The Gromov-Hausdorff Distance and Groups of Polynomial Growth

Eric Vaz de Souza Santos

Monografia apresentada ao Departamento de Matemática da Universidade Federal de Sergipe como parte dos requisitos para obtenção do título de Bacharel em Matemática Aplicada e Computacional.

Área de concentração: Matemática

Orientador: Prof. Dr. Cayo Rodrigo Felizardo Dória

Universidade Federal de Sergipe
São Cristóvão - Sergipe - Brasil
2024

Resumo

Esta monografia tem como objetivo a demonstração do Teorema de Gromov dos Grupos de Crescimento Polinomial, que caracteriza grupos de crescimento polinomial como aqueles que têm subgrupos nilpotentes de índice finito. Para tal, introduzimos os conceitos essenciais para a compreensão da demonstração, como espaços métricos, espaços topológicos, variedades e grupos. Também desenvolvemos a teoria da distância e da convergência de Gromov-Hausdorff, ferramentas da Geometria Métrica que utilizamos para estudar grupos como objetos geométricos. Finalmente, estudamos o crescimento de grupos e as principais propriedades dos grupos de crescimento polinomial, culminando em uma demonstração do Teorema de Gromov.

Abstract

This undergraduate thesis aims to present a proof of Gromov's Theorem on Groups of Polynomial Growth, which characterizes groups of polynomial growth as those that have nilpotent subgroups of finite index. For such, we introduce basic concepts that are essential in understanding the proof, namely metric spaces, topological spaces, manifolds, and groups. We also explore the theory of the Gromov-Hausdorff distance and convergence, which are tools from Metric Geometry that prove to be useful in the study of groups as geometric objects. Lastly, we examine group growth and the essential properties of groups of polynomial growth, culminating in a proof of Gromov's Theorem.

À minha mãe, Sofia, por tudo. Ao meu gato, Adolfo, pela inspiração.

Contents

1 Introduction

This text aims to serve as a gentle introduction to Geometric Group Theory. Specifically, we offer the needed background to tackle what is now a classic result: Gromov's Theorem on Groups of Polynomial Growth. The growth function of a finitely generated group G associated with a finite generated set S (we assume that the set is symmetric, i.e., $x \in S \iff x^{-1} \in S$) is defined by taking $n \in \mathbb{N}$ to the number $b_S(n)$ of elements of G that can be written as a string of, at most, n elements of S. A group is said to be of polynomial growth when $b_S(n)$ is bounded above by a polynomial function of n.

The study of the asymptotic behaviour of the growth function can be traced back to Albert Schwarz in [Sch55] (unfortunately, only available in Russian), and was rediscovered in the west by John Milnor in [Mil68]. Early motivation came due to results regarding the growth function of the fundamental groups of Riemannian manifolds (see Proposition 4.11, for instance).

It is possible to show that the growth function of a finitely generated group grows, at most, exponentially (see Corollary 4.21). It was also known since the late 1960s [Wol68] that finitely generated nilpotent groups G are groups of polynomial growth, and that if $H \subset G$ is a subgroup of finite index generated by some finite $R \subset H$, then b_S and b_R are equivalent in their growth. Indeed, all known groups of polynomial growth were examples of these two facts: They were groups that had a nilpotent group of finite index (and were aptly named *almost nilpotent groups*).

Some early results characterized some classes of groups as either polynomial or exponential: It was shown, for instance, by Joseph Wolf in [Wol68], that finitely generated solvable groups are either of exponential growth or are almost nilpotent. Jacques Tits [Tit72] proved that finitely generated subgroups of Lie groups that have finitely many components must have exponential growth or have a solvable subgroup of finite index. Together, these results imply that, for finitely generated subgroups of Lie groups that have finitely many components, having polynomial growth is equivalent to being almost nilpotent. Two questions remained unanswered through the 1970s: Whether there exists a group of neither polynomial nor exponential growth, and whether there exists a group of polynomial growth without a nilpotent group of finite index.

The solution to the first question was given by Rostislav Grigorchuk in [Gri85]: Yes, there are groups that are of *intermediate* growth, i.e., neither exponential nor polynomial, today known as the Grigorchuk groups. The answer of the second question, offered by Mikhael Gromov in [Gro81], was a negative: All groups of polynomial growth are almost nilpotent. This is Gromov's Theorem on Groups of Polynomial Growth (Theorem 4.36), and it is our objective to acquire the tools necessary to prove it.

Gromov's remarkable proof was geometric in nature. If we define the norm $|g|$ of every element $g \in G$ (with respect to a finite generating set $S \subset G$) as the length of the shortest string of elements of S that is equal to g, then the function $d : G \times G \to \mathbb{R}$ defined as $d(g, h) = |g^{-1}h|$ turns G into a metric space. Note that $b_S(n)$ is precisely the cardinality of the closed ball with radius n centered at the identity. The group operation $h \mapsto gh$ for any $g \in G$ defines an isometry of the space to itself.

Therefore, we can study groups as geometric objects in their own right. The fundamental observation is that one can associate to every group of polynomial growth G a very nice (connected, locally path connected, locally compact, homogeneous, finite dimensional) space Y_G. Consider, for instance, the group $\mathbb{Z}^2$ and the set of generators $S = \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$. The group norm induced by this generating set is precisely the sum metric $d((a, b), (a', b')) = |a - a'| + |b - b'|$.

Consider shrinking $\mathbb{Z}^2$ towards the origin, obtaining finer and finer grids of points (more formally, defining a sequence of metrics d_n on $\mathbb{Z}^2$ as $d((a, b), (a', b')) = \varepsilon_n(|a - a'| + |b - b'|)$, where $\lim_{n\to\infty} \varepsilon_n = 0$). We obtain a sequence of spaces that start to become almost indistinguishable to the plane $\mathbb{R}^2$, which is the very nice space we want to associate $\mathbb{Z}^2$ to. A form of distance between metric spaces originally defined by David Edwards in [Edw75], and today known as the Gromov-Hausdorff distance, can be used to make this idea precise.

This distance will be our objective of study in Section 3. In particular, we obtain a condition (Proposition 3.35) for an isometry from G to itself to be associated with an isometry from Y_G to itself. The group operation then induces an action of G on Y_G. It turns out that the group of isometries of Y_G can receive the structure of a Lie group. The less general version of the result regarding subgroups of Lie groups can then be leveraged into a proof of Gromov's Theorem.

2 Spaces and their symmetries

In this section, we are going to introduce the geometrical concepts we are going to work with. We begin with metric spaces and their basic properties. In particular, we explore the Hausdorff distance, a historical predecessor of the Gromov-Hausdorff distance that is defined only for subsets of a fixed metric space. This distance is very well behaved in the set of all closed, bounded subsets of a metric space, where it is a true metric. We also give a brief overview of the Hausdorff measure and dimension, two other concepts introduced by Felix Hausdorff that are going to be vital in the proof of Gromov's Theorem.

In the sequence, we lay the essential definitions of topological spaces and manifolds. We also discuss groups, giving special attention to some useful results involving finitely generated groups. In particular we prove the theorem of classification of finitely generated abelian groups. Finally, we introduce topological and Lie groups. We provide a proof that the group of isometries of a metric space is a topological group and also comment on conditions that ensure that a topological group has a compatible smooth structure that makes it into a Lie group.

2.1 Metric spaces

Definition 2.1. A metric space on a set X is an ordered pair (X, d), where $d : X \times X \to \mathbb{R}^{\geq 0}$ is a function satisfying the following (where a, b and c are arbitrary elements of X)

1. $d(a, b) = 0 \iff a = b$ (Positivity);
2. $d(a, b) = d(b, a)$ (Symmetry);
3. $d(a, c) \leq d(a, b) + d(c, d)$ (Triangle inequality).

Spaces that fail axiom 1 but satisfy the weaker condition $d(a, a) = 0$ are often called pseudometric spaces.

Remark 2.2. Let (X, d) be a pseudometric space. Define a relation on X as $a \sim b \iff d(a, b) = 0$. This can be shown to be an equivalence relation. We can also show that d will have this nice property

$$(a_1 \sim a_2 \text{ and } b_1 \sim b_2) \implies d(a_1, b_1) = d(a_2, b_2)$$

Thus there is a function $d_\sim : {}^X/_\sim \to \mathbb{R}^{\geq 0}$ that satisfies $d_\sim([a]_\sim, [b]_\sim) = d(a, b)$. The pair $({}^X/_\sim, d_\sim)$ forms a metric space.

Example 2.3. Let X be the set of all real-valued Lebesgue-integrable functions on the $\mathbb{R}^n$. Define a pseudometric on X by $d(f, g) = \int |f - g|$. It is a standard exercise of measure theory to show that $d(f, g) = 0 \iff f = g$ (almost everywhere). Thus applying Remark 2.2 on this space will result on the L^1 Lebesgue space.

Remark 2.4. Let (X, d) be a metric space and $f : S \to X$ be any function. There is a natural pseudometric d_f on S induced by f: Given $a, b \in S$ define $d_f(a, b) = d(f(a), f(b))$. It will satisfy all axioms except for positivity, since $f(a) = f(b)$ is a possibility. This is an actual metric if and only if f is injective.

In particular, any subset of a metric space can be seen with the metric space as induced by the inclusion function, which is always injective.

