset \mathbb{R}^n \to U' \subset \mathbb{R}^n$ is called $\eta$-quasi-symmetric if for all $x, y, z \in \mathbb{R}^n$ we have

$$\frac{|f(x) - f(y)|}{|f(x) - f(z)|} \leqslant \eta \left( \frac{|x - y|}{|x - z|} \right) \tag{20.2}$$

Thus, if we take $c = \eta(1)$, then every $\eta$-quasi-symmetric map is also $c$-weakly quasi-symmetric. It is a nontrivial theorem (see e.g. [**Hei01**]) that for $U = U' = \mathbb{R}^n$, the two concepts are equivalent.

Exercise 20.7. Show that:

1. Every invertible affine transformation $L : \mathbb{R}^n \to \mathbb{R}^n$ is $\eta$-quasi-symmetric for $\eta(t) = H(L)$.
2. $L$-Lipschitz homeomorphisms are $\eta$-quasi-symmetric with $\eta(t) = L^2$.

As in the case of quasi-isometries, we will say that a homeomorphism is (weakly) quasi-symmetric if it is $\eta$-quasi-symmetric (respectively $c$-weakly quasi-symmetric) for some $\eta$ or $c < \infty$.

The following exercise requires nothing by definition of quasi-symmetry:

Exercise 20.8. Show that composition of quasi-symmetric maps is again quasi-symmetric. Show that the inverse of a quasi-symmetric map is also quasi-symmetric.

Recall that we think of $S^n$ as the 1-point compactification of $\mathbb{R}^n$. The drawback of the definition of quasi-symmetric maps is that we are restricted to the maps of $\mathbb{R}^n$ rather than $S^n$. In particular, we cannot apply this definition to Moebius transformations.

Definition 20.9. A homeomorphism of $S^n$ is called *quasi-moebius* if it is a composition of a Moebius transformation with a quasi-symmetric map.

Note that Moebius transformations of $S^n$ can be characterized by the property that they preserve the cross-ratios

$$[x, y, z, w] := \frac{|x - y| \cdot |z - w|}{|y - z| \cdot |w - x|}, x, y, z \in S^n.$$

Similarly, a homeomorphism $f$ of $S^n$ is quasi-Moebius if and only if there exists a homeomorphism $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$[f(x), f(y), f(z), f(w)] \leqslant \kappa([x, y, z, w])$$

for all $x, y, z, w$. See [**Väi85**].

## 20.3. Quasiconformal maps

The idea of quasiconformality is very natural: We take the definition of weakly quasiconformal maps *via* the ratio (20.1) and then take the limit in this ratio as $\rho = |x-y| = |y-z| \to 0$.

For a homeomorphism $f : U \subset \mathbb{R}^n \to U' \subset \mathbb{R}^n$ between two domains in $\mathbb{R}^n$ and $x \in U$ we define the quantity

$$\forall x, \quad H_x(f) := \limsup_{\rho \to 0} \left( \sup_{y,z} \frac{|f(x)-f(y)|}{|f(x)-f(z)|} \right), \tag{20.3}$$

where, for each $\rho > 0$, the supremum is taken over $y, z$ so that $\rho = |x-y| = |x-z|$. For instance, if $f$ is $c$-weakly quasi-symmetric, then $H_x(f) \leqslant c$ for every $x \in U$.

Definition 20.10. A homeomorphism $f : U \to U'$ is called *quasiconformal* if $\sup_{x \in U} H_x(f)$ finite.

The function $H_x(f)$ is called *(linear) dilatation function* of $f$; a quasiconformal map $f$ is said to have *dilatation* $\leqslant H$ if the essential supremum of $H_x(f)$ in $U$ is $\leqslant H$. Note that the essential supremum is the $L^\infty$-norm, so it ignores subsets of measure zero. We will see the reason for this discrepancy between definition of quasiconformality (where $H_x$ is required to be uniformly bounded) and the definition of dilatation of $f$, in the next section.

Thus, the intuitive meaning of quasiconformality is that *quasiconformal maps send infinitesimal spheres to infinitesimal ellipsoids of uniformly bounded eccentricity.*

Exercise 20.11. Let $f : S^n \to S^n$ be a Moebius transformation, $p = f^{-1}(\infty)$. Then $f|\mathbb{R}^n \setminus \{p\}$ is 1-quasiconformal, i.e., conformal. Clearly, it suffices to verify conformality only for the inversion in the unit sphere.

Note that here and in what follows *we do not assume that conformal maps preserve orientation.* For instance, in this terminology, complex conjugation is a conformal map $\mathbb{C} \to \mathbb{C}$.

Exercise 20.12. 1. Suppose that $f : U \to U'$ is diffeomorphism so that $\|D_x(f)\|$ is uniformly bounded above and $|J_x(f)|$ is uniformly bounded below. Show, using definition of differentiability, that $f$ is quasiconformal. Namely, verify that $H_x(f) = H(D_x(f))$ for every $x \in U$.

2. Show that every diffeomorphism $S^n \to S^n$ is quasiconformal.

## 20.4. Analytical properties of quasiconformal mappings

We begin with some preliminary material from real analysis. For a subset $E \subset \mathbb{R}^n$ we let $mes(E)$ denote the $n$-dimensional Lebesgue measure of $E$. In what follows, $\Omega$ is an open subset in $\mathbb{R}^n$.

1. **Derivatives of measures.** Let $\mu$ be a measure on $\Omega$ of the Lebesgue class, i.e., $\mu$-measurable sets are in the Borel $\sigma$-algebra. The *derivative* of $\mu$, denoted $\mu'(x)$ at $x \in \Omega$ is defined as

$$\mu'(x) := \limsup \frac{\mu(B)}{mes(B)}$$

where the limit is taken over all balls $B$ containing $x$ whose radii tend to zero. The key fact that we will need is the following theorem (see e.g. [**Fol99**, Theorem 3.22]):

THEOREM 20.13 (Lebesgue–Radon–Nikodym differentiation theorem). *The function $\mu'(x)$ is Lebesgue–measurable and is finite a.e. in $\Omega$. (Furthermore, $\mu'(x)$ is the Radon–Nikodym derivative of the component of $\mu$ which is absolutely continuous with respect to the Lebesgue measure.)*

For a continuous map $f : \Omega \to \mathbb{R}^m$ we define the pull-back measure $\mu = \mu_f$ by

$$\mu(E) := mes(f(E)).$$

2. **Rademacher–Stepanov Theorem.** Let $f : \Omega \to \mathbb{R}^m$. Recall that the map $f$ is called *differentiable* at $x \in \Omega$ with derivative at $x$ equal $D_x f = A \in Lin(\mathbb{R}^n, \mathbb{R}^m)$ if

$$\lim_{h \to 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0.$$

It follows directly from the definition that, for $n = m$, at every point $x$ of differentiability of $f$, the measure derivative of $\mu_f$ equals the absolute value of the Jacobian of $f$:

$$\mu'_f(x) = |\det(A)| = |J_f(x)|.$$

The other key result that we will use is:

THEOREM 20.14 (Rademacher and Stepanov, see e.g. Theorem 3.4 in [**Hei05**]). *Let $f : \Omega \to \mathbb{R}^m$. For every $x \in \Omega$ define*

$$|D_x^+(f)| := \limsup_{h \to 0} \frac{|f(x+h) - f(x)|}{|h|}.$$

*Let $E := \{x \in \Omega : |D_x^+(f)| < \infty\}$. Then $f$ is differentiable a.e. in $E$.*

A special case of this theorem is Rademacher's theorem 1.40, since for $L$-Lipschitz maps

$$|D_x^+(f)| \leqslant L.$$

We now return to quasiconformal maps. Recall that dilatation $H_x(f)$ of a homeomorphism $f$ at a point $x$ is defined as

$$H_x(f) := \limsup_{\rho \to 0} \frac{R(\rho)}{r(\rho)}$$

where

$$R(\rho) = \max\{|f(x+h) - f(x)| : |h| = \rho\}, \quad r(\rho) = \min\{|f(x+h) - f(x)| : |h| = \rho\}.$$

3. **Differentiability a.e. of quasiconformal homeomorphisms.**

THEOREM 20.15 (F. Gehring, see [**Väi71**]). *Every quasiconformal map is differentiable a.e. in $\Omega$ and*

$$\|D_x f\| \leqslant H_x(f) |J_x(f)|^{1/n}$$

*for a.e. $x$ in $\Omega$.*

Proof. By the definition of $|D_x^+(f)|$ and $H_x(f)$:

$$|D_x^+(f)| = \limsup_{\rho\to 0} \frac{R_\rho}{\rho} = H_x(f) \limsup_{\rho\to 0} \frac{r_\rho}{\rho}$$

Notice that for $r = r_\rho$, $B(f(x), r) \subset f(B(x,\rho)$, which implies that

$$\omega_n r^n = mes\,(B(f(x), r)) \leqslant mes\,(f(B(x,\rho))$$

where $\omega_n$ is the volume of the unit $n$-ball. Therefore,

$$\frac{mes(f(B(x,\rho)))}{mes(B(x,\rho))} \geqslant \frac{r^n}{\rho^n}$$

and, thus,

$$\mu'_f(x) = \limsup_{\rho\to 0} \frac{mes(f(B(x,\rho)))}{mes(B(x,\rho))} \geqslant \limsup_{\rho\to 0} \frac{r^n}{\rho^n} = \left(\frac{1}{H_x(f)}|D_x^+(f)|\right)^n.$$

It follows that

$$|D_x^+(f)| \leqslant H_x(f)(\mu'_f(x))^{1/n}.$$

The right-hand side of this inequality is finite for a.e. $x$ (by Borel's theorem, Theorem 20.13). Thus, $f$ is differentiable at a.e. $x$ by Rademacher-Stepanov theorem. We also obtain (for a.e. $x \in \Omega$)

$$|D_x(f)| = |D_x^+(f)| \leqslant H_x(f)(\mu'_f(x))^{1/n} = H_x(f)|J_x(f)|^{1/n} \quad \square$$

This differentiability theorem is strengthened as follows:

Theorem 20.16 (F. Gehring, J. Vaisala, see **[Väi71]**). *For $n \geqslant 2$, quasiconformal maps $f : U \to \mathbb{R}^n$ belong to the Sobolev class $W^{1,n}_{loc}$, i.e., their 1st partial distributional derivatives are locally in $L^n(U)$. This, in particular, implies that quasiconformal maps are absolutely continuous on almost every coordinate line segment (this property is called ACL).*

This theorem has an important corollary

Corollary 20.17 (F. Gehring, J. Vaisala, see **[Väi71]**). *For $n \geqslant 2$, every quasiconformal mapping $f : U \to \mathbb{R}^n$, has a.e. nonvanishing Jacobian: $J_x(f) \neq 0$ a.e. in $U$.*

Proof. We will prove a weaker property that will suffice for our purposes, i.e. that $J_x(f) \neq 0$ on a subset of a positive measure. Suppose to the contrary that $J_x(f) = 0$ a.e. in $U$. The inequality

$$|D_x f| \leqslant H_x(f)|J_x(f)|^{1/n}$$

then implies that $D_x f = 0$ a.e. in $U$, i.e., all partial derivatives vanish a.e. in $\Omega$. Let $J = [p, q = p + Te_1]$ be a nondegenerate coordinate line segment (parallel to the $x_1$-axis), connecting $p$ to $q$, on which $f$ is absolutely continuous. This means that the Fundamental Theorem of Calculus applies to $f|J$:

$$f(q) - f(p) = \int_J \frac{\partial}{\partial x_1} f(x) dx_1 = \int_0^T \frac{d}{dt} f(p_1 + te_1, p_2, ..., p_n) dt = 0$$

Hence, $f(p) = f(q)$ which contradicts injectivity of $f$. $\square$

4. Analytical definition of quasiconformality. Since quasiconformal maps are differentiable a.e., it is natural to ask if quasiconformality of a map could be defined analytically, in terms of its derivatives. Below are two equivalent analytical definitions of quasiconformality.

DEFINITION 20.18. Suppose that $U \subset \mathbb{R}^n$ is a domain, $f : U \to U' \subset \mathbb{R}^n$ is a homeomorphism. Then $f$ is quasiconformal provided that $D_x(f)$ is in $W^{1,n}_{loc}$ and

$$K(f) := ess \sup_{x \in U} K(D_x(f)) < \infty. \tag{20.4}$$

Equivalently, the homeomorphism $f$ is ALC and satisfies (20.4). Here $K(A)$ is the maximal dilatation of a linear transformation $f$. A homeomorphism $f$ is called $K$-quasiconformal if $K(f) \leqslant K$. Furthermore,

$$H_x(f) = H(D_x f)$$

a.e. in $U$.

We refer the reader to **[Väi71]** for the proof of equivalence of this definition of quasiconformality to the one in Definition 20.10.

REMARK 20.19. 1. The reason for defining $K$-quasiconformality in terms of maximal dilatation is that it is equivalent to yet another, more geometric, definition, in terms of extremal length (modulus) of families of curves; the latter definition, for historic reasons, is the main definition of quasiconformality, see **[Väi71]**.

2. We can now explain the discrepancy in the definition of maps with bounded dilatation: The condition that $H_x(f)$ is bounded is need in order to ensure that $f$ belongs to $W^{1,n}_{loc}$; one the other hand, the actual bound on dilatation is computed only almost everywhere in $U$. This makes sense since derivatives of $f$ exist only a.e..

In view of Exercise 20.2, the two key measures of quasiconformality, $H(f)$ and $K(f)$ are log-comparable, so using one or the other is only the matter of convenience. What's most important, is that $K(f) = 1$ if and only if $H(f) = 1$. If $n = 2$, then, of course, $K_x(f) = H_x(f)$ and $K(f) = H(f)$.

5. **Liouville theorem.**

THEOREM 20.20 (F. Gehring, see **[Väi71]**). *Every 1-quasiconformal homeomorphism of an open connected domain in $\mathbb{S}^n$ $(n \geqslant 3)$ is the restriction of a Moebius transformation.*

This theorem is a generalization of the classical Liouville's theorem which states that smooth conformal maps between domains in $S^n, n \geqslant 3$, are restrictions of Moebius transformations.

Liouville's theorem fails, of course, in dimension 2. We will see, however, in Section 21.5.1 that an orientation-preserving quasiconformal homeomorphism $f : U \to U'$ of two domains in $S^2 = \mathbb{C} \cup \{\infty\}$, is 1-quasiconformal if and only if it is conformal. Composing with the complex conjugation, we see that every 1-quasiconformal map is either holomorphic or antiholomorphic. In particular:

THEOREM 20.21. *$f : S^2 \to S^2$ is 1-quasiconformal if and only if $f$ is a Moebius transformation.*

6. Quasiconformal and quasi-symmetric maps. So far, we had the implications

$$\text{quasi-symmetric} \Rightarrow \text{weakly quasi-symmetric} \Rightarrow \text{quasiconformal}$$

for maps between domains in $\mathbb{R}^n$. It turns out that these arrows can be reversed:

THEOREM 20.22 (See [**Hei01**].). *Every quasiconformal homeomorphism of* $\mathbb{R}^n$ *is quasiconformal if and only if it is quasi-symmetric.*

7. **Convergence property.** Let $x, y, z \in S^n$ be three distinct points. A sequence of quasiconformal maps $f_i : S^n \to S^n$ is said to be *normalized at* $\{x, y, z\}$ if the limits $\lim_i f_i(x), \lim_i f_i(y), \lim f_i(z)$ exist and are all distinct. Since Moebius transformations act transitively on triples of distinct points, the above condition could be replaced by the requirement that $f_i$'s fix the points $x, y, z$.

THEOREM 20.23 (See [**Väi71**]). *Let* $U \subset S^n$ *be a (connected) domain and* $f_i : U \to f_i(U) \subset S^n$ *be a sequence of* $K$*-quasiconformal homeomorphisms normalized at three points in* $U$*. Then* $(f_i)$ *contains a subsequence which converges to a quasiconformal map.*

The same theorem holds for $n = 1$, except one replaces *quasiconformal* with *quasi-moebius.*

**Historical remark.** Quasiconformal mappings for $n = 2$ were introduced in 1920-s by Groetch as a generalization of conformal mappings. Quasiconformal mappings in higher dimensions were introduced by Lavrentiev in 1930-s for the purposes of application to hydrodynamics. The discovery of relation between quasi-isometries of hyperbolic spaces and quasiconformal mappings was made by Efremovich and Tihomirova [**ET64**] and Mostow [**Mos73**] in 1960-s.

## 20.5. Quasiconformal maps and hyperbolic geometry

Below we will be using the upper half-space model of the hyperbolic space $\mathbb{H}^{n+1}, n \geqslant 1$. Let $f : \mathbb{H}^{n+1} \to \mathbb{H}^{n+1}$ be an $(L, A)$-quasi-isometry and let $f_\infty : \mathbb{S}^n = \mathbb{R}^n \cup \{\infty\} \to \mathbb{S}^n$ be the homeomorphic extension of $f$ to the boundary sphere of the hyperbolic space given by Theorem 9.83. To simplify the notation, we retain the notation $f$ for $f_\infty$. After compositing $f$ with an isometry of $\mathbb{H}^{n+1}$, we can assume that $f(\infty) = \infty$.

THEOREM 20.24. *There exists* $C = C(L, A)$*, so that for the function*

$$\eta(t) = e^{2C+A} t^L,$$

*the mapping* $f : \mathbb{S}^n \to \mathbb{S}^n$ *is* $\eta$*-quasi-symmetric.*

PROOF. Consider an annulus $\mathbb{A} \subset \mathbb{R}^n$ given by

$$\mathbb{A} = \{x : R_1 \leqslant |x| \leqslant R_2\}$$

where $0 < R_1 \leqslant R_2 < \infty$. We will refer to the ratio $t = \frac{R_2}{R_1}$ as the *eccentricity* of $\mathbb{A}$. Of course, for points $y, z$ which belong to the outer and inner boundaries of $\mathbb{A}$, the ratio

$$\frac{|y - x|}{|z - x|}$$

is just the eccentricity of $\mathbb{A}$. Consider the smallest annulus $\mathbb{A}'$ centered at $x' = f(x)$ which contains $f(\mathbb{A})$. Let $t'$ be the eccentricity of $\mathbb{A}'$. Then

$$\frac{|f(y) - f(x)|}{|f(z) - f(x)|} \leqslant \epsilon'.$$

Our goal is, then, to show that $t' \leqslant \eta(t)$ for the function $\eta$ as above.

After composing $f$ with translations of $\mathbb{R}^n$, we can assume that $x = x' = f(x) = 0$. Let $\alpha \subset \mathbb{H}^{n+1}$ denote vertical geodesic, connecting 0 to $\infty$, i.e., $\alpha$ is the $x_{n+1}$-axis in $\mathbb{H}^{n+1}$. Let $\pi_\alpha : \mathbb{H}^{n+1} \to \alpha$ denote the orthogonal projection to $\alpha$: For every $p \in \mathbb{H}^{n+1}$, $\pi_\alpha(p) = q \in \alpha$, such that the geodesic $\overline{pq}$ is orthogonal to $\alpha$. The map $\pi_\alpha$ is the nearest-point projection to $\alpha$. This map, obviously, extends to $\mathbb{H}^{n+1} \cup S^n$. Then, $\pi_\alpha(\mathbb{A})$ is the interval $\sigma[R_1 e_{n+1}, R_2 e_{n+1}] \subset \alpha$, whose hyperbolic length is $\ell = \log(R_2/R_1)$, see Exercise 8.10.

By Lemma 9.80, the $(L, A)$-quasi-geodesic $f(\alpha)$ lies within distance $\theta(L, A, \delta)$ from the $\alpha \subset \mathbb{H}^{n+1}$, since we are assuming that $f(0) = 0, f(\infty) = \infty$. Here $\delta$ is the hyperbolicity constant for $\mathbb{H}^{n+1}$. Recall that, by Proposition 9.82, quasi-isometries "almost commute" with nearest-point projections and, thus:

$$d(f(\pi_\alpha(x)), \pi_\alpha f(x)) \leqslant C = C(L, A, \delta), \quad \forall x \in X.$$

Thus, $\pi_\alpha(\mathbb{A})$ is an interval of the hyperbolic length $\leqslant c' := 2C + Ld + A$

Lemma 20.25.

$$\operatorname{diam}(\pi_\alpha(f(\mathbb{A}))) \leqslant 2C + L\ell + A$$

*where $\delta$ is the hyperbolicity constant of $\mathbb{H}^{n+1}$.*

Proof. The ideal boundary of the preimage $\tilde{A} = \pi_\alpha^{-1}(\sigma) \subset \mathbb{H}^{n+1}$ is the annulus $\mathbb{A}$. Thus, it suffices to work with the spherical half-shell $\tilde{\mathbb{A}}$.

Let $p, q \in \tilde{\mathbb{A}}$. Then

$$d(f\pi_\alpha(p), \pi_\alpha f(p)) \leqslant C, \quad d(f\pi_\alpha(q), \pi_\alpha f(q)) \leqslant C.$$

Since $d(\pi_\alpha(p), \pi_\alpha(q)) \leqslant \ell$,

$$d(f\pi_\alpha(p), f\pi_\alpha(q)) \leqslant L\ell + A,$$

and, by triangle inequality we obtain

$$d(\pi_\alpha f(p), \pi_\alpha f(q)) \leqslant 2C + (L\ell + A). \quad \square$$

Now we can finish the proof of the theorem: The above lemma implies that

$$f(\mathbb{A}) \subset \pi_\alpha^{-1}(\sigma')$$

where $\sigma' \subset \alpha$ has length $\leqslant \ell' = 2C + L\ell + A$. The ideal boundary of $\pi_\alpha^{-1}(\sigma')$ is an annulus of eccentricity $\leqslant e^{\ell'}$. Thus, eccentricity of the annulus $\mathbb{A}'$ is at most

$$e^{\ell'} = e^{2C+A} \cdot e^{L\ell} = e^{2C+A} \cdot e^{L\log(t)} = e^{2C+A} t^L,$$

where $t = R_2/R_1$. $\square$

The following converse theorem was first proven by Tukia in the case of hyperbolic spaces and then extended by Paulin in the case of more general Gromov-hyperbolic spaces.

THEOREM 20.26 (P. Tukia [**Tuk94**], F. Paulin [**Pau96**]). *Every $\eta$-quasi-symmetric homeomorphism $f : \mathbb{R}^n \to \mathbb{R}^n$ extends to an $(A, A)$-quasi-isometric map $F$ of the hyperbolic space $\mathbb{H}^{n+1}$, where*

$$A = \eta(e) + 2\log(\eta(e)) + \log(\eta(e + 1)).$$

PROOF. We define a extension $F$ as follows. For every $p \in \mathbb{H}^{n+1}$, let $\alpha = \alpha_p$ be the complete vertical geodesic through $p$. This geodesic limits to points $\infty$ and $x = x_p \in \mathbb{R}^n$. Let $y \in \mathbb{R}^n$ be a point so that $\pi_\alpha(y) = p$ (the point $y$ is non-unique, of course). Let $x' := f(x), y' := f(y)$, let $\alpha' \subset \mathbb{H}^{n+1}$ be the vertical geodesic through $x'$ and let $p' := \pi_{\alpha'}(y')$. Lastly, set $F(p) := p'$.

We will prove only that $F$ is an $(A, A)$ coarse Lipschitz, where $A = A(\eta)$. The quasi-inverse to $F$ will be a map $\bar{F}$ defined *via* extension of the map $f^{-1}$ following the same procedure. We will leave it as an exercise to verify that $\bar{F}$ is indeed a quasi-inverse to $F$ and estimate $d(\bar{F} \circ F, id)$.

Suppose that $d(p_1, p_2) \leqslant 1$. We would like to bound $d(p'_1, p'_2)$ from above. Without loss of generality, we may assume that $p_1 = e_{n+1} \in \mathbb{H}^{n+1}$. It suffices to consider two cases:

1. $p_1, p_2$ belong to the common vertical geodesic $\alpha$, $x_1 = x_2 = x$ and $d(p_1, p_2) \leqslant 1$. I will assume, for concreteness, that $y_1 \leqslant y_2$. Hence,

$$d(p_1, p_2) = \log\left(\frac{|y_2 - x|}{|y_1 - x|}\right) \leqslant 1$$

Since the map $f$ is $\eta$-quasi-symmetric,

$$\frac{1}{\eta(e)} \leqslant \left(\eta\left(\frac{|y_2 - x|}{|y_1 - x|}\right)\right)^{-1} \leqslant \frac{|y'_2 - x'|}{|y'_1 - x'|} \leqslant \eta\left(\frac{|y_2 - x|}{|y_1 - x|}\right) \leqslant \eta(e).$$

In particular,

$$d(p'_1, p'_2) \leqslant C_1 = \log(\eta(e)).$$

2. Suppose that the points $p_1, p_2$ have the same last coordinate, which equals 1 since $p_1 = e_{n+1}$, and $t = |p_1 - p_2| \leqslant e$. The points $p'_1, p'_2$ belong to vertical lines $\alpha'_1, \alpha'_2$ limit to points $x'_1, x'_2 \in \mathbb{R}^n$. Without loss of generality (by postcomposing $f$ with an isometry of $\mathbb{H}^{n+1}$) we may assume that $|x'_1 - x'_2| = 1$. Let $y_i \in \mathbb{R}^n, y'_i \in \mathbb{R}^n$ be points so that

$$\pi_{\alpha_i}(y_i) = p_i, \pi_{\alpha'_i}(y'_i) = p'_i.$$

Then

$$|y_i - x_i| = |p_i - x_i| = R_i = 1, \quad i = 1, 2,$$
$$|y'_i - x'_i| = |p'_i - x'_i| = R'_i \quad i = 1, 2.$$

We can assume that $R'_1 \leqslant R'_2$. Then

$$d(p'_1, p'_2) \leqslant \frac{1}{R'_1} + \log\left(\frac{R'_2}{R'_1}\right),$$

since we can first travel from $p'_1$ to the line $\alpha'_2$ horizontally (along path of the length $\frac{1}{R'_1}$) and then vertically, along $\alpha'_2$ (along path of the length $\log(R'_2/R'_1)$). We then apply the $\eta$-quasi-symmetry condition to the triple of points $x_1, y_1, x_2$ and get:

$$\frac{1}{R'_1} \leqslant \eta\left(\frac{t}{R_1}\right) \leqslant \eta(e).$$

Setting $R_3 := |x_1 - y_2|, R'_3 := |x'_1 - y'_2|$ and applying $\eta$-quasi-symmetry condition to the triple of points $x_1, y_1, y_2$, we obtain

$$\frac{R'_3}{R'_1} \leqslant \eta\left(\frac{R_3}{R_1}\right) \leqslant \eta\left(\frac{t+1}{1}\right) \leqslant \eta(e+1).$$

Since $R'_2 \leqslant R'_3 + 1$, we get:

$$\frac{R'_2}{R'_1} \leqslant \frac{R'_3 + 1}{R'_1} \leqslant \eta(e+1) + \eta(e).$$

Putting it all together, we obtain that in Case 2:

$$d(p'_1, p'_2) \leqslant \eta(e) + \log(\eta(e+1) + \eta(e)) = C_2.$$

Thus, in general, for $p_1, p_2 \in \mathbb{H}^{n+1}, d(p_1, p_2) \leqslant 1$, we get:

$$d(F(p_1), F(p_2)) \leqslant C_1 + C_2 = A. \quad \square$$

Now, for points $p, q \in \mathbb{H}^{n+1}$, so that $d(p_1, p_2) \geqslant 1$, we find a chain of points $p_0 = p, ..., p_{k+1} = q$, where $k = \lfloor d(p,q) \rfloor$ and $d(p_i, p_{i+1}) \leqslant 1, i = 0, ..., k$. Hence,

$$d(F(p), F(q)) \leqslant A(k+1) \leqslant Ad(p,q) + A.$$

Hence, the map $F$ is $(A, A)$ coarse Lipschitz, where

$$A = C_1 + C_2 \leqslant \eta(e) + 2\log(\eta(e)) + \log(\eta(e+1)). \quad \square$$

CHAPTER 21

# Groups quasi-isometric to $\mathbb{H}^n$

The main result of this chapter is the following theorem which is due to P. Tukia, see [**Tuk86**] and [**Tuk94**]:

THEOREM 21.1 (P. Tukia). *If $G$ is a finitely generated group QI to $\mathbb{H}^{n+1}$ (with $n \geqslant 2$), then $G$ acts geometrically on $\mathbb{H}^{n+1}$. In particular, $G$ is virtually isomorphic to a uniform lattice in the Lie group* $\mathrm{Isom}(\mathbb{H}^{n+1})$.

Recall that if a group $G$ is QI to $\mathbb{H}^{n+1}$ then it quasi-acts on $\mathbb{H}^{n+1}$, see Lemma 5.60. Furthermore (by Theorem 9.107), every such quasi-action $\varphi$ determines an action $G \curvearrowright S^n$ on the boundary sphere of $\mathbb{H}^{n+1}$. Since the quasi-action of $G$ was by uniform quasi-isometries, the action of $G \curvearrowright S^n$ is by uniformly quasiconformal homeomorphisms, see Theorem 20.24. According to Lemma 5.60, the quasi-action $G \curvearrowright \mathbb{H}^{n+1}$ is geometric and, by Lemma 9.91, every point $\xi \in S^n$ is a conical limit point of $G \curvearrowright S^n$. Lastly, according to Theorem 9.107, the fact that the quasi-action $G \curvearrowright \mathbb{H}^{n+1}$ is geometric translates to:

The action $G \curvearrowright Trip(S^n)$ is properly discontinuous and cocompact. In particular, $G < Homeo(S^n)$ is a discrete subgroup, where $Homeo(S^n)$ is equipped with the topology of uniform convergence.

Our goal (and this is the main result of Sullivan and Tukia), under the above hypothesis, there exists a quasiconformal homeomorphism $f : S^n \to S^n$ which conjugates $\Gamma$ to a group of Moebius transformations, whose action on $\mathbb{H}^{n+1}$ is geometric. Once the existence of such $f$ is proven, Theorem 21.1 would follow.

Thus, in order to prove Theorem 21.1 one is naturally lead to study *uniformly quasiconformal group actions on $S^n$*.

## 21.1. Uniformly quasiconformal groups

Let $G < Homeo(S^n)$ be a group of consisting of quasiconformal homeomorphisms. The group $G$ is called *uniformly quasiconformal* (abbreviated as *uq*) if there exists $K < \infty$ so that $K(g) \leqslant K$ for all $g \in G$, where $K(g)$ is the quasiconformality constant. Trivial examples of uniformly quasiconformal groups are given by groups $\Gamma < Mob(S^n)$ of Moebius transformations and their quasiconformal conjugates

$$\Gamma^f = f\Gamma f^{-1},$$

where $f$ is $k$-quasiconformal. Then for every $g \in \Gamma^f$,

$$K(g) = K(f\gamma f^{-1}) \leqslant k^2 = K.$$

We say that a uniformly quasiconformal subgroup $G < Homeo(S^n)$ is *exotic* if it is not quasiconformally conjugate to a group of Moebius transformations. It is a fundamental fact of quasiconformal analysis in dimension 2 observed first by D. Sullivan in [**Sul81**] that

THEOREM 21.2. *There are no exotic uniformly quasiconformal subgroups in* $Homeo(S^2)$.