Remark 2.5. Let (X, d) be a metric space and S a set that contains X. It is possible to endow S with a metric that agrees with d when restricted to X. Let $x_0 \in X$ be an arbitrary but fixed point.

Provide $S \setminus X$ with the following metric (which we call the discrete metric):

$$d(a, b) = \begin{cases} 0, & \text{if } a = b \\ 1, & \text{if } a \neq b \end{cases}$$

By having two metrics, one defined on X and another defined on $S \setminus X$, we must choose a value of $d(a, b)$ when $a \in X$ and $b \notin X$. The reader may check that the following is an adequate choice:

$$d(a, b) = d(a, x_0) + 1.$$

Example 2.6. Let V be a real vector space. A norm $|-|: V \to \mathbb{R}$ on V induces a metric, defined as $d(a,b) = |a-b|$. In particular, an inner product $\langle -,- \rangle : V \times V \to \mathbb{R}$ induces a norm $v \mapsto \langle v, v \rangle$, which in turns induces a metric. The usual inner product on $\mathbb{R}^n$ induces the Euclidean metric.

Example 2.7 (The Product Metric). Let (X, d_X), (Y, d_Y) be two metric spaces. The set $X \times Y$ can be endowed with the following metric:

$$d((a_1,b_1),(a_2,b_2)) = \max\{d(a_1,a_2), d(b_1,b_2)\}$$

This metric, when given to $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, does not agree with the Euclidean metric (although, as we will see later on, a metric on a set induces a topology on it. The product metric and the Euclidean metric induce identical topologies on $\mathbb{R}^2$).

Here are some definitions we will need:

Definition 2.8. In what follows (X, d_X) and (Y, d_Y) are metric spaces:

1. A sequence $\{x_n\}_{n=1}^\infty$ taking points in X is said to *converge* to $x \in X$ if $\lim\limits_{n\to\infty} d(x_n, x) = 0$ as a sequence of real numbers. Then x is called the *limit* of the sequence, which is unique if it exists. We use the notation $\lim\limits_{n\to\infty} x_n = x$,
2. The *closure* of a subset $S \subset X$ is the set $\overline{S} \subset X$ of all the points of X that are limits of sequences that are entirely contained in S. We say that S is *closed* if $S = \overline{S}$.
3. A subset $S \subset X$ is called *dense* in X if $\overline{S} = X$.
4. A metric space is called *separable* if it has a countable, dense subset.
5. A function $f : X \to Y$ is *continuous* if for every convergent sequence $\{x_n\}_{n=1}^\infty$ taking points in X we have $\lim\limits_{n\to\infty} f(x_n) = f\left(\lim\limits_{n\to\infty} x_n\right)$.
6. A function $f : X \to Y$ is *uniformly* continuous if for all sequences $\{x_n\}_{n=1}^\infty$, $\{x'_n\}_{n=1}^\infty$ such that $\lim\limits_{n\to\infty} d(x_n, x'_n) = 0$ we have $\lim\limits_{n\to\infty} d(f(x_n), f(x'_n)) = 0$.
7. A function $f : X \to Y$ is an *isometry* if $d(f(x), f(x')) = d(x, x')$ for every $x, x' \in X$ (note that isometries are uniformly continuous).
8. A metric space X is called *homogeneous* if for all $x_1, x_2 \in X$ there is a bijective isometry $f : X \to X$ with $f(x_1) = x_2$.
9. A sequence $\{x_n\}_{n=1}^\infty$ is a *Cauchy sequence* if of every $\varepsilon > 0$ there is some $n_0 \in \mathbb{N}$ such that $d(x_n, x_{n'}) < \varepsilon$ if $n, n' > n_0$ (note that convergent sequences are Cauchy).
10. A metric space is *complete* if all of its Cauchy sequences are convergent.
11. A metric space X is *bounded* if the image of the metric function d is bounded as a subset of $\mathbb{R}$. In this case its supremum is called the *diameter* of X, which is denoted $\operatorname{diam}(X)$.
12. Let $S \subset X$ be any subset and $x \in X$. We define the *distance between S and x* as $d(S, x) = \inf\{d(s,x) \ : \ s \in S\}$.
13. Let $S \subset X$ and $r > 0$. We define the *open and closed r-neighbourhood* of S respectively as $B_r(S) = \{x \in X \ : \ d(S,x) < r\}$ and $\overline{B}_r(S) = \{x \in X \ : \ d(S,x) \leq r\}$. If $S = \{s\}$ is a singleton, we use $B_r(s)$ and $\overline{B}_r(s)$ and call these the *open and closed balls* centered at s with a radius r.
14. A subset of X is called *open* if it is an union of open balls. A subset of $S \subset X$ is called *closed* if $X \setminus S$ is open.
15. Let $\varepsilon > 0$. A subset $S \subset X$ is a ε-net if $\overline{B}_\varepsilon(S) = X$.
16. A metric space is *totally bounded* if it admits a finite ε-net for every $\varepsilon > 0$. Totally bounded spaces are separable.
17. A metric space is *compact* if it is complete and totally bounded.

18. A metric space is *proper* if all of its closed balls are compact. All proper spaces are complete.

Remark 2.9. A metric, as a function between metric spaces $d : X \times X \to \mathbb{R}$ (where the metric on $X \times X$ is as in Example 2.7) is always continuous.

Remark 2.10. Useful examples of ε-net in a metric space X are the maximal ε-separated sets. A ε-separated set is a subset $S \subset X$ such that $d(s_1, s_2) \geq \varepsilon$ for all $s_1, s_2 \in S$. A maximal ε-separated set is a ε-separated set S such that if $S \subsetneq S' \subset X$, then S' is not ε-separated. Such a set will always be a ε-net: if there is some $x \in X$ with $d(x, s) > \varepsilon$ for all $s \in S$, then $S \cup \{x\}$ will be ε-separated, contradicting the maximally.

Assuming Zorn's Lemma, all metric spaces have such a set: The set P of all ε-separated sets of X is a partially ordered set under inclusion. If $P' \subset P$ is totally ordered subset, then the union of all elements of P' will be an upper bound to P' with respect to inclusion while remaining a ε-separated set.

Remark 2.11. Compact subsets of a metric space are closed and bounded, while the contrary implication holds if and only if the space is proper.

A technique of manipulating sequences in totally bounded spaces that will come up later is employed in the proof of the following Proposition:

Proposition 2.12. *A space X is totally bounded if and only if every sequence admits a Cauchy subsequence.*

Proof. Assume X is totally bounded and let $\{x_n\}_{n=1}^{\infty}$ be any sequence. For all $n \in \{1, 2, \dots\}$, let S_n be a finite $1/n$-net of X. Note that there is some $s_1 \in S_1$ such that infinitely many terms of $\{x_n\}_{n=1}^{\infty}$ are contained in $\overline{B}_1(s_1)$, simply because there are infinitely many terms in the sequence and each S_n is finite (and a net). Let $\{x_n^1\}_{n=1}^{\infty}$ be the subsequence of all terms of the original sequence that are contained in this closed ball.

Inductively, let $\{x_n^m\}_{n=1}^{\infty}$ be a subsequence of $\{x_n^{m-1}\}_{n=1}^{\infty}$ such that all terms are contained in a ball of radius $\frac{1}{i}$ centered at some $s_i \in S_i$, for all $i \in \{1, 2, \dots, m\}$. Then there are only finitely many sets $\overline{B}_{1/m}(s_m) \cap \overline{B}_{1/m+1}(s)$ for $s \in S_{m+1}$ and one of them must contain infinitely many terms of $\{x_n^m\}_{n=1}^{\infty}$. Construct $\{x_n^{m+1}\}_{n=1}^{\infty}$ accordingly. The sequence $\{x_n^n\}_{n=1}^{\infty}$ is the needed Cauchy subsequence.[1]

Conversely let X be a space that is not totally bounded, that is, there is some $\varepsilon > 0$ such that no finite subset $S \subset X$ satisfies $\overline{B}_\varepsilon(S) = X$. Let $x_1 \in X$ be any point. There must be some $x_2 \in X \setminus \overline{B}_\varepsilon(x_1)$. Having constructed a finite sequence $x_1, \dots, x_N$ such that $x_n \notin \bigcup_{m=1}^{n-1} \overline{B}_\varepsilon(x_m)$, chose x_{N+1} similarly. The resulting sequence $\{x_n\}_{n=1}^{\infty}$ cannot contain a Cauchy sequence, since $d(x_n, x_{n+1}) > \varepsilon$. ■

2.1.1 Continuous extension and complete metric spaces

A classic problem of metric topology is to extend a continuous function $f : X' \to Y$ defined on a subset $X' \subset X$ of a metric space into a continuous function $f : X \to Y$ defined on the whole space. This is not always possible: choose your favourite continuous function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ such that $\lim_{x\to 0} f(x)$ doesn't exist. This limit will also fail to exist on any function defined on the entire line that agrees with f.

Indeed this will characterize the problem of extending continuous functions defined on dense subsets: a continuous function $f : X' \to Y$ defined on a dense subset $X' \subset X$ will have a continuous extension on X if and only for all $a \in X$ and every sequence $\{a_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} a_n = a$, the limit $\lim_{n\to\infty} f(a_n)$ exists. This should also show that in this case the continuous extension will be unique, since limits are unique.

Proposition 2.13. *An uniformly continuous function $f : X' \to Y$ defined on a dense subset $X' \subset X$, taking values on a complete metric space Y can be extended uniquely to a continuous function $f : X \to Y$. This extension will also be uniformly continuous.*

[1] Here we proved the existence of a sequence of sequences, each having some needed property, and then constructed a new sequence by taking the first term of the first sequence, the second term of the second sequence, and so on, resulting in a sequence that has all of the infinitely many needed properties. This is known as the diagonal principle (or Cantor's diagonal argument, after Georg Cantor's proof of the non-countability of the real numbers).

Proof. For every $x \in X$, let $\{x_n\}_{n=1}^{\infty}$ be a sequence of points in X' that converges to x. The sequence $\{f(x_n)\}_{n=1}^{\infty}$ is Cauchy and must have a limit $y \in Y$ due to completeness. Define $f(x) = y$. Continuity on X holds by construction, while uniqueness follows from the uniqueness of limits.

To show that f is uniformly continuous, choose sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ taking values on X, such that $\lim_{n\to\infty} d(a_n, b_n) = 0$. Each pair a_n, b_n can be seen as the limits of respective sequences $\{a_n^m\}_{m=1}^{\infty}$, $\{b_n^m\}_{m=1}^{\infty}$, both taking values on X'. Due to the continuity of the metric, we have

$$0 = \lim_{n\to\infty} d(a_n, b_n) = \lim_{n\to\infty} d(\lim_{m\to\infty} a_n^m, \lim_{m\to\infty} b_n^m) = \lim_{n\to\infty} d(a_n^n, b_n^n).$$

In turn, the uniform continuity of f on X' implies

$$0 = \lim_{n\to\infty} d(f(a_n^n), f(b_n^n)) = \lim_{n\to\infty} d(\lim_{m\to\infty} f(a_n^m), \lim_{m\to\infty} f(b_n^m)).$$

Finally, due to the continuity of f on X,

$$0 = \lim_{n\to\infty} d(\lim_{m\to\infty} f(a_n^m), \lim_{m\to\infty} f(b_n^m)) = \lim_{n\to\infty} d(f(a_n), f(b_n)).$$

Thus, we have uniform continuity. ■

In particular we have the following

Corollary 2.14. *An isometry $f : X' \to Y'$ between dense subsets $X' \subset X$, $Y' \subset Y$ of metric spaces X, Y can be extended to a unique continuous function $f : X \to Y$. This extension will also be an isometry.*

Also, by assuming that the domain of f is totally bounded we can weaken the hypothesis of Proposition 2.13 by using the following result.

Proposition 2.15. *Let $f : X \to Y$ be a function that takes Cauchy sequences to Cauchy sequences. If X is totally bounded, then f is uniformly continuous.*

Proof. Assume that X is totally bounded but f is not uniformly continuous. There are sequences $\{x_n\}_{n=1}^{\infty}$ and $\{x'_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} d(x_n, x'_n) = 0$ but for some $\varepsilon > 0$ there are infinitely many terms of the sequence $\{d(f(x_n), f(x'_n))\}_{n=1}^{\infty}$ that are greater than ε. We may pass to a subsequence and assume that all terms are greater than ε.

Proposition 2.12 tells us that $\{x_n\}_{n=1}^{\infty}$ has a Cauchy subsequence, and we might as well assume that $\{x_n\}_{n=1}^{\infty}$ is Cauchy. Consider the interwoven sequence $\{z_n\}_{n=1}^{\infty}$ given by $(x_1, x'_1, x_2, x'_2, \ldots)$. Since we have $\lim_{n\to\infty} d(x_n, x'_n) = 0$ it follows that $\{z_n\}_{n=1}^{\infty}$ is also Cauchy. Note that we arrive at a contradiction if we assume f takes Cauchy sequences to Cauchy sequences, for we have assumed that the distance between subsequent terms of $\{f(z_n)\}_{n=1}^{\infty}$ is greater than ε. ■

2.1.2 Compact metric spaces

The following facts will prove to be useful:

Lemma 2.16. *Let $f : X \to X$ be an isometry from a compact metric space into itself. Then it is necessarily a bijection.*

Proof. Injectivity is clear. For surjectivity, suppose there is some $x_0 \in X \setminus f(X)$. Since $f(X)$ is isometric to X, a compact set, it must also be compact (note that isometries preserve the diameters of sets, which implies that $f(X)$ is bounded. Also, isometries take ε-nets of their domain to ε-nets of their image, implying that $f(X)$ is totally bounded). Then $d(x_0, f(X)) = \varepsilon$ is strictly positive (if it were not, one could build a sequence $\{y_k\}_{k=1}^{\infty}$ of points in $f(X)$ with $d(x_0, y_k) < 1/k$. This would be a Cauchy sequence contained in $f(X)$ that converges in X to x_0 instead of a point in $f(X)$, contradicting completeness). Consider the sequence $(x_k)_{k=1}^{\infty}$ defined as $x_k = f^k(x_0)$ (that is the function f applied k times on x_0). Let $i < j$ be any natural numbers. We have

$$\begin{aligned} d(x_i, x_j) &= d(f^i(x_0), f^j(x_0)) \\ &= d(f^i(x_0), f^i(f^{j-i}(x_0))) \\ &= d(x_0, f^{j-i}(x_0)) \geq \varepsilon \end{aligned}$$

Therefore we have found a sequence on X with no Cauchy subsequence, contradicting its compactness. ■

Proposition 2.17. *Let $f : X \to Y$ and $g : Y \to X$ be isometries between metric spaces. If X is compact then g is an isometric bijection.*[2]

Proof. The composition $g \circ f$ is an isometry from a compact metric space into itself and must be a bijection. Then g is surjective, and so it must be a bijection. ■

2.1.3 The Hausdorff distance

Definition 2.18. Given a metric space (X, d) and two non-empty subsets $Y, Z \subset X$, we define the (Hausdorff) distance between Y and Z as

$$d_H(Y, Z) = \max\{\sup_{a\in Y} d(a, Z), \sup_{b\in Z} d(b, Y)\}.$$

Here's an equivalent and perhaps more clarifying definition:

$$d'_H(Y, Z) = \inf\{r \geq 0 : Z \subset \overline{B}_r(Y) \text{ and } Y \subset \overline{B}_r(Z)\}.$$

Proposition 2.19. *The two definitions above are equivalent*

Proof. First note that $d'_H(Y, Z)$ can be seen an an infimum of the intersection of two sets, hence

$$d'_H(Y, Z) \geq \max\{\inf\{r \geq 0 : Z \subset \overline{B}_r(Y)\}, \inf\{r > 0 : Y \subset \overline{B}_r(Z)\}\}$$

We wish to show that $\inf\{r \geq 0 : Y \subset \overline{B}_r(Z)\} = \sup_{a\in Y} d(a, Z)$. In order to do so, simply note that

$$r < d(a, Z) \iff a \notin \overline{B}_r(Z)$$

That is, $\{r \geq 0 \ : \ r = d(a, Z) \text{ for some } a \in Y\}$ and $\{r \geq 0 : Y \subset \overline{B}_r(Z)\}$ are complementary intervals and thus share a common extremity.

We have shown that $d'_H(Y, Z) \geq d_H(Y, Z)$. To show the inverse inequality, let $s = d_H(Y, Z) + \varepsilon$ for some $\varepsilon > 0$. That is, for each $a \in Y$ and $b \in Z$ we have $d(a, Z) \leq s$ and $d(b, Y) \leq s$. Then $s \in \{r \geq 0 : Z \subset \overline{B}_r(Y) \text{ and } Y \subset \overline{B}_r(Z)\}$. If we set $\varepsilon = \frac{d'_H(Y,Z)-d_H(Y,Z)}{2}$, assuming it is a positive number, we would arrive at the following contradiction:

$$s = d_H(Y, Z) + \frac{d'_H(Y, Z) - d_H(Y, Z)}{2} = \frac{d'_H(Y, Z) + d_H(Y, Z)}{2} < \frac{d'_H(Y, Z) + d'_H(Y, Z)}{2}$$

■

The second definition is useful because it allows us to easily show that the Hausdorff distance is a (potentially infinite) pseudometric. It is clearly symmetric and non-negative, leaving us with the task of proving the triangle inequality.