This theorem fails rather badly for $n \geqslant 3$: First, P. Tukia [**Tuk81**] constructed examples of uniformly quasiconformal subgroups $G < Homeo(S^n), n \geqslant 3$, which are isomorphic to connected solvable Lie groups, but are not isomorphic to subgroups of $\mathrm{Isom}(\mathbb{H}^m)$ for any $m$. Algebraically, Tukia's examples are semidirect products $\mathbb{R}^k \rtimes \mathbb{R}^2$, where $(a, b) \in \mathbb{R}^2$ acts on $\mathbb{R}^k$ *via* a diagonal matrix $D(a, b)$ that has (generically) two distinct eigenvalues $\neq \pm 1$. In particular, such groups have to be exotic. Further examples of discrete exotic uniformly quasiconformal subgroups of $Homeo(S^3)$ were constructed in [**FS88**], [**Mar86**] (these groups have torsion) and in [**Kap92**] (these are certain surface groups acting on $S^3$). An example of a discrete uniformly quasiconformal subgroup of $Homeo(S^3)$ which is not *isomorphic* to subgroup of $Mob(S^4)$ was constructed in [**Isa90**].

PROBLEM 21.3. Suppose that $G < Homeo(S^n)$ is a discrete uniformly quasiconformal subgroup. Is it true that $G$ is isomorphic to a subgroup of $\mathrm{Isom}(\mathbb{H}^n)$ for some $m$?

The answer to this questions is probably negative. One can, nevertheless, ask which algebraic properties of discrete groups of Moebius transformations are shared by discrete uniformly quasiconformal subgroups, e.g.:

PROBLEM 21.4. Is it true that discrete infinite uniformly quasiconformal subgroups of $Homeo(S^n)$ never have Property (T)? Even stronger, one can ask if all discrete uniformly quasiconformal subgroups of $Homeo(S^n)$ have Haagerup property.

It is unclear what the answer to this question is. In view of Theorem 9.167 it is conceivable that every hyperbolic group $G$ acts on *some* $S^n$ as a uniformly quasiconformal group.

PROBLEM 21.5. Suppose that $G < Homeo(S^n)$ is a discrete uniformly quasiconformal subgroup. Is it true that the action $G \curvearrowright S^n$ extends to a uniformly quasiconformal action $G \curvearrowright \mathbb{H}^{n+1}$? $G$ is isomorphic to a subgroup of $\mathrm{Isom}(\mathbb{H}^n)$ for some $m$?

Note that in view of theorems 21.6 and 5.64, there exists a hyperbolic space $X$ quasi-isometric to $\mathbb{H}^{n+1}$, so that $G$ acts *isometrically* on $X$ and the actions on $G$ on $\partial_\infty X$ and $S^n$ are topologically conjugate.

## 21.2. Hyperbolic extension of uniformly quasiconformal groups

As we saw, every quasi-action $G \curvearrowright \mathbb{H}^{n+1}$ extends to a uq action $G \curvearrowright S^n$. Our first goal is to prove the converse:

THEOREM 21.6 (P. Tukia, [**Tuk94**]). *Action of every uq group* $G < Homeo(S^n)$ *on* $S^n$ *extends to a quasi-action* $\varphi : G \curvearrowright \mathbb{H}^{n+1}$.

PROOF. For every $g \in G$ we let $\varphi(g) : \mathbb{H}^{n+1} \to \mathbb{H}^{n+1}$ denote the quasi-isometric extension of $g$ constructed in Theorem 20.26. Since every $g \in G$ is $K$-quasiconformal, every $\varphi(g)$ is an $(L, A)$-quasi-isometry, where $L$ and $A$ depend only on $K$. We need to show that the extension defines a quasi-action. We will only show that

$$\mathrm{dist}(\varphi(g_1) \circ \varphi(g_2), \varphi(g_1 g_2)) \leqslant C = C(L, A)$$

for all $g_1, g_2$, since the proof that $d(\varphi(1), id) \leqslant C = C(L, A)$ is similar.

Note that $f_1 = \varphi(g_1) \circ \varphi(g_2)$ is an $(L^2, LA + A)$ quasi-isometry, while $f_2 = \varphi(g_1 g_2)$ is an $(L, A)$ quasi-isometry and $(f_1)_\infty = (f_2)_\infty$.

By homogeneity of $\mathbb{H}^{n+1}$, every point of the hyperbolic space is a centroid of an ideal triangle. Therefore, by Lemma 9.86,

$$\operatorname{dist}(f_1, f_2) \leqslant C(L, A) = D(L, A, 0, \delta)$$

where $\delta$ is the hyperbolicity constant of $\mathbb{H}^{n+1}$. □

Therefore, study of uq groups is equivalent to study of quasi-actions on $\mathbb{H}^{n+1}$. In particular, this allows one to define *conical limit points* for uq subgroups $\Gamma < Homeo(S^n)$ as conical limit points of the extended quasi-actions.

Our goal, thus, is to prove the following theorem which was first established by D. Sullivan [**Sul81**] for $n = 2$ (without restrictions on conical limit points) and then by Tukia in full generality:

Theorem 21.7 (P. Tukia, [**Tuk86**]). *Suppose that $G$ is a countable uq group acting on $S^n, n \geqslant 2$, so that (almost) every point of $S^n$ is a conical limit point of $G$. Then $G$ is quasiconformally conjugate to a subgroup of $Mob(S^n)$.*

Before proving Tukia's theorem, we will need few technical tools.

### 21.3. Least volume ellipsoids

Observe that a closed ellipsoid centered at 0 in $\mathbb{R}^n$ can be described as

$$E = E_A = \{x \in \mathbb{R}^n : \varphi_A(x) = x^T A x \leqslant 1\}$$

where $A$ is some positive-definite symmetric $n \times n$ matrix. Volume of such ellipsoid is given by the formula

$$Vol(E_A) = \omega_n \left(\det(A)\right)^{-1/2}$$

where $\omega_n$ is the volume of the unit ball in $\mathbb{R}^n$. A subset $X \subset \mathbb{R}^n$ is *centrally-symmetric* if $X = -X$.

Theorem 21.8 (F. John, [**Joh48**]). *For every compact centrally-symmetric subset $X \subset \mathbb{R}^n$ with nonempty interior, there exists unique ellipsoid $E(X)$ of the least volume containing $X$. The ellipsoid $E(X)$ is called the* John–Loewner *ellipsoid of $X$.*

Proof. The existence of $E(X)$ is clear by compactness. We need to prove uniqueness. Consider the function $f$ on the space $P_n$ of positive definite symmetric $n \times n$ matrices, given by

$$f(A) = -\frac{1}{2} \ln\left(\det(A)\right).$$

Lemma 21.9. *The function $f$ is strictly convex.*

Proof. Take $A, B \in P_n$ and consider the family of matrices $C_t = tA + (1-t)B$, $0 \leqslant t \leqslant 1$. Strict convexity of $f$ is equivalent to strict convexity of $f$ on such line segments of matrices. Since $A$ and $B$ can be simultaneously diagonalized by a matrix $M$, we obtain:

$$f(D_t) = f(MC_tM^T) = -\ln\det(M) - \frac{1}{2}\ln\det(C_t) = -\ln\det(M) + f(C_t),$$

where $D_t$ is a segment in the space of positive-definite diagonal matrices. Thus, it suffices to prove strict convexity of $f$ on the space of positive-definite diagonal matrices $D = Diag(x_1, ..., x_n)$. Then,

$$f(D) = -\frac{1}{2}\sum_{i=1}^{n} \ln(x_i)$$

is strictly convex since ln is strictly concave. □

In particular, whenever $V \subset P_n$ is a convex subset and $f|V$ is proper, $f$ attains unique minimum on $V$. Since log is a strictly increasing function, the same uniqueness assertion holds for the function $\det^{-1/2}$ on $P_n$. Let $V = V_X$ denote the set of matrices $C \in P_n$ so that $X \subset E_C$. Since $\varphi_A(x)$ is linear as a function of $A$ for any fixed $x \in X$, it follows that $V$ convex. Thus, the least volume ellipsoid containing $X$ is unique. □

### 21.4. Invariant measurable conformal structure

Recall (see Section 2.1.3) that a measurable Riemannian metric on $S^n = \mathbb{R}^n \cup \{\infty\}$ is a measurable map $g$ from $S^n$ to the space $P_n$ of positive definite symmetric $n \times n$ matrices. (Since we are working in measurable category, we can and will ignore the point $\infty$.)

A *measurable conformal structure* on $S^n$ is a measurable Riemannian metric defined up to multiplication by a positive measurable function. In order to avoid the ambiguity with the choice of the conformal factor, one can normalize the measurable metric $g$ so that $\det(g(x)) = 1$ for all $x \in S^n$.

If $f : S^n \to S^n$ is a quasiconformal mapping, it acts on measurable Riemannian metrics *via* the pull-back by the usual formula:

$$f^*(g) = h, \quad h(x) = (D_x f)\ g(f(x))\ (D_x f)^T.$$

If we consider normalized Riemannian metrics, then the appropriate action is given by the formula:

$$f^\bullet(g) = h, \quad h(x) = (J_x)^{-2n} D_x f g(f(x))(D_x f)^T$$

in order for $h$ to be normalized as well. Here $J_x$ is the Jacobian of $f$ at $x$. We will think of normalized measurable Riemannian metrics as measurable conformal structures.

A measurable conformal structure $\mu$ on $S^n$ is called *bounded* if it is represented by a bounded normalized measurable Riemannian metric, i.e., a bounded map $S^n \to P_n \cap \{det = 1\}$. Below, we interpret boundedness of $\mu$ in terms of eigenvalues.

Given a measurable Riemannian metric $\mu(x) = A_x$, we define its *linear dilatation* $H(\mu)$ as the essential supremum of the ratios

$$H(x) := \frac{\sqrt{\lambda_n(x)}}{\sqrt{\lambda_1(x)}},$$

where $\lambda_1(x) \le ... \leqslant \lambda_n(x)$ are the eigenvalues of $A_x$. Geometrically speaking, if $E_x \subset T_x\mathbb{R}^n$ is the unit ball with respect to $A_x$, then $H(x)$ is the eccentricity of the ellipsoid $E_x$, i.e., the ratio of the largest to the smallest axis of $E_x$. In particular, $H(x)$ and $H(\mu)$ are conformal invariants of $\mu$.

EXERCISE 21.10. A measurable conformal structure $\mu$ is bounded if and only if $H(\mu) < \infty$.

The following was first observed by Sullivan in [**Sul81**] for $n = 2$ and, then, by Tukia [**Tuk86**] for arbitrary $n$:

We say that a *measurable conformal structure* $\mu(x) = A_x$ on $\mathbb{R}^n$ is invariant under a quasiconformal group $G$ if

$$g^\bullet \mu = \mu, \forall g \in G.$$

In detail:

$$\forall g \in G, \quad (J_{g,x})^{-\frac{1}{2n}} (D_x g)^T \cdot A_{gx} \cdot D_x g = A_x$$

a.e. in $\mathbb{R}^n$.

PROPOSITION 21.11. *Let $G < Homeo(S^n)$ be a countable uniformly quasiconformal group. Then $G$ admits an invariant measurable conformal structure $\lambda$ on $S^n$.*

PROOF. Let $\mu_0$ be the Euclidean metric, it is given by the constant matrix function $x \mapsto I$. Consider the orbit $G \cdot \mu_0$ in the space of normalized measurable Riemannian metrics. The ideal is to "average" all the measurable conformal structures in this orbit.

Since $G$ is countable, for a.e. $x \in \mathbb{R}^n$, we have well-defined matrix-valued functions corresponding to $g^\bullet(\mu_0)$ on $T_x\mathbb{R}^n$:

$$A_{g,x} := (J_{g,x})^{-\frac{1}{2n}} (D_x g)^T \cdot D_x g.$$

Thus $H(A_{g,x}) = H_g(x)$, is the linear dilatation of $g$ at $x$, see 20.10. Therefore, the assumption that $G$ is uq is equivalent to the assumption that the family of measurable conformal structures $G \cdot \mu_0$ is uniformly bounded:

$$\sup_{g \in G} H(g^\bullet \mu_0) = H < \infty.$$

Geometrically, one can think of this as follows. For a.e. $x$ we let $E_{g,x}$ denote the unit ball in $T_x\mathbb{R}^n$ with respect to $g^\bullet(\mu_0)$. From the Euclidean viewpoint, $E_{g,x}$ is just an ellipsoid of the volume $\omega_n$ (since $g^\bullet(\mu_0)$ is normalized). This ellipsoid (up to scaling) is the image of the unit ball under the inverse of the derivative $D_x g$. Then uniform boundedness of conformal structures $g^\bullet(\mu_0)$ simply means that eccentricities of the ellipses $E_{g,x}$ are bounded by the number $H$, which is independent of $g$ and $x$. Since volume of each $E_{g,x}$ is fixed, it follows that the diameter of the ellipsoid is uniformly bounded above and below: There exists $0 < R < \infty$ so that

$$B(0, R^{-1}) \subset E_{g,x} \subset B(0, R), \forall g \in G$$

for a.e. $x \in \mathbb{R}^n$.

Let $U_x$ denote the union of the ellipsoids

$$\bigcup_{g \in G} E_{g,x}.$$

Since each ellipsoid $E_{g,x}$ is centrally-symmetric, so is $U_x$. By construction, the family of sets $\{U_x, x \in \mathbb{R}^n\}$ is invariant under the group $G$:

$$(J_{g,x})^{-1/n} D_x g(U_x) = U_{g(x)}, \quad \forall g \in G.$$

We then let $E_x$ denote the John-Loewner ellipsoid of $U_x$. Since the group $G$ preserves the family of sets $U_x$ and since, after normalization, the action of $D_x g$ on the tangent space is volume-preserving, it follows (by uniqueness of John-Loewner ellipsoid, see Theorem 21.8) that $G$ also preserves the family of ellipsoids $E_x$.

Clearly,

$$B(0, R^{-1}) \subset E_x \subset B(0, R)$$

a.e. in $\mathbb{R}^n$, and, hence, eccentricities of the ellipsoids $E_x$ are bounded from above and below. Let $\mu(x)$ denote the (a.e. defined) function $\mathbb{R}^n \to P_n$ which sends $x$ to the matrix $A_x$ such that $E_x$ is the unit ball with respect to the quadratic form defined by $A_x$. Then, $H(A_x) \leqslant R^2$ a.e..

LEMMA 21.12. *The function $x \to A_x$ is measurable.*

PROOF. Since $G$ is countable, it is an increasing union of finite subsets $G_i \subset G$. For each $i$ we define the sets

$$U_{x,i} = \bigcup_{g \in G_i} E_{g,x}$$

and the corresponding John-Loewner ellipsoids $E_{x,i}$. We leave it to the reader to check that since each ellipsoid $E_{g,x}$ is measurable as a function of $y$, then $E_{x,i}$ is also measurable. Note also that

$$E_y = \bigcup_{i \in \mathbb{N}} E_{x,i}.$$

Let $\mu_i : \mathbb{R}^n \to P_n$ denote the measurable functions defining by the ellipsoids $E_{x,i}$. We will think of this function as a function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$,

$$(x, v) \mapsto v^T \mu_i(x) v \in \mathbb{R}_+.$$

Then the fact that $E_i \subset E_{i+1}$ means that

$$\mu_i(x, v) \geqslant \mu_{i+1}(x, v).$$

Furthermore,

$$g = \lim_i g_i.$$

Now, lemma follows from Lebesgue monotone convergence theorem (Beppo Levi's theorem), see e.g. [**SS05**]. □

This also concludes the proof of the proposition. □

The above theorem also holds for uncountable uq groups, see [**Tuk86**], but we will not need this fact.

## 21.5. Quasiconformality in dimension 2

In this section we reformulate quasiconformality of a map in the 2-dimensional case in terms of the *Beltrami equation* and explain the relation between measurable conformal structures on domains in $\mathbb{C}$ and *Beltrami differentials*. We refer to [**Ahl06**] and [**Leh87**] for further details.

**21.5.1. Beltrami equation.** For computational purposes, we will use the complex differentials $dz = dx + idy$ and $d\bar{z} = dx - idy$. These differentials define coordinates on the complexification of the real tangent space $T_zU$, $U \subset \mathbb{R}^2$. Accordingly,

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right),$$

$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

To simplify the notation, we let $\partial f$ denote $\frac{\partial f}{\partial z} = f_z$ and let $\bar\partial f$ denote $\frac{\partial f}{\partial \bar z} = f_{\bar z}$, the holomorphic and antiholomorphic derivatives.

Consider a function $f(z)$ which is differentiable at a point $z \in \mathbb{C}$. Writing $f = u + iv$, we obtain the formula for the (real) Jacobian of $f$:

$$J_f = u_x v_y - u_y v_x = |\partial f|^2 - |\bar\partial f|^2.$$

We will assume from now on that $f$ is orientation-preserving at $z$, i.e., $|\partial f(z)| > |\bar\partial f(z)|$.

For $\alpha \in [0, 2\pi]$, the directional derivative of $f$ at $z$ in the direction $e^{i\alpha}$ equals

$$\partial_\alpha f = \partial f + e^{-2i\alpha}\bar\partial f.$$

We now can compute lengths of major and minor semi-axes of the ellipsoid which is the image of the unit tangent circle under $D_z f$:

$$\max_\alpha |\partial_\alpha f| = |\partial f| + |\bar\partial f|,$$

$$\min_\alpha |\partial_\alpha f| = |\partial f| - |\bar\partial f|.$$

Thus,

$$H_z(f) = \max_{\alpha,\beta} \frac{|\partial_\alpha f|}{|\partial_\beta f|} = \frac{|\partial f| + |\bar\partial f|}{|\partial f| - |\bar\partial f|}$$

is the linear dilatation of $f$ at $z$. Setting $\mu(z) = \frac{\bar\partial f}{\partial f}$, we obtain

$$H_z(f) = \frac{1 + |\mu(z)|}{1 - |\mu(z)|}.$$

Suppose now that $f : U \to \mathbb{C}$ and $f \in W^{1,2}_{loc}(U)$; in particular, $f$ is differentiable a.e. in $U$, its derivatives are locally square-integrable in $U$ and $J_z(f) > 0$ in $U$, i.e., $f$ is orientation-preserving. Then, we have a measurable function

$$\mu = \mu(z) = \frac{f_{\bar z}}{f_z}, \tag{21.1}$$

called the *Beltrami differential* of $f$; the equation (21.1) is called the *Beltrami equation.* Let $k = k_f = \|\mu\|$ be the $L^\infty$-norm of $\mu$ in $U$. Then,

$$K(f) = \sup_{z\in U} H_z(f) = \frac{1+k}{1-k}.$$

Thus, the following are equivalent for a function $f$:

1. $f$ is $K$-quasiconformal, where $K = \frac{1+k}{1-k}$.
2. $f$ satisfies the Beltrami equation and $k = \|\mu\| < 1$.

In particular, an (orientation-preserving) quasiconformal map is 1-quasiconformal if and only if $k_f = 0$, i.e., $\mu = 0$, equivalently, $\bar\partial f = 0$ (almost everywhere). A theorem of Weyl (see e.g. **[Ahl06]** then states that such maps are holomorphic.

**21.5.2. Measurable Riemannian metrics.** Let $f : U \to U' \subset \mathbb{C}$ be an orientation-preserving quasiconformal homeomorphism, $w = f(z)$, with the Beltrami differential $\mu$. For $w = u + iv$ it is useful to compute the pull-back of the Euclidean metric $du^2 + dv^2 = |dw|^2$ by the map $f$:

$$|dw|^2 = |\partial f dz + \bar\partial f d\bar z|^2 =$$

$$|\partial f|^2 \cdot |dz + \frac{\bar\partial f}{\partial f} d\bar z|^2 = |\partial f|^2 \cdot |dz + \mu(z) d\bar z|^2.$$

Therefore, let $ds^2_\mu$ denote the measurable Riemannian metric $|dz + \mu(z)d\bar{z}|^2$.

Our next goal is to show that an arbitrary measurable Riemannian metric in a domain $U$ in $\mathbb{C}$ is conformal to a metric of the form $|dz + \mu(z)d\bar{z}|^2$ for some $\mu$. Let

$$ds^2 = Edx^2 + 2Fdxdy + Gdy^2.$$

We will consider only the case of diagonal form $ds^2$, $F = 0$, since the general case is obtained by change of variables $z = e^{i\theta}w$, which converts the form $|dz+\mu(z)d\bar{z}|^2$ to $|dw + \mu(z)e^{-2i\theta}d\bar{w}|^2$. For $F = 0$, $\mu$ is real and the condition that $ds^2_\mu$ is a multiple of $ds^2$ translates to

$$1 + \mu = t\sqrt{E}, \quad 1 - \mu = t\sqrt{G}$$

for some $t \in (0, \infty)$. Solving this, we obtain

$$\mu = \frac{\sqrt{E} - \sqrt{G}}{\sqrt{E} + \sqrt{G}}.$$

Clearly, $|\mu| < 1$. Furthermore, $\lim_{z\to z_0} |\mu(z)| = 1$ if and only if

$$\lim_{z\to z_0} \frac{E(z)}{G(z)} \in \{0, \infty\}.$$

Thus, the condition that the measurable conformal structure $[ds^2]$ defined by $ds^2$ is bounded is equivalent to the condition that $\|\mu\| < 1$.

The conclusion, therefore, is that the correspondence $\mu \mapsto ds^2_\mu$ establishes an equivalence of Beltrami differentials $\mu$ with norm $< 1$ and bounded measurable conformal structures. Furthermore, if $f$ is a quasiconformal map solving the *Beltrami equation* (21.1), then $f^*(|dz|^2)$ is conformal to the metric $ds^2_\mu$.

THEOREM 21.13 (Measurable Riemann Mapping Theorem). *For every measurable function $\mu(z)$ on $U$ satisfying $\|\mu\|_\infty < 1$, there exists a quasiconformal homeomorphism $f : U \to U' \subset S^2$ with the Beltrami differential $\mu$. Equivalently, every bounded measurable conformal structure $[ds^2]$ on $U$ is equivalent to the standard conformal structure on a domain $U' \subset S^2$ via a quasiconformal map $f : U' \to U$.*

**Historic Remarks.** In the case of smooth Riemannian metric $ds^2$, a local version of this theorem was proven by Gauss, it is called *Gauss' theorem on isothermal coordinates*. In full generality it was established by Morrey [**Mor38**]. Modern proofs could be found, for instance, in [**Ahl06**] and [**Leh87**].

## 21.6. Approximate continuity and strong convergence property

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *approximately continuous* at a point $x \in \mathbb{R}^n$ if for every $\epsilon > 0$

$$\lim_{r\to 0} \frac{mes(\{y \in B(x, r) : |f(x) - f(y)| > \epsilon\})}{mes(B(x, r))} = 0. \tag{21.2}$$

(Here, as before, $mes$ denotes the Lebesgue measure.) In other words, as we "zoom into" the point $x$, "most" points $y \in B(x, r)$, have value $f(y)$ close to $f(x)$, i.e., the rescaled functions $f_r(x) := f(rx)$ converge (with $r \to 0$) in measure to the constant function.

LEMMA 21.14. *Every $L_\infty$-function $f : \mathbb{R}^n \to \mathbb{R}$, is almost continuous at almost every point.*

PROOF. The proof is an application of the *Lebesgue Density theorem* (see e.g. [**SS05**, p. 106]): For every measurable function $h$ on $\mathbb{R}^n$ and almost every $x$,

$$\lim_{r\to 0}\frac{1}{mes(B_r)}\int_{B_r}|h(y)-h(x)|dy=0.$$

Here and below, we let $B_r = B(x, r)$.

Fix $\epsilon > 0$ and let $E_r \subset B_r$ denote the subset consisting of $y \in B_r$ so that

$$|f(y)-f(x)|>\epsilon.$$

If the equality (21.2) fails, then

$$\lim_{r\to 0}\frac{mes(E_r)}{mes(B_r)}>0.$$

Then

$$\frac{1}{mes(B_r)}\int_{B_r}|f(y)-f(x)|dy\geqslant \epsilon\frac{mes(E_r)}{mes(B_r)}.$$

Since

$$\lim_{r\to 0}\frac{mes(E_r)}{mes(B_r)}>0,$$

we conclude that

$$\liminf_{r\to 0}\frac{1}{mes(B_r)}\int_{B_r}|f(y)-f(x)|dy\neq 0,$$

contradicting Lebesgue Density Theorem. □

The key analytical ingredient that is needed for the proof of Tukia's theorem is:

THEOREM 21.15 (Tukia's strong convergence property, [**Tuk86**]; see also [**IM01**] for a stronger version). *Let $(f_i)$ be a sequence of $K$-quasiconformal maps converging to a quasiconformal map $f$. Let $E_i$ is a sequence of subsets in the common domain of $f_i$'s so that*

$$\limsup_i H(f_i|E_i^c)=H$$

*and $\lim_i mes(E_i) = 0$. Then $f$ satisfies $H(f) \leqslant f$.*

## 21.7. Proof of Tukia's theorem on uniformly quasiconformal groups

THEOREM 21.16 (P. Tukia, [**Tuk86**]). *Let $G < Homeo(S^n)$ be a uniformly quasiconformal group. Assume also that $\mu$ is a $G$-invariant bounded measurable conformal structure on $S^n$ which is almost continuous at a conical limit point $\xi$ of $G$. Then there exists a quasiconformal homeomorphism $f : S^n \to S^n$ which sends $\mu$ to the standard conformal structure on $S^n$ and conjugates $G$ to a group of Moebius transformations.*

PROOF. As before, we will identify $S^n$ with $\widehat{\mathbb{R}}^n = \mathbb{R}^n \cup \infty$. We first explain Sullivan's proof of this theorem in the case $n = 2$ since it is easier and does not use conical limit points assumption.

In view of the Measurable Riemann Mapping theorem for $S^2$, the bounded measurable conformal structure $\mu$ on $S^2$ is equivalent to the standard conformal structure $\mu_0$ on $S^2$, i.e., there exists a quasiconformal map $f : S^2 \to S^2$ which sends $\mu$ to $\mu_0$:

$$f^{\bullet}\mu_0=\mu.$$

Since a quasiconformal group $G$ preserves the conformal $\mu$ on $S^2$, it follows that the conjugate group $G^f = fGf^{-1}$ preserves the conformal structure $\mu_0$. Therefore, each $h \in G^f$ is 1-quasiconformal homeomorphism of $S^2$, hence, a Moebius transformation, see Section 20.4. Thus, $G^f$ acts as a group of Moebius automorphisms of the round sphere, which proves theorem for $n = 2$.

We now consider the general case. The *zooming* argument below will be used again in the proofs of Schwarz and Mostow Rigidity Theorems (§22.4 and §22.8). The idea is that the fraction appearing in the definition of derivative (at a point $p$) of a function $f$ of several real variables is nothing but pre- and post-composition of $f$ with some Moebius transformations. In the case when $p$ is a conical limit point of a quasiconformal group $\Gamma$, we can approximate the pre-composition by using a pre-composition using elements of $\Gamma$.

Without loss of generality, we may assume that the conical limit point $\xi$ is the origin in $\mathbb{R}^n$ and (by conjugating $G$ *via* an affine transformation if necessary) that $\mu(0) = \mu_0(0)$ is the standard conformal structure on $\mathbb{R}^n$. We will identify $\mathbb{H}^{n+1}$ with the upper half-space $\mathbb{R}^{n+1}_+$. Let $e = e_{n+1} = (0, ..., 0, 1) \in \mathbb{H}^{n+1}$. Let $\phi : G \curvearrowright \mathbb{H}^{n+1}$ be the quasi-action, extending the action $G \curvearrowright S^n$, see Theorem 20.26. Let $(L, A)$ be the quasi-isometry constants for this quasi-action.

By definition of a conical limit point, there exists a sequence $g_i \in G$, a number $c \in \mathbb{R}$, so that $\lim_{i\to\infty} \phi(g_i)(e) = 0$ and

$$d(\phi(g_i)(e), t_i e) \leqslant c$$

where $d$ is the hyperbolic metric on $\mathbb{H}^{n+1}$ and $t_i > 0$ is a sequence converging to zero.

Let $\gamma_i$ denote the hyperbolic isometry (the Euclidean dilation) given by

$$x \mapsto t_i x, x \in \mathbb{R}^{n+1}.$$

Set

$$\tilde{g}_i := g_i^{-1} \circ \gamma_i.$$

Then

$$d(\phi(\tilde{g}_i)(e), e) \leqslant Lc + A$$

for all $i$. Furthermore, each $\tilde{g}_i$ is an $(L, A)$-quasi-isometry of $\mathbb{H}^{n+1}$. Let $f_i := (\tilde{g}_i)_\infty$ denote the extensions of $\tilde{g}_i$ to $S^n$. By Theorem 20.24, each $f_i$ is $K$-quasiconformal for some $K$ independent of $i$.

Let $T$ be an ideal hyperbolic triangle with centroid $e$ with the set $\zeta$ of ideal vertices. then, by Morse Lemma, the quasi-geodesic triangles $\tilde{g}_i(T)$ are uniformly close to ideal geodesic triangles $T_i$ in $\mathbb{H}^{n+1}$, so that $\mathrm{dist}(center(T_i), e) \leqslant Const$. Therefore, after passing to a subsequence which we suppress) geodesic triangles $T_i$ converge to an ideal triangle $T_\infty$ in $\mathbb{H}^{n+1}$. In particular, the $K$-quasiconformal maps $f_i$ restricted to the set $\xi$ converge to a bijection $\xi \to \xi' \subset S^n$. Therefore, by the convergence property of quasiconformal maps (see Section 20.4), after passing again to a subsequence, the sequence $(f_i)$ converges to a quasiconformal map $f : S^n \to S^n$.

We also have:

$$\mu_i := f_i^\bullet(\mu) = (\gamma_i)^\bullet (g_i)^{-1\bullet}(\mu) = (\gamma_i)^\bullet \mu,$$

since $g^\bullet(\mu) = \mu, \forall g \in G$. Thus,

$$\mu_i(x) = \mu(\gamma_i x) = \mu(t_i x),$$

in other words, the measurable conformal structure $\mu_i$ is obtained by "zooming into" the point 0. Since $x$ is an approximate continuity point for $\mu$, the functions $\mu_i(x)$ converge (in measure) to the constant function $\mu_0 = \mu(0)$. Thus, we have the diagram:

$$\begin{array}{ccc} \mu & \xrightarrow{f_i} & \mu_i \\ & & \downarrow \\ \mu & \xrightarrow{f} & \mu_0 \end{array}$$

If we knew that the derivatives $Df_i$ subconverge (in measure) to the derivative of $Df$, then we would conclude that

$$f^\bullet \mu = \mu_0.$$

Then $f$ would conjugate the group $G$ (preserving $\mu$) to a group $G^f$ preserving $\mu_0$ and, hence, acting conformally on $S^n$.

However, derivatives of quasiconformal maps (in general), converge only in the "biting" sense (see [**IM01**]), which does not suffice for our purposes. Thus, we have to use a less direct argument below.

We claim that every element of $G^f$ is 1-quasiconformal. It suffices to verify this property locally. Thus, restrict to a certain round ball $B$ in $\mathbb{R}^n$. Since $\mu$ is approximately continuous at 0, for every $\epsilon \in (0, \frac{1}{2})$,

$$\|\mu_i(x) - \mu(0)\| < \epsilon$$

away from a subset $W_i \subset B$ of measure $< \epsilon_i$, where $\lim_i \epsilon_i = 0$. Thus, for $x \in W_i$,

$$1 - \epsilon < \lambda_1(x) \leqslant ... \leqslant \lambda_n(x) < 1 + \epsilon,$$

where $\lambda_k(x)$ are the eigenvalues of the matrix $A_{i,x}$ of the normalized metric $\mu_i(x)$. Thus,

$$H_x(\mu_i) < \frac{\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}} \leqslant \sqrt{1+4\epsilon} \leqslant 1 + 2\epsilon.$$

away from subsets $W_i$. For every $g \in G$, each map $h_i := f_i g f_i^{-1}$ is conformal with respect to the structure $\mu_i$ and, hence $(1 + 2\epsilon)$-quasiconformal away from the set $W_i$. Since measures of the sets $W_i$ converge to zero, we conclude, by the strong convergence property, that each $h := \lim h_i$ is $(1 + 2\epsilon)$-quasiconformal. Since this holds for arbitrary $\epsilon > 0$ and arbitrary round ball $B$, we conclude that each $h$ is 1-quasiconformal (with respect to the standard conformal structure on $S^n$). By Liouville's Theorem for quasiconformal maps (see Section 20.4), it follows that $h$ is Moebius.

Thus, the group $G^f = fGf^{-1}$ consists of Moebius transformations. This concludes the proof of Theorem 21.16. □

**Proof of QI rigidity of groups acting geometrically on $\mathbb{H}^{n+1}$**

We now can conclude the proof of Theorem 21.1. Let $G$ be a finitely generated group quasi-isometric to $\mathbb{H}^{n+1}, n \geqslant 2$. Then there exists a quasi-action $G \curvearrowright \mathbb{H}^{n+1}$ and this quasi-action extends to a uq action $G \curvearrowright S^n$. By Lemma 9.91, every point of $S^n$ is a conical limit point for this action. Since the quasi-action $G \curvearrowright \mathbb{H}^{n+1}$ is geometric, the action $G \curvearrowright S^n$ is a *uniform convergence action*, see Theorem 9.107. Furthermore, by Proposition 21.11, there exists a $G$-invariant bounded measurable conformal structure $\mu$ on $S^n$. Note that the action $G \curvearrowright S^n$ need not be faithful, but it has to have finite kernel, which we will ignore from now.

By Theorem 21.16, the action $G \curvearrowright S^n$ is quasiconformally conjugate to a Moebius action $G^f \curvearrowright S^n$. Being a uniform convergence group is a purely topological concept invariant under homeomorphic conjugation. Thus, the group $G^f$ also acts on $S^n$ as a uniform convergence group. Recall that the Moebius group $Mob(S^n)$ is isomorphic to the isometry group $\mathrm{Isom}(\mathbb{H}^{n+1})$ *via* the extension map from hyperbolic space to the boundary sphere, see Corollary 8.17. Therefore, by applying Theorem 9.104, we conclude that the isometric action $G^f \curvearrowright \mathbb{H}^{n+1}$ is again geometric. Therefore, the group $G$ admits a geometric action on $\mathbb{H}^{n+1}$ which finishes the proof of Theorem 21.1. □

## 21.8. QI rigidity for surface groups

Note that the proof of Tukia's theorem fails in the case $n = 1$, i.e., groups quasi-isometric to the hyperbolic plane. However, Theorem 9.107 still implies that $G$ acts on $S^1$ as a uniform convergence group. It was proven as the result of combined efforts of Tukia, Gabai, Casson and Jungreis in 1988—1994 (see [**Tuk88, Gab92, CJ94**]) that every uniform convergence group acting on $S^1$ is a Fuchsian group, i.e., a discrete cocompact subgroup of $\mathrm{Isom}(\mathbb{H}^2)$. Below we outline an alternative argument, which relies, however, on Perelman's proof of Thurston's Geometrization conjecture for 3-dimensional manifolds.

THEOREM 21.17. *If a group $G$ is QI to the hyperbolic plane, then $G$ admits a geometric action on $\mathbb{H}^2$.*

PROOF. Let $\tilde{M} \subset Trip(S^1)$ denote is set of positively oriented ordered triples of distinct points on $S^1$, i.e., points $\xi_1, \xi_2, \xi_3$ on $S^1$ which appear in the counter-clockwise order on the circle. Thus, $\tilde{M}$ is a connected 3-dimensional manifold, open subset of $S^1 \times S^1 \times S^1$.

LEMMA 21.18. *If $g \in G$ fixes point in $\tilde{M}$, then it fixes the entire $\tilde{M}$.*

PROOF. Let $g \in G$ and assume that $g$ fixes three distinct points $\xi_1, \xi_2, \xi_3$ in $S^1$. In particular, $g$ preserves each component of $S^1 \setminus \{\xi_1, \xi_2, \xi_3\}$. These components are arcs $\alpha_i, i = 1, 2, 3$. Since $g$ fixes points $\xi_i$, it also preserves orientation on each $\alpha_i$. Since the action $G \curvearrowright \tilde{M}$ is properly discontinuous, the element $g$ has finite order. We claim that $g$ fixes each $\alpha_i$ pointwise. We identify each $\alpha_i$ with $\mathbb{R}$, then $g : \mathbb{R} \to \mathbb{R}$ is an orientation-preserving homeomorphism of finite order. Pick a point $x \in \mathbb{R}$ not fixed by $g$ and suppose that $y = g(x) > x$. Then, since $g$ preserves orientation, $g(y) > y$; similarly, $g^i(x) > g^{i-1}(x)$ for every $i \in \mathbb{Z}$. Thus, $g$ cannot have finite order. Contradiction. The same argument applies if $y < x$. □

Let, therefore, $\bar{G}$ denote the quotient of $G$ by the (finite) kernel of the action $G \curvearrowright S^1$. Then $\bar{G}$ acts freely on $\tilde{M}$.

LEMMA 21.19. *$\tilde{M}$ is homeomorphic to $\mathbb{H}^2 \times S^1$.*

PROOF. Given an ordered triple $\xi = (\xi_1, \xi_2, \xi_3)$ of distinct points in $S^1$, there exist a unique ideal hyperbolic triangle $T_\xi$ with the ideal vertices $\xi_i$. Let $p_\xi$ denote the center of this triangle, i.e. the center of the inscribed circle. This point can be also defined as the fixed point of the isometry of order 3 in $\mathrm{Isom}(\mathbb{H}^2)$ which cyclically permutes the points $\xi_i$.

Clearly, the map $\xi \to p_\xi$ is continuous as a map $\tilde{M} \to \mathbb{H}^2$. Furthermore, let $\rho_i$ denote the geodesic rays emanating from $p_\xi$ and asymptotic to $\xi_i, i = 1, 2, 3$. Note that the rays $\rho_i$ meet at the angle $2\pi/3$ at $p_\xi$. Thus, the ray $\rho_1$ uniquely determines the rays $\rho_2, \rho_3$ (since the triple $\xi$ is positively oriented). Let $v_\xi$ be the derivative of $\rho_1$ at $p_\xi$. Thus, we obtain a map

$$c : \tilde{M} \to U\mathbb{H}^2,$$

where $U\mathbb{H}^2$ is the unit tangent bundle of $\mathbb{H}^2$. Clearly, this map is continuous. It also has continuous inverse: Given $(p, v) \in U\mathbb{H}^2$, we let $\rho_1$ be the geodesic ray emanating from $p$ with the derivative $v$; from this we construct rays $\rho_2, \rho_3$ and, therefore, the points $\xi_i, i = 1, 2, 3 \in S^1$. Since $\mathbb{H}^2$ is contractible, the unit tangent bundle $U\mathbb{H}^2$ is trivial and, hence, $\tilde{M}$ is homeomorphic to $U\mathbb{H}^2 \cong \mathbb{H}^2 \times S^1$. $\square$

In particular, $\pi_i(\tilde{M}) = 0, i \geqslant 2$, and $\pi_1(\tilde{M}) \cong \mathbb{Z}$. We now consider the quotient $M = \tilde{M}/G$. Since the action $\bar{G} \curvearrowright S^1$ is free, properly discontinuous and cocompact, $M$ is a compact 3-dimensional manifold. Furthermore, $Z = \pi_1(\tilde{M}) < \pi_1(M)$ is a normal infinite cyclic subgroup and

$$\bar{G} \cong \pi_1(M)/Z.$$

Since $\pi_i(\tilde{M}) = 0, i \geqslant 2$, the same holds for $M$. One can then look at the classification of closed 3-dimensional aspherical manifolds $M$ given by Perelman's geometrization theorem. Every such manifold $M$ is obtained by gluing of hyperbolic and *Seifert* manifolds along boundary tori and Klein bottles. If $M$ is hyperbolic, $\pi_1(M)$ is also hyperbolic and, hence, cannot contain a normal infinite cyclic subgroup. If $M$ has nontrivial splitting, then it is *Haken* and, hence, Seifert, see e.g. [**Hem78**]. Now, one considers the classification of Seifert manifolds $M$ (see e.g. [**Hem78**]) and concludes that if $Z < \pi_1(M)$ is an infinite normal cyclic subgroup, then the quotient group $\pi_1(M)/Z$ is either Fuchsian or is virtually isomorphic to $\mathbb{Z}^2$. However, the group $G$ we are interested in is quasi-isomeric to $\mathbb{H}^2$ and, hence, is Gromov-hyperbolic. This eliminates the second possibility. Thus, $\bar{G}$ is a Fuchsian group. Since kernel of $G \to \bar{G}$ is finite, we conclude that $G$ is virtually isomorphic to a Fuchsian group. $\square$

CHAPTER 22

# Quasi-isometries of nonuniform lattices in $\mathbb{H}^n$

## 22.1. Lattices

Recall that a *lattice* in a Lie group $G$ is a discrete subgroup $\Gamma$ such that the quotient $\Gamma \setminus G$ has finite volume. Here, the left-invariant volume form on $G$ is defined by taking a Riemannian metric on $G$ which is left-invariant under $G$ and right-invariant under $K$, the maximal compact subgroup of $G$. Thus if $X := G/K$, then this quotient manifold has a Riemannian metric which is (left) invariant under $G$. Hence, $\Gamma$ is a lattice iff $\Gamma$ acts on $X$ properly discontinuously so that $Vol(\Gamma \setminus X)$ is finite. Note that the action of $\Gamma$ on $X$ need not be free. Recall also that a lattice $\Gamma$ is *uniform* if $\Gamma \setminus X$ is compact and $\Gamma$ is *nonuniform* otherwise.

Each lattice is finitely-generated (this is clear for uniform lattices but is not at all obvious otherwise); in the case of the hyperbolic spaces finite generation follows from the thick-thin decomposition discussed below. Thus, if $\Gamma$ is a lattice in a linear Lie group, then, by Selberg lemma 3.51, $\Gamma$ contains a torsion-free subgroup of finite index. In particular, if $\Gamma$ is a lattice in $PO(n,1)$ (which is isomorphic to the isometry group of the hyperbolic $n$-space) then $\Gamma$ is virtually torsion-free. We also note that a finite-index subgroup in a lattice is again a lattice. Passing to a finite-index subgroup, of course, does not affect uniformity of a lattice.

EXAMPLE 22.1. Consider the group $G = PO(2,1)$ and a non-uniform lattice $\Gamma < G$. After passing to a finite-index subgroup in $\Gamma$, we may assume that $\Gamma$ is torsion-free. Then the quotient $\mathbb{H}^2/\Gamma$ is a non-compact surface with the fundamental group $\Gamma$. Therefore, $\Gamma$ is a free group of finite rank.

EXERCISE 22.2. Show that groups $\Gamma$ in the above example cannot be cyclic.

Recall that a *horoball* in $\mathbb{H}^n$ (in the unit ball model) is a domain bounded by a round Euclidean ball $B \subset \mathbb{H}^n$, whose boundary is tangent to the boundary of $\mathbb{H}^n$ in a single point (called the *center* or *footpoint* of the horoball). The boundary of a horoball in $\mathbb{H}^n$ is called a *horosphere*. In the upper half-space model, the horospheres with the footpoint $\infty$ are horizontal hyperplanes

$$\{(x_1, ..., x_{n-1}, t) : (x_1, ..., x_{n-1}) \in \mathbb{R}^{n-1}\},$$

where $t$ is a positive constant.

LEMMA 22.3. *Suppose that $\Gamma < PO(n,1)$ is a torsion-free discrete group containing a parabolic element $\gamma$. Then $\Gamma$ is a non-uniform lattice.*

PROOF. Recall that every parabolic isometry of $\mathbb{H}^n$ has unique fixed point in the ideal boundary sphere $S^{n-1}$. By conjugating $\Gamma$ by an isometry of $\mathbb{H}^n$, we can assume that $\gamma$ fixes the point $\infty$ in the upper half-space model $\mathbb{R}^n_+$ of $\mathbb{H}^n$. Therefore, $\gamma$ acts on as a Euclidean isometry on $\mathbb{R}^n_+$. After conjugating $\gamma$ by a

Euclidean isometry, $\gamma$ has the form

$$x \mapsto Ax + v,$$

where $v \in \mathbb{R}^{n-1} \setminus \{0\}$ and $A$ is an orthogonal transformation fixing the vector $v$. Hence, $\gamma$ preserves the Euclidean line $L \subset \mathbb{R}^{n-1}$ (spanned by $v$) and the restriction of $\gamma$ to $L$ is the translation $x \mapsto x + v$. Let $H$ denote the hyperbolic plane in $\mathbb{H}^n$, which is the vertical Euclidean half-plane above the line $L$. Again, $\gamma$ acts on $H$ as the translation $x \mapsto x + v$. We introduce the coordinates $(x, y)$ on $H$, where $x \in \mathbb{R}$ and $y > 0$. Then for every $z = (x, y) \in H$,

$$d(z, \gamma z) < \frac{|v|}{y}$$

where $|v|$ is the Euclidean norm of the vector $v$. Let $c_z$ denote the projection of the geodesic $[z, \gamma z]$ to the hyperbolic manifold $M = \mathbb{H}^n/\Gamma$. By sending $y$ to infinity, we conclude that the (nontrivial) free homotopy class $[\gamma]$ in $M = \mathbb{H}^n/\Gamma$ represented by $\gamma \in \Gamma$, contains loops $c_z$ of arbitrarily short length. This is impossible if $M$ were a compact Riemannian manifold. □

The converse to the above lemma is much less trivial and follows from

THEOREM 22.4 (Thick-thin decomposition). *Suppose that $\Gamma$ is a nonuniform lattice in $Isom(\mathbb{H}^n)$. Then there exists an (infinite) collection $C$ of open horoballs $C := \{B_j, j \in J\}$, with pairwise disjoint closures, so that*

$$\Omega := \mathbb{H}^n \setminus \bigcup_{j \in J} B_j$$

*is $\Gamma$-invariant and $M_c := \Omega/\Gamma$ is compact. Furthermore, every parabolic element $\gamma \in \Gamma$ preserves (exactly) one of the horoballs $B_j$.*

The proof of this theorem is based on a mild generalization of the Zassenhaus theorem due to Kazhdan and Margulis, see e.g. [**BP92**], [**Kap01**], [**Rat94**], [**Thu97**].

The quotient $M_c$ is called the *thick part* of $M = \mathbb{H}^n/\Gamma$ and its (noncompact) complement in $M$ is called the *thin part* of $M$. If $\Gamma$ is torsion-free, then it acts freely on $\mathbb{H}^n$ and $M$ has natural structure of a hyperbolic manifold of finite volume. If $\Gamma$ is not torsion-free, then $M$ is a *hyperbolic orbifold*. Clearly, when $\Gamma < PO(n, 1)$ is a lattice, the quotient $M = \mathbb{H}^n/\Gamma$ is compact if and only if $C = \emptyset$.

The set $\Omega$ is called a *truncated hyperbolic space*. The boundary horospheres of $\Omega$ are called *peripheral horospheres*. Since each closed horoballs used to define $\Omega$ are pairwise disjoint, $\Omega$ is contractible. In particular, if $\Gamma$ is torsion-free, then it has finite type. In general, $\Gamma$ is of type $\mathbf{F}_\infty$.

Note that the stabilizer $\Gamma_j$ of each horosphere $\partial B_j$ acts on this horosphere cocompactly with the quotient $T_j := \partial B_j/\Gamma_j$. The quotient $B_j/\Gamma_j$ is naturally homeomorphic to $T_j \times \mathbb{R}_+$, this product decomposition is inherited from the foliation of $B_j$ by the horospheres with the common footpoint $\xi_j$ and the geodesic rays asymptotic to $\xi_j$. If $\Gamma$ is torsion-free, orientation preserving and $n = 3$, the quotients $T_j$ are 2-tori.

Observe that a hyperbolic horoball cannot be stabilized by a hyperbolic isometry. Indeed, by working with the upper half-space model of $\mathbb{H}^n$, we can assume that the (open) horoball in question is given by

$$B = \{(x_1, \ldots, x_n) : x_n > 1\}.$$

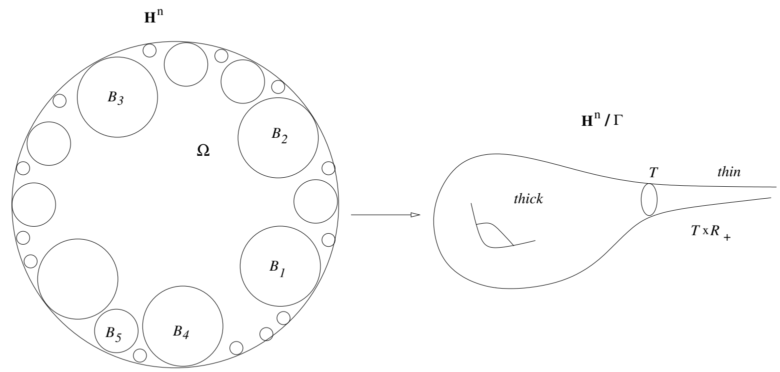



FIGURE 22.1. Truncated hyperbolic space and thick-thin decomposition.

Every hyperbolic isometry $\gamma$ stabilizing $B$ would have to fix $\infty$ and act and a Euclidean isometry on the boundary horosphere of $B$. Thus, $\gamma$ is either elliptic or parabolic. In particular, stabilizers of the horoballs $B_j$ in Theorem 22.4 contain no hyperbolic elements. Since we can assume that $\Gamma$ is torsion-free, we obtain

COROLLARY 22.5. *A lattice in $PO(n,1)$ is uniform if and only if it does not contain parabolic elements.*

*Arithmetic groups* provide a general source for lattices in Lie groups. Recall that two subgroups $\Gamma_1, \Gamma_2$ of a group $G$ are called *commensurable* if $\Gamma_1 \cap \Gamma_2$ has finite index in $\Gamma_1, \Gamma_2$. Let $G$ be a Lie group with finitely many components.

DEFINITION 22.6. An *arithmetic subgroup* in $G$ is a subgroup of $G$ commensurable to the subgroup of the form $\Gamma := \phi^{-1}(GL(N,\mathbb{Z}))$ for a (continuous) homomorphism $\phi : G \to GL(N,\mathbb{R})$ with compact kernel.

It is clear that every arithmetic subgroup is discrete in $G$. It is a much deeper theorem that every arithmetic subgroup is a lattice in a Lie subgroup $H \leqslant G$, see e.g. **[Mar91, Rag72]**.

**Bianchi groups.** We now describe a concrete class of non-uniform arithmetic lattices in the isometry group of hyperbolic 3-space, called *Bianchi groups*. Let $D$ denote a *square-free negative integer*, i.e., an integer which is not divisible by the square of a prime number. Consider the *imaginary quadratic field*

$$\mathbb{Q}(\sqrt{D}) = \{a + \sqrt{D}b : a, b \in \mathbb{Q}\}$$

in $\mathbb{C}$. Set

$$\omega := \sqrt{D}, \text{ if } D \equiv 2, 3, \mod 4$$

$$\omega := \frac{1+\sqrt{D}}{2}, \text{ if } D \equiv 1, \mod 4$$

Then the ring of integers of $\mathbb{Q}(\sqrt{D})$ is

$$O_D = \{a + \omega b : a, b \in \mathbb{Z}\}.$$

For instance, if $D = -1$, then $O_D$ is the ring of *Gaussian integers*

$$\{a + ib : a, b \in \mathbb{Z}\}.$$

A *Bianchi group* is the group of the form

$$SL(2, O_D) < SL(2, \mathbb{C})$$

for some $D$. Since the ring $O_D$ is discrete in $\mathbb{C}$, it is immediate the every Bianchi subgroup is discrete in $SL(2, \mathbb{C})$. By abusing terminology, one also refers to the group $PSL(2, O_D)$ as a Bianchi subgroup of $PSL(2, \mathbb{C})$.

Bianchi groups $\Gamma$ are *arithmetic lattices* in $SL(2, \mathbb{C})$; in particular, quotients $\mathbb{H}^3/\Gamma$ has finite volume. Furthermore, every arithmetic lattice in $SL(2, \mathbb{C})$ is *commensurable* to a Bianchi group. We refer the reader to **[MR03]** for the detailed discussion of these and other facts about Bianchi groups.

**Commensurators of lattices.**

Recall (see §3.4) that the commensurator of a subgroup $\Gamma$ in a group $G$ is the subgroup $Comm_G(\Gamma) < G$ consisting of elements $g \in G$ such that the groups $g\Gamma g^{-1}$ and $\Gamma$ are commensurable, i.e. $|\Gamma : g\Gamma g^{-1} \cap \Gamma| < \infty$, $|g\Gamma g^{-1} : g\Gamma g^{-1} \cap \Gamma| < \infty$.

Below we consider commensurators in the situation when $\Gamma$ is a lattice in a Lie group $G$.

EXERCISE 22.7. Let $\Gamma := SL(2, O_D) \subset G := SL(2, \mathbb{C})$ be a Bianchi group.

1. Show that $Comm_G(\Gamma) \subset SL(2, \mathbb{Q}(\omega))$. In particular, $Comm_G(\Gamma)$ is dense in $G$.

2. Show that the set of fixed points of parabolic elements in $\Gamma$ (in the upper half-space model of $\mathbb{H}^3$) is

$$\mathbb{Q}(\omega) \cup \{\infty\}.$$

3. Show that $Comm_G(\Gamma) = SL(2, \mathbb{Q}(\omega))$.

G. Margulis proved (see **[Mar91]**, Chapter IX, Theorem B and Lemma 2.7; see also **[Zim84]**, Theorem 6.2.5) that a lattice in a semisimple real Lie group $G$ is *arithmetic* if and only if its commensurator is dense in $G$.

Consider now the case when $G$ is either a Lie group or a finitely-generated group and $\Gamma \leqslant G$ is a finitely-generated subgroup. We note that each element $g \in Comm_G(\Gamma)$ determines a quasi-isometry $f : \Gamma \to \Gamma$. Indeed, the Hausdorff distance between $\Gamma$ and $g\Gamma g^{-1}$ is finite. Hence the quasi-isometry $f$ is given by composing $g : \Gamma \to g\Gamma g^{-1}$ with the nearest-point projection to $\Gamma$.

The main goal of the remainder of the chapter is to prove the following

THEOREM 22.8 (R. Schwartz **[Sch96b]**). *Let $\Gamma \subset G = Isom(\mathbb{H}^n)$ be a nonuniform lattice, $n \geqslant 3$. Then:*

*(a) For each quasi-isometry $f : \Gamma \to \Gamma$ there exists $\gamma \in Comm_G(\Gamma)$ which is within finite distance from $f$. The distance between these maps depends only on $\Gamma$ and on the quasi-isometry constants of $f$.*

*(b) Suppose that $\Gamma, \Gamma'$ are non-uniform lattices which are quasi-isometric to each other. Then there exists an isometry $g \in Isom(\mathbb{H}^n)$ such that the groups $\Gamma'$ and $g\Gamma g^{-1}$ are commensurable.*

*(c) Suppose that $\Gamma'$ is a finitely-generated group which is quasi-isometric to a nonuniform lattice $\Gamma$ above. Then the groups $\Gamma, \Gamma'$ are virtually isomorphic*

Our proof will mostly follow **[Sch96b]**.

Note that this theorem fails in the case of the hyperbolic plane (except for the last part). Indeed, every free group $F_r$ of rank $\geqslant 2$ can be realized as a non-uniform lattice $\Gamma$ acting on $\mathbb{H}^2$. In view of thick-thin decomposition of the hyperbolic surface $M = \mathbb{H}^2/\Gamma$, $\Gamma$ contains only finitely many $\Gamma$-conjugacy classes of maximal parabolic subgroups: Every such class corresponds to a component of $M \setminus M_c$. Suppose now that $r \geqslant 3$. Then there are *atoroidal* automorphisms $\phi$ of $F_r$, so that for every nontrivial cyclic subgroup $C \subset F_n$ and every $m$, $\phi^m(C)$ is not conjugate to $C$, see e.g. [**BFH97**]. Therefore, such $\phi$ cannot send parabolic subgroups of $\Gamma$ to parabolic subgroups of $\Gamma$. Hence, the quasi-isometry of $F_n$ given by $\phi$ cannot extend to a quasi-isometry $\mathbb{H}^2 \to \mathbb{H}^2$. It follows that (a) fails for $n = 2$. Similarly, one can show that (b) fails, since commensurability preserves arithmeticity and there are both arithmetic and non-arithmetic lattices in $Isom(\mathbb{H}^2)$. All these lattices are virtually free, hence, virtually isomorphic.

## 22.2. Coarse topology of truncated hyperbolic spaces

On each truncated hyperbolic space $\Omega$ we put the path-metric $d$ which is induced by the restriction of the Riemannian metric of $\mathbb{H}^n$ to $\Omega$. This metric is invariant under $\Gamma$ and, since the quotient $\Omega/\Gamma$ is compact, $(\Omega, d)$ is quasi-isometric to the group $\Gamma$. We will use the notation dist for the hyperbolic distance function in $\mathbb{H}^n$.

LEMMA 22.9. *The identity map* $\iota : (\Omega, d) \to (\Omega, \mathrm{dist})$ *is 1-Lipschitz and uniformly proper.*

PROOF. If $p$ is a path in $\Omega$ then $p$ has the same length with respect to the metrics $d$ and dist. This immediately implies that $\iota$ is 1-Lipschitz. Uniform properness follows from the fact that the group $\Gamma$ acts geometrically on both $(\Omega, d), (\Omega, \mathrm{dist})$ and that the map $\iota$ is $\Gamma$-equivariant, see Lemma 5.34. □

LEMMA 22.10. *The restriction of $d$ to each peripheral horosphere $\Sigma \subset \partial\Omega$ is a flat metric.*

PROOF. Without loss of generality, we may assume that (in the upper half-space model of $\mathbb{H}^n$), $\Sigma = \{(x_1, \ldots, x_n) : x_n = 1\}$. Hence

$$\Omega \subset \{(x_1, \ldots, x_n) : 0 < x_n \leqslant 1\}.$$

The hyperbolic Riemannian metric restricted to $\Sigma$ equals the flat metric on $\Sigma$. Therefore, it is enought to show that for every path $p$ in $\Omega$ there exists a path $q$ in $\Sigma$ so that $\mathrm{length}(q) \leqslant \mathrm{length}(p)$ and the end-points of $p$ and $q$ are the same. Consider the vertical projection

$$\pi : \Omega \to \Sigma, \quad \pi(x_1, \ldots, x_{n-1}, x_n) = (x_1, \ldots, x_{n-1}, 1).$$

We leave it to the reader to check that $\|d\pi\| \leqslant 1$ (with respect to the hyperbolic metric). Therefore, for every path $p$ as above,

$$\mathrm{length}(p) \leqslant \mathrm{length}(q), \quad q = \pi \circ p. \quad \square$$

LEMMA 22.11. *For every horoball $B \subset \mathbb{H}^n$ the $R$-neighborhood $\mathcal{N}_R(B)$ of $B$ in $\mathbb{H}^n$ is also a horoball $B'$.*

PROOF. We again work in the upper half-space model so that

$$B = \{(x_1, \ldots, x_n) : x_n > 1\}$$

and $\Sigma$ is the boundary of $B$. Let $\pi : \mathbb{H}^n \setminus B \to \Sigma$ denote the vertical projection as in the proof of the previous lemma. We leave it to the reader to check that

$$\mathrm{dist}(x, \Sigma) = \mathrm{dist}(x, \pi(x)).$$

It follows, in view of Exercise 8.10, that $\mathcal{N}_R(B)$ is the horoball given by

$$\{(x_1, \ldots, x_n) : x_n > e^{-R}\}. \quad \square$$

We refer the reader to Section 6.6 for the notion of *coarse separation* used below. The following lemma is the key for distinguishing the case of the hyperbolic plane from the higher-dimensional hyperbolic spaces (of dimension $\geqslant 3$):

LEMMA 22.12. *Let $\Omega$ is a truncated hyperbolic space of dimension $\geqslant 3$. Then each peripheral horosphere $\Sigma \subset \Omega$ does not coarsely separate $\Omega$.*

PROOF. Let $R < \infty$ and let $B$ be the horoball bounded by $\Sigma$. Then the union of $\mathcal{N}_R(\Sigma) \cup B$ is a horoball $B'$ in $\mathbb{H}^n$ (where the metric neighborhood is taken in $\mathbb{H}^n$). We claim that $B'$ does not separate $\Omega$. Indeed, the horoball $B'$ does not separate $\mathbb{H}^n$. Therefore, for each pair of points $x, y \in \Omega \setminus B'$, there exists a piecewise-geodesic path $p$ connecting them within $\mathbb{H}^n \setminus B'$. If the path $p$ is entirely contained in $\Omega$, we are done. Otherwise, it can be subdivided into finitely many subpaths, each of which is either contained in $\Omega$ or connects a pair of points on the boundary of a complementary horoball $B_j \subset \mathbb{H}^n \setminus \Sigma$.