Proposition 2.20. *For any Y, Z, C subsets of X we have*

$$d_H(Y, C) \leq d_H(Y, Z) + d_H(Z, C)$$

Proof. First, note that r-neighbourhoods display a similar property:

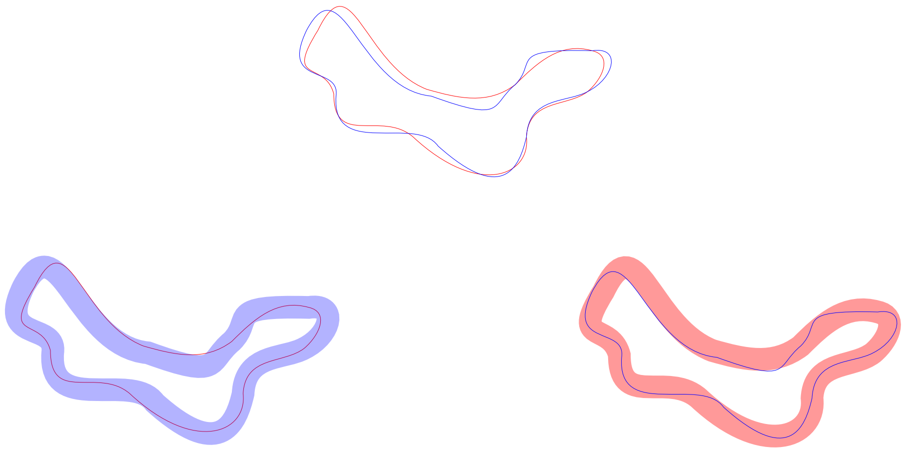
$$\overline{B}_r(\overline{B}_s(Y)) \subset \overline{B}_{r+s}(Y)$$

The first set is the collection of points $x \in X$ such that $d(x, x') \leq r$ for some $x' \in \overline{B}_s(Y)$, which in turn satisfies $d(x', a) \leq s$ for some $a \in Y$. Thus $d(x, a) \leq d(x, x') + d(x', a) \leq r + s$ and $x \in \overline{B}_{r+s}(Y)$.

Thus we can apply this observation to the second definition of the Hausdorff distance. Let $r = d_H(Y, Z)+\varepsilon$ and $s = d_H(Z, C)+\varepsilon$. By definition we have $Z \subset \overline{B}_r(Y)$, $Y \subset \overline{B}_r(Z)$, $C \subset \overline{B}_s(Z)$ and $Z \subset \overline{B}_s(C)$.

By combining $Y \subset \overline{B}_r(Z)$ and $Z \subset \overline{B}_s(C)$ we have $Y \subset \overline{B}_r(\overline{B}_s(C)) \subset \overline{B}_{r+s}(C)$. Analogously we obtain $C \subset \overline{B}_s(\overline{B}_r(Y)) \subset \overline{B}_{s+r}(Y)$. Thus, $d_H(Y, C) \leq d_H(Y, Z) + d_H(Z, C) + 2\varepsilon$ for arbitrary ε. ■

[2]This is a neat metric analogue of the Schröder–Bernstein Theorem for sets, the principle that states that if A and B are sets and there are injective functions $A \to B$ and $B \to A$, then there must be a bijection between A and B. Proposition 2.17 is not true if we do not require X to be compact: Consider the inclusion $f : [0, 1] \cup [2, \infty) \to [0, \infty)$ and $g : [0, \infty) \to [0, 1] \cup [2, \infty)$ given by $g(x) = x + 2$, which are isometries.



Two similar curves in the plane. The width of the blue and red bands gives an estimate of the Hausdorff distance between the curves.

It may not come as a surprise that the Hausdorff metric behaves even better when restricted to closed and bounded sets, compacts in particular:

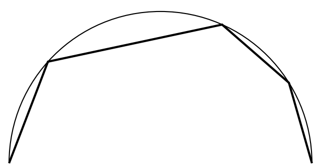
- For any bounded subsets $Y, Z \subset X$ the distance $d_H(Y, Z)$ must always be finite: The value of $\sup_{a\in Y} d(a, Z) = \sup_{a\in Y} \inf_{b\in Z} d(a, b)$ is bounded above by $\sup_{a\in Y, b\in Z} d(a, b)$. By fixing any $a' \in Y$ and $b' \in Z$, we get $d(a, b) < d(a, a') + d(a', b') + d(b', b) < \operatorname{diam}(Y) + d(a', b') + \operatorname{diam}(Z)$.
- For any subset $Y \subset X$ we must have $d_H(Y, \overline{Y}) = 0$: for any, we have $a \in Y$ $d(a, \overline{Y}) = 0$, since $a \in \overline{Y}$. Also, by definition, $d(a', Y) = 0$ for any $a' \in \overline{Y}$.
- Any closed subsets $Y, Z \subset$ with $d_H(Y, Z) = 0$ are equal. For, assuming $Y \not\subset Z$, there must be $a \in Y$ such that $d(a, Z) > 0$ (Otherwise, $a \in Z$). Since $d(Y, Z) \geq d(a, Z)$, we arrive at a contradiction.

Thus we then have the Hausdorff Space of a metric space:

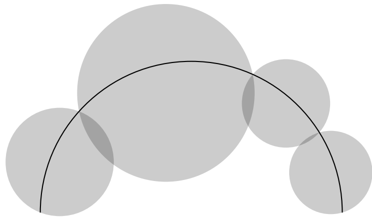
Theorem 2.21. *The set of all closed and bounded subsets of a metric space is a metric space with respect to the Hausdorff metric.*

2.1.4 Measures and dimension

We wish to generalize the notions of lengths, areas and volumes to general metric spaces. A difficulty lies in the fact that whatever generalization we choose will be deeply tied to the concept of *dimension*. To see that, recall the usual definition of the length of a curve in $\mathbb{R}^2$, by approximating a curve by polygonal paths, whose length we can calculate directly. The approximate length obtained this way increases (more accurately, is non-decreasing) as we choose finer and finer approximations and, for nice enough curves, approaches a limit, which we call its length.



Alternatively, we could achieve the same result by covering the curve by disks. Assuming that a given approximation of the curve by a polygonal path is fine enough, the section of the curve contained between any two adjacent vertices of the polygon will be entirely contained within the disk that has the corresponding edge as its diameter. Then by adding the diameters of all the disks we recover the estimate for the length. We obtain progressively better approximations by covering the curve by disks of progressively smaller radii, which is analogous to choosing finer polygonal approximations. This is something we could easily generalize to metric spaces.



Consider, though, what would happen once we attempted to measure the *length* of an open region of the plane of area A using this method: Since the area of a disk of diameter D is proportional to D^2, the minimum number N of disks with diameter not greater then D required to cover the region will be bounded below by a multiple of $\frac{A}{D^2}$. Adding the diameters of the N disks we get something that grows at least as quickly as $N \cdot D \geq \frac{A}{D^2} \cdot D = \frac{A}{D}$ when D approaches zero. That is, the *one dimensional size* of an open region of the plane is always infinite.

This does gives us a clue as to how to put areas in terms of covers and diameters: Instead of adding the diameters of all the disks we might want to add something proportional to the square of each diameter, assuring that the total sum does not increase asymptotically when $D \to 0$. Note that the same argument as above shows that if we define areas in this way, curves that have positive lengths will have no area.

Definition 2.22. Let X be a metric space and α a non-negative real number. We define an auxiliary quantity

$$H^\alpha_\delta(X) = \inf \left\{ \sum_{i=1}^{\infty} [\operatorname{diam}(S_i)]^\alpha \ : \ X = \bigcup_{i=1}^{\infty} S_i, \ \operatorname{diam}(S_i) \leq \delta \right\}$$

Where we use the non-standard conventions that $0^0 = 1$ (for when $\alpha = 0$ and each $\operatorname{diam}(S_i) = 0$) and that $\inf \varnothing = \infty$ (for when X admits no appropriate countable cover). We allow H^α_δ to take values on the extended number line $\mathbb{R} \cup \{\infty\}$ or $[0, \infty]$.

We define the α-dimensional measure of X as

$$H^\alpha(X) = \lim_{\delta \to 0} H^\alpha_\delta(X),$$

where again we allow the limit to assume the value ∞.

Remark 2.23. The α-dimensional measure of any subset S of a metric space X is not greater than the α-dimensional measure of X, for if $X = \bigcup_{i=1}^{\infty} S_i$ and $\operatorname{diam}(S_i) \leq \delta$, then $S = \bigcup_{i=1}^{\infty} (S_i \cap S)$ and $\operatorname{diam}(S_i \cap S) \leq \delta$.

Remark 2.24. The α-dimensional measure of two isometric spaces is equal for all $\alpha \geq 0$. Indeed, if $f : X \to Y$ is an isometry, then $X = \bigcup_{i=1}^{\infty} S_i$ and $\operatorname{diam}(S_i) \leq \delta$ if and only if $Y = \bigcup_{i=1}^{\infty} f(S_i)$ and $\operatorname{diam}(f(S_i)) \leq \delta$, which implies $H^\alpha_\delta(X) \geq H^\alpha_\delta(Y)$ and thus $H^\alpha(X) \geq H^\alpha(Y)$. The opposite inequality follows from symmetry.

Remark 2.25. If $X = \bigcup_{j=1}^{\infty} X_j$, then $H^\alpha(X) \leq \sum_{j=1}^{\infty} H^\alpha(X_j)$. For, if each $X_j = \bigcup_{i=1}^{\infty} S_{i,j}$ with each $\operatorname{diam}(S_{i,j}) \leq \delta$, then $X = \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} S_i$, which implies $H^\alpha_\delta(X) \leq \sum_{j=1}^{\infty} H^\alpha_\delta(X_j)$.

Note that we had no particular reason to restrict α to be an integer. In fact, many interesting spaces have a positive α-dimensional measure for fractional α.