EXERCISE 22.13. The intersection of $B'$ with $\Sigma_j = \partial B_j$ is isometric to a metric ball in the Euclidean space $(\Sigma_j, d)$. Hint: Use the upper half-space model so that $B_j$ is given by

$$\{(x_1, \ldots, x_n) : x_n > 1\}.$$

Then $B' \cap \Sigma_j$ is the intersection of a Euclidean hyperplane with a Euclidean metric ball.

Note that a metric ball cannot not separate $\mathbb{R}^{n-1}$, provided that $n - 1 \geqslant 2$. Thus we can replace $p_j = p \cap B_j$ with a new path $p'_j$ which connects the end-points of $p_j$ within the complement $\Sigma_j \setminus B'$. By making these replacements for each $j$ we get a path $q$ connecting $x$ to $y$ within $\Omega \setminus B'$. Therefore, $B'$ does not separate $\Omega$.

We are now ready to show that $\Sigma$ cannot coarsely separate $(\Omega, d)$. Suppose to the contrary that for some $R$, $Y := \Omega \setminus \mathcal{N}_R^{(\Omega)}(B)$ contains at least two deep components $C_1, C_2$. Let $x_i \in C_i, i = 1, 2$. By the definition of a deep component of $Y$, there are continuous proper paths $\alpha_i : \mathbb{R}_+ \to C_i, \alpha_i(0) = x_i$, $i = 1, 2$. Thus,

$$\lim_{t \to \infty} \operatorname{dist}(\alpha_i(t), \Sigma) = \infty.$$

By Lemma 22.9, there exists $T$, so that $y_i := \alpha_i(T) \notin B'$, $i = 1, 2$. Therefore, as we proved above, we can connect $y_1$ to $y_2$ by a path in $\Omega \setminus B' \subset Y$. Therefore, $C_1 = C_2$. Contradiction. $\square$

EXERCISE 22.14. Show that Lemma 22.12 fails for $n = 2$. Hint: Note that Cayley graph of a free nonabelian group of finite rank has infinitely many ends.

Now, suppose that $\Omega, \Omega'$ are truncated hyperbolic spaces for lattices $\Gamma, \Gamma' < \operatorname{Isom}(\mathbb{H}^n)$, and $f : \Omega \to \Omega'$ is an $(L, A)$–quasi-isometry. Let $\Sigma$ be a peripheral horosphere of $\Omega$.

PROPOSITION 22.15. *There exists a peripheral horosphere $\Sigma' \subset \partial\Omega'$ which is within finite Hausdorff distance from $f(\Sigma)$.*

Proof. Since $\Omega'/\Gamma'$ is compact, there exists $D < \infty$ so that for every $x \in \Omega'$,

$$\mathrm{dist}(x, \partial\Omega') \leqslant D.$$

Note that $\Omega$, being isometric to $\mathbb{R}^{n-1}$, has bounded geometry and is uniformly contractible. Therefore, according to Theorem 6.55, $f(\Sigma)$ coarsely separates $\mathbb{H}^n$; however it cannot coarsely separate $\Omega'$, since $f$ is a quasi-isometry and $\Sigma$ does not coarsely separate $\Omega$. Let $r < \infty$ be such that $\mathcal{N}_r(f(\Sigma))$ separates $\mathbb{H}^n$ into (two) deep components $X_1, X_2$. Suppose that for each complementary horoball $B'_j$ of $\Omega'$ (bounded by the horosphere $\Sigma'_j$),

$$\mathcal{N}_{-r}(B'_j) := B'_j \setminus \mathcal{N}_r(\Sigma'_j) \subset X_1.$$

Then, for every $x \in \Omega'$,

$$\mathrm{dist}(x, X_1) \leqslant r + D.$$

It follows that for every $x \in \mathbb{H}^n$,

$$\mathrm{dist}(x, X_1) \leqslant 2r + D,$$

which means that the component $X_2$ cannot be deep, a contradiction.

Thus, there are complementary horoballs $B'_1, B'_2$ for $\Omega'$ such that

$$\mathcal{N}_{-r}(B'_1) \subset X_1, \quad \mathcal{N}_{-r}(B'_2) \subset X_2.$$

Set $\Sigma_i := \partial B'_i$. If both intersections

$$\Sigma_i \cap X_i, i = 1, 2,$$

are unbounded, then $f(\Sigma)$ coarsely separates $\Omega'$, which is again a contradiction. Therefore, say, for $i = 1$, there exists $r' < \infty$ so that $\Sigma' := \Sigma'_1$ satisfies

$$\Sigma' \subset \mathcal{N}_{r'}(f(\Sigma)).$$

Our goal is to show that $f(\Sigma) \subset \mathcal{N}_\rho(\Sigma')$ for some $\rho < \infty$. The nearest-point projection $\Sigma' \to f(\Sigma)$ defines a quasi-isometric embedding $h : \Sigma' \to \Sigma$. However, Lemma 7.71 proves that a quasi-isometric embedding between two Euclidean spaces of the same dimension is a quasi-isometry. Thus, there exists $\rho < \infty$ such that $f(\Sigma) \subset \mathcal{N}_\rho(\Sigma')$. □

Exercise 22.16. Show that the horosphere $\Sigma'$ in Proposition 22.15 is unique.

We now improve Proposition 22.15 to get a uniform control on distance from $f(\Sigma)$ to a boundary horosphere of $\Omega'$.

Lemma 22.17. *In the above proposition,* $\mathrm{dist}_{Haus}(f(\Sigma), \Sigma')) \leqslant R = R(L, A)$, *where $R$ is independent of* $\Sigma$.

Proof. The proof is by inspection of the arguments in the proof of Proposition 22.15. First of all, the constant $r$ depends only on the quasi-isometry constants of the mapping $f$ and the uniform geometry/uniform contractibility bounds for $\mathbb{R}^{n-1}$ and $\mathbb{H}^n$. The inradii of the shallow complementary components of $\mathcal{N}_r(f(\Sigma))$ again depend only on the above data. Similarly, inradii of (necessarily shallow) components of $\Omega' \setminus f(\Sigma)$ also depend only on quasi-isometry constants of $f$. Therefore, there exists a uniform constant $r'$ such that $\Sigma_1$ or $\Sigma_2$ is contained in $\mathcal{N}_{r'}(f(\Sigma))$. Finally, the upper bound on $\rho$ such that $\mathcal{N}_\rho(Image(h)) = \Sigma'$ (coming from Lemma 7.71) again depends only on the quasi-isometry constants of the projection $h : \Sigma' \to \Sigma$. □

REMARK 22.18. Proposition 22.15 and Lemma 22.15 together form the *Quasi-Flat Lemma* from [**Sch96b**], §3.2 . This lemma can be seen as a version of the Morse Lemma 9.38 for truncated hyperbolic spaces. These spaces are, in fact, relatively hyperbolic in the strong sense. See Chapter 9.21 for further details.

### 22.3. Hyperbolic extension

Let $\Omega, \Omega' \subset \mathbb{H}^n$ be truncated hyperbolic spaces ($n \geqslant 3$) for lattices $\Gamma, \Gamma' <$ Isom($\mathbb{H}^n$), let $C, C'$ denote the sets whose elements are peripheral horospheres of $\Omega, \Omega'$ respectively. Suppose that $f : \Omega \to \Omega'$ is a quasi-isometry. The main result of this section is

THEOREM 22.19 (Horoball QI extension theorem). *The quasi-isometry $f$ admits a quasi-isometric extension $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$. Moreover, the extension $\tilde{f}$ satisfies the following equivariance property:*

*Suppose that $f : X \to X'$ is equivariant with respect to an isomorphism*

$$\rho : \Gamma \to \Gamma'.$$

*Then the extension $\tilde{f}$ is also $\rho$–equivariant.*

PROOF. By Proposition 22.15 and and Lemma 22.15, for every peripheral horosphere $\Sigma \subset \Omega$ there exists a peripheral horosphere $\Sigma'$ of $\Omega'$ so that $\text{dist}(f(\Sigma), \Sigma') \leqslant R < \infty$, where $R$ depends only on quasi-isometry constants of $f$. By uniqueness of the horosphere $\Sigma'$, the map

$$\theta : C \to C', \theta : \Sigma \mapsto \Sigma'$$

is $\rho$-equivariant, provided that $f$ was $\rho$-equivariant.

We first alter $f$ on $\partial\Omega$ by postcomposing $f|\Sigma$ with the nearest-point projection to $\Sigma'$ for every $\Sigma \in C$. The new map is again a quasi-isometry; this modification clearly perserves $\rho$-equivariance. We retain the notation $f$ for the new quasi-isometry, which now satisfies

$$f(\Sigma) \subset \Sigma', \forall \Sigma \in C.$$

We will construct the extension $\tilde{f}$ to each complementary horoball $B \subset \mathbb{H}^n \setminus \Omega$. We will use the upper half-space model of $\mathbb{H}^n$ so that the horoballs $B$ and $B'$ bounded by $\Sigma, \Sigma'$ are both given by

$$\{(x_1, ..., x_{n-1}, 1) : (x_1, ..., x_{n-1}) \in \mathbb{R}^{n-1}\}.$$

For each vertical (unit speed) geodesic ray $\rho(t), t \in \mathbb{R}_+$, in $B$ we define the (unit speed) geodesic ray $\rho'(t)$ in $B'$ to be the vertical geodesic ray in $B'$ with the initial point $f(\rho(0))$. This gives the extension of $f$ into $B$:

$$\tilde{f}(\rho(t)) = \rho'(t).$$

We will now verify that this extension is coarsely Lipschitz. It suffices to consider points $x, y \in B$ within unit distance from each other. If $x, y \in B$ belong to the same vertical ray $\rho$, then, clearly, $d(f(x), f(y)) = d(x, y)$. Therefore, by the triangle inequality, the problem reduces to estimation of $d(f(x), f(y))$ for $x, y$ such that $x_n = y_n = t$, i.e., $x, y$ belong to the same horosphere $H_t$ with the footpoint $\infty$. It follows from Exercise 8.36 that the Euclidean distance $|x - y|$ between $x$ and $y$ is at most

$$\sqrt{2(e-1)}t.$$

Therefore, the distance between $x$ and $y$ along the horosphere $H_t$ is at most $\epsilon := \sqrt{2(e-1)}$. Let $\bar{x}, \bar{y} \in \Sigma$ denote the vertical projections of the points $x, y$ respectively. Then the distance $\mathrm{dist}_\Sigma$ along the horosphere $\Sigma$ is estimated as

$$\mathrm{dist}_\Sigma(\bar{x}, \bar{y}) = t\mathrm{dist}_{H_t}(x, y) \leqslant \epsilon t.$$

Since $f$ is $(L, A)$–coarse Lipschitz,

$$\mathrm{dist}_\Sigma(f(\bar{x}), f(\bar{y})) \leqslant tL\epsilon + A.$$

It follows that

$$d(\tilde{f}(x), \tilde{f}(y)) \leqslant \mathrm{dist}_{H_t}(\tilde{f}(x), \tilde{f}(y)) \leqslant L\epsilon + \frac{A}{t} \leqslant L\epsilon + A.$$

This proves that the extension $\tilde{f}$ is coarse Lipschitz in the horoball $B$. Since being coarse Lipschitz is a local property, the mapping $\tilde{f}$ is coarse Lipschitz on $\mathbb{H}^n$. The same argument applies to the hyperbolic extension $\tilde{f}'$ of the coarse inverse $f'$ to the mapping $f$. It is clear that the mapping $\tilde{f} \circ \tilde{f}'$ and $\tilde{f}' \circ \tilde{f}$ have bounded displacement. Thus, $\tilde{f}$ is a quasi-isometry. The equivariance property of $f$ is clear from the construction. □

Since $\tilde{f}$ is a quasi-isometry of $\mathbb{H}^n$, it admits a quasiconformal extension $h : \partial_\infty \mathbb{H}^n \to \partial_\infty \mathbb{H}^n$ (see Theorems 9.83 and 20.24). By continuity of the quasiconformal extension, if $f$ were $\rho$-equivariant, so is $h$. Let $\Lambda, \Lambda'$ denote the sets of the footpoints of the peripheral horospheres of $\Omega, \Omega'$ respectively. Since $f(\Sigma) = \Sigma'$ for every peripheral horosphere of $\Omega$, continuity of the extension also implies that $h(\Lambda) = \Lambda'$.

## 22.4. Zooming in

In the previous section we constructed a quasi-isometry $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ and a quasiconformal mapping $h : \partial_\infty \mathbb{H}^n \to \partial_\infty \mathbb{H}^n$ which is the boundary extension of $\tilde{f}$. Our main goal is to show that $h$ is Moebius. By the Liouville's theorem for quasiconformal mappings (Theorem 20.20), $h$ is Moebius if $h$ is 1-quasiconformal, i.e., for a.e. point $\xi \in S^{n-1}$, the derivative of $h$ at $\xi$ belongs to the group of *similarities* $\mathbb{R} \cdot O(n-1)$.

We will continue to work with the upper half-space of the hyperbolic space $\mathbb{H}^n$.

Proposition 22.20. *Suppose that $h$ is not Moebius. Then there exists a quasi-isometry $F : \Omega \to \Omega'$ whose extension to the sphere at infinity is a linear map which is not a similarity.*

Proof. Since $h$ is differentiable a.e. and is not Moebius, there exists a point $\xi \in S^{n-1} \setminus \Lambda$ such that $Dh(\xi)$ exists, is invertible but is not a similarity. By pre- and post-composing $f$ with isometries of $\mathbb{H}^n$ we can assume that $\xi = 0 = h(\xi)$. Let $L \subset \mathbb{H}^n$ denote the vertical geodesic through $\xi$. Since $\xi$ is not a footpoint of a complementary horoball to $\Omega$, there exists a sequence of points $z_j \in L \cap \Omega$ which converges to $\xi$. For each $t > 0$ define $\alpha_t : z \mapsto t \cdot z$, a hyperbolic translation along $L$. Let $t_j$ be such that $\alpha_{t_j}(z_1) = z_j$. Set

$$\tilde{f}_j := \alpha_{t_j}^{-1} \circ \tilde{f} \circ \alpha_{t_j};$$

the quasiconformal extensions of these mappings to $\partial_\infty \mathbb{H}^n$ are given by

$$h_j(z) = \frac{h(t_j z)}{t_j}.$$

By the definition of differentiability,

$$\lim_{j\to\infty} h_j = A = Dh(0),$$

where convergence is uniform on compact sets in $\mathbb{R}^{n-1}$.

Let us verify that the sequence of quasi-isometries $\tilde{f}_j$ coarsely subconverges to a quasi-isometry of $\mathbb{H}^n$. Indeed, since the quasi-isometry constants of all $\tilde{f}_j$ are all the same, in view of coarse Arzela-Ascoli theorem (Theorem 5.26), it suffices to show that $\{\tilde{f}_j(z_1)\}$ is a bounded sequence in $\mathbb{H}^n$. Let $L_1, L_2$ denote a pair of distinct geodesics in $\mathbb{H}^n$ through $z_1$, so that the point $\infty$ does not belong to $L_1 \cup L_2$. Then, by Morse Lemma 9.38, for a certain uniform constant $c$, the quasi-geodesics $\tilde{f}_j(L_i)$ are within distance $\leqslant c$ from geodesics $L^*_{1j}, L^*_{2j}$ in $\mathbb{H}^n$. Note that the geodesics $L^*_{1j}, L^*_{2j}$ subconverge to geodesics in $\mathbb{H}^n$ with distinct end-points (since the mapping $A$ is 1-1). The point $\tilde{f}_j(z_1)$ is within distance $\le c$ from $L^*_{1j}, L^*_{2j}$. If the sequence $\tilde{f}_j(z_1)$ were unbounded, we would get that $L^*_{1j}, L^*_{2j}$ subconverge to geodesics with a common end-point at infinity. Contradiction.

We thus pass to a subsequence such that $(\tilde{f}_j)$ coarsely converges to a quasi-isometry $f_\infty : \mathbb{H}^n \to \mathbb{H}^n$. Note, however, that $f_\infty$ need not send $\Omega$ to $\Omega'$. Recall that quotients $\Omega/\Gamma, \Omega'/\Gamma'$ are compact. Therefore, there exist sequences $\gamma_j \in \Gamma, \gamma'_j \in \Gamma'$ such that $\gamma_j(z_j)$, $\gamma'_j(\tilde{f}(z_j))$ belong to a compact subset of $\mathbb{H}^n$. Hence, the sequences $\beta_j := \alpha_{t_j}^{-1} \circ \gamma_j^{-1}, \beta'_j := \alpha_{t_j}^{-1} \circ \gamma_j'^{-1}$ is precompact in $Isom(\mathbb{H}^n)$ and, therefore, they subconverge to isometries $\beta_\infty, \beta'_\infty \in \mathrm{Isom}(\mathbb{H}^n)$. Set

$$\Omega_j := \alpha_{t_j}^{-1}\Omega = \alpha_{t_j}^{-1} \circ \gamma_j^{-1}\Omega = \beta_j\Omega,$$

and

$$\Omega'_j := \alpha_{t_j}^{-1}\Omega' = \beta'_j\Omega',$$

then $\tilde{f}_j : \Omega_j \to \Omega'_j$.

On the other hand, the sequences of sets $(\Omega_j), (\Omega'_j)$ subconverge (in the Chabauty topology, see §1.4) to the sets $\beta_\infty\Omega, \beta'_\infty\Omega'$; the limiting map $\tilde{f}_\infty$ is a quasi-isometry between $\beta_\infty\Omega$ and $\beta'_\infty\Omega'$. Since $\beta_\infty\Omega$ and $\beta'_\infty\Omega'$ are isometric copies of $\Omega$ and $\Omega'$, the assertion follows. □

The situation when we have a linear mapping (which is not a similarity) sending $\Lambda$ to $\Lambda'$ seems, at the first glance, impossible. Here, however, is an example:

EXAMPLE 22.21. Let $\Gamma := SL(2, \mathbb{Z}[i]), \Gamma' := SL(2, \mathbb{Z}[\sqrt{-2}])$. Then

$$\Lambda = \mathbb{Q}(i) \cup \{\infty\}, \quad \Lambda' = \mathbb{Q}(\sqrt{-2}) \cup \{\infty\}.$$

Take the real linear mapping $A : \mathbb{C} \to \mathbb{C}$ by sending 1 to 1 and $i$ to $\sqrt{-2}$. Then $A$ is invertible, is not a similarity, but $A(\Lambda) = \Lambda'$.

Thus, in order to get a contradiction, we have to exploit the fact that the linear map in question is the quasiconformal extension of an quasi–isometry between truncated hyperbolic spaces. We will show (Theorem 22.27) the following:

For every peripheral horosphere $\Sigma \subset \partial\Omega$ whose footpoint is not $\infty$, there exists a family of peripheral horospheres $\Sigma_k \subset \partial\Omega$ so that:

$$\mathrm{dist}(\Sigma, \Sigma_k) \leqslant Const, \quad \lim_{k\to\infty} \mathrm{dist}(\Sigma', \Sigma'_k) = \infty.$$

(Recall that $\mathrm{dist}(\cdot,\cdot)$ denotes the minimal distance between the horospheres and $\theta(\Sigma) = \Sigma', \theta(\Sigma_k) = \Sigma'_k$, $k \in \mathbb{N}$.)

Of course, this means that $\tilde{f}$ cannot be coarse Lipschitz. We will prove the above statement by conjugating $\tilde{f}$ by an inversion which interchanges the horospheres with the footpoint at $\infty$ and the horospheres $\Sigma, \Sigma'$ above. This will amount to replacing the affine map $A$ with above an *inverted linear map*, such maps are defined and analyzed in the next section.

## 22.5. Inverted linear mappings

Let $A : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be an (invertible) linear mapping. Recall that the be the inversion $J$ in the unit sphere about the origin is given by the formula

$$J(x) = \frac{x}{|x|^2}.$$

Definition 22.22. An *inverted linear map* is the conjugate of $A$ by the inversion $J$, i.e., the composition

$$h := J \circ A \circ J.$$

Equivalently,

$$h(x) = \frac{|x|^2}{|Ax|^2} A(x).$$

Lemma 22.23. *The function* $\phi(x) = \frac{|x|^2}{|Ax|^2}$ *is* asymptotically constant, *in the sense that*

$$|\nabla\phi(x)| = O(|x|^{-1}), \quad \|Hess\,(\phi(x))\| = O(|x|^{-2})$$

*as* $|x| \to \infty$.

Proof. The function $\phi$ is a rational vector-function of degree zero, hence, its gradient is a rational vector-function of degree $-1$, while every component of its Hessian is a rational function of degree $-2$. $\square$

Note, however, that $\phi$ is not a constant mapping unless $A$ is a similarity. Hence, $h$ is linear if and only if $A$ is a similarity.

Corollary 22.24. *Fix a fixed positive real number* $R$*, and let* $(v_k)$ *be a sequence diverging to infinity in* $\mathbb{R}^{n-1}$*. Then the sequence of maps*

$$h_k(x) := h(x + v_k) - h(v_k)$$

*subconverges (uniformly on the* $R$*-ball* $B = B(0,R) \subset \mathbb{R}^n$*) to an affine map, as* $k \to \infty$*.*

Proof. We have:

$$h(x + v_k) - h(v_k) = \phi(x + v_k)A(x + v_k) - \phi(v_k)A(v_k) =$$

$$\phi(x + v_k)A(x) - (\phi(x + v_k) - \phi(v_k))\, A(v_k).$$

Since $\phi(y)$ is asymptotically constant, $\lim_{k\to\infty} \phi(x + v_k)A(x) = c \cdot A(x)$ for some constant $c$ (uniformly on $B(0,R)$). Since $(\phi(x + v_k) - \phi(v_k)) = O(|v_k|^{-1})$ (as $k \to \infty$), the sequence of vectors

$$(\phi(x + v_k) - \phi(x_k))A(x_k)$$

is uniformly bounded for $x \in B(0,R)$. Furthermore, for every pair of induces $1 \leqslant i,j \leqslant n-1$

$$\frac{\partial^2}{\partial x_i \partial x_j}\left(\phi(x+v_k)-\phi(v_k)\right)A(v_k) = \frac{\partial^2}{\partial x_i \partial x_j}\phi(x+v_k)\cdot A(v_k) =$$

$$O(|v_k|^{-2})A(v_k) = O(|v_k|^{-1}).$$

Therefore, Hessians of $h_k|_B$ uniformly converge to zero as $k \to \infty$. ☐

We would like to strengthen the assertion that $\phi$ is not constant (unless $A$ is a similarity). Let $G$ be a group of Euclidean isometries acting geometrically on $E = \mathbb{R}^{n-1}$. Fix a vector $v \in \mathbb{R}^{n-1}$.

Lemma 22.25. *There exists a number $R$ and a sequence of points $v_k \in Gv$ diverging to infinity, such that the restrictions $\phi$ on $B(v_k,R)\cap Gv$ are not constant for all $k$.*

Proof. Let $R$ be such that

$$\bigcup_{g\in G} B(gv,R) = E.$$

Suppose that the sequence $v_k$ as required does not exist. This means that there exists $r < \infty$ such that the restriction of $\phi$ to $B(v_k,R)\cap Gv$ is constant for each $v_k \in Y_k := Gv \setminus B(v,r)$. Since $\dim(E) = n-1 \geqslant 2$, the union

$$\bigcup_{v_k\in Y_k} B(v_k,R)$$

is connected. Therefore, the function $\phi$ is constant on $Gv \setminus B(v,r)$. Note that the set

$$\left\{\frac{y}{|y|} : y \in Gx \setminus B(v,r)\right\}$$

is dense in the unit sphere. Since $\phi(y/|y|) = \phi(y)$, it follows that $\phi$ is a constant function. ☐

## 22.6. Scattering

We now return to the discussion of quasi-isometries. We continue with notation of Section 22.4. In particular, we have an invertible affine mapping (which is not a similarity) $A : E = \mathbb{R}^{n-1} \to E = \mathbb{R}^{n-1}$, $A(\Lambda) = \Lambda'$, where $\Lambda, \Lambda' \subset E = \mathbb{R}^{n-1}$ be the sets of footpoints of peripheral horospheres of truncated hyperbolic spaces $\Omega, \Omega'$.

By composing $A$ with Euclidean translations we can assume that $0 = A(0)$ belongs to both $\Lambda$ and $\Lambda'$: Indeed, let $u \in \Lambda \setminus \{\infty\}, v := A(u)$, define $L_u, L_v$ to be the translations by $u, v$ respectively. Consider

$$A_2 := L_v^{-1}\circ A \circ L_u, \quad \Lambda_1 := L_u^{-1}\Lambda, \quad \Lambda_1' := L_v^{-1}\Lambda',$$

$$\Omega_1 := L_u^{-1}\Omega, \quad \Omega_1' := L_v^{-1}\Omega'.$$

Then $A_2(\Lambda_1) = \Lambda_1'$, $A_2(0) = 0$, $0 \in \Lambda_1 \cap \Lambda_1'$.

We retain the notation $A, \Lambda, \Lambda', \Omega, \Omega'$ for the linear map and the new sets of footpoints of horoballs and truncated hyperbolic spaces.

Let $J : E \cup \{\infty\} \to E \cup \{\infty\}$ be the inversion in the unit sphere centered at the origin. Then $\infty = J(0)$ belongs to both $J(\Lambda)$ and $J(\Lambda')$. The quasi-isometry

$$J \circ \tilde{f} \circ J$$

sends $J(\Omega)$ to $J(\Omega')$ and $J \circ A \circ J$ is the boundary extension of this quasi-isometry. Since $\{0, \infty\} \subset J(\Lambda) \cap J(\Lambda')$, we still have two horoballs $B_\infty, B'_\infty$ (with footpoints at $\infty$) in the complements of $J(\Omega), J(\Omega')$.

In order to simplify the notation, we set

$$\Gamma := J\Gamma J, \quad \Gamma' := J\Gamma' J, \quad \Omega := J(\Omega), \quad \Omega' := J(\Omega'), \quad \lambda := J(\Lambda), \quad \Lambda' := J(\Lambda')$$

and use $h$ for the inverted linear map $J \circ A \circ J$.

Let $\Gamma_\infty$, $\Gamma'_\infty$ be the stabilizers of $\infty$ in $\Gamma, \Gamma'$ respectively. Then $\Gamma_\infty$, $\Gamma'_\infty$ act geometrically on the Euclidean space $E = \mathbb{R}^{n-1}$. Given $x \in \mathbb{R}^{n-1}$ define the set $h_*(x) := h(\Gamma_\infty x)$.

Lemma 22.26 (Scattering lemma). *Suppose that $A$ is not a similarity. Then for each $x \in E$, $h_*(x)$ is not contained in the union of finitely many $\Gamma'_\infty$–orbits.*

Proof. If $h_*(x)$ were contained in the union of finitely many $\Gamma'_\infty$–orbits, then, in view of discreteness if $\Gamma'_\infty$, for every metric ball $B = B(x, R) \subset E$, the intersection

$$(\Gamma'_\infty \cdot h_*(x)) \cap B$$

would be finite. We will show that this is not the case.

Let $x_k = \gamma_k x \in \Gamma_\infty x, k \in \mathbb{N}$ and $R < \infty$ be as in Lemma 22.25, where $G := \Gamma_\infty$. Since $\Gamma'_\infty$ acts geometrically on $E$, there exists a sequence of elements $\gamma'_k \in \Gamma'_\infty$ such that the set $\{\gamma'_k h(x_k), k \in \mathbb{N}\}$ is relatively compact in $E$. By Lemma 22.25, the mapping

$$h|B(x_k, R) \cap \Gamma_\infty x$$

is not linear for each $k$. Therefore, the maps

$$\gamma'_k \circ h \circ \gamma_k := h_k$$

cannot be affine on $B(x, R) \cap \Gamma_\infty x$. On the other hand, the sequence of maps $h_k$ subconverges to an affine mapping $h_\infty$ on $B(x, R)$ (Corollary 22.24). This implies that the sequence of images of these maps (restricted to $B(x, R) \cap \Gamma_\infty x$) cannot be finite. We conclude that the union

$$\bigcup_{k \in \mathbb{N}} h_k \left(\Gamma_\infty x \cap B(x, R)\right) \subset (\Gamma'_\infty \cdot h_*(x)) \cap B$$

is an infinite set. Lemma follows. □

Theorem 22.27. *Suppose that $h$ is an inverted linear map which is not a similarity. Then $h$ admits no quasi-isometric extension $\Omega \to \Omega'$.*

Proof. Let $x$ be the footpoint of a complementary horoball $B$ to $\Omega$, $B \neq B_\infty$. Then, by the scattering lemma, $h_*(x)$ is not contained in a finite union of $\Gamma'_\infty$–orbits. Let $\gamma_k \in \Gamma_\infty$ be a sequence such that the $\Gamma'_\infty$–orbits of the points $h\gamma_k(x)$ are all distinct. Since $\Gamma'_\infty$ acts on $E$ geometrically, there exist elements $\gamma'_k \in \Gamma'_\infty$, so that all points $x'_k := \gamma'_k h \gamma_k(x)$ belong to a certain compact $K \subset E$. Noe that all the points $x'_k$ are distinct. Let $B'_k$ denote the complementary horoball to $\Omega'$ whose footpoint is $x'_k$. If, for an infinite sequence $k_i$, the Euclidean diameters of the balls $B'_{k_i}$ are bounded from below, then these (distinct!) horoballs will eventually intersect. However, all complementary horoballs of $\Omega'$ are pairwise disjoint. Thus,

$$\lim_{k\to\infty} \operatorname{diam}_E(B'_k) = 0.$$

Let $B_k$ be the complementary horoball to $\Omega$ whose footpoint is $\gamma_k x$. Then

$$\operatorname{dist}(B_k, B_\infty) = \operatorname{dist}(B_1, B_\infty) = -\log(\operatorname{diam}_E(B_1)) =: D,$$

while

$$\operatorname{dist}(B'_k, B'_\infty) = -\log(\operatorname{diam}(B'_k)) \to \infty.$$

If $f : \Omega \to \Omega'$ is an $(L, A)$ quasi-isometry whose quasiconformal extension is $h$ then (according to Lemma 22.17)

$$\operatorname{dist}(B'_j, B'_\infty) \leqslant R(L, A) + LD + A.$$

Contradiction. □

Therefore, we have proven

THEOREM 22.28. *Suppose that $f : \Omega \to \Omega'$ is a quasi-isometry of truncated hyperbolic spaces, $n \geqslant 3$. Then $f$ admits an (unique) extension to $S^{n-1}$ which is Moebius.*

## 22.7. Schwartz Rigidity Theorem

Before proving Theorem 22.8 we will need two technical assertions concerning isometries of $\mathbb{H}^n$ which "almost preserve" truncated hyperbolic spaces.