Example 2.26. Consider the Cantor set: define $C_0 = [0, 1] \subset \mathbb{R}$ and $C_n = {}^1\!/_3 C_{n-1} \cup ({}^2\!/_3 + {}^1\!/_3 C_{n-1})$ for all $n \in \{1, 2, \ldots\}$. The Cantor set is defined as the intersection $C = \bigcap_{n=0}^{\infty} C_n$. Note that each C_n is composed of 2^n intervals of length ${}^1\!/_{3^n}$, which forms a finite cover of C. Thus, $H^\alpha_{1/3^n}(C) \leq \left(\frac{2}{3^\alpha}\right)^n$ for all $n \in \{0, 1, \ldots\}$, which implies that $H^\alpha(C) = 0$ for all $\alpha > \log_3(2)$.

We now formalize the discussion in the beginning of this subsection about curves having no area and open regions of the plane have infinite length:

Proposition 2.27. *Let X be a metric space. There is a unique number $d \in [0, \infty]$ such that $H^\alpha(X) = \infty$ for all $\alpha < d$ and $H^\alpha(X) = 0$ for all $\alpha > d$.*

Proof. The set $\{\alpha \in [0, \infty) \ : \ H^\alpha(X) \neq \infty\}$ is bounded below so it has an infimum $d \in [0, \infty]$ (since we have set $\inf \varnothing = \infty$). Of course, $H^\alpha(X) = \infty$ for all $\alpha < d$. If $d = \infty$, we are done. Otherwise, let $\alpha > d$. By the definition of inf, there is some $\beta \in (d, \alpha)$ such that $H^\beta(X) = M \neq \infty$.

Let $\varepsilon > 0$. By the definition of H^β, there is some $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$ we have $H^\beta_\delta(X) < M + \varepsilon$. By the definition of H^β_δ, there is some cover $\{S_i\}_{i=1}^\infty$ of X with $\operatorname{diam}(S_i) < \delta$ such that

$$\sum_{i=1}^\infty [\operatorname{diam}(S_i)]^\beta < H^\beta_\delta(X) + \varepsilon < M + 2\varepsilon$$

By considering that $\beta < \alpha$, we can see that:

$$\begin{aligned}
\sum_{i=1}^\infty [\operatorname{diam}(S_i)]^\beta \leq \sum_{i=1}^\infty [\operatorname{diam}(S_i)]^\alpha &= \sum_{i=1}^\infty [\operatorname{diam}(S_i)]^\beta \, [\operatorname{diam}(S_i)]^{\alpha-\beta} \\
&< \sum_{i=1}^\infty [\operatorname{diam}(S_i)]^\beta \, \delta^{\alpha-\beta} \\
&= \delta^{\alpha-\beta} \sum_{i=1}^\infty [\operatorname{diam}(S_i)]^\beta \\
&< \delta^{\alpha-\beta}(M + 2\varepsilon)
\end{aligned}$$

Since both ε and δ are arbitrary (δ only needs to be in $(0, \delta_0)$) the cover can be chosen such that this sum becomes arbitrarily small. Thus $H^\beta(X) = 0$. ■

We can then unequivocally define the dimension of any metric space:

Definition 2.28. Let X be a metric space. The number d from Proposition 2.27 is called the *Hausdorff dimension* of X, or $\dim(X)$.

Example 2.29. The Hausdorff dimension of n-dimensional Euclidean space is n (even for $\mathbb{R}^0$, if we define it as a one point space).

Example 2.30. Example 2.26 shows that the Hausdorff dimension of the Cantor set is not greater than $\log_3(2)$. It is possible show that it is precisely $\log_3(2)$.

These are fundamental properties of the dimension:

Remark 2.31. Let X be a metric space and let $\{S_i\}_{i=1}^\infty$ be a cover of X by subsets. We have

$$\dim(X) = \sup\{\dim(S_i) \ : \ i \in \{1, 2, \ldots\}\}$$

Indeed, from Remark 2.25, we know that

$$\dim(X) = \sup\{\alpha \geq 0 \ : \ H^\alpha(X) = 0\} \leq \sup\{\dim(S_i) \ : \ i \in \{1, 2, \ldots\}\}$$

But Remark 2.23 implies that $\dim(X) \geq \dim(S_i)$.

Remark 2.32. Isometric metric spaces have the same dimension, which follows directly from the definition and from Remark 2.24.

Corollary 2.33. *Let X be a homogeneous, separable metric space. Suppose there is an open set $U \subset X$ such that $\dim(U) = d$. Then $\dim(X) = d$.*

Proof. Let $B \subset U$ be an open ball of radius ε centered at some $x \in U$. Let $\{x_i\}_{i=1}^\infty$ a countable and dense subset of X. There is a family of isometries $f_i : X \to X$ such that $f_i(x) = x_i$. The sets $f_i(U)$ all have the same dimension, by Remark 2.32, and cover all of X, since each one of them contains a ball of radius ε centered around some x_i, and thus the result follows from Remark 2.31. ■

2.2 Topological spaces

Let us discuss open and closed subsets of metric spaces. Open subsets have the familiar properties of open subsets of $\mathbb{R}$: The intersection of two open subsets is again open and the union of any family of open subsets is again open. The empty set and the entire metric space are both open.

Also similarly to subsets of the line, the complement of a subset is open if and only if the subset is closed. Note that by De Morgan's laws we infer that the intersection of any family of closed subsets is closed and the union of two closed subsets is closed. These are the features of metric spaces that we generalize into topological spaces:

Definition 2.34. Let X be any set. A set τ of subsets of X is called a *topology* on X if it satisfies the following set identities:

1. The empty set and X are in τ.
2. Any union of elements of τ is in τ.
3. The intersection of two elements of τ is in τ.

In such case the pair (X, τ) is called a *topological space* and an element of τ is called an *open* subset of X with respect to this topology. The complement of any open subset is called a *closed* subset.

Of course, metric spaces are topological spaces and the two definitions of closed and open sets are consistent.

Example 2.35. The set of all open sets of a metric space forms a topology.

Example 2.36. Let X be any set. We may always endow it with a topology, known as the discrete topology. Let τ be the set of all subsets of X (equivalently, let τ be the topology on X such that all unitary sets be open). Also equivalently, this is the topology induced on X by the metric

$$d(a,b) = \begin{cases} 1 \text{ if } a \neq b \\ 0 \text{ if } a = b. \end{cases}$$

Example 2.37. Let (X, τ) be a topological space. Any subset $S \subset X$ can be made a topological space, endowed with the topology $\{A \cap S \ : \ A \in \tau\}$, known as the subspace topology or the induced topology. Whenever we refer to some subset of a topological space, it may be assumed that it carries the subspace topology.

Below are some basic definitions regarding topological spaces we are going to need:

Definition 2.38. In what follows (X, τ) and (Y, υ) are topological spaces.

1. The *closure* $\overline{S}$ of a subset $S \subset X$ is the intersection of all closed subsets of X that contain X (for metric spaces, this is equivalent to the definition given previously).
2. A subset $S \subset X$ is called *dense* in X if $\overline{S} = X$.
3. A function $f : X \to Y$ is *continuous* w.r.t. τ and υ if for all $S \in \upsilon$ we have $f^{-1}(S) \in \tau$. For topological spaces that are metric, this is equivalent to the definition given before.
4. A continuous function $f : X \to Y$ is a *homeomorphism* if it is a bijection and its inverse $f^{-1} : Y \to X$ is continuous. In this case X and Y are *homeomorphic* to each other (this is an equivalence relation).
5. If a property of a topological space is shared by all topological spaces homeomorphic to it, we call this property a *topological property*, or a *topological invariant.* Properties 6 to 12 below are all examples of topological invariants.
6. The minimal cardinal number $\mathfrak{c}$ such that X admits a dense subset with cardinality $\mathfrak{c}$ is called the *density* of X. A topological space is called *separable* if it has the density $\#\mathbb{N}$, i.e., has a countable, dense subset.
7. We say that (X, τ) is a Hausdorff space if for all $a, b \in X$ there are two disjoint open sets $A, B \in \tau$ such that $a \in B$ and $b \in B$ (note that every metric space is Hausdorff).

8. A topological space is called *connected* if it has exactly two subsets that are both open and closed: The empty set and the entire space. The maximal (w.r.t. inclusion) connected subsets of a topological space are called its *connected components*

9. X is called *path connected* if for all $a, b \in X$ there is a continuous function $f : [0,1] \to X$ (where $[0,1]$ is given the usual metric topology as a subset of $\mathbb{R}$) such that $f(0) = a$ and $f(1) = b$.

10. X is called *locally connected* (resp. *locally path connected*) if for all $a \in X$ and open set $A \subset X$, there is an open set $B \subset A$ such that $x \in B$ and B is connected (resp. path connected).

11. X is called *compact* if for all families $\{A_\lambda\}_{\lambda \in L}$ with each $A_\lambda \in \tau$ and $X = \bigcup_{\lambda \in L} A_\lambda$ we have a finite subset $L' \subset L$ with $X = \bigcup_{\lambda \in L'} A_\lambda$. For metric spaces, this is equivalent to the previous definition.

12. X is called *locally compact* if for all $a \in X$ there is an open subset $A \subset X$ and compact subset $K \subset X$ with $a \in A \subset K$.

Example 2.39. Let X be any set. On the opposite direction to Example 2.36, we may give it the set $\{\varnothing, X\}$ as a topology. This is very much not a Hausdorff space, since all points belong to the same open set and to none other. This is the only example we will give of a topology that does not come from an underlying metric.