Let $\Omega$ be the truncated hyperbolic space corresponding to a non-uniform lattice $\Gamma < G = \operatorname{Isom}(\mathbb{H}^n)$. We will say that a subset $A \subset G = \operatorname{Isom}(\mathbb{H}^n)$ *almost preserves* $\Omega$ if there exists $C < \infty$ so that

$$\operatorname{dist}_{Haus}(\Omega, \alpha\Omega) \leqslant C, \forall \alpha \in A.$$

LEMMA 22.29. *Suppose that $\beta_k \in G$ is a converging sequence almost preserving $\Omega$. Then the sequence $(\beta_k)$ has to be finite.*

PROOF. Suppose to the contrary that the sequence $(\beta_k)$ consists of distinct elements. Then, after passing to a subsequence, there exists a peripheral horosphere $\Sigma$ of $\Omega$ with footpoint $\xi$ so that all elements of the sequence $\xi_k := \beta_k(\xi)$ are distinct. Therefore, the sequence of horospheres $\Sigma_k := \beta_k(\Sigma)$ converges to $\Sigma$ in Chabauty topology. Since the sequence $(\beta_k)$ almost preserves $\Omega$, for every $k$ there exists a peripheral horosphere $\Sigma_k$ of $\Omega$ so that $\operatorname{dist}_H(\Sigma_k, \beta_k(\Sigma)) \leqslant C$, for some $C$ independent of $k$. Then, however, horospheres $\Sigma_k$ will intersect $\Sigma$ for all large $k$. On the other hand, the footpoints $\xi_k$ of these horospheres are distinct from $\xi$, which implies that $\Sigma_k \neq \Sigma$ for all $k$. This contradicts the fact that all peripheral horospheres of $\Sigma$ are disjoint. □

PROPOSITION 22.30. *Let $\Gamma, \Gamma'$ be nonuniform lattices in $G = \operatorname{Isom}(\mathbb{H}^n)$ such that $\Gamma'$ almost preserves $\Omega$, the truncated hyperbolic space of $\Gamma$. Then the groups $\Gamma, \Gamma'$ are commensurable.*

PROOF. Suppose the assertion fails. Then the projection of $\Gamma'$ to $\Gamma \setminus G$ is infinite. Therefore, there exists an infinite sequence $(\psi_k)$ of elements of $\Gamma'$ whose projections to $G/\Gamma$ are all distinct. Let $\Lambda$ denotes the set of footpoints of peripheral horospheres of $\Sigma$. Since $\Lambda/\Gamma$ is finite, after passing to a subsequence in $(\psi_k)$, we can assume that for some horosphere $\Sigma \subset \partial\Omega$, for every $k$, the footpoints of all the horospheres $\psi_k(\Sigma)$ lie in the same $\Gamma$-orbit. Therefore, there are elements $\gamma_k \in \Gamma$ so that every $\alpha_k := \gamma_k \psi_k$ fixes the footpoint $\xi$ of $\Sigma$. Since every $\gamma_k$ preserves $\Omega$, the infinite set $A = \{\alpha_k : k \in \mathbb{N}\}$ still almost preserves $\Omega$. Furthermore, $A$ still projects injectively to $\Gamma \backslash G$. Without loss of generality, we may assume that $\xi = \infty$ in the upper half-space model of $\mathbb{H}^n$. Therefore, the elements of $A$ are Euclidean similarities. Since the stabilizer $\Gamma_\infty$ of $\infty$ in $\Gamma$ acts cocompactly on the Euclidean space $E = \mathbb{R}^{n-1}$, there exists a constant $C'$ and a sequence $\tau_k \in \Gamma_\infty$ so that for $\beta_k := \tau_k \alpha_k$,

$$|\beta_k(0)| \leqslant C'.$$

Set $B := \{\beta_k : k \in \mathbb{N}\}$. Again, the set $B$ is infinite, almost preserves $\Omega$ and projects injectively to $\Gamma \setminus G$. In particular, for every $\beta \in B$, $\mathrm{dist}_{Haus}(\Sigma, \beta\Sigma) \leqslant C$. Thus, $B$ is contained in the compact set of similarities

$$\{\beta(x) = \lambda U x + v, e^{-C} \leqslant |\lambda| \leqslant e^C, U \in O(n-1), |v| \leqslant C'\}.$$

Therefore, the set $B$ is infinite and precompact in $G$. This contradicts Lemma 22.29. ☐

*Proof of Theorem 22.8.*

Suppose that $\Gamma < G = \mathrm{Isom}(\mathbb{H}^n), n \geqslant 3$, is a non-uniform lattice.

(a) For each $(L, A)$–quasi-isometry $f : \Gamma \to \Gamma$, there exists $\gamma \in Comm_G(\Gamma)$ which is within finite distance from $f$.

Proof. The quasi-isometry $f$ extends to a quasi-isometry of the hyperbolic space $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$ (Theorem 22.19). The latter quasi-isometry extends to a quasiconformal mapping $h : \partial_\infty \mathbb{H}^n \to \partial_\infty \mathbb{H}^n$. This quasiconformal mapping has to be Moebius according to Theorem 22.28. Therefore, $\tilde{f}$ is within finite distance from an isometry $\alpha$ of $\mathbb{H}^n$ (which is an isometric extension of $h$ to $\mathbb{H}^n$), see Lemma 9.86.

Exercise 22.31. Verify that $\mathrm{dist}(\tilde{f}, \alpha)$ depends only on $\Gamma$ and quasi-isometry constants $(L, A)$ of $f$.

It remains to verify that $\alpha$ belongs to $Comm_G(\Gamma)$. We note that $f$ (and, hence, $\alpha$) sends the peripheral horospheres of $\Omega$ within (uniformly) bounded distance of peripheral horospheres of $\Omega$. Therefore, the group

$$\Gamma' := \alpha \Gamma \alpha^{-1}$$

almost preserves $\Omega$. Hence, by Proposition 22.30, the groups $\Gamma, \Gamma'$ are commensurable. Thus, $\alpha \in Comm_G(\Gamma)$.

(b) Suppose that $\Gamma, \Gamma' < G = \mathrm{Isom}(\mathbb{H}^n)$ are non-uniform lattices which are quasi-isometric to each other. Then there exists an isometry $\alpha \in Isom(\mathbb{H}^n)$ such that the groups $\Gamma$ and $\alpha \Gamma' \alpha^{-1}$ are commensurable.

Proof. The proof is analogous to (a): The quasi-isometry $f : \Omega' \to \Omega$ of truncated hyperbolic spaces corresponding to the lattices $\Gamma', \Gamma$ is within finite distance from an isometry $\alpha$. The group $\Gamma'' := \alpha \Gamma' \alpha^{-1}$ again almost preserves $\Omega$. Thus, by Proposition 22.30, the groups $\Gamma, \Gamma''$ are commensurable.

(c) Suppose that $\Gamma'$ is a finitely-generated group which is quasi-isometric to a nonuniform lattice $\Gamma$ above. Then the groups $\Gamma, \Gamma'$ are virtually isomorphic, more precisely, there exists a finite normal subgroup $K \subset \Gamma'$ such that the groups $\Gamma, \Gamma'/K$ contain isomorphic subgroups of finite index.

Proof. Let $f : \Gamma \to \Gamma'$ be a quasi-isometry and let $f' : \Gamma' \to \Gamma$ be its quasi-inverse. Then, by Lemma 5.60, we have a quasi-action $\Gamma' \curvearrowright \Omega$ via

$$\gamma' \mapsto \rho(\gamma') := f' \circ \gamma' \circ f \in QI(\Omega).$$

According to Part (a), each quasi-isometry $g = \rho(\gamma')$ is within (uniformly) bounded distance from a quasi-isometry of $\Omega$ induced by an element $g^*$ of $Comm_G(\Gamma)$. We get a map

$$\psi : \gamma' \mapsto \rho(\gamma') = g \mapsto g^* \in Comm_G(\Gamma).$$

We claim that this map is a homomorphism with finite kernel. Let $h_\infty$ denote the extension to $\partial_\infty \mathbb{H}^n$ for quasi-isometries $h : \mathbb{H}^n \to \mathbb{H}^n$. Then $\psi$ induces a homomorphism

$$\psi_\infty : \gamma' \mapsto g_\infty = g^*_\infty, \quad \psi_\infty : \Gamma' \to Comm_G(\Gamma).$$

Since the quasi-action $\rho : \Gamma' \curvearrowright \Omega'$ is geometric (see Lemma 5.60), by Lemma 9.90 kernel $K$ of the quasi-action $\Gamma' \curvearrowright \Omega'$ is quasi-finite. The subgroup $K \lhd \Gamma'$ is also the kernel of the homomorphism $\psi_\infty$; by Lemma 9.86, the subgroup $K$ is finite.

The rest of the proof is the same as for (a) and (b): The group $\Gamma'' := \psi(\Gamma')$ almost preserves $\Omega$, hence, it is commensurable to $\Gamma$. □

## 22.8. Mostow Rigidity Theorem

Theorem 22.32 (Mostow Rigidity Theorem). *Suppose that $n \geqslant 3$ and $\Gamma, \Gamma' <$ $\mathrm{Isom}(\mathbb{H}^n)$ are lattices and $\rho : \Gamma \to \Gamma'$ is an isomorphism. Then $\rho$ is induced by an isometry, i.e. there exists an isometry $\alpha \in \mathrm{Isom}(\mathbb{H}^n)$ such that*

$$\alpha \circ \gamma = \rho(\gamma) \circ \alpha$$

*for all $\gamma \in \Gamma$.*

Proof. **Step 1.** We first observe that $\Gamma$ is uniform if and only if $\Gamma'$ is uniform. Indeed, if $\Gamma$ is uniform, it is Gromov-hyperbolic and, hence, cannot contain a noncyclic free abelian group. On the other hand, if $\Gamma$ is non-uniform then thick-thin decomposition implies that $\Gamma$-stabilizers of peripheral horospheres of the corresponding truncated hyperbolic space contain free abelian subgroups of rank $n - 1$.

Proposition 22.33. *There exists a $\rho$–equivariant quasi-isometry $f : \mathbb{H}^n \to \mathbb{H}^n$.*

Proof. As in the proof of Theorem 22.8 we choose truncated hyperbolic spaces $\Omega \subset \mathbb{H}^n, \Omega' \subset \mathbb{H}^n$ which are invariant under $\Gamma$ and $\Gamma'$ respectively. (In case if $\Gamma$ acts cocompactly on $\mathbb{H}^n$ we would take of course $\Omega = \Omega' = \mathbb{H}^n$.)

Then Lemma 5.35 implies that there exists a $\rho$–equivariant quasi-isometry

$$f : \Omega \to \Omega'.$$

According to Theorem 22.19, $f$ admits a $\rho$-equivariant extension to a quasi-isometry $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$.

Remark 22.34. The most difficult part of the proof of Theorem 22.19 was to show that $f$ sends peripheral horospheres uniformly close to peripheral horospheres. In the equivariant setting the proof is much easier: Note that $\rho$ sends maximal abelian subgroups to maximal abelian subgroups. Stabilizers of peripheral horospheres are virtually $\mathbb{Z}^{n-1}$. Therefore, $\rho$ sends stabilizers of peripheral horospheres to stabilizers of peripheral horospheres. From this, it is immediate that peripheral horospheres map uniformly close to peripheral horospheres.

Therefore, according to Theorem 22.19, $f$ extends to a $\rho$–equivariant quasi-isometry $\tilde{f} : \mathbb{H}^n \to \mathbb{H}^n$.

**Step 2.** Let $h$ denote the $\rho$–equivariant quasi-conformal homeomorphism $S^{n-1} \to S^{n-1}$ which is the extension of $f$ guaranteed by Theorem 20.24. Our goal is to show that $h$ is Moebius. We argue as in the proof of Theorem 22.28. We will identify $S^{n-1}$ with the extended Euclidean space $\hat{\mathbb{R}}^{n-1} = \mathbb{R}^{n-1} \cup \{\infty\}$. Accordingly, we will identify $\mathbb{H}^n$ with the upper half-space. The key for the proof is the fact that $h$ is differentiable almost everywhere on $\mathbb{R}^{n-1}$ so that its Jacobian is nonzero for almost every $z \in \mathbb{R}^{n-1}$. (The latter fails for quasi-Moebius homeomorphisms of the circle: Although they are differentiable almost everywhere, their derivatives vanish almost everywhere as well.)

Suppose that $z \in S^{n-1}$ is a point of differentiability of $h$ so that $J_z(h) \neq 0$. Since only countable number points in $S^{n-1}$ are fixed points of parabolic elements, we can choose $z$ to be a conical limit point of $\Gamma$. By applying a Moebius change of coordinates, we can assume that $z = h(z) = 0 \in \mathbb{R}^{n-1}$ and that $h(\infty) = \infty$.

The following proof is yet another version of the *zooming argument*. Let $L \subset \mathbb{H}^n$ be the vertical geodesic emanating from 0; pick a base-point $y_0 \in L$. Since $z$ is a conical limit point, there is a sequence of elements $\gamma_i \in \Gamma$ so that

$$\lim_{i\to\infty} \gamma_i(y_0) \to z$$

and

$$\mathrm{dist}(\gamma_i(y_0), L) \leqslant C$$

for each $i$. Let $y_i$ denote the nearest-point projection of $\gamma_i(y_0)$ to $L$. Take the sequence of hyperbolic translations $T_i : x \mapsto \lambda_i x$ with the axis $L$, so that $T_i(y_0) = y_i$. Then the sequence $k_i := \gamma_i^{-1} T_i$ is relatively compact in $\mathrm{Isom}(\mathbb{H}^n)$ and lies in a compact $K \subset \mathrm{Isom}(\mathbb{H}^n)$. Now we form the sequence of quasiconformal homeomorphisms

$$h_i(x) := \lambda_i^{-1} h(\lambda_i x) = T_i^{-1} \circ h \circ T_i(x).$$

Note that $\lambda_i \to 0$ as $i \to \infty$. Since the function $h$ is assumed to be differentiable at zero, there is a linear transformation $A \in GL(n-1, \mathbb{R})$ so that

$$\lim_{i\to\infty} h_i(x) = Ax$$

for all $x \in \mathbb{R}^{n-1}$. Since $h(\infty) = \infty$, it follows that

$$\lim_{i\to\infty} h_i = A$$

pointwise on $S^{n-1}$.

By construction, $h_i$ conjugates the group $\Gamma_i := T_i^{-1} \Gamma T_i \subset \mathrm{Isom}(\mathbb{H}^n)$ into the group of Moebius transformations. We have

$$\Gamma_i = T_i^{-1} \Gamma T_i = (k_i^{-1} \gamma_i) \Gamma (k_i^{-1} \gamma_i)^{-1} = k_i^{-1} \Gamma k_i.$$

After passing to a subsequence, we can assume that

$$\lim_{i\to\infty} k_i = k \in \mathrm{Isom}(\mathbb{H}^n)\,.$$

Therefore the *sequence of sets* $\Gamma_i$ converges to $\Gamma_\infty := k^{-1}\Gamma k$ (in the Chabauty topology on $\mathrm{Isom}(\mathbb{H}^n)$). For each sequence $\beta_i \in \Gamma_i$ which converges to some $\beta \in \mathrm{Isom}(\mathbb{H}^n)$ we have

$$\lim_{i\to\infty} h_i\beta_i h_i^{-1} = A\beta A^{-1}.$$

Since $h_i\beta_i h_i^{-1} \in \mathrm{Isom}(\mathbb{H}^n)$ for each $i$, it follows that $A\beta A^{-1} \in \mathrm{Isom}(\mathbb{H}^n)$ for each $\beta \in \Gamma_\infty$. Thus

$$A\Gamma_\infty A^{-1} \subset \mathrm{Isom}(\mathbb{H}^n)\,.$$

Since the group $\Gamma_\infty$ is nonelementary, the orbit $\Gamma_\infty \cdot (\infty)$ is infinite. Hence $\Gamma_\infty$ contains an element $\gamma$ such that $\gamma(\infty) \notin \{\infty, 0\}$.

LEMMA 22.35. *Suppose that* $\gamma \in \mathrm{Isom}(\mathbb{H}^n)$ *is such that* $\gamma(\infty) \neq \infty, 0$, $A \in GL(n-1,\mathbb{R})$ *is an element which conjugates* $\gamma$ *to* $A\gamma A^{-1} \in \mathrm{Isom}(\mathbb{H}^n)$. *Then* $A$ *is a Euclidean similarity, i.e it belongs to* $\mathbb{R}_+ \times O(n-1)$.

PROOF. Suppose that $A$ is not a similarity. Let $P$ be a hyperplane in $\mathbb{R}^{n-1}$ which contains the origin 0 but does not contain $A\gamma^{-1}(\infty)$. Then $\gamma \circ A^{-1}(P)$ does not contain $\infty$ and hence is a round sphere $\Sigma$ in $\mathbb{R}^{n-1}$.

Since $A$ is not a similarity, the image $A(\Sigma)$ is an ellipsoid which is not a round sphere. Hence the composition $A\gamma A^{-1}$ does not send planes to round spheres and therefore it is not Moebius. Contradiction. □

We, therefore, conclude that the derivative of $h$ at 0 is Moebius transformation $A \in \mathbb{R}_+ \times O(n-1)$. Thus, $h$ is conformal at a.e. point of $\mathbb{R}^n$. One option now is to use Liouville's theorem for quasiconformal maps (Theorem 20.20). Instead, we will give a direct argument.

**Step 3.** We will be using the notation of Step 2.

Let $G := \mathrm{Isom}(\mathbb{H}^n)$, identified with the group of Moebius transformations of $S^{n-1}$. Consider the quotient

$$Q = G\backslash \mathrm{Homeo}(S^{n-1})$$

consisting of the cosets $[f] = \{g \circ f : g \in G\}$. Give this quotient the quotient topology, where we endow $\mathrm{Homeo}(S^{n-1})$ with the topology of pointwise convergence. Since $G$ is a closed subgroup in $\mathrm{Homeo}(S^{n-1})$, it follows that every point in $Q$ is closed. (Actually, $Q$ is Hausdorff, but we will not need this.) The group $\mathrm{Homeo}(S^{n-1})$ acts on $Q$ by the formula

$$[f] \mapsto [f \circ g], g \in \mathrm{Homeo}(S^{n-1}).$$

It is clear from the definition of the quotient topology on $Q$ that this action is continuous, i.e. the map

$$Q \times \mathrm{Homeo}(S^{n-1}) \to Q$$

is continuous.

Since $h$ is a $\rho$–equivariant homeomorphism, we have

$$[h] \circ \gamma = [h], \quad \forall \gamma \in \Gamma.$$

Recall that we have a sequence of dilations $T_i$ (fixing the origin), a sequence $\gamma_i \in \Gamma$ and a sequence $k_i \in G$ which converges to $k \in G$, so that

$$T_i = \gamma_i \circ k_i,$$

and

$$\lim_i h_i = A \in \mathbb{R}_+ \times O(n-1) \subset G,$$

where

$$h_i = T_i^{-1} \circ h \circ T_i.$$

Therefore

$$[h_i] = [h\gamma_i k_i] = [h] \circ k_i,$$
$$[1] = [A] = \lim_i [h_i] = \lim_i ([h_i] \circ k_i) = [h] \circ \lim_i k_i = [h] \circ k.$$

(Recall that every point in $Q$ is closed.) Thus $[h] = [1] \circ k^{-1} = [1]$, which implies that $h$ is the restriction of an element $\alpha \in \mathrm{Isom}(\mathbb{H}^{\mathrm{n}})$. Since

$$\rho(\gamma) \circ \alpha(x) = \alpha \circ \gamma(x), \quad \forall x \in S^{n-1},$$

it follows that $\alpha : \mathbb{H}^n \to \mathbb{H}^n$ is $\rho$–equivariant. □

CHAPTER 23

# A survey of quasi-isometric rigidity

This survey covers three types of problems within the theme of quasi-isometric rigidity:

(1) Description of the group of quasi-isometries $QI(X)$ of specific metric spaces $X$ (and finitely generated groups $G$). In some cases, $QI(X)$ coincides with the subgroup of isometries of $X$ or the subgroup of virtual automorphisms of $G$ or with the commensurator of $G$, either abstract or considered in a larger group. To describe these different flavors of QI rigidity, we introduce the following terminology.

A metric space $X$ or a group $G$ is called *strongly QI rigid* if each $(L, A)$–quasi-isometry $f : X \to X$ is within finite distance from an isometry $\phi : X \to X$ or an element $\phi$ of $Comm(G)$ and, moreover, $\text{dist}(f, \phi) \leqslant C(L, A)$.

(2) Identification of the classes of groups that are QI rigid. A class of groups $\mathcal{G}$ is *QI rigid* if each group $G$ which is quasi-isometric to a member of $\mathcal{G}$ is virtually isomorphic to a member of $\mathcal{G}$. This problem was formulated for the first time (with a slightly different terminology) by M. Gromov in [**Gro83**]. It is sometimes related to the first problem. Indeed if a group $G'$ is quasi-isometric to $G$ then there exists a homomorphism $G' \to QI(G)$ which in many cases has finite kernel (see Lemma 5.62). If $QI(G)$ is either very close to $G$ or very close to an ambient group in which all groups in the class $\mathcal{G}$ lie, then one is halfway through a proof of QI rigidity of $\mathcal{G}$.

(3) Quasi-isometric classification within a given class of groups. In other words, given a class of groups $\mathcal{G}$, to describe the equivalence classes up to quasi-isometry contained in it. This can be achieved either by a complete description of the equivalence classes or by using QI invariants. An extreme case is when the QI class of a group contains only its finite index subgroups, their quotients by finite normal subgroups and finite extensions of these quotients. With that in view, we call a group $G$ *QI rigid* if any group $G'$ which is quasi-isometric to $G$ is virtually isomorphic to $G$.

## 23.1. Rigidity of symmetric spaces, lattices, hyperbolic groups

THEOREM 23.1 (P. Pansu, [**Pan89**]). *Let $X$ be a quaternionic hyperbolic space $\mathbf{H}\mathbb{H}^n$ $(n \geqslant 2)$ or the octonionic hyperbolic plane $\mathbf{O}\mathbb{H}^2$. Then $X$ is strongly QI rigid.*

Theorem 23.1, the work of P. Tukia [**Tuk86**] for the real-hyperbolic spaces $\mathbb{H}^n, n \geqslant 3$, and that of R. Chow [**Cho96**] for complex-hyperbolic spaces $\mathbf{C}\mathbb{H}^n, n \geqslant 2$, yield the following result:

Theorem 23.2. *Let $X$ be a symmetric space of negative curvature which is not the hyperbolic plane $\mathbb{H}^2$. Then the class of uniform lattices in $X$ is QI rigid.*

QI rigidity also holds for the uniform lattices in $\mathbb{H}^2$, see §21.8:

Theorem 23.3. *The class of fundamental groups of closed hyperbolic surfaces is QI rigid.*

For higher rank symmetric spaces strong QI rigidity follows from a series of results of Kleiner and Leeb, which were independently (although a bit later) obtained by Eskin and Farb in [**EF97b**].

Theorem 23.4 (B. Kleiner, B. Leeb, [**KL98b**]). *Let $X$ be a symmetric space of nonpositive curvature such that each de Rham factor of $X$ is a symmetric space of rank $\geqslant 2$. Then $X$ is strongly QI rigid.*

As an application of this rigidity theorem, Kleiner and Leeb obtained:

Theorem 23.5 (B. Kleiner, B. Leeb, [**KL98b**]). *Let $X$ be a symmetric space of nonpositive curvature without Euclidean de Rham factors. Then the class of uniform lattices in $X$ is QI rigid.*

Kleiner and Leeb also established strong QI rigidity for Euclidean buildings:

Theorem 23.6 (B. Kleiner, B. Leeb, [**KL98b**]). *Let $X$ be a Euclidean building such that each de Rham factor of $X$ is a Euclidean building of rank $\geqslant 2$. Then $X$ is strongly QI rigid.*

Turning to non-uniform lattices, one should first note that Theorem 22.8 of R. Schwartz (see Chapter 22) in its most general form holds with the space $\mathbb{H}^n, n \geqslant 3$, replaced by any negatively curved symmetric space different from $\mathbb{H}^2$. This theorem answers the three types of problems described in the beginning of the section, and can be formulated as follows.

Theorem 23.7 (R. Schwartz [**Sch96a**]).

1. *Let $\Gamma$ be a non-uniform lattice of a negatively curved symmetric space $X$ different from $\mathbb{H}^2$. Then $QI(\Gamma)$ coincides with $Comm_G(\Gamma)$, where $G = \mathrm{Isom}(X)$.*
2. *The class of non-uniform lattices of negatively curved symmetric spaces different from $\mathbb{H}^2$ is QI rigid.*
3. *If $\Gamma$ and $\Gamma'$ are quasi-isometric non-uniform lattices of negatively curved symmetric spaces $X$ and respectively $X'$ then $X = X'$. If moreover $X \neq \mathbb{H}^2$ then $\Gamma$ and $\Gamma'$ are commensurable in $G = \mathrm{Isom}(X)$.*

A non-uniform lattice of $\mathbb{H}^2$ contains a finite index subgroup which is free non-abelian. In that case we may therefore apply QI rigidity of virtually free groups (Theorem 18.38) and conclude:

Theorem 23.8. *Each non-abelian free group is QI rigid. Thus, each nonuniform lattice in $\mathbb{H}^2$ is QI rigid.*

Also, in the special case when $X$ is the 3-dimensional hyperbolic space, $\mathrm{Isom}(X)$ can be identified with $SL(2,\mathbb{C})$, Schwartz's result has the following arithmetic version. Let $K_1$ and $K_2$ be imaginary quadratic extensions of $\mathbb{Q}$ and let $\mathcal{O}_1$ and $\mathcal{O}_2$ their respective rings of integers. The arithmetic lattices $SL(2,\mathcal{O}_1)$ and $SL(2,\mathcal{O}_2)$ are quasi-isometric if and only if $K_1$ and $K_2$ are isomorphic.

When instead of imaginary quadratic extensions, one takes totally real quadratic extensions, the corresponding groups $SL(2, \mathcal{O}_i)$ become non-uniform $\mathbb{Q}$-rank one lattices of isometries of the space $\mathbb{H}^2 \times \mathbb{H}^2$, that is of a rank two symmetric space. In general, when $K_i$ are algebraic extensions of $\mathbb{Q}$, $SL(2, \mathcal{O}_i)$ are non-uniform $\mathbb{Q}$-rank one lattices of isometries of a product with factors $\mathbb{H}^2$ and $\mathbb{H}^3$.

THEOREM 23.9 (R. Schwartz [**Sch96a**]). (1) *Let $K$ be an algebraic extension of $\mathbb{Q}$, let $\mathcal{O}$ be its ring if integers, let $\Gamma = PSL(2, \mathcal{O})$ and let $X$ be the product of hyperbolic spaces on which $\Gamma$ acts with finite covolume. The group $QI(\Gamma)$ coincides with $Comm_G(\Gamma)$, where $G = \mathrm{Isom}(X)$.*

(2) *Let $K_i$ be two algebraic extensions of $\mathbb{Q}$, and $\mathcal{O}_i$ their corresponding rings if integers, $i = 1, 2$.*

*The lattices $PSL(2, \mathcal{O}_1)$ and $PSL(2, \mathcal{O}_2)$ are quasi-isometric if and only if the fields $K_1$ and $K_2$ are isomorphic.*

Note that every irreducible lattice in a semisimple group having $\mathbb{R}$-rank at least 2 and at least a factor of $\mathbb{R}$-rank one is an arithmetic $\mathbb{Q}$-rank one lattice [**Pra73**, Lemma 1.1], though the lattice may in general be quite different from the example of $PSL(2, \mathcal{O})$.

The higher $\mathbb{Q}$-rank case was settled by A. Eskin.

THEOREM 23.10 (A. Eskin [**Esk98**]). (1) *Let $\Gamma$ be a non-uniform irreducible lattice of a symmetric space $X$ with all factors of rank at least 2. Then $QI(\Gamma)$ coincides with $Comm_G(\Gamma)$, where $G = \mathrm{Isom}(X)$.*

(2) *The class of non-uniform irreducible lattices of symmetric spaces with all factors of rank at least 2 is QI rigid.*

(3) *If $\Gamma$ and $\Gamma'$ are quasi-isometric non-uniform irreducible lattices of symmetric spaces with all factors of rank at least 2, $X$ and respectively $X'$, then $X = X'$ and $\Gamma$ and $\Gamma'$ are commensurable in $G = \mathrm{Isom}(X)$.*

Theorems 23.7, 23.9 and 23.10 imply that given an arithmetic irreducible non-uniform lattice $\Gamma$ in a symmetric space either of rank one $\neq \mathbb{H}^2$ or with all factors of rank $\geqslant 2$ or $\Gamma$ equal to $PSL(2, \mathcal{O})$ for the ring of integers $\mathcal{O}$ of some algebraic field $K$, the group of quasi-isometries $QI(\Gamma)$ equals $Comm_G(\Gamma)$, which is a countable dense subgroup of $G$. On the other hand, when $\Gamma$ is a non-arithmetic lattice in a symmetric space $X \neq \mathbb{H}^2$ of rank one, $QI(\Gamma) = Comm_G(\Gamma)$ is also a lattice in $X$, moreover it is maximal with respect to inclusion. Thus in this case each QI equivalence class consists of a maximal non-uniform lattice and all its finite index subgroups.

One may ask if similar results also hold for uniform lattices $\Gamma$. In that case $QI(\Gamma)$ contains $G$, it is equal to it when $X$ is either of higher rank or $X = \mathbf{H}\mathbb{H}^n$ ($n \geqslant 2$) or $X = \mathbf{O}\mathbb{H}^2$. In the real and complex case $QI(\Gamma) = QI(X)$ is much larger than $G$.

Concerning the classification statement (3) in Theorems 23.7, 23.9 and 23.10, only the first part holds for uniform lattices. Indeed, any two uniform lattices $\Gamma, \Gamma'$ in a space $X$ are quasi-isometric but not all of them are commensurable. If $\Gamma, \Gamma'$ were commensurable in $G$ then $Comm_G(\Gamma) = Comm_G(\Gamma')$, where $G = \mathrm{Isom}(X)$. But if one takes $X = \mathbb{H}^n, n \geqslant 3$, $\Gamma$ arithmetic and $\Gamma'$ non-arithmetic, then $Comm_G(\Gamma)$ is dense in $G$, while $Comm_G(\Gamma')$ is discrete in $G$.

Restricting to the sub-class of arithmetic uniform lattices does not grant commensurability either, as the following example emphasizes. Consider a quadratic form of signature $(1, n)$, $L(x_1, ..., x_{n+1}) = \sqrt{2}x_{n+1}^2 - a_1x_1^2 - \cdots - a_nx_n^2$, where $a_i$ are positive rational numbers. The set

$$\mathbb{H}_L = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \; ; \; L(x_1, \ldots, x_{n+1}) = 1 \, , \, x_{n+1} > 0\}$$

is a model of the hyperbolic $n$-dimensional space. Its group of isometries is $SO_0(L)$, the connected component containing the identity of the stabilizer of the form $L$ in $SL(n+1, \mathbb{R})$. The discrete subgroup $\Gamma_L = SO_0(L) \cap SL(n+1, \mathbb{Z}(\sqrt{2}))$ is a uniform lattice.