Often a topology on a set is not defined directly, but from a special collection of open sets called a basis:

Definition 2.40. Let (X, τ) be a topological space. A subset $\mathcal{B} \subset \tau$ is called a basis for this topology if it satisfies

1. Every element of X belongs to an element of $\mathcal{B}$.

2. For every $B_1, B_2 \in \mathcal{B}$ and $a \in B_1 \cap B_2$ there is some $B_3 \in \mathcal{B}$ such that $a \in B_3 \subset B_1 \cap B_2$.

Conversely, if X is any set and $\mathcal{B}$ is a set of subset of X satisfying 1. and 2., then the set of all subset of X that can be built from an union of elements of $\mathcal{B}$ is a topology and is called the *topology generated by* $\mathcal{B}$. In this case, $\mathcal{B}$ is a basis for this topology.

The minimum cardinal number $\mathfrak{c}$ such that a topological space (X, τ) admits a basis of cardinality $\mathfrak{c}$ is called its *weight*. A topological space is called second countable if its weight is $\#\mathbb{N}$, i.e., if it has a countable basis.

Bases themselves are often not constructed directly but formed from smaller sets called subbases. A *subbasis* $\mathcal{S}$ of (X, τ) is any subset of τ such that the set of all finite intersections of elements of $\mathcal{S}$ is a basis for X. If X is any set and $\mathcal{S}$ is a collection of its subsets we can construct the minimal topology on X that contains $\mathcal{S}$ as the collection of all arbitrary unions of finite intersections of elements of $\mathcal{S}$. Then $\mathcal{S}$ is a subbasis of this topology, which we call the topology *generated by* $\mathcal{S}$.

Since homeomorphisms take bases of a topological space to bases of the image, the weight is also a topological invariant. Second countability is (assuming the axiom of choice) a stronger property than separability: Take a countable basis and from each set in the basis take one element, making a countable set of points. Given some point in the second countable space and open set containing it, there is a basis element containing the point and contained in the open set. Thus one of the countably many points chosen will intersect the open set.

Example 2.41. Let X be any set, $\{(Y_\lambda, \upsilon_\lambda)\}$ be a family of topological spaces and $\{f_\lambda : X \to Y_\lambda\}_{\lambda \in L}$ be a family of functions. In order to f_λ to be continuous with respect to some topology τ on X it is necessary that $f_\lambda^{-1}(B) \in \tau$ for all $B \in \upsilon_\lambda$. We may endow X with the topology generated by the set

$$\{f_\lambda^{-1}(B) \ : \ \lambda \in L, \ B \in \upsilon_\lambda\}$$

This is called the initial or coarsest topology on X with respect to. $\{f_\lambda\}_{\lambda \in L}$. It is the minimal topology (w.r.t. inclusion) that makes each f_λ continuous.

Example 2.42. Let $\{X_\lambda\}_{\lambda\in L}$ be a family of topological spaces. The topology usually given to the Cartesian product $X = \prod_{\lambda\in L} X_\lambda$ (i.e., to the set of all functions $f : L \to \bigcup_{\lambda\in L} X_\lambda$ satisfying $f(\lambda) \in X_\lambda$) is the initial topology w.r.t. to all of the projection maps $\pi_\lambda : X \to X_\lambda$ defined as $\pi_\lambda(f) = f(\lambda)$. This is called the product topology.

Note that if $\{A_\lambda\}_{\lambda\in L}$ is a collection such that each A_λ is an open subset of X_λ and $A_\lambda = X_\lambda$ for all but finitely many $\lambda \in L$, then $\prod_{\lambda\in L} A_\lambda$ is an open subset of X (in fact, subsets of this kind form a basis for the product topology).

Example 2.43. If $\{X_n\}_{n=1}^{\infty}$ is a countable collection of second countable spaces, each with a countable basis $\mathcal{B}_n$, then the set of all products $A_1 \times \ldots \times A_N \times \prod_{n=N+1}^{\infty} X_n$ with each $A_n \in \mathcal{B}_n$ is a countable basis for $\prod_{n=1}^{\infty} X_n$. This countable products of second countable spaces are second countable.

Example 2.44. If each space in $\{X_\lambda\}_{\lambda\in L}$ is a Hausdorff space, then so is $X = \prod_{\lambda\in L} X_\lambda$. Given distinct $f_1, f_2 \in X$, we have by definition $f_1(\lambda') \neq f_2(\lambda')$ for some $\lambda' \in L$. We take distinct open sets $A_{\lambda'}, B_{\lambda'} \subset X_\lambda$, respectively containing $f_1(\lambda')$ and $f_2(\lambda')$ and note that the subsets $A_\lambda \times \prod_{\lambda\in L\setminus\{\lambda'\}} X_\lambda$ and $A_\lambda \times \prod_{\lambda\in L\setminus\{\lambda'\}} X_\lambda$ are open, distinct and contain f_1 and f_2, respectively.

The product topology described above behaves as expected for finite families of topological spaces. The topological product of two metric spaces has the same topology as the metric product described previously. A property of the product topology that we can not neglect (for it characterizes it) is the following:

Proposition 2.45. *Let $X = \prod_{\lambda\in L} X_\lambda$ be a product of topological spaces and and $\{f_\lambda\}_{\lambda\in L}$ be a family of continuous functions $f_\lambda : Y \to X_\lambda$ for some fixed topological space Y. Then there is exactly one continuous function $f : Y \to X$ such that $\pi_\lambda \circ f = f_\lambda$ (where $\pi_\lambda : X \to X_\lambda$ are the projections). In particular, a function $f : Y \to X$ is continuous if and only if each $\pi_\lambda \circ f : Y \to X_\lambda$ is continuous.*

Proof. There is only one way to define f: As the the function $f : Y \to X$ such that $f(b)(\lambda) = f_\lambda(b)$ for all $b \in Y$. We must show that it is continuous. We know that X has a basis composed by all products $A = \prod_{\lambda\in L} A_\lambda$, with all $A_\lambda \subset X_\lambda$ open and only finitely many of them satisfying $A_\lambda \subsetneq X_\lambda$. It then suffices to show that if each $B_\lambda = f_\lambda^{-1}(A_\lambda) \subset Y$ is open, then $B = f^{-1}(A)$ must also be open. The set B is composed of all of the elements $b \in Y$ such that $f_\lambda(b) \in B_\lambda$. Thus $B = \bigcap_{\lambda\in L} B_\lambda$. Since we have $B_\lambda = Y$ for all but finitely many $\lambda \in L$, we have that B is a finite union of open subsets of Y and therefore it is itself open. ∎

2.2.1 Manifolds

The Euclidean spaces $\mathbb{R}^n$ are very nice. They are metric spaces and thus Hausdorff spaces. They have a countable dense set, namely $\mathbb{Q}^n$. They are second countable, as one can check that the set of all open balls with rational radii and centered at an element of $\mathbb{Q}^n$ is a basis for its topology. They are connected and locally path connected. The following class of spaces tries to be very general, while maintaining these and other nice properties of $\mathbb{R}^n$:

Definition 2.46. A topological space X is called a n-dimensional topological manifold if it is a second countable Hausdorff space such that every point $x \in X$ is contained in some open subset of X that is homeomorphic to an open subset of $\mathbb{R}^n$.

One usually writes "let X^n be a topological manifold" as short for "let X be a n-dimensional topological manifold", even though this convention is a bit confusing.

Proposition 2.47. *Let X_1^m, X_2^n be two topological manifolds. Then $X_1 \times X_2$ is a $(m+n)$-dimensional topological manifold. If $m = n$, then $X_1 \sqcup X_2$ is a n-dimensional topological manifold.*

Proof. The product and disjoint union are both Hausdorff and second countable spaces (see Examples 2.44 and 2.43). If $(a, b) \in X_1 \times X_2$, take $U_1 \subset X_1$ and and $U_2 \subset X_2$, open sets that contain a and b, respectively, and let $\phi_1 : U_1 \to V_1$ and $\phi_2 : U_2 \to V_2$ be homeomorphisms with

$V_1 \subset \mathbb{R}^m$ and $V_2 \subset \mathbb{R}^n$ both open. As we have seen, the product $U_1 \times U_2$ is open and contains (a, b) and $V_1 \times V_2 \subset \mathbb{R}^{m+n}$ is also open. The function $\phi_1 \times \phi_2 : U_1 \times U_2 \to V_1 \times V_2$, defined as $(x_1, x_2) \mapsto (\phi_1(x_1), \phi_2(x_2))$ is the needed homeomorphism.

For the disjoint union it is simpler: Let $a \in X_1 \sqcup X_2$. We may assume without loss of generality that $a \in X_1$. Then an open set in X_1 containing a that is homeomorphic to a subset of $\mathbb{R}^n$ is also homeomorphic to a corresponding open subset of $X_1 \cup X_2$ containing a. ■

The definition above is often put in terms of *charts* and *atlases*. If X is a n-dimensional topological manifold and $a \in X$, then there is some open set $U \subset X$ containing x, some open set $V \subset \mathbb{R}^n$ and some homeomorphism $\phi : U \to V$. Then ϕ is called a *coordinate chart*, or a *local coordinate system*. The definition really asks for a collection of charts $\{(U_\lambda, \phi_\lambda)_{\lambda \in L}$ such that $X = \bigcup_{\lambda \in L} U_\lambda$. Such collection is called an atlas.

A chart is simply a way to endow each point in U a set of real coordinates, allowing one to treat this open set as if it where a subset of $\mathbb{R}^n$. The catch is that, unless a single chart covers the entire manifold, a particular coordinate system surrounding a point can only be used in its domain. Besides, a single point will typically be contained in the domains of several charts and thus it will have many representations through coordinates. This leads to the importance of studying change of coordinates functions.

If $\phi_1 : U_1 \to V_1$ and $\phi_2 : U_2 \to V_2$ are two charts and $a \in U_1 \cap U_2 \neq \varnothing$, the following functions allow us to go back and forth between the two systems of coordinates:

$$\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$$

$$\phi_1 \circ \phi_2^{-1} : \phi_2(U_2 \cap U_1) \to \phi_1(U_2 \cap U_1)$$

These are of course homeomorphisms and are called change of coordinate functions. Further restrictions can be added to Definition 2.46 by requiring that these functions are as nice as one would like.

Definition 2.48. Let (X, τ) be a n-dimensional topological manifold. An atlas $\mathfrak{A} = \{(U_\lambda, \phi_\lambda)_{\lambda \in L}$ on X is called *smooth* if all of its change of coordinates functions are smooth (as in C^∞ functions between open subsets of $\mathbb{R}^n$).

Two smooth atlases $\mathfrak{A}$ and $\mathfrak{B}$ on X are called compatible if $\mathfrak{A} \cup \mathfrak{B}$ is again smooth. This is an equivalence relation on the class of all smooth atlases of X. A particular equivalence class $\overline{\mathfrak{A}}$ with respect to this relation is called a *smooth structure* on X. We then call the triple $(X, \tau, \mathfrak{A})$ a *smooth manifold*.

If $(X, \tau, \overline{\mathfrak{A}})$ and $(Y, \upsilon, \overline{\mathfrak{B}})$ are two smooth manifolds, we say that a function $f : X \to Y$ is *smooth* at some point $a \in X$ if for some $(U_1, \phi) \in \mathfrak{A}$ and $(U_2, \psi) \in \mathfrak{B}$ with $a \in U_1$ and $f(a) \in U_2$ and $f(U_1) \subset U_2$ we have that $\psi \circ f \circ \phi^{-1} : \phi(U_1) \to \psi(U_2)$ is smooth. We say that f is smooth if it is smooth at every $a \in X$. Furthermore, f is called a *diffeomorphism* if it is a bijection and its inverse is smooth. Then X and Y are said to *diffeomorphic* smooth manifolds.

2.3 Groups and actions

Definition 2.49. A group structure on a set G is an ordered pair $(G, \cdot)$, where $\cdot : G \times G \to G$ is a function (evaluated as $g_1 \cdot g_2$) that satisfies the following properties

1. $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$ for all $g_1, g_2, g_3 \in G$ (associativity);
2. There exists $e \in G$ such that $e \cdot g = g$ and $g \cdot e = g$ for all $g \in G$ (existence of an identity);
3. For all $g \in G$ there is some $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$ (existence of an inverse).

The function $\cdot$ is called the group operation. When the operation is implicit we often call the set G itself a group. When two groups are being dealt with simultaneously it is common to use the same symbol for the operation of both groups. It is also common to omit the symbol completely, writing $a \cdot b$ as ab.

Here are some essential definitions:

Definition 2.50.

1. A function $f : G \to H$ between groups is a *group homomorphism* when it preserves the group operation, i.e.,

$$f(g_1 \cdot g_2) = f(g_1) \cdot f(g_2)$$

Note that homomorphisms can be deduced to take the identity of G to the identity of H.

2. A group homomorphism is called a *group isomorphism* when it is bijective.

3. A subset $H \subset G$ is a subgroup when the restriction of the group operation to $H \times H$ turns H into a group. Then the inclusion $i : H \to G$ is a group homomorphism.

4. If $f : G \to H$ is a homomorphism and $e_H \in H$ is the identity, then the set $f^{-1}(e_H)$ is known as the *kernel* of f, or $\ker f$. This is a subgroup of G.

5. A subgroup $N \subset G$ is *normal* if there is some homomorphism $f : G \to H$ such that $N = \ker f$.

6. The subgroup of G generated by a subset $S \subset G$ (denoted $\langle S \rangle$) is the minimal subgroup of G with respect to inclusion that contains S. Equivalently, this subgroup consists of all elements of G that can be written as a finite string $s_1 \cdot \ldots \cdot s_n$ of elements of S and their inverses.

7. A group G is called *finitely generated* when $G = \langle S \rangle$ for some finite subset $S \subset G$. A group is called *cyclic* if it can be generated by a unitary set.

8. Let $H \subset G$ be a subgroup. A left coset of H in G is given for a fixed $g \in G$, defined as $Hg = \{hg \ : \ h \in H\}$. The corresponding right coset is defined as $gH = \{gh \ : \ h \in H\}$. A subgroup H is normal if and only if $gH = Hg$ for all $g \in G$.

9. Let $H \subset G$ be a subgroup. The cardinality of the set ${}^{G}\!/_{H} = \{Hg \ : \ g \in G\}$ is called the *index* of H in G. It is denoted $|G : H|$.

10. If H is normal, we can give ${}^{G}\!/_{H}$ a group structure with the operation $H(g_1) \cdot H(g_2) = H(g_1 g_2)$. The function $\pi : G \to {}^{G}\!/_{H}$ given by $g \mapsto Hg$ is then a surjective group homomorphism, known as the *quotient map*, and H is its kernel.

11. The set $Z(G) = \{z \in G \ : \ zg = gz \ \forall g \in G\}$ is known as the *center* of the group. This is a normal subgroup of G. Note that $Z(G) = G$ if and only if the group operation is commutative. In this case we say that G is an *abelian* group.

12. Let $g_1, g_2 \in G$. The element $[g_1, g_2] = g_1^{-1} g_2^{-1} g_1 g_2$ is known as the *commutator* of g_1 and g_2. If $H_1, H_2 \subset G$ are two subgroups, their *commutator subgroup* $[H_1, H_2] \subset G$ is the subgroup generated by the commutators $[g_1, g_2]$ with $g_1 \in H_1$ and $g_2 \in H_2$. In particular, $[G, G]$ is called the commutator subgroup of G, and it its always normal. The quotient group $G^{\text{ab}} = \frac{G}{[G,G]}$ is known as the *abelianization* of G.

13. The *derived series* of G is the sequence of subgroups $\{G^{(i)}\}_{i=0}^{\infty}$ with $G^{(0)} = G$ and $G^{(i+1)} = [G^{(i)}, G^{(i)}]$. A group is called *solvable* if it has a derived series with $G^{(i)} = \{e\}$ for all i greater than some N.

14. The *lower central series* of G is the sequence of subgroups $\{G_i\}_{i=0}^{\infty}$ with $G_0 = G$ and $G_{i+1} = [G_i, G]$. A group is called *nilpotent* if it has a lower central series with $G_i = \{e\}$ for all i greater than some N. All nilpotent groups are solvable and all abelian groups are nilpotent.

Example 2.51. The integers $\mathbb{Z}$ form a group under addition. All of its subgroups are in the form $n\mathbb{Z} = \langle \{n\} \rangle$ for some $n \in \mathbb{Z}$, i.e., the set of all multiples of n. The quotients $\frac{\mathbb{Z}}{n\mathbb{Z}}$ are, up to isomorphism, all of the cyclic groups (note that $\mathbb{Z}$ is isomorphic to $\frac{\mathbb{Z}}{0\mathbb{Z}}$).

Example 2.52. Let S be any set and let $e \notin S$ some element not in S. Let $W(S)$ be the set of all functions $f : \mathbb{N} \to (S \cup \{e\}) \times \{1, -1\}$ that satisfy $f(n) = (e, 1)$ for all but finitely many $n \in \mathbb{N}$. One could picture the elements of $W(S)$ as strings $f = s_{\lambda_1}^{\varepsilon_1} s_{\lambda_2}^{\varepsilon_2} \ldots s_{\lambda_k}^{\varepsilon_k}$ with each $s_{\lambda_i} \in S \cup \{e\}$ and $\varepsilon_i \in \{1, -1\}$ (that is, $f(n) = (s_{\lambda_n}, \varepsilon_n)$ if $n \in \{1, \ldots, k\}$ and $f(n) = (e, 1)$ otherwise).