If two such lattices $\Gamma_{L_1}$ and $\Gamma_{L_2}$ are commensurable then there exist $g \in GL(n+1, \mathbb{Q}[\sqrt{2}])$ and $\lambda \in \mathbb{Q}[\sqrt{2}]$ such that $L_1 \circ g = \lambda L_2$. In particular, if $n$ is odd then the ratio between the discriminant of $L_1$ and the discriminant of $L_2$ is a square in $\mathbb{Q}[\sqrt{2}]$. It suffices to take two forms such that this is not possible, for instance (like in [**GPS88**]):

$$L_1 = \sqrt{2}x_{n+1}^2 - x_1^2 - x_2^2 - \cdots - x_n^2 \text{ and } L_2 = \sqrt{2}x_{n+1}^2 - 3x_1^2 - x_2^2 - \cdots - x_n^2 \, .$$

Theorem 23.11 (B. Kleiner, B. Leeb, [**KL01**]). *Suppose that $\Gamma$ is a finitely-generated group which is quasi-isometric to a Lie group $G$ with center $C$ and semisimple quotient $G/C = H$. Then $\Gamma$ fits into a short exact sequence*

$$1 \to K \to \Gamma \to Q \to 1,$$

*where $K$ is quasi-isometric to $C$ and $Q$ is virtually isomorphic to a uniform lattice in $H$.*

Problem 23.12. Prove an analogue of the above theorem for all Lie groups $G$ (without assuming that the sol-radical of $G$ is central).

Theorem 23.13 (M. Bourdon, H. Pajot [**BP00**]). *Let $X$ be a thick hyperbolic building of rank 2 with right-angled fundamental polygon and whose links are complete bipartite graphs. Then $X$ is strongly QI rigid.*

Theorem 23.14 (X. Xie [**Xie06**]). *Let $X$ be a thick hyperbolic building of rank 2. Then $X$ is strongly QI rigid.*

Problem 23.15. Construct an example of a hyperbolic group with Menger curve boundary, which is QI rigid.

Problem 23.16. Let $G$ be a *random* $k$-generated group, $k \geqslant 2$. Is $G$ QI rigid?

Randomness can be defined for instance as follows. Consider the set $B(n)$ of presentations

$$\langle x_1, \ldots, x_k | R_1, \ldots, R_l \rangle$$

where the total length of the words $R_1, ..., R_l$ is $\leqslant n$. Then a class $C$ of $k$-generated groups is said to consist of random groups if

$$\lim_{n\to\infty} \frac{\text{card } (B(n) \cap C)}{\text{card } B(n)} = 1.$$

Here is another notion of randomness: Fix the number $l$ of relators, assume that all relators have the same length $n$; this defines a class of presentations $S(k, l, n)$. Then require

$$\lim_{n\to\infty} \frac{\text{card } (S(k, l, n) \cap C)}{\text{card } S(k, l, n)} = 1.$$

See [**KS08**] for a comparison of various notions of randomness for groups.

Theorem 23.17 (M. Kapovich, B. Kleiner, [**KK00**]). *There is a 3-dimensional hyperbolic group which is strongly QI rigid.*

Example 23.18. Let $S$ be a closed hyperbolic surface, $M$ is the unit tangent bundle over $S$. Then we have an exact sequence

$$1 \to \mathbb{Z} \to G = \pi_1(M) \to Q := \pi_1(S) \to 1.$$

This sequence does not split even after passage to a finite index subgroup in $G$, hence $G$ is not virtually isomorphic to $Q \times \mathbb{Z}$. However, since $Q$ is hyperbolic, the group $G$ is quasi-isometric to $Q \times \mathbb{Z}$. See Theorem 9.113.

Example 23.19. The product of free groups $G = F_n \times F_m$, $(n, m \geqslant 2)$ is not QI rigid.

Proof. The group $G$ acts discretely, cocompactly, isometrically on the product of simplicial trees $X := T \times T'$. However, there are examples [**Wis96**], [**BM00**], of groups $G'$ acting discretely, cocompactly, isometrically on $X$ so that $G'$ contains no proper finite index subgroups. Then $G$ is quasi-isometric to $G'$ but these groups are clearly not virtually isomorphic. □

## 23.2. Rigidity of relatively hyperbolic groups

Theorem 23.20 (C. Druţu, [**Dru09**]). *The class of relatively hyperbolic groups is QI rigid.*

Other theorems appearing in this section emphasize that various subclasses of relatively hyperbolic groups are likewise QI rigid: Non-uniform lattices in rank one symmetric spaces by Theorem 23.7, fundamental groups of non-geometric Haken manifolds with at least one hyperbolic component by Theorem 23.47, fundamental groups of graphs of groups with finite edge groups [**PW02**].

Theorem 23.20 suggests the following natural question.

Problem 23.21 (P. Papazoglou). Is there a geometric criterion allowing to recognize whether a finitely generated group is relatively hyperbolic (without any reference to peripheral subgroups or subsets)?

Concerning the proof of Theorem 23.20, it is not difficult to see that if $\mathfrak{q} : X \to Y$ is a quasi-isometry between two metric spaces and $X$ is hyperbolic relative to $\mathcal{A}$ then $Y$ is hyperbolic relative to $\{\mathfrak{q}(A) \ ; \ A \in \mathcal{A}\}$. Thus the main step is to prove that if a group $G$ is hyperbolic relative to some collection of subsets $\mathcal{A}$ then it is hyperbolic relative to some collection of subgroups $H_1, ..., H_n$, such that each $H_i$ is contained in the tubular neighborhood of some $A_i$ [**Dru09**]. A variation of the same argument, appears in [**BDM09**]:

Theorem 23.22. *Let $X$ be a metric space which is hyperbolic relative to a collection of subsets $\mathcal{A}$. Suppose that $\mathfrak{q} : G \to X$ is a quasi-isometric embedding. Then $G$ is hyperbolic relative to the collection of pre-images $\mathfrak{q}^{-1}(\mathcal{N}_C(A)), A \in \mathcal{A}$ for some $C < \infty$.*

This result and the above argument implies that either $\mathfrak{q}(G)$ is in an $M$–tubular neighborhood of some set $A \in \mathcal{A}$ or $G$ is hyperbolic relative to finitely many (proper) subgroups $H_i$ with each $\mathfrak{q}(H_i)$ in an $M$–tubular neighborhood of some set $A_i \in \mathcal{A}$. Thus we have the following generalization of the Quasi-Flat Lemma of R. Schwartz [**Sch96b**] (see Remark 22.18). Below by a group that is *not relatively hyperbolic*

*(NRH)* we mean a group that contains no finite collection of infinite index subgroups with respect to which it is relatively hyperbolic.

The following theorem was proven by J. Behrstock, C. Druţu, L. Mosher in [**BDM09**], Theorem 4.1:

THEOREM 23.23 (NRH subgroups are always peripheral). *Let $X$ be a metric space hyperbolic relative to a collection $\mathcal{A}$ of subsets. For every $L \geqslant 1$ and $C \geqslant 0$ there exists $R = R(L, C, X, \mathcal{A})$ such that the following holds.*

*If $G$ is a finitely generated group with a word metric* dist *and $G$ is not relatively hyperbolic then any $(L, C)$–quasi-isometric embedding $\mathfrak{q} : G \to X$ has its image $\mathfrak{q}(G)$ contained in the $M$–neighborhood of some set $A \in \mathcal{A}$.*

In the theorem above the constant $R$ *does not* depend on the group $G$. In [**DS05**] the same theorem was proved under the stronger hypothesis that the group $G$ has one asymptotic cone without global cut-points.

As in the case of R. Schwartz's argument, Theorem 23.23 is a step towards classification of relatively hyperbolic groups.

THEOREM 23.24 (J. Behrstock, C. Druţu, L. Mosher, [**BDM09**], Theorem 4.8). *Let $G$ be a finitely generated group hyperbolic relative to a finite collection of finitely generated subgroups $\mathcal{H}$ such that each $H \in \mathcal{H}$ is not relatively hyperbolic.*

*If $G'$ is a finitely generated group quasi-isometric to $G$ then $G'$ is hyperbolic relative to a finite collection of finitely generated subgroups $\mathcal{H}'$ where each subgroup in $\mathcal{H}'$ is quasi-isometric to one of the subgroups in $\mathcal{H}$.*

It is impossible to establish a relation between peripheral subgroups of QI relatively hyperbolic groups, when working in full generality, hence there is no mention of this in Theorem 23.20. For instance when $G = G' = A * B * C$, the group $G$ is hyperbolic relative to $\{A, B, C\}$, and also hyperbolic relative to $\{A * B, C\}$.

By results in [**PW02**], the classification of relatively hyperbolic groups reduces to the classification of one-ended relatively hyperbolic groups. Theorem 23.24 points out a fundamental necessary condition for the quasi-isometry of two one-ended relatively hyperbolic groups with NRH peripheral subgroups: that the peripheral subgroups define the same collection of quasi-isometry classes.

Related to this, one may ask whether every relatively hyperbolic group admits a list of peripheral subgroups that are NRH? The answer is negative in general, a counter-example is Dunwoody's inaccessible group [**Dun93**]. Since finitely presented groups are accessible, this raises the following natural question.

PROBLEM 23.25 (J. Behrstock, C. Druţu, L. Mosher, [**BDM09**]). Is there an example of a finitely presented relatively hyperbolic group such that every list of peripheral subgroups contains a relatively hyperbolic group?

### 23.3. Rigidity of classes of amenable groups

The class of amenable groups is QI rigid, see Theorem 16.10. Recall that by Corollary 16.63 the set of finitely generated groups splits into amenable groups and paradoxical groups. This implies that the class of paradoxical groups is also QI rigid. Since this latter class is characterized by the fact that the Cheeger constant is positive (Theorem 16.3), it follows that having a positive Cheeger constant is a QI invariant property. As noted, QI invariance of the property of having positive Cheeger constant is true not only when the two quasi-isometric objects are groups, but also when they are graphs or manifolds of locally bounded geometry or when one is a manifold and one is a finitely generated group.

Various sub-classes of amenable groups behave quite differently with respect to QI rigidity, and relatively little is known about the QI classification and the description of groups of quasi-isometries.

The class of virtually nilpotent groups is QI rigid by Corollary . Concerning the QI classification the following is known.

THEOREM 23.26 (P. Pansu [**Pan83**]). *If $G$ and $N$ are finitely generated quasi-isometric nilpotent groups then the graded Lie groups associated to $G/\mathrm{Tor}\, G$ and $H/\mathrm{Tor}\, H$ are isomorphic.*

One of the steps in the proof of Theorem 23.26 is that all asymptotic cones of a finitely generated nilpotent group $G$ with a canonically chosen word metric are isometric to the graded Lie group associated to $G/\mathrm{Tor}\, G$, endowed with a Carnot-Caratheodory metric, i.e. Theorem 14.30.

Theorem 23.26 points out new quasi-isometry invariants in the class of nilpotent groups: the nilpotency class of $\bar{G} = G/\mathrm{tor}(G)$ and the rank of each of the Abelian groups $C^i\bar{G}/C^{i+1}\bar{G}$, where $C^i\bar{G}$ is the $i$-th group in the lower central series of $\bar{G}$.

Other QI invariants in the class of nilpotent group that may help to distinguish nilpotent groups with the same associated nilpotent graded Lie groups are the Betti numbers [**Sha04**, Theorem 1.2].

Pansu's theorem 23.26 implies the following.

THEOREM 23.27. (1) *Each finitely generated abelian group is QI rigid.*
(2) *Two finitely generated abelian groups are quasi-isometric if and only if their ranks are equal (see Theorem 10.8).*

Unlike the Abelian groups, the nilpotent groups are not completely classified up to QI. In particular the following remains an open problem.

PROBLEM 23.28. Is it true that two nilpotent simply-connected Lie groups (endowed with left-invariant metrics) are quasi-isometric if and only if they are isomorphic?

The statement is true for graded nilpotent Lie groups, according to Theorem 23.26.

The group of quasi-isometries is very large already for Abelian groups. Indeed, an example of quasi-isometry of $\mathbb{R}^2$ endowed with the polar coordinates $(\rho, \theta)$ is a map defined on $\mathbb{R}^2 \setminus \{(0,0)\}$ by $(\rho, \theta) \mapsto (f(\rho), \theta + o(\rho))$, where $f : (0, \infty) \to (0, \infty)$ is a quasi-isometry.

In view of Theorem 11.36, the class of (virtually) solvable groups is not QI rigid. However, the groups $G_A$ and $G_B$ constructed in the proof are both elementary amenable.

PROBLEM 23.29. Is the class of elementary amenable groups QI rigid?

THEOREM 23.30 (B. Farb and L. Mosher [**FM00**]). *The class of finitely presented non-polycyclic abelian-by-cyclic groups is QI rigid.*

The starting point in the proof of B. Farb and L. Mosher is to consider torsion-free finite index subgroups and to apply a result of R. Bieri and R. Strebel [**BS78**], stating that for every torsion-free finitely presented abelian-by-cyclic group there exists $n \in \mathbb{N}$ and a matrix $M = (m_{ij}) \in M(n, \mathbb{Z})$ with non-zero determinant such that the group has the presentation

$$\left\langle a_1, a_2, \ldots, a_n, t \; ; \; [a_i, a_j], t a_i t^{-1} a_1^{m_{1i}} a_2^{m_{2i}} \cdots a_n^{m_{ni}} \right\rangle . \tag{23.1}$$

Let $\Gamma_M$ be the group with the presentation in (23.1) for the integer matrix $M$. The group $\Gamma_M$ is polycyclic if and only if $|\det M| = 1$ [**BS80**].

In [**FM00**], Farb and Mosher prove that if a finitely generated group $G$ is quasi-isometric to the group $\Gamma_M$, for an integer matrix $M$ with $|det\ M| > 1$, then a quotient $G/F$ of $G$ by a finite normal subgroup $F$ is virtually isomorphic to a group $\Gamma_N$ defined by an integer matrix $N$ with $|\det N| > 1$.

THEOREM 23.31 (B. Farb, L. Mosher, [**FM00**]). *Let $M_1$ and $M_2$ be integer matrices with $|det\ M_i| > 1,\ i = 1, 2$. The groups $\Gamma_{M_1}$ and $\Gamma_{M_2}$ are quasi-isometric if and only if there exist two positive integers $k_1$ and $k_2$ such that $M_1^{k_1}$ and $M_2^{k_2}$ have the same absolute Jordan form.*

The absolute Jordan form of a matrix is obtained from the Jordan form over $\mathbb{C}$ by replacing the diagonal entries with their absolute values and arranging the Jordan blocks in a canonical way.

In the case of the Baumslag-Solitar solvable groups, which form a subclass in the previous class, more can be said.

THEOREM 23.32 (B. Farb, L. Mosher, [**FM98**], [**FM99**]). *Each solvable Baumslag–Solitar group*

$$BS(1, q) = \left\langle x, y : xyx^{-1} = y^q \right\rangle$$

*is QI rigid.*

This theorem is complemented by

THEOREM 23.33 (K. Whyte, [**Why01**]). *All non-solvable Baumslag–Solitar groups*

$$BS(p, q) = \left\langle x, y : xy^p x^{-1} = y^q \right\rangle ,$$

$|p| \neq 1, |q| \neq 1$ *are QI to each other.*

Since non-solvable Baumslag–Solitar groups are all nonamenable, these results complete QI classification of Baumslag–Solitar groups.

PROBLEM 23.34.
(1) Is the class of finitely generated polycyclic groups QI rigid?
(2) What is the QI classification of finitely generated (finitely presented) polycyclic groups ?

PROBLEM 23.35. Is the class of finitely presented solvable groups QI rigid? Is the class of finitely presented metabelian groups (i. e. solvable groups of derived length 2) QI rigid?

Even the problem of QI rigidity for finitely presented polycyclic abelian-by-cyclic groups has remained open for some time. The recent papers of Eskin, Fisher and Whyte [**EFW06, EFW07**] made a major progress in the direction of this conjecture. In particular, they prove:

THEOREM 23.36. *Consider the class $Poly_3$ of groups $G$ which are not virtually nilpotent and fit in short exact sequences*

$$1 \to \mathbb{Z}^2 \to G \to \mathbb{Z} \to 1.$$

*Then the class $Poly_3$ is QI rigid.*

The groups in $Poly_3$ are the fundamental groups $\Gamma$ of certain 3-dimensional manifolds, which are called $Sol$-manifolds, they are quotients of the 3-dimensional solvable Lie group $Sol_3$ by lattices $\Gamma < Sol_3$.

In view of 23.34, (1), one may ask what is the QI classification of solvable Lie groups. Note that this would settle the QI classification of all connected Lie groups, since any connected Lie group is quasi-isometric to a closed connected group of real upper triangular matrices [**dC08**, Lemma 6.7].

Problem 23.28 can be extended to the following, the positive answer of which would be a major progress in the QI classification of Lie groups.

PROBLEM 23.37 (Y. Cornulier [**dC09**]). Suppose that $G_1, G_2$ are closed connected subgroups of the group of real upper triangular matrices endowed with left-invariant metrics. Is it true that $G_1, G_2$ are quasi-isometric if and only if are they isomorphic?

**Groups QI to abelian-by-abelian solvable groups.** Generalizing the results of [**EFW06, EFW07**], I. Peng in [**Pen11a, Pen11b**] considered quasi-isometries of lattices in solvable Lie groups $G$ of the type

$$G = G_\varphi = \mathbb{R}^n \ltimes_\varphi \mathbb{R}^m$$

where $\varphi : \mathbb{R}^m \to GL(n, \mathbb{R})$ is an action of $\mathbb{R}^m$ on $\mathbb{R}^n$. The number $m$ is called the *rank* of $G$. The group $G$ is clearly a Lie group and we equip $G$ with left-invariant Riemannian metric. Then $G$ admits *horizontal* and *vertical* foliations by the left translates of $\mathbb{R}^n$ and $\mathbb{R}^m$.

DEFINITION 23.38. 1. $\varphi$ is *diagonalizable* if it is conjugate by $GL(n, \mathbb{R})$ to a representation $\mathbb{R}^m \to GL(n, \mathbb{R})$ with diagonal image.

2. $G_\varphi$ is *unimodular* if $\varphi(\mathbb{R}^m) \subset SL(n, \mathbb{R})$.

3. $G_\varphi$ is *nondegenerate* if $\varphi$ is injective.

(Parts 2 and 3 are given in [**Pen11b**] in terms of root systems, but it is easy to see that our definitions are equivalent to hers.)

The main result of [**Pen11a, Pen11b**] is to control quasi-isometries $G_\varphi \to G_\psi$ between the solvable groups, they key being that quasi-isometries send leaves of horizontal and vertical foliations uniformly close to leaves of horizontal and vertical foliations. The precise result is too technical to be stated here (see [**Pen11b**, Theorem 5.2]; below is its main corollary, which is a combination of the work of I. Peng and T. Dymarz (Corollary 1.0.2 in [**Pen11b**]):

THEOREM 23.39 (I. Peng, T. Dymarz). *Suppose that $G_\varphi$ is non-degenerate, unimodular and $\varphi$ is diagonalizable. Then for every finitely generated group $\Gamma$ quasi-isometric to $G$, the group $\Gamma$ is virtually polycyclic.*

As a special case, consider $G$ of rank 1. It is proven by T. Dymarz in [**Dym10**] that

THEOREM 23.40 (Dymarz rank 1 QI rigidity theorem). *Every finitely-generated group QI to $G$ is virtually isomorphic to a lattice in $G$.*

## 23.4. Various other QI rigidity results

We note that the most progress in establishing QI rigidity was achieved in the context of lattices in Lie groups or certain solvable groups. Below we review some QI rigidity results for groups which do not belong to these classes.

The following rigidity theorem was proven by J. Behrstock, B. Kleiner, Y. Minsky and L. Mosher in [**BKMM12**]:

THEOREM 23.41. *Let $S$ be a closed surface of genus $g$ with $n$ punctures, so that $3g-3+n \geqslant 2$ and $(g,n) \neq (1,2)$. Then the Mapping Class group $\Gamma = Map(S)$ of $S$ is strongly QI rigid. Moreover, quasi-isometries of $\Gamma$ are uniformly close to automorphisms of $\Gamma$.*

Note that for a closed surface $S$, the group $Map(S)$ is isomorphic to the group of outer automorphisms $Out(\pi)$, where $\pi = \pi_1(S)$, see [**FM11**]. Furthermore, N. Ivanov [**Iva88**] proved that $Out(Map(S))$ is trivial if $3g-3+n \geqslant 2, (g,n) \neq (2,0)$ and $Out(Map(S)) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ if $(g,n) = (2,0)$,

Recall that for a group $\pi$, the group of outer automorphisms $Out(\pi)$ is the quotient

$$Out(\pi) = Aut(\pi)/Inn(\pi)$$

where $Inn(\pi)$ consists of automorphisms of $\pi$ given by conjugations *via* elements of $\pi$.

PROBLEM 23.42. Is the group $Out(F_n)$ QI rigid?

Artin groups and Coxeter groups are prominent classes of groups which appear frequently in geometric group theory. Note that some of these groups are not QI rigid, e.g., the group $F_2 \times F_2$, see [**BM00**]. In particular, if $G$ is a Coxeter or Artin group which splits as the fundamental group of graph of groups with finite edge groups, where one of the vertex groups $G_v$ is virtually $F_2 \times F_2$, then $G$ cannot be QI rigid. The same applies if one takes a direct product of such $G$ with a Coxeter/Artin group. Also, there are many of Coxeter groups which appear as uniform lattices in $O(n,1)$ (for relatively small $n$). Such Coxeter groups are QI to non-Coxeter lattices in $O(n,1)$. This leads to

PROBLEM 23.43. (a) Suppose that $G$ is an Artin group, which does not contain $F_2 \times F_2$. Is such $G$ QI rigid?

(b) Suppose that $G$ is a non-hyperbolic 1-ended Coxeter group, which does not contain $F_2 \times F_2$. Is $G$ QI rigid?

Note that QI rigidity was proven in [**BKS08**] for certain classes of Artin groups, while Artin braid group (modulo its center) is isomorphic to $Map(S)$, where $S$ is a sphere with several punctures.

Examples of Burger and Moses [**BM00**] show that residual finiteness is not QI invariant, however, the following question seems to be open:

PROBLEM 23.44. Is the Hopfian property geometric?

Gromov and Thurston [**GT87**] constructed some interesting examples of closed negatively curved manifolds, whose fundamental group are not isomorphic to lattices in rank 1 Lie groups. We will refer to these manifolds as *Gromov-Thurston manifolds*. Some of these manifolds are obtained as ramified covers over closed hyperbolic $n$-manifolds ($n \geqslant 4$), ramified over totally-geodesic submanifolds.

Problem 23.45. Are the fundamental groups of Gromov-Thurston $n$-manifolds QI rigid?

The reason one is hopeful that these groups $\Gamma$ are QI rigid is the following. Each $\Gamma$ is associated with a uniform lattice $\Gamma' < O(n,1)$, a sublattice

$$\Gamma'' = \Gamma' \cap O(n-2,1)$$

which, thus, yields a $\Gamma'$-invariant collection of $n-2$-dimensional hyperbolic subspaces $X_i \subset \mathbb{H}^n$, where $X_1$ is $\Gamma''$-invariant. (For instance, Gromov-Thurston manifold could appear as a ramified cover over $\mathbb{H}^n/\Gamma'$ which is ramified over the submanifold $X_1/\Gamma''$.) While the entire hyperbolic $n$-space is highly non-rigid, the pair $(\mathbb{H}^n, \cup_i X_i)$ is QI rigid, see [**Sch97**].

Problem 23.46. Let $S$ be a closed hyperbolic surface. Let $M$ be the 4-dimensional manifold obtained by taking the 2-fold ramified cover over $S \times S$, which is ramified over the diagonal, see [**BGS85**, Exercise 1]. Is $\pi_1(M)$ QI rigid?

Theorem 23.47 (M. Kapovich, B. Leeb, [**KL97**]). *The class of fundamental groups $G$ of closed 3-dimensional Haken 3-manifolds, which are not $Sol_3$-manifolds*[1]*, is QI rigid.*

In view of rigidity for $Sol_3$-groups proven by Eskin, Fisher and Whyte (Theorem 23.36), the above rigidity theorem also holds for the fundamental groups of arbitrary closed 3-manifolds.

On the other hand, it was proven by Behrstock and Neumann [**BN08**] that fundamental groups of all *nongeometric* 3-dimensional graph-manifolds are QI to each other.

Problem 23.48. Classify fundamental groups of non-geometric irreducible 3-dimensional manifolds up to quasi-isometry.

Partial progress towards this problem is achieved in another paper of Behrstock and Neumann [**BN12**]

Theorem 23.49 (P. Papasoglu, [**Pap05**]). *The class of finitely presented groups which split over $\mathbb{Z}$ is QI rigid. Moreover, quasi-isometries of 1-ended groups $G$ preserve the JSJ decomposition of $G$*

Theorem 23.50 (M. Kapovich, B. Kleiner, B. Leeb, [**KKL98**]). *Quasi–isometries preserve de Rham decomposition of the universal covers of closed nonpositively curved Riemannian manifolds.*

D. Burago and B. Kleiner [**BK02**] constructed examples of separated nets in $\mathbb{R}^2$ which are not bi-Lipschitz equivalent.

Problem 23.51 (M. Gromov, [**Gro93**], p. 23; D. Burago, B. Kleiner, [**BK02**]). If two finitely generated groups $G$ and $H$ endowed with word metrics are quasi-isometric, are they bi-Lipschitz equivalent?

Is this at least true when $H = G \times \mathbb{Z}/2\mathbb{Z}$ or when $H$ is a finite index subgroup in $G$?

[1] I.e. excluding $G$ which are polycyclic but not nilpotent.

The following theorem answers a question of Gromov [**Gro93**, § 1.A0]. (See Theorem 16.15 for a version of this result in the context of graphs.) The following theorem is implicitly contained in the paper [**DSS95**] of W. A. Deuber, M. Simonovits and V. T. Sós:

THEOREM 23.52 (K. Whyte, [**Why99**]). *Suppose that $G_1, G_2$ are non-amenable finitely generated groups which are quasi-isometric. Then $G_1, G_2$ are bi-Lipschitz equivalent.*

Prior to this theorem, in was proved by P. Papasoglou in [**Pap95a**] that two free groups are bi-Lipschitz equivalent. This result has some implications in $L^2$–cohomology of groups.

The case of amenable groups was settled (in negative) by T. Dymarz [**Dym10**]. She constructed lamplighter group examples which are quasi-isometric but not bi-Lipschitz homeomorphic. Her examples, however, are commensurable. Hence, one can ask the following:

PROBLEM 23.53. Generate an equivalence relation CLIP on groups by combining commensurability and bi-Lipschitz homeomorphisms. Is CLIP equivalent to the quasi-isometry equivalence relation?

THEOREM 23.54 (R. Sauer [**Sau06**]). *The cohomological dimension $cd_{\mathbb{Q}}$ of a group (over $\mathbb{Q}$) is a QI invariant. Moreover, if $G_1, G_2$ are groups and $f : G_1 \to G_2$ is a quasi-isometric embedding, then $cd_{\mathbb{Q}}(G_1) \leqslant cd_{\mathbb{Q}}(G_2)$.*

We refer to [**Bro82b**] for the definitions of cohomological dimension. Note that partial results on QI invariance of cohomological dimension were proven earlier by S. Gersten, [**Ger93b**] and Y. Shalom, [**Sha04**].

Recall that Property (T) is not a QI invariant (Theorem 17.52). However, the following question is still open:

PROBLEM 23.55. Is a-T-menability a QI invariant?

# Bibliography


[AB56] L. V. Ahlfors and A. Beurling, *The boundary correspondence under quasiconformal mappings*, Acta Math. **96** (1956), 125–142.

[Ady79] S.I. Adyan, *The Burnside problem and identities in groups*, Springer, 1979.

[Ady82] ———, *Random walks on free periodic groups*, Izv. Akad. Nauk SSSR Ser. Mat. **46** (1982), no. 6, 1139–1149.

[Aea91] J. M. Alonso and et al., *Notes on word hyperbolic groups*, Group theory from a geometrical viewpoint (Trieste, 1990), World Sci. Publ., River Edge, NJ, 1991, Edited by H. Short, pp. 3–63.

[Ahl06] L. V. Ahlfors, *Lectures on quasiconformal mappings*, second ed., University Lecture Series, vol. 38, American Mathematical Society, Providence, RI, 2006, With supplemental chapters by C. J. Earle, I. Kra, M. Shishikura and J. H. Hubbard.

[AK11] Y. Algom-Kfir, *Strongly contracting geodesics in outer space*, Geom. Topol. **15** (2011), no. 4, 2181–2233.

[And05] J. W. Anderson, *Hyperbolic geometry*, Springer, 2005.

[Att94] O. Attie, *Quasi-isometry classification of some manifolds of bounded geometry*, Math. Z. **216** (1994), 501–527.

[AVS57] G.M. Adelson-Velskii and Yu. A. Sreider, *The banach mean on groups*, Uspehi Mat. Nauk. (N.S.) **126** (1957), 131–136.

[AW81] C.A. Akemann and M.E. Walter, *Unbounded negative definite functions*, Canad. J. Math. **33** (1981), 862–871.

[Bal95] W. Ballmann, *Lectures on spaces of nonpositive curvature*, DMV Seminar, Band 25, Birkhäuser, 1995.

[Bas72] H. Bass, *The degree of polynomial growth of finitely generated nilpotent groups*, Proc. London Math. Soc. **25** (1972), 603–614.

[Bat99] M. Batty, *Groups with a sublinear isoperimetric inequality*, IMS Bulletin **42** (1999), 5–10.

[Bav91] C. Bavard, *Longueur stable des commutateurs*, Enseign. Math. (2) **37** (1991), no. 1-2, 109–150.

[BC01] R. Bishop and R. Crittenden, *Geometry of manifolds*, AMS Chelsea Publishing, Providence, RI, 2001, Reprint of the 1964 original.

[BC12] J. Behrstock and R. Charney, *Divergence and quasimorphisms of right-angled Artin groups*, Math. Ann. **352** (2012), no. 2, 339–356.

[BDCK66] J. Bretagnolle, D. Dacunha-Castelle, and J.-L. Krivine, *Lois stables et espaces $L^p$*, Ann. IHP, section B **2** (1966), no. 3, 231–259.

[BdlHV08] B. Bekka, P. de la Harpe, and A. Valette, *Kazhdan's property (T)*, New Mathematical Monographs, vol. 11, Cambridge University Press, Cambridge, 2008.

[BDM09] J. Behrstock, C. Druţu, and L. Mosher, *Thick metric spaces, relative hyperbolicity, and quasi-isometric rigidity*, Math. Annalen **344** (2009), 543–595.

[BE12] L. Bartholdi and A. Erschler, *Growth of permutational extensions*, Invent. Math. **189** (2012), no. 2, 431–455.

[Bea83] A.F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, vol. 91, Springer, 1983.

[Ber68] G. Bergman, *On groups acting on locally finite graphs*, Annals of Math. **88** (1968), 335–340.

[BFGM07] U. Bader, A. Furman, T. Gelander, and N. Monod, *Property (T) and rigidity for actions on Banach spaces*, Acta Math. **198** (2007), 57–105.

[BFH97] M. Bestvina, M. Feighn, and M. Handel, *Laminations, trees, and irreducible automorphisms of free groups*, Geom. Funct. Anal. **7** (1997), 215–244.

[BFH00] ______, *The Tits alternative for* $\mathrm{Out}(F_n)$*. I. Dynamics of exponentially-growing automorphisms.*, Ann. of Math. **151** (2000), no. 2, 517–623.

[BFH04] ______, *Solvable subgroups of* $\mathrm{Out}(F_n)$ *are virtually Abelian*, Geom. Dedicata **104** (2004), 71–96.

[BFH05] ______, *The Tits alternative for* $\mathrm{Out}(F_n)$*. II. A Kolchin type theorem*, Ann. of Math. (2) **161** (2005), no. 1, 1–59.

[BG03] E. Breuillard and T. Gelander, *On dense free subgroups of Lie groups*, J. Algebra **261** (2003), no. 2, 448–467.

[BGS85] W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of non-positive curvature*, Progress in Math., vol. 61, Birkhauser, 1985.

[BH99] M. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Springer-Verlag, Berlin, 1999.

[BH05] M. Bridson and J. Howie, *Conjugacy of finite subsets in hyperbolic groups*, Internat. J. Algebra Comput. **5** (2005), 725–756.

[Bie76a] R. Bieri, *Homological dimension of discrete groups*, Mathematical Notes, Queen Mary College, 1976.

[Bie76b] ______, *Normal subgroups in duality groups and in groups of cohomological dimension* 2, J. Pure Appl. Algebra **1** (1976), 35–51.

[Bie79] ______, *Finitely presented soluble groups*, Séminaire d'Algèbre Paul Dubreil 31ème année (Paris, 1977–1978), Lecture Notes in Math., vol. 740, Springer, Berlin, 1979, pp. 1–8.

[Big01] S. Bigelow, *Braid groups are linear*, J. Amer. Math. Soc. **14** (2001), no. 2, 471–486 (electronic).

[BK98] D. Burago and B. Kleiner, *Separated nets in Euclidean space and Jacobians of bi-Lipschitz maps*, Geom. Funct. Anal. **8** (1998), no. 2, 273–282.

[BK02] ______, *Rectifying separated nets*, Geom. Funct. Anal. **12** (2002), 80–92.

[BKMM12] J. Behrstock, B. Kleiner, Y. Minsky, and L. Mosher, *Geometry and rigidity of mapping class groups*, Geom. Topol. **16** (2012), no. 2, 781–888.

[BKS08] M. Bestvina, B. Kleiner, and M. Sageev, *The asymptotic geometry of right-angled Artin groups. I*, Geom. Topol. **12** (2008), no. 3, 1653–1699.

[BM00] M. Burger and S. Mozes, *Lattices in product of trees*, Inst. Hautes Études Sci. Publ. Math. (2000), no. 92, 151–194.

[BN08] J. Behrstock and W. Neumann, *Quasi-isometric classification of graph manifold groups*, Duke Math. J. **141** (2008), 217–240.

[BN12] ______, *Quasi-isometric classification of non-geometric 3-manifold groups*, Crelle Journal **669** (2012), 101–120.

[Bol79] B. Bollobás, *Graph theory, an introductory course*, Graduate Texts in Mathematics, vol. 63, Springer, 1979.

[Boo57] W. Boone, *Certain simple, unsolvable problems of group theory. V, VI*, Nederl. Akad. Wetensch. Proc. Ser. A. 60 = Indag. Math. **19** (1957), 22–27, 227–232.

[Bou63] N. Bourbaki, *Éléments de mathématique. Fascicule XXIX. Livre VI: Intégration. Chapitre 7: Mesure de Haar. Chapitre 8: Convolution et représentations*, Actualités Scientifiques et Industrielles, No. 1306, Hermann, Paris, 1963.

[Bou65] ______, *Topologie générale*, Hermann, Paris, 1965.

[Bou02] ______, *Lie groups and Lie algebras. Chapters 4–6*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 2002, Translated from the 1968 French original by Andrew Pressley.

[Bow91] B. Bowditch, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, Group theory from a geometrical viewpoint (Trieste, 1990), World Sci. Publ., River Edge, NJ, 1991, pp. 64–167.

[Bow94] ______, *Some results on the geometry of convex hulls in manifolds of pinched negative curvature*, Comment. Math. Helv. **69** (1994), no. 1, 49–81.

[Bow95] ______, *A short proof that a sub-quadratic isoperimetric inequality implies a linear one*, Mich. J. Math. **42** (1995), 103–107.

[Bow97] ______, *Relatively hyperbolic groups*, preprint, University of Southampton `http://www.maths.soton.ac.uk/pure/preprints.phtml`, 1997.

[Bow98] ———, *A topological characterisation of hyperbolic groups*, J. Amer. Math. Soc. **11** (1998), 643–667.

[Bow06] ———, *A course on geometric group theory*, MSJ Memoirs, vol. 16, Mathematical Society of Japan, Tokyo, 2006.

[Boż80] M. Bożejko, *Uniformly amenable discrete groups*, Math. Ann. **251** (1980), 1–6.

[BP92] R. Benedetti and C. Pertonio, *Lectures on hyperbolic geometry*, Springer, 1992.

[BP00] M. Bourdon and H. Pajot, *Rigidity of quasi-isometries for some hyperbolic buildings*, Comment. Math. Helv. **75** (2000), no. 4, 701–736.

[BP03] ———, *Cohomologie $l_p$ et espaces de Besov*, J. Reine Angew. Math. **558** (2003), 85–108.

[Bre92] L. Breiman, *Probability*, Classics in Applied Mathematics, vol. 7, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, Corrected reprint of the 1968 original.

[Bro81a] R. Brooks, *The fundamental group and the spectrum of the Laplacian*, Comment. Math. Helv. **56** (1981), no. 4, 581–598.

[Bro81b] ———, *Some remarks on bounded cohomology*, Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud., vol. 97, Princeton Univ. Press, Princeton, N.J., 1981, pp. 53–63.

[Bro82a] ———, *Amenability and the spectrum of the Laplacian*, Bull. Amer. Math. Soc. (N.S.) **6** (1982), no. 1, 87–89.

[Bro82b] K.S. Brown, *Cohomology of Groups*, Graduate Texts in Mathematics, vol. 87, Springer, 1982.

[BS66] A. I. Borevich and I. R. Shafarevich, *Number theory*, Translated from the Russian by Newcomb Greenleaf. Pure and Applied Mathematics, Vol. 20, Academic Press, New York, 1966.

[BS69] J. Bell and A. Slomson, *Models and Ultraproducts*, North-Holland, Amsterdam, 1969.

[BS78] R. Bieri and R. Strebel, *Almost finitely presented soluble groups*, Comm. Math. Helv. **53** (1978), 258–278.

[BS80] ———, *Valuations and finitely presented metabelian groups*, Proc. London Math. Soc. **41** (1980), no. 3, 439–464.

[BS97a] W. Ballmann and J. Świątkowski, *On $l^2$-cohomology and property (T) for automorphism groups of polyhedral cell complexes*, Geom. Funct. Anal. **7** (1997), no. 4, 615–645.

[BS97b] I. Benjamini and O. Schramm, *Every graph with a positive Cheeger constant contains a tree with a positive Cheeger constant*, Geom. Funct. Anal. **7** (1997), 403–419.

[BS00] M. Bonk and O. Schramm, *Embeddings of Gromov hyperbolic spaces*, Geom. Funct. Anal. **10** (2000), no. 2, 266–306.

[BT24] S. Banach and A. Tarski, *Sur la décomposition des ensembles de points en parties respectivement congruentes*, Fundamenta Mathematicae **6** (1924), 244–277.

[BT02] J. Burillo and J. Taback, *Equivalence of geometric and combinatorial Dehn functions*, New York J. Math. **8** (2002), 169–179 (electronic).

[Bur99] J. Burillo, *Dimension and fundamental groups of asymptotic cones*, J. London Math. Soc. (2) **59** (1999), 557–572.

[Bus65] H. Busemann, *Extremals on closed hyperbolic space forms*, Tensor (N.S.) **16** (1965), 313–318.

[Bus82] P. Buser, *A note on the isoperimetric constant*, Ann. Sci. École Norm. Sup. (4) **15** (1982), no. 2, 213–230.

[Bus10] ———, *Geometry and spectra of compact Riemann surfaces*, Modern Birkhäuser Classics, Birkhäuser Boston Inc., Boston, MA, 2010, Reprint of the 1992 edition.

[BW11] M. Bridson and H. Wilton, *On the difficulty of presenting finitely presentable groups*, Groups Geom. Dyn. **5** (2011), no. 2, 301–325.

[Cal08] D. Calegari, *Length and stable length*, Geom. Funct. Anal. **18** (2008), no. 1, 50–76. MR MR2399095

[Can84] J. Cannon, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata **16** (1984), no. 2, 123–148.

[CCJ$^+$01] P.-A. Cherix, M. Cowling, P. Jolissaint, P. Julg, and A. Valette, *Groups with the Haagerup property*, Progress in Mathematics, vol. 197, Birkhäuser, 2001.

[CCS04] P.-A. Cherix, M. Cowling, and B. Straub, *Filter products of $C_0$-semigroups and ultraproduct representations for Lie groups*, J. Funct. Anal. **208** (2004), no. 1, 31–63.

[CDH10] I. Chatterji, C. Druţu, and F. Haglund, *Kazhdan and Haagerup properties from the median viewpoint*, Adv. Math. **225** (2010), no. 2, 882–921.

[CDP90] M. Coornaert, T. Delzant, and A. Papadopoulos, *Géométrie et théorie des groupes*, Lec. Notes Math., vol. 1441, Springer-Verlag, 1990.

[CGT82] J. Cheeger, M. Gromov, and M. Taylor, *Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds*, J. Differential Geom. **17** (1982), no. 1, 15–53.

[Cha01] I. Chavel, *Isoperimetric inequalities*, Cambridge Tracts in Mathematics, vol. 145, Cambridge University Press, Cambridge, 2001, Differential geometric and analytic perspectives.

[Cha06] ———, *Riemannian geometry*, second ed., Cambridge Studies in Advanced Mathematics, vol. 108, Cambridge University Press, Cambridge, 2006, A modern introduction.

[Che70] J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, Problems in analysis (Papers dedicated to Salomon Bochner, 1969), Princeton Univ. Press, Princeton, N. J., 1970, pp. 195–199.

[Cho80] C. Chou, *Elementary amenable groups*, Illinois J. Math. **24** (1980), no. 3, 396–407.

[Cho96] R. Chow, *Groups quasi-isometric to complex hyperbolic space*, Trans. Amer. Math. Soc. **348** (1996), no. 5, 1757–1769.

[CJ94] A. Casson and D. Jungreis, *Convergence groups and Seifert-fibered 3-manifolds*, Inventiones Mathematicae **118** (1994), 441–456.

[CM70] P. Chernoff and J. Marsden, *On continuity and smoothness of group actions*, Bull. Amer. Math. Soc. **76** (1970), 1044–1049.

[Com84] W. W. Comfort, *Topological groups*, Handbook of set-theoretic topology, North-Holland, Amsterdam, 1984, pp. 1143–1263.

[CSC93] Th. Coulhon and L. Saloff-Coste, *Isopérimétrie pour les groupes et variétés*, Rev. Math. Iberoamericana **9** (1993), 293–314.

[CSGdlH98] T. Ceccherini-Silberstein, R. Grigorchuk, and P. de la Harpe, *Décompositions paradoxales des groupes de Burnside*, C. R. Acad. Sci. Paris Sér. I Math. **327** (1998), no. 2, 127–132.

[Cut01] N. J. Cutland, *Loeb measures in practice: Recent advances*, Lecture Notes in Mathematics, vol. 1751, Springer-Verlag, Berlin, 2001.

[CY75] S. Y. Cheng and S. T. Yau, *Differential equations on Riemannian manifolds and their geometric applications*, Comm. Pure Appl. Math. **28** (1975), no. 3, 333–354.

[Dah03a] F. Dahmani, *Combination of convergence groups*, Geometry and Topology **7** (2003), 933–963.

[Dah03b] ———, *Les groupes relativement hyperboliques et leurs bords*, Ph.D. thesis, University of Strasbourg, 2003.

[Dav08] M. Davis, *The geometry and topology of Coxeter groups*, London Mathematical Society Monographs Series, vol. 32, Princeton University Press, Princeton, NJ, 2008.

[Day57] M.M. Day, *Amenable semigroups*, Illinois J. Math. **1** (1957), 509–544.

[dC92] M. P. do Carmo, *Riemannian geometry*, Birkhäuser Boston Inc., Boston, MA, 1992.

[dC06] Y. de Cornulier, *Strongly bounded groups and infinite powers of finite groups*, Comm. Algebra **34** (2006), 2337–2345.

[dC08] ———, *Dimension of asymptotic cones of lie groups*, J. Topology **1** (2008), no. 2, 343–361.

[dC09] ———, *Asymptotic cones of Lie groups and cone equivalences*, preprint, arXiv:0907.2546, 2009.

[dCTV08] Y. de Cornulier, R. Tessera, and A. Valette, *Isometric group actions on Banach spaces and representations vanishing at infinity*, Transform. Groups **13** (2008), no. 1, 125–147.

[DD89] W. Dicks and M. J. Dunwoody, *Groups acting on graphs*, Cambridge University Press, Cambridge-New York, 1989.

[dDW84] L. Van den Dries and A. Wilkie, *Gromov's theorem on groups of polynomial growth and elementary logic*, J. Algebra **89** (1984), 349–374.

[Del77] P. Delorme, 1-*cohomologie des représentations unitaires des groupes de Lie semi-simples et résolubles. Produits tensoriels et représentations*, Bull. Soc. Math. France **105** (1977), 281–336.

[Del96] T. Delzant, *Sous-groupes distingués et quotients des groupes hyperboliques*, Duke Math. J. **83** (1996), no. 3, 661–682.

[DG] T. Delzant and M. Gromov, *Groupes de Burnside et géométrie hyperbolique*, preprint.

[DG11] F. Dahmani and V. Guirardel, *The isomorphism problem for all hyperbolic groups*, Geom. Funct. Anal. **21** (2011), 223–300.

[Dix50] J. Dixmier, *Les moyennes invariantes dans les semi-groups et leurs applications*, Acta Sci. Math. Szeged **12** (1950), 213–227.

[dlH73] P. de la Harpe, *Moyennabilité de quelques groupes topologiques de dimension infinie*, C. R. Acad. Sci. Paris Sér. I Math. **277** (1973), no. 14, 1037–1040.

[dlH00] ———, *Topics in geometric group theory*, Chicago Lectures in Mathematics, University of Chicago Press, 2000.

[dlHGCS99] P. de la Harpe, R. I. Grigorchuk, and T. Chekerini-Sil′berstaĭn, *Amenability and paradoxical decompositions for pseudogroups and discrete metric spaces*, Proc. Steklov Inst. Math. **224** (1999), 57–97.

[dlHV89] P. de la Harpe and A. Valette, *La propriété (T) de Kazhdan pour les groupes localemant compacts*, Astérisque, vol. 175, Société Mathématique de France, 1989.

[Dod84] J. Dodziuk, *Difference equations, isoperimetric inequality and transience of certain random walks*, Trans. Amer. Math. Soc. **284** (1984), no. 2, 787–794.

[DP98] A. Dyubina and I. Polterovich, *Structures at infinity of hyperbolic spaces*, Russian Math. Surveys **53** (1998), no. 5, 1093–1094.

[DP01] ———, *Explicit constructions of universal* $\mathbb{R}$*–trees and asymptotic geometry of hyperbolic spaces*, Bull. London Math. Soc. **33** (2001), 727–734.

[DR09] M. Duchin and K. Rafi, *Divergence of geodesics in Teichmüller space and the Mapping Class group*, GAFA **19** (2009), 722–742.

[Dru01] C. Druţu, *Cônes asymptotiques et invariants de quasi-isométrie pour des espaces métriques hyperboliques*, Ann. Inst. Fourier (Grenoble) **51** (2001), no. 1, 81–97.

[Dru09] ———, *Relatively hyperbolic groups: geometry and quasi-isometric invariance*, Comment. Math. Helv. **84** (2009), no. 3, 503–546.

[DS84] P. G. Doyle and J. L. Snell, *Random walks and electric networks*, Carus Mathematical Monographs, vol. 22, Mathematical Association of America, Washington, DC, 1984.

[DS05] C. Druţu and M. Sapir, *Tree-graded spaces and asymptotic cones of groups*, Topology **44** (2005), 959–1058, with an appendix by D. Osin and M. Sapir.

[DSS95] W. A. Deuber, M. Simonovits, and V. T. Sós, *A note on paradoxical metric spaces*, Studia Sci. Math. Hungar. **30** (1995), no. 1-2, 17–23.

[Dun85] M. J. Dunwoody, *The accessibility of finitely presented groups*, Inventiones Mathematicae **81** (1985), 449–457.

[Dun93] ———, *An inaccessible group*, Geometric group theory, Vol. 1 (Sussex, 1991) (G. A. Niblo and M. A. Roller, eds.), London Math. Soc. Lecture Note Ser., vol. 181, Cambridge Univ. Press, Cambridge, 1993, pp. 75–78.

[Dym10] T. Dymarz, *Bilipschitz equivalence is not equivalent to quasi-isometric equivalence for finitely generated groups*, Duke Math. J. **154** (2010), no. 3, 509–526.

[Dyu00] A. Dyubina, *Instability of the virtual solvability and the property of being virtually torsion-free for quasi-isometric groups*, Internat. Math. Res. Notices **21** (2000), 1097–1101.

[Ebe72] P. Eberlein, *Geodesic flow on certain manifolds without conjugate points*, Transaction of AMS **167** (1972), 151– 170.

[ECH+92] D.B.A. Epstein, J. Cannon, D.F. Holt, S. Levy, M.S. Paterson, and W.P. Thurston, *Word Processing and Group Theory*, Jones and Bartlett, 1992.

[EF97a] D. B. A. Epstein and K. Fujiwara, *The second bounded cohomology of word-hyperbolic groups*, Topology **36** (1997), no. 6, 1275–1289.

[EF97b] A. Eskin and B. Farb, *Quasi-flats and rigidity in higher rank symmetric spaces*, J. Amer. Math. Soc. **10** (1997), no. 3, 653–692.

[Efr53] V. A. Efremovič, *The proximity geometry of riemannian manifolds*, Uspehi Matem. Nauk (N.S.) **8** (1953), 189.

[EFW06] A. Eskin, D. Fisher, and K. Whyte, *Coarse differentiation of quasi-isometries I: Spaces not quasi-isometric to Cayley graphs*, math/0607207, 2006.

[EFW07] ———, *Coarse differentiation of quasi-isometries II: Rigidity for Sol and Lamplighter groups*, arXiv:0706.0940, 2007.

[EO73] P. Eberlein and B. O'Neill, *Visibility manifolds*, Pacific J. Math. **46** (1973), 45–109.

[Ers03] A. Erschler, *On Isoperimetric Profiles of Finitely Generated Groups*, Geometriae Dedicata **100** (2003), 157–171.

[Ers04] ———, *Not residually finite groups of intermediate growth, commensurability and non-geometricity*, J. Algebra **272** (2004), no. 1, 154–172.

[Ers06] ———, *Piecewise automatic groups*, Duke Math. J. **134** (2006), no. 3, 591–613.

[Esk98] A. Eskin, *Quasi-isometric rigidity of nonuniform lattices in higher rank symmetric spaces*, J. Amer. Math. Soc. **11** (1998), no. 2, 321–361.

[ET64] V. Efremovich and E. Tihomirova, *Equimorphisms of hyperbolic spaces*, Izv. Akad. Nauk SSSR **28** (1964), 1139– 1144.

[Far98] B. Farb, *Relatively hyperbolic groups*, Geom. Funct. Anal. **8** (1998), no. 5, 810–840.

[Fed69] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.

[FJ03] R. Fleming and J. Jamison, *Isometries on Banach spaces: Function spaces*, Monographs and Surveys in Pure and Applied Mathematics, CRC Press, 2003.

[FM98] B. Farb and L. Mosher, *A rigidity theorem for the solvable Baumslag–Solitar groups*, Invent. Math. **131** (1998), no. 2, 419–451, With an appendix by Daryl Cooper.

[FM99] ———, *Quasi-isometric rigidity for the solvable Baumslag–Solitar groups, II*, Invent. Math. **137** (1999), no. 3, 613–649.

[FM00] ———, *On the asymptotic geometry of abelian-by-cyclic groups*, Acta Math. **184** (2000), no. 2, 145–202.

[FM11] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton University Press, 2011.

[Fø55] E. Følner, *On groups with full banach mean value*, Math. Scand. **3** (1955), 243–254.

[Fol99] G. Folland, *Real analysis: Modern techniques and their applications*, Pure and Applied Mathematics, Wiley–Interscience, 1999.

[Fox53] R. H. Fox, *Free differential calculus. I. Derivation in the free group ring*, Ann. of Math. (2) **57** (1953), 547–560.

[Fri60] A. A. Fridman, *On the relation between the word problem and the conjugacy problem in finitely defined groups*, Trudy Moskov. Mat. Obšč. **9** (1960), 329–356.

[FS88] M. Freedman and R. Skora, *Strange actions of groups on spheres. II*, "Holomorphic functions and moduli", Vol. II (Berkeley, CA, 1986), Springer, New York, 1988, pp. 41–57.

[FS96] B. Farb and R. Schwartz, *The large-scale geometry of Hilbert modular groups*, J. Differential Geom. **44** (1996), no. 3, 435–478.

[FW91] M. Foreman and F. Wehrung, *The Hahn-Banach theorem implies the existence of a non-Lebesgue measurable set*, Fundam. Math. **138** (1991), 13–19.

[G6̈0] P. Günther, *Einige Sätze über das Volumenelement eines Riemannschen Raumes*, Publ. Math. Debrecen **7** (1960), 78–93.

[Gab92] D. Gabai, *Convergence groups are Fuchsian groups*, Annals of Math. **136** (1992), 447–510.

[Gau73] C. F. Gauß, *Werke. Band VIII*, Georg Olms Verlag, Hildesheim, 1973, Reprint of the 1900 original.

[GdlH90] E. Ghys and P. de la Harpe, *Sur les groupes hyperbolic d'apres Mikhael Gromov*, Progress in Mathematics, vol. 83, Birkhäuser, 1990.

[Geo08] R. Geoghegan, *Topological methods in group theory*, Graduate Texts in Mathematics, vol. 243, Springer, New York, 2008.

[Ger87] S. Gersten, *Reducible diagrams and equations over groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 15–73.

[Ger92] ———, *Bounded cocycles and combings of groups*, Internat. J. Algebra Comput. **2** (1992), no. 3, 307–326.

[Ger93a] ______, *Isoperimetric and isodiametric functions of finite presentations*, Geometric group theory, Vol. 1 (Sussex, 1991), London Math. Soc. Lecture Note Ser., vol. 181, Cambridge Univ. Press, Cambridge, 1993, pp. 79–96.

[Ger93b] ______, *Quasi-isometry invariance of cohomological dimension*, C. R. Acad. Sci. Paris Sér. I Math. **316** (1993), no. 5, 411–416.

[Ger94] ______, *Quadratic divergence of geodesics in CAT(0)-spaces*, Geom. Funct. Anal. **4** (1994), no. 1, 37–51.

[GHL04] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, third ed., Universitext, Springer-Verlag, Berlin, 2004.

[Ghy04] E. Ghys, *Groupes aléatoires (d'après Misha Gromov,. . . )*, Astérisque (2004), no. 294, viii, 173–204.

[GM07] Y. Glasner and N. Monod, *Amenable actions, free products and fixed-point property*, Bull. London Math. Society **39** (2007), 138–150.

[Gol98] R. Goldblatt, *Lectures on the hyperreals*, Graduate Texts in Mathematics, vol. 188, Springer-Verlag, New York, 1998, An introduction to nonstandard analysis.

[GP10] V. Guillemin and A. Pollack, *Differential topology*, AMS Chelsea Publishing, Providence, RI, 2010, Reprint of the 1974 original.

[GPS88] M. Gromov and I. Piatetski-Shapiro, *Non-arithmetic groups in Lobacevski spaces*, Publ. Math. IHES **66** (1988), 93–103.

[Gri83] R. I. Grigorchuk, *On the Milnor problem of group growth*, Dokl. Akad. Nauk SSSR **271** (1983), no. 1, 30–33.

[Gri84a] ______, *Degrees of growth of finitely generated groups and the theory of invariant means*, Izv. Akad. Nauk SSSR Ser. Mat. **48** (1984), no. 5, 939–985.

[Gri84b] ______, *The growth rate of finitely generated groups and the theory of invariant means*, Inv. Akad. Nauk. **45** (1984), 939–986.

[Gri87] ______, *Superamenability and the problem of occurence of free semigroups*, Functional Analysis and its Applications **21** (1987), 64–66.

[Gro81] M. Gromov, *Groups of polynomial growth and expanding maps*, Publ. Math. IHES **53** (1981), 53–73.

[Gro82] ______, *Volume and bounded cohomology*, Publ. Math. IHES (1982), no. 56, 5–99 (1983).

[Gro83] ______, *Infinite groups as geometric objects*, Proceedings of the International Congress of Mathematicians, Warsaw, Amer. Math. Soc., 1983, pp. 385–392.

[Gro86] ______, *Isoperimetric inequalities in Riemannian manifolds*, "Asymptotic Theory of Finite Dimensional Normed Spaces", Lecture Notes Math., vol. 1200, Springer-Verlag, Berlin, 1986, pp. 114–129.

[Gro87] ______, *Hyperbolic groups*, "Essays in group theory", Math. Sci. Res. Ins. Publ., vol. 8, Springer, 1987, pp. 75–263.

[Gro93] ______, *Asymptotic invariants of infinite groups*, "Geometric groups theory", volume 2, Proc. of the Symp. in Sussex 1991 (G.A. Niblo and M.A. Roller, eds.), Lecture Notes series, vol. 182, Cambridge University Press, 1993.

[Gro03] ______, *Random walk in random groups*, Geom. Funct. Anal. **13** (2003), no. 1, 73–146.

[Gro07] ______, *Metric structures for Riemannian and non-Riemannian spaces*, Modern Birkhäuser Classics, Birkhäuser Boston Inc., Boston, MA, 2007, based on the 1981 French original, with appendices by M. Katz, P. Pansu and S. Semmes.

[GS90] S. Gersten and H. Short, *Small cancellation theory and automatic groups*, Invent. Math. **102** (1990), 305–334.

[GT87] M. Gromov and W. Thurston, *Pinching constants for hyperbolic manifolds*, Inventiones Mathematicae **89** (1987), 1–12.

[Gui70] Y. Guivarc'h, *Groupes de Lie à croissance polynomiale*, C. R. Acad. Sci. Paris, Ser. A—B **271** (1970), A237–A239.

[Gui73] ______, *Croissance polynomiale et périodes des fonctions harmonique*, Bull. Soc. Math. France **101** (1973), 333–379.

[Gui77] A. Guichardet, *Étude de la* 1*-cohomologie et de la topologie du dual pour les groupes de Lie à radical abélien*, Math. Ann. **228** (1977), no. 1, 215–232.

[Hal64] J. Halpern, *The independence of the axiom of choice from the boolean prime ideal theorem*, Fund. Math. **55** (1964), 57–66.

[Hal76] M. Hall, *The theory of groups*, Chelsea Publishing Co., New York, 1976, Reprinting of the 1968 edition.

[Hat02] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[Hei01] J. Heinonen, *Lectures on analysis on metric spaces*, Universitext, Springer-Verlag, New York, 2001.

[Hei05] ———, *Lectures on Lipschitz analysis*, Report. University of Jyväskylä Department of Mathematics and Statistics, vol. 100, University of Jyväskylä, Jyväskylä, 2005.

[Hel01] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Graduate Studies in Mathematics, vol. 34, Amer. Math. Soc., 2001.

[Hem78] J. Hempel, *3-manifolds*, Annals of Mathematics Studies, Princeton University Press, 1978.

[Hig40] G. Higman, *The units of group-rings*, Proc. London Math. Soc. (2) **46** (1940), 231–248.

[Hir76] M. Hirsch, *Differential topology*, Graduate Texts in Mathematics, vol. 33, Springer, 1976.

[Hir90] J. Hirschfeld, *The nonstandard treatment of Hilbert's fifth problem*, Trans. Amer. Math. Soc. **321** (1990), no. 1, 379–400.

[HJ99] K. Hrbacek and T. Jech, *Introduction to set theory*, third ed., Monographs and Textbooks in Pure and Applied Mathematics, vol. 220, Marcel Dekker Inc., New York, 1999.

[HNN49] G. Higman, B. H. Neumann, and H. Neumann, *Embedding theorems for groups*, J. London Math. Soc. **24** (1949), 247–254. MR MR0032641 (11,322d)

[HP74] E. Hille and R. Phillips, *Functional analysis and semi-groups*, American Mathematical Society, Providence, R. I., 1974, Third printing of the revised edition of 1957, American Mathematical Society Colloquium Publications, Vol. XXXI.

[Hum75] J. E. Humphreys, *Linear algebraic groups*, Springer-Verlag, New York, 1975, Graduate Texts in Mathematics, No. 21.

[Hun80] T. W. Hungerford, *Algebra*, Graduate Texts in Mathematics, vol. 73, Springer-Verlag, New York, 1980.

[HW41] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, 1941.

[IM01] T. Iwaniec and G. Martin, *Geometric function theory and non-linear analysis*, Oxford Mathematical Monographs, Oxford University Press, New York, 2001.

[IS98] S. V. Ivanov and P. E. Schupp, *On the hyperbolicity of small cancellation groups and one-relator groups*, Trans. Amer. Math. Soc. **350** (1998), no. 5, 1851–1894.

[Isa90] N. A. Isachenko, *Uniformly quasiconformal discontinuous groups that are not isomorphic to Möbius groups*, Dokl. Akad. Nauk SSSR **313** (1990), no. 5, 1040–1043.

[Iva88] N. Ivanov, *Automorphisms of Teichmüller modular groups*, Topology and geometry—Rohlin Seminar, Lecture Notes in Math., vol. 1346, Springer, Berlin, 1988, pp. 199–270.

[Iva92] ———, *Subgroups of Teichmüller Modular Groups*, Translations of Math. Monographs, vol. 115, AMS, 1992.

[Iva94] S. V. Ivanov, *The free Burnside groups of sufficiently large exponents*, Internat. J. Algebra Comput. **4** (1994), no. 1-2, ii+308 pp.

[Jec03] Th. Jech, *Set Theory: The Third Millennium Edition, Revised and Expanded*, Springer, 2003.

[Joh48] F. John, *Extremum problems with inequalities as subsidiary conditions*, Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, Interscience Publishers, Inc., New York, N. Y., 1948, pp. 187–204.

[JR88] W. Jaco and J. H. Rubinstein, *PL minimal surfaces in 3-manifolds*, J. Diff. Geom. **27** (1988), no. 3, 493–524.

[JR89] ———, *PL equivariant surgery and invariant decomposition of 3-manifolds*, Advances in Math. **73** (1989), 149–191.

[JvN35] P. Jordan and J. von Neumann, *On inner products in linear metric spaces*, Ann. of Math. **36** (1935), 719–723.

[Kak41] S. Kakutani, *Concrete representation of abstract* $(L)$*-spaces and the mean ergodic theorem*, Ann. of Math. (2) **42** (1941), 523–537.

[Kap92] M. Kapovich, *Intersection pairing on hyperbolic 4-manifolds*, preprint, 1992.

[Kap96] I. Kapovich, *Detecting quasiconvexity: algorithmic aspects*, Geometric and computational perspectives on infinite groups (Minneapolis, MN and New Brunswick, NJ, 1994), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 25, Amer. Math. Soc., Providence, RI, 1996, pp. 91–99.

[Kap01] M. Kapovich, *Hyperbolic manifolds and discrete groups*, Birkhäuser Boston Inc., Boston, MA, 2001.

[Kap05] ———, *Representations of polygons of finite groups*, Geom. Topol. **9** (2005), 1915–1951.

[Kap07] ———, *Energy of harmonic functions and Gromov's proof of the Stallings' theorem*, Preprint, arXiv:0707.4231, 2007.

[Kei76] J. Keisler, *Foundations of Infinitesimal Calculus*, Prindel-Weber-Schmitt, Boston, 1976.

[Kel72] G. Keller, *Amenable groups and varieties of groups*, Illinois J. Math. **16** (1972), 257–268.

[Kes59] H. Kesten, *Full Banach mean values on countable groups*, Math. Scand. **7** (1959), 146–156.

[KK00] M. Kapovich and B. Kleiner, *Hyperbolic groups with low-dimensional boundary*, Ann. Sci. École Norm. Sup. (4) **33** (2000), no. 5, 647–669.

[KK05] ———, *Coarse Alexander duality and duality groups*, Journal of Diff. Geometry **69** (2005), 279–352.

[KK07] ———, *Weak hyperbolization conjecture for 3-dimensional CAT(0) groups*, Geometry, Groups and Dynamics **1** (2007), 67–79.

[KKL98] M. Kapovich, B. Kleiner, and B. Leeb, *Quasi-isometries and the de Rham decomposition*, Topology **37** (1998), 1193–1212.

[KL95] M. Kapovich and B. Leeb, *On asymptotic cones and quasi-isometry classes of fundamental groups of nonpositively curved manifolds*, Geom. Funct. Anal. **5** (1995), no. 3, 582–603.

[KL97] ———, *Quasi-isometries preserve the geometric decomposition of Haken manifolds*, Inventiones Mathematicae **128** (1997), 393–416.

[KL98a] ———, *3-manifold groups and nonpositive curvature*, Geom. Funct. Anal. **8** (1998), no. 5, 841–852.

[KL98b] B. Kleiner and B. Leeb, *Rigidity of quasi-isometries for symmetric spaces and Euclidean buildings*, Math. Publ. of IHES **86** (1998), 115–197.

[KL01] ———, *Groups quasi-isometric to symmetric spaces*, Comm. Anal. Geom. **9** (2001), no. 2, 239–260.

[KL09] ———, *Induced quasi-actions: A remark*, Proc. Amer. Math. Soc. **137** (2009), no. 5, 1561–1567.

[Kle10] B. Kleiner, *A new proof of Gromov's theorem on groups of polynomial growth*, J. Amer. Math. Soc. **23** (2010), no. 3, 815–829.

[KS08] I. Kapovich and P. Schupp, *On group-theoretic models of randomness and genericity*, Groups Geom. Dyn. **2** (2008), no. 3, 383–404.

[KSTT05] L. Kramer, S. Shelah, K. Tent, and S. Thomas, *Asymptotic cones of finitely presented groups*, Adv. Math. **193** (2005), no. 1, 142–173.

[Kun80] K. Kunen, *Set Theory: An Introduction to Independence Proofs*, Elsevier, 1980.

[KW92] V. Kaimanovich and W. Woess, *The Dirichlet problem at infinity for random walks on graphs with a strong isoperimetric inequality*, Probab. Theory Related Fields **91** (1992), 445–466.

[Lan64] S. Lang, *Algebraic numbers*, Addison-Wesley Publishing Co., Inc., Reading, Mass.-Palo Alto-London, 1964.

[Leh87] O. Lehto, *Univalent functions and Teichmüller spaces*, Springer, 1987.

[Li04] P. Li, *Harmonic functions and applications to complete manifolds*, Preprint, 2004.

[ŁRN51] J. Łoś and C. Ryll-Nardzewski, *On the application of Tychonoff's theorem in mathematical proofs*, Fund. Math. **38** (1951), 233–237.

[LS77] R.C. Lyndon and P.E. Schupp, *Combinatorial group theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 89, Springer, 1977.

[LT79] J. Lindenstrauss and L. Tzafriri, *Classical Banach spaces. II*, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], vol. 97, Springer-Verlag, Berlin, 1979, Function spaces.

[LT92] P. Li and L. Tam, *Harmonic functions and the structure of complete manifolds*, J. Differential Geom. **35** (1992), 359–383.

[Lux62] W. A. J. Luxemburg, *Two applications of the method of construction by ultrapowers to anaylsis*, Bull. Amer. Math. Soc. **68** (1962), 416–419.

[Lux67] ______, *Beweis des satzes von Hahn-Banach*, Arch. Math (Basel) **18** (1967), 271–272.

[Lux69] ______, *Reduced powers of the real number system and equivalents of the Hahn-Banach extension theorem*, Applications of Model Theory to Algebra, Analysis, and Probability (Internat. Sympos., Pasadena, Calif., 1967), Holt, Rinehart and Winston, New York, 1969, pp. 123–137.

[Lys96] I.G. Lysenok, *Infinite burnside groups of even exponent*, Izvestiya Akad. Nauk. SSSR Ser., Mat **60** (1996), no. 3, 453–654.

[Mac] N. Macura, *CAT(0) spaces with polynomial divergence of geodesics*, preprint 2011.

[Mag39] W. Magnus, *On a theorem of Marshall Hall*, Ann. of Math. (2) **40** (1939), 764–768.

[Mal40] A. I. Mal'cev, *On isomorphic matrix representations of infinite groups*, Mat. Sb. **8** (1940), 405–422.

[Mal49a] ______, *Generalized nilpotent algebras and their associated groups*, Mat. Sbornik N.S. **25(67)** (1949), 347–366.

[Mal49b] ______, *On a class of homogeneous spaces*, Izvestiya Akad. Nauk. SSSR Ser., Mat **13** (1949), 9–32, Amer. Math. Soc. Translation No. 39 (1951).

[Mar86] G. Martin, *Discrete quasiconformal groups which are not quasiconformal conjugates of Mobius groups*, Ann. Ac. Sci. Fenn. **11** (1986), 179–202.

[Mar91] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Springer-Verlag, Berlin, 1991.

[Mas91] W.S. Massey, *A basic course in algebraic topology*, Graduate Texts in Mathematics, vol. 127, Springer, 1991.

[Mil68] J. Milnor, *A note on curvature and fundamental group*, J. Diff. Geom. **2** (1968), 1–7.

[Mil12] J.S. Milne, *Group theory*, http://www.jmilne.org/math/CourseNotes/GT.pdf, 2012.

[Min01] I. Mineyev, *Straightening and bounded cohomology of hyperbolic groups*, Geom. Funct. Anal. **11** (2001), no. 4, 807–839.

[MNLGO92] J. C. Mayer, J Nikiel, and L. G. L. G. Oversteegen, *Universal spaces for* $\mathbb{R}$*-trees*, Trans. Amer. Math. Soc. **334** (1992), no. 1, 411–432.

[MO10] N. Monod and N. Ozawa, *The Dixmier problem, lamplighters and Burnside groups*, J. Funct. Anal. **258** (2010), no. 1, 255–259.

[Mor24] M. Morse, *A fundamental class of geodesics on any closed surface of genus greater than one*, Transactions of AMS **26** (1924), 25–60.

[Mor38] C.B. Morrey, *On the solution of quasilinear elliptic partial differential equations*, Transactions of AMS **43** (1938), 126–166.

[Mos57] G. D. Mostow, *On the fundamental group of a homogeneous space*, Ann. of Math. **66** (1957), 249–255.

[Mos73] ______, *Strong rigidity of locally symmetric spaces*, Annals of mathematical studies, vol. 78, Princeton University Press, Princeton, 1973.

[Mou88] G. Moussong, *Hyperbolic Coxeter groups*, Ph.D. thesis, The Ohio State University, 1988.

[MR03] C. Maclachlan and A. Reid, *The arithmetic of hyperbolic 3-manifolds*, Graduate Texts in Mathematics, vol. 219, Springer-Verlag, New York, 2003.

[MSW03] L. Mosher, M. Sageev, and K. Whyte, *Quasi-actions on trees. I. Bounded valence*, Ann. of Math. (2) **158** (2003), no. 1, 115–164.

[Mun75] J. R. Munkres, *Topology: a first course*, Prentice-Hall Inc., 1975.

[MY81] W. Meeks and S.T. Yau, *The equivariant Dehn's lemma and equivariant loop theorem*, Comment. Math. Helv. **56** (1981), 225– 239.

[MZ74] D. Montgomery and L. Zippin, *Topological transformation groups*, Robert E. Krieger Publishing Co., Huntington, N.Y., 1974, Reprint of the 1955 original.

[N6́4] A. Néron, *Modèles minimaux des variétes abèliennes sur les corps locaux et globaux*, Math. Publ. of IHES **21** (1964), 5–128.

[Nag83] J. Nagata, *Modern dimension theory*, revised ed., Sigma Series in Pure Mathematics, vol. 2, Heldermann Verlag, Berlin, 1983.

[Nag85] ———, *Modern general topology*, second ed., North Holland Mathematical Library, vol. 33, North Holland, 1985.

[New68] B. B. Newman, *Some results on one-relator groups*, Bull. Amer. Math. Soc. **74** (1968), 568–571.

[Nib04] G. Niblo, *A geometric proof of Stallings' theorem on groups with more than one end*, Geometriae Dedicata **105** (2004), no. 1, 61–76.

[Nic] B. Nica, *Linear groups: Mal'cev's theorem and Selberg's lemma*, Preprint, http://www.uni-math.gwdg.de/nica/linear_groups.pdf.

[Nov58] P. S. Novikov, *On the algorithmic insolvability of the word problem in group theory*, American Mathematical Society Translations, Ser 2, Vol. 9, American Mathematical Society, Providence, R. I., 1958, pp. 1–122.

[NR97] W. Neumann and L. Reeves, *Central extensions of word hyperbolic groups*, Ann. of Math. **145** (1997), 183–192.

[NY11] P. Nowak and G. Yu, *Large scale geometry*, European Mathematical Society, 2011.

[Ol'80] A. Yu. Ol'shanskii, *On the question of the existence of an invariant mean on a group*, Uspekhi Mat. Nauk **35** (1980), no. 4, 199–200.

[Ol'91a] ———, *Geometry of defining relations in groups*, Mathematics and its Applications (Soviet Series), vol. 70, Kluwer Academic Publishers Group, Dordrecht, 1991.

[Ol'91b] ———, *Hyperbolicity of groups with subquadratic isoperimetric inequalities*, Intl. J. Alg. Comp. **1** (1991), 282–290.

[Ol'91c] ———, *Periodic quotient groups of hyperbolic groups*, Mat. Sb. **182** (1991), no. 4, 543–567.

[Ol'92] ———, *Almost every group is hyperbolic*, Internat. J. Algebra Comput. **2** (1992), no. 1, 1–17.

[Ol'95] ———, SQ-*universality of hyperbolic groups*, Mat. Sb. **186** (1995), no. 8, 119–132.

[Oll04] Y. Ollivier, *Sharp phase transition theorems for hyperbolicity of random groups*, Geom. Funct. Anal. **14** (2004), no. 3, 595–679.

[OOS09] A. Yu. Ol'shanskii, D. V. Osin, and M. V. Sapir, *Lacunary hyperbolic groups*, Geom. Topol. **13** (2009), no. 4, 2051–2140, With an appendix by M. Kapovich and B. Kleiner.

[Ore51] O. Ore, *Some remarks on commutators*, Proc. Amer. Math. Soc. **2** (1951), 307–314.

[OS01] A. Yu. Ol'shanskii and M. V. Sapir, *Length and area functions on groups and quasi-isometric Higman embeddings*, International Journal of Algebra and Computation **11** (2001), 137–170.

[OS02] ———, *Non-amenable finitely presented torsion-by-cyclic groups*, Publ. Math. IHES **96** (2002), 43–169.

[Osi01] D. Osin, *Subgroup distortions in nilpotent groups*, Comm. Algebra **29** (2001), no. 12, 5439–5463.

[Osi06] ———, *Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems*, Mem. Amer. Math. Soc. **179** (2006), no. 843, vi+100pp.

[Osi10] ———, *Small cancellations over relatively hyperbolic groups and embedding theorems*, Ann. of Math. **172** (2010), 1–39.

[OV90] A. Onishchik and E. Vinberg, *Lie groups and algebraic groups*, Springer, 1990.

[OW11] Y. Ollivier and D. Wise, *Cubulating random groups at density less than* 1/6, Trans. Amer. Math. Soc. **363** (2011), no. 9, 4701–4733.

[Pan83] P. Pansu, *Croissance des boules et des géodésiques fermées dan les nilvariétés*, Ergodic Th. & Dyn. Sys. **3** (1983), 415–455.

[Pan89] ———, *Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un*, Ann. of Math. (2) **129** (1989), no. 1, 1–60.

[Pan95] ———, *Cohomologie $L^p$: invariance sous quasi-isométrie*, preprint, `http://www.math.u-psud.fr/~pansu/liste-prepub.html`, 1995.

[Pan07] ———, *Cohomologie $L^p$ en degré 1 des espaces homogènes*, Potential Anal. **27** (2007), no. 2, 151–165.

[Pap] P. Papasoglu, *Notes On Hyperbolic and Automatic Groups*, Lecture Notes, based on the notes of M. Batty, http://www.math.ucdavis.edu/~kapovich/280-2009/hyplectures_papasoglu.pdf.

[Pap95a] ———, *Homogeneous trees are bi-Lipschitz equivalent*, Geom. Dedicata **54** (1995), no. 3, 301–306.

[Pap95b] ———, *On the subquadratic isoperimetric inequality*, Geometric group theory, vol. 25, de Gruyter, Berlin-New-York, 1995, R. Charney, M. Davis, M. Shapiro (eds), pp. 193–200.

[Pap95c] ———, *Strongly geodesically automatic groups are hyperbolic*, Invent. Math. **121** (1995), no. 2, 323–334.

[Pap96] ———, *An algorithm detecting hyperbolicity*, Geometric and computational perspectives on infinite groups (Minneapolis, MN and New Brunswick, NJ, 1994), vol. 25, Amer. Math. Soc., Providence, RI, 1996, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., pp. 193–200.

[Pap98] ———, *Quasi-flats in semihyperbolic groups*, Proc. Amer. Math. Soc. **126** (1998), 1267–1273.

[Pap00] ———, *Isodiametric and isoperimetric inequalities for complexes and groups*, J. London Math. Soc. (2) **62** (2000), no. 1, 97–106.

[Pap05] ———, *Quasi-isometry invariance of group splittings*, Ann. of Math. **161** (2005), no. 2, 759–830.

[Par08] J. Parker, *Hyperbolic spaces*, vol. 2, Jyväskylä Lectures in Mathematics, 2008.

[Pau91] F. Paulin, *Outer automorphisms of hyperbolic groups and small actions on* **R***-trees*, Arboreal group theory (Berkeley, CA, 1988), Math. Sci. Res. Inst. Publ., vol. 19, Springer, New York, 1991, pp. 331–343. MR 1105339 (92g:57003)

[Pau96] ———, *Un groupe hyperbolique est déterminé par son bord*, J. London Math. Soc. (2) **54** (1996), no. 1, 50–74.

[Paw91] J. Pawlikowski, *The Hahn-Banach theorem implies the Banach-Tarski paradox*, Fundamenta Mathematicae **138** (1991), no. 1, 21–22.

[PB03] P. Papasoglu and M. Batty, *Notes on hyperbolic and automatic groups*, Warwick University Lecture Notes, 2003.

[Ped95] E.K. Pedersen, *Bounded and continuous control*, Proceedings of the conference "Novikov conjectures, index theorems and rigidity" volume II, Oberwolfach 1993, LMS Lecture Notes Series, vol. 227, Cambridge University Press, 1995, pp. 277–284.

[Pen11a] I. Peng, *Coarse differentiation and quasi-isometries of a class of solvable Lie groups I*, Geom. Topol. **15** (2011), no. 4, 1883–1925.

[Pen11b] ———, *Coarse differentiation and quasi-isometries of a class of solvable Lie groups II*, Geom. Topol. **15** (2011), no. 4, 1927–1981.

[Pin72] D. Pincus, *Independence of the prime ideal theorem from the Hahn Banach theorem*, Bull. Amer. Math. Soc. **78** (1972), 766–770.

[Pin74] ———, *The strength of the Hahn-Banach theorem*, Victoria Symposium on Nonstandard Analysis (Univ. Victoria, Victoria, B.C., 1972), Springer, Berlin, 1974, pp. 203–248. Lecture Notes in Math., Vol. 369.

[Poi95] F. Point, *Groups of polynomial growth and their associated metric spaces*, J. Algebra **175** (1995), no. 1, 105–121.

[Pra73] G. Prasad, *Strong rigidity of Q–rank one lattices*, Invent. Math. **21** (1973), 255–286.

[PW02] P. Papasoglu and K. Whyte, *Quasi-isometries between groups with infinitely many ends*, Comment. Math. Helv. **77** (2002), no. 1, 133–144.

[Rab58] M. O. Rabin, *Recursive unsolvability of group theoretic problems*, Ann. of Math. (2) **67** (1958), 172–194.

[Rag72] M. Raghunathan, *Discrete subgroups of lie groups*, Springer, 1972.

[Rat94] J.G. Ratcliffe, *Foundations of hyperbolic manifolds*, Springer, 1994.

[Res89] Yu. G. Reshetnyak, *Space mappings with bounded distortion*, Translations of Mathematical Monographs, vol. 73, American Mathematical Society, Providence, RI, 1989, Translated from the Russian by H. H. McFaden.

[Rin61] W. Rinow, *Die innere Geometrie der metrischen Räume*, Die Grundlehren der mathematischen Wissenschaften, Bd. 105, Springer-Verlag, Berlin, 1961.

[Rip82] E. Rips, *Subgroups of small cancellation groups*, Bull. London Math. Soc. **14** (1982), no. 1, 45–47.

[Rob47] R.M. Robinson, *On the decomposition of spheres*, Fund. Math. **34** (1947), 246–260.

[Roe03] J. Roe, *Lectures on coarse geometry*, University Lecture Series, vol. 31, American Mathematical Society, Providence, RI, 2003.

[Ros74] J. Rosenblatt, *Invariant measures and growth conditions*, Trans. Amer. Math. Soc. **193** (1974), 33–53.

[Ros76] ———, *Invariant means for the bounded measurable functions on a non-discrete locally compact group*, Math. Ann. **220** (1976), no. 3, 219–228.

[Roy68] H. Royden, *Real Analysis*, Macmillan, New York, 1968.

[RR85] H. Rubin and J. E. Rubin, *Equivalents of the axiom of choice. II*, Studies in Logic and the Foundations of Mathematics, vol. 116, North-Holland Publishing Co., Amsterdam, 1985.

[RS94] E. Rips and Z. Sela, *Structure and rigidity in hyperbolic groups. I*, Geom. Funct. Anal. **4** (1994), no. 3, 337–371.

[Rud87] W. Rudin, *Real and complex analysis*, McGraw-Hill International editions, 1987.

[Sal12] A. Sale, *Short conjugators and compression exponents in free solvable groups*, preprint, arXiv:1202.5343v1, 2012.

[Sap13] M. Sapir, *Non-commutative combinatorial algebra*, preprint, 2013.

[Sau06] R. Sauer, *Homological invariants and quasi-isometry*, Geom. Funct. Anal. **16** (2006), no. 2, 476–515.

[Sch38] I.J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), 522–536.

[Sch96a] R. Schwartz, *Quasi-isometric rigidity and Diophantine approximation*, Acta Math. **177** (1996), no. 1, 75–112.

[Sch96b] ———, *The quasi-isometry classification of rank one lattices*, Publ. Math. IHES **82** (1996), 133–168.

[Sch97] ———, *Symmetric patterns of geodesics and automorphisms of surface groups*, Invent. Math. **128** (1997), no. 1, 177–199.

[Sel60] A. Selberg, *On discontinuous groups in higher-dimensional symmetric spaces*, Contributions to Function Theory (K. Chandrasekhadran, ed.), Tata Inst. of Fund. Research, Bombay, 1960, pp. 147–164.

[Sel95] Z. Sela, *The isomorphism problem for hyperbolic groups. I*, Ann. of Math. (2) **141** (1995), no. 2, 217–283.

[Sel97] ———, *Structure and rigidity in (Gromov) hyperbolic groups and discrete groups in rank* 1 *Lie groups. II*, Geom. Funct. Anal. **7** (1997), no. 3, 561–593.

[Sel99] ———, *Endomorphisms of hyperbolic groups. I. The Hopf property*, Topology **38** (1999), no. 2, 301–321.

[Ser80] J. P. Serre, *Trees*, Springer, New York, 1980.

[Sha00] Y. Shalom, *Rigidity of commensurators and irreducible lattices*, Invent. Math. **141** (2000), no. 1, 1–54.

[Sha04] ———, *Harmonic analysis, cohomology, and the large-scale geometry of amenable groups*, Acta Math. **192** (2004), no. 2, 119–185.

[She78] S. Shelah, *Classification theory and the number of non-isomorphic models*, Studies in Logic and the Foundations of Mathematics, vol. 92, North-Holland Publishing Co., 1978.

[Sil] L. Silbermann, *Equivalence between properties (T) and (FH)*, unpublished notes.

[Šir76] V. L. Širvanjan, *Imbedding of the group $B(\infty, n)$ in the group $B(2, n)$*, Izv. Akad. Nauk SSSR Ser. Mat. **40** (1976), no. 1, 190–208, 223.

[SS05] E. Stein and R. Shakarchi, *Real analysis*, Princeton Lectures in Analysis, III, Princeton University Press, 2005, Measure theory, integration, and Hilbert spaces.

[ST10] Y. Shalom and T. Tao, *A finitary version of Gromov's polynomial growth theorem*, Geom. Funct. Anal. **20** (2010), no. 6, 1502–1547.

[Sta68] J. Stallings, *On torsion-free groups with infinitely many ends*, Ann. of Math. **88** (1968), 312–334.

[Str74] R. A. Struble, *Metrics in locally compact groups*, Compositio Math. **28** (1974), 217–222.

[Sul81] D. Sullivan, *On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions*, Riemann surfaces and related topics, Proceedings of the 1978 Stony Brook Conference, Ann. Math. Studies, vol. 97, Princeton University Press, 1981, pp. 465–496.

[Šva55] A. S. Švarc, *A volume invariant of coverings*, Dokl. Akad. Nauk SSSR (N.S.) **105** (1955), 32–34.

[SW79] P. Scott and T. Wall, *Topological methods in group theory*, "Homological Group Theory", London Math. Soc. Lecture Notes, vol. 36, 1979, pp. 137– 204.

[Swa93] G. A. Swarup, *A note on accessibility*, Geometric group theory, Vol. 1 (Sussex, 1991), London Math. Soc. Lecture Note Ser., vol. 181, Cambridge Univ. Press, Cambridge, 1993, pp. 204–207. MR 1238527 (94m:57004)

[SY94] R. Schoen and S.-T. Yau, *Lectures on differential geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology, I, International Press, Cambridge, MA, 1994, Lecture notes prepared by Wei Yue Ding, Kung Ching Chang, Jia Qing Zhong and Yi Chao Xu, Translated from the Chinese by Ding and S. Y. Cheng, Preface translated from the Chinese by Kaising Tso.

[Tar38] A. Tarski, *Algebraische fassung des massproblems*, Fund. Math. **31** (1938), 47–66.

[Tar86] ———, *Collected papers. Vol. 1*, Contemporary Mathematicians, Birkhäuser Verlag, Basel, 1986, 1921–1934, Edited by Steven R. Givant and Ralph N. McKenzie.

[Thu97] W. Thurston, *Three-dimensional geometry and topology, I*, Princeton Mathematical Series, vol. 35, Princeton University Press, 1997.

[Tit72] J. Tits, *Free subgroups in linear groups*, Journal of Algebra **20** (1972), 250–270.

[Tuk81] P. Tukia, *A quasiconformal group not isomorphic to a Möbius group*, Ann. Acad. Sci. Fenn. Ser. A I Math. **6** (1981), 149–160.

[Tuk86] ———, *On quasiconformal groups*, J. Analyse Math. **46** (1986), 318–346.

[Tuk88] ———, *Homeomorphic conjugates of Fuchsian groups*, J. Reine Angew. Math. **391** (1988), 1–54.

[Tuk94] ———, *Convergence groups and Gromov's metric hyperbolic spaces*, New Zealand J. Math. **23** (1994), no. 2, 157–187.

[TW93] C. Thomassen and W. Woess, *Vertex-transitive graphs and accessibility*, J. Combin. Theory Ser. B **58** (1993), no. 2, 248–268.

[Väi71] J. Väisälä, *Lectures on n-dimensional quasiconformal mappings*, Lecture Notes in Math., vol. 229, Springer, 1971.

[Väi05] ———, *Gromov hyperbolic spaces*, Expo. Math. **23** (2005), no. 3, 187–231. MR 2164775 (2006j:53055)

[Väi85] ———, *Quasi-Möbius maps*, J. Analyse Math. **44** (1984/85), 218–234.

[Var99] N. Varopoulos, *Distance distortion on Lie groups*, Random walks and discrete potential theory (Cortona, 1997), Sympos. Math., XXXIX, Cambridge Univ. Press, Cambridge, 1999, pp. 320–357.

[Vav] N. Vavilov, *Concrete group theory*, http://www.math.spbu.ru/user/valgebra/groubook.pdf.

[VdDW84] L. Van den Dries and A.J. Wilkie, *Gromov's theorem on groups of polynomial growth and elementary logic*, Journ. of Algebra **89** (1984), 349– 374.

[Ver82] A. Vershik, *Amenability and approximation of infinite groups*, Selecta Math. Soviet. **2** (1982), no. 4, 311–330, Selected translations.

[Ver11] R. Vershynin, *Lectures in geometric functional analysis*, Preprint, University of Michigan, 2011.

[vN28] J. von Neumann, *Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre*, Math. Ann. **99** (1928), 373–391.

[vN29] ———, *Zur allgemeinen theorie des masses*, Fund. math. **13** (1929), 73–116.

[VSCC92] N. Th. Varopoulos, L. Saloff-Coste, and T. Coulhon, *Analysis and geometry on groups*, Cambridge Tracts in Mathematics, vol. 100, Cambridge University Press, Cambridge, 1992.

[Vuo88] M. Vuorinen, *Conformal geometry and quasiregular mappings*, Lecture Notes in Math, vol. 1319, Springer, 1988.

[Wag85] S. Wagon, *The Banach-Tarski paradox*, Cambridge Univ. Press, 1985.

[War83] F. W. Warner, *Foundations of differentiable manifolds and lie groups*, Graduate Texts in Mathematics, 94, Springer-Verlag, 1983.

[War12] E. Warner, *Ultraproducts and foundations of higher order Fourier analysis*, Thesis, Princeton University, 2012.

[Wei84] B. Weisfeiler, *On the size and structure of finite linear groups*, Preprint, math. arxiv, 1203.1960, 1984.

[Wen08] S. Wenger, *Gromov hyperbolic spaces and the sharp isoperimetric constant*, Invent. Math. **171** (2008), no. 1, 227–255.

[Why99] K. Whyte, *Amenability, bilipschitz equivalence, and the von Neumann conjecture*, Duke Math. J. **99** (1999), 93–112.

[Why01] ———, *The large scale geometry of the higher Baumslag-Solitar groups*, Geom. Funct. Anal. **11** (2001), no. 6, 1327–1343.

[Wis96] D. Wise, *A non-Hopfian automatic group*, J. Algebra **180** (1996), no. 3, 845–847.

[Woe00] W. Woess, *Random walks on infinite graphs and groups*, Cambridge University Press, 2000.

[WW75] J.H. Wells and L.R. Williams, *Embeddings and extensions in analysis*, Springer-Verlag, New York-Heidelberg, 1975.

[Wys88] J. Wysoczánski, *On uniformly amenable groups*, Proc. Amer. Math. Soc. **102** (1988), no. 4, 933–938.

[Xie06] X. Xie, *Quasi-isometric rigidity of Fuchsian buildings*, Topology **45** (2006), 101–169.

[Yam04] A. Yaman, *A topological characterisation of relatively hyperbolic groups*, J. Reine Angew. Math. (Crelle's Journal) **566** (2004), 41–89.

[Yom83] Y. Yomdin, *The geometry of critical and near-critical values of differentiable mappings*, Math. Ann. **264** (1983), no. 4, 495–515.

[Yu05] G. Yu, *Hyperbolic groups admit proper isometric actions on $\ell^p$-spaces*, Geom. Funct. Anal. **15** (2005), no. 5, 1144–1151.

[Ż03] A. Żuk, *Property (T) and Kazhdan constants for discrete groups*, Geom. Funct. Analysis **13** (2003), 643–670.

[Zer04] E. Zermelo, *Beweis, dass jede Menge wohlgeordnet werden kann*, Math. Ann. **59** (1904), 514–516.

[Zim84] R. Zimmer, *Ergodic theory and semisimple groups*, Monographs in Math, vol. 81, Birkhauser, 1984.

# Index