Define an equivalence relation of $W(S)$ by equating strings that should be equal according to the group axioms by taking e to be the identity:

$$s_{\lambda_1}^{\varepsilon_1} \ldots s_{\lambda_i}^{\varepsilon_i} e^1 s_{\lambda_{i+1}}^{\varepsilon_{i+1}} \ldots s_{\lambda_k}^{\varepsilon_k} = s_{\lambda_1}^{\varepsilon_1} \ldots s_{\lambda_i}^{\varepsilon_i} e^{-1} s_{\lambda_{i+1}}^{\varepsilon_{i+1}} \ldots s_{\lambda_k}^{\varepsilon_k} = s_{\lambda_1}^{\varepsilon_1} \ldots s_{\lambda_i}^{\varepsilon_i} s_{\lambda_{i+1}}^{\varepsilon_{i+1}} \ldots s_{\lambda_k}^{\varepsilon_k}$$

$$s_{\lambda_1}^{\varepsilon_1} \dots s_{\lambda_{i-1}}^{\varepsilon_{i-1}} s_{\lambda_i}^{\varepsilon_i} s_{\lambda_i}^{-\varepsilon_i} s_{\lambda_{i+2}}^{\varepsilon_{i+2}} \dots s_{\lambda_k}^{\varepsilon_k} = s_{\lambda_1}^{\varepsilon_1} \dots s_{\lambda_{i-1}}^{\varepsilon_{i-1}} e^1 s_{\lambda_{i+2}}^{\varepsilon_{i+2}} \dots s_{\lambda_k}^{\varepsilon_k}$$

The quotient set $F(S)$ is then a group under the operation of concatenation of strings. We call this group the *free group* over S, or the group of words with letters in S. A group that is isomorphic to a free group over some set is also called a free group.

The group of integers is free, as it is isomorphic to a free group over a set with one element. Up to isomorphism, this is the only free, abelian group.

We may identify S with strings with only one letter (different from the identity) and under this identification we have $F(S) = \langle S \rangle$.

Example 2.53. Let X be a set. The group $\mathrm{Sym}(X)$ of all bijections $f : X \to X$ is a group under composition, which is known as the *group of permutations*, or the *symmetric group*. When X has additional structure, some particular subgroups of the group of permutations become very important:

- When X is a metric space, the subset of $\mathrm{Sym}(X)$ consisting of all bijective isometries is a subgroup.
- When X is a topological space, the subset of $\mathrm{Sym}(X)$ consisting of all homeomorphisms is a subgroup.
- When X is a smooth manifold, the subset of $\mathrm{Sym}(X)$ consisting of all diffeomorphisms is a subgroup.
- When X is a group, the subset of $\mathrm{Sym}(X)$ consisting of all isomorphisms is a subgroup. These are known as automorphisms groups.

Definition 2.54. Let X be a set and let $H \subset \mathrm{Sym}(X)$ be some subgroup of the group of permutations. An *action* of a group G on X is a group homomorphism $f : G \to H$. We may think of an action as a way to multiply an element of the group by an element of the set, often writing $g \cdot a$ instead of $f(g)(a)$ when $g \in G$ and $a \in X$.

We say that two elements $a, b \in X$ share the same orbit with respect to f when there is some $g \in G$ with $g \cdot a = b$. This is an equivalence relation on X and we then call $O_G(a)$, the equivalence class of $a \in X$, the *orbit* of a. The set of all equivalence classes X/G the *orbit space* of X with respect to f.

We say that an element $g \in G$ stabilizes $a \in X$ if $g \cdot a = a$. The set $G_a \subset G$ of all such elements is called the *stabilizer* of a, and it is a subgroup of G.

We say that the action $f : G \to H$ is *transitive* if for all $a, b \in X$ there is some $g \in G$ such that $g \cdot a = b$. Note that the orbit space is not very interesting in this case: There is only one equivalence class.

An action $f : G \to H$ is called *faithful* if it is injective.

Remark 2.55. Let f be an action of a group G on a set X. For all $a \in X$, there is a natural bijection between $O_a(G)$ and the set G/G_a (which is not necessarily a group, as G_a is often not normal). The map $\phi : G \to O_a(G)$ given by $g \mapsto g \cdot a$ is not injective in general, but we can characterize its fibers:

$$g_1 \cdot a = g_2 \cdot a \iff (g_2^{-1} g_1) \cdot a = a \iff g_2^{-1} g_2 \in G_a$$

That is, the partition of G in fibers with respect with ϕ is the same as the partition of G in cosets of G_a, which means the induced map $G/G_a \to O_a(G)$ is a bijection. In particular, the orbit of an element of X is finite if and only if its stabilizer is a subgroup of finite index in G.

Example 2.56. Let G be a group and $H \subset G$ be a subgroup. Let $\mathrm{Aut}(G)$ be the group of all isomorphisms $G \to G$. The function $H \to \mathrm{Aut}(G)$ that takes an element of H to the isomorphism $g \mapsto hg$ is an action. The orbit space G/H is the same as the set of left cosets defined previously. This is an example of an orbit space that inherits a structure from the original space (as long as H is normal).

Example 2.57. Let G be a group. The action $G \to \mathrm{Aut}(G)$ that takes an element of $g \in G$ to the automorphism $g' \mapsto g g' g^{-1}$ is known as *conjugation* action. It's kernel is precisely the center of G. For all $g' \in G$, its orbit is known in this case as its *conjugacy class*. Its stabilizer, known as the *centralizer* of g', consists of all $g \in G$ such that $g g' g^{-1} = g'$, or equivalently, $g g' = g' g$: It is the group of all elements of G that commute with g'.

Example 2.58. We have defined a metric space X as being homogeneous if for all $a, b \in X$ there is an isometry $f : X \to X$ such that $f(a) = b$. This is precisely the same as saying that the group of bijective isometries of X acts transitively on the space.

Definition 2.59. Let $\{G_i\}_{i=1}^n$ be a finite family of groups. A group structure usually given to the Cartesian product $G_1 \times \ldots \times G_n$ is defined by just applying the group operations entrywise:

$$(g_1, \ldots, g_n) \cdot (g_1', \ldots, g_n') := (g_1 \cdot g_1', \ldots, g_n \cdot g_n')$$

When each of the groups G_i is abelian, the product is often denoted $G_1 \oplus \ldots \oplus G_n$ and called the direct sum of the family.

Note that each of the projections $\pi_i : G_1 \times \ldots \times G_n \to G_i$ defined as $(g_1, \ldots, g_n) \mapsto g_i$ is a surjective group homomorphism.

The following results concerning finitely generated groups will be necessary later. In what follows, the greatest common divisor of a finite set of integers $S = \{m_1, \ldots, m_n\}$ is, among all nonnegative common divisors of all elements of S, the greatest with respect to the partial order of divisibility, denoted $\gcd(m_1, \ldots, m_n)$. This means, for instance, that $\gcd(0, \ldots, 0) = 0$, since zero is greater than all nonnegative integers with respect to divisibility.

Lemma 2.60. *Let G be a finitely generated abelian group with a generating set $\{s_1, \ldots, s_n\}$. For every element in the form $m_1 s_1 + \ldots m_n s_n$ with each $m_i \in \{0, 1, \ldots\}$ (where $m_i s_i$ is just s_i added to itself m_i times and $0 s_i$ is the identity), there is a generating set $\{r_1, \ldots, r_n\}$ with*

$$m_1 s_1 + \ldots + m_n s_n = r_1 gcd(m_1, \ldots, m_n)$$

Proof. If $m_1 + \ldots + m_n = 0$ we have $0s_1 + \ldots + 0s_n = e = s_1 \gcd(0, \ldots, 0)$ and we are done. Otherwise we may assume that $\gcd(m_1, \ldots, m_n) = 1$. Let us argue by induction on $m = m_1 + \ldots + m_n$. If $m = 1$ we have $m_1 s_1 + \ldots m_i s_i = r_i$ for some $i \in \{1, \ldots, n\}$, so we obtain the next generating set by reordering the current.

If $m > 1$ we have at least two of $\{m_1, \ldots, m_n\}$ different from zero. Without loss of generality, $m_1 \geq m_2 > 0$.

Note that the set $\{s_1, s_1 + s_2, s_3, \ldots, s_n\}$ generates G. A little number theory shows that

$$\gcd(m_1 - m_2, m_2, m_3, \ldots, m_n) = \gcd(m_1, m_2, m_3, \ldots, m_n) = 1.$$

Also note that

$$(m_1 - m_2) + m_2 + m_3 + \ldots m_n = m_1 + m_3 + \ldots + m_n < m.$$

Thus we are allowed to use the inductive hypothesis on the generating set $\{s_1, s_1 + s_2, s_3, \ldots, s_n\}$ and the natural numbers $\{m_1 - m_2, m_2, m_3, \ldots, m_n\}$. But we also have

$$(m_1 - m_2)s_1 + m_2(s_1 + s_2) + m_3 s_3 + \ldots + m_n s_n = m_1 s_1 + \ldots + m_n s_n$$

■

Proposition 2.61. *Every finitely generated abelian group is isomorphic to a direct sum of finitely many cyclic groups.*

Proof. First we note that the existence of such isomorphism is equivalent to the existence of a generating set $\{s_1, \ldots, s_n\} \subset G$ such that for every sequence $m_1, \ldots m_n \in \mathbb{Z}$ we have

$$m_1 s_1 + \ldots + m_n s_n = 0 \implies m_1 = \ldots m_n = 0,$$

where again the multiplication of $s_i \in G$ by $m_i \in \mathbb{Z}$ simply means adding s_i to itself m_i times if m_i is positive or adding $-s_1$ to itself m_i times otherwise. Now we argue by induction: For every finitely generated group there is some $n \in \{1, 2, \ldots\}$ that is the minimum possible cardinality of a generating set (by the well-ordering principle). Note that if $n = 1$ the group is itself cyclic and thus the first inductive step is complete. Furthermore, among all members of all of the generating sets with n elements, there is some $s \in G$ such that the cardinality of $\langle\{s\}\rangle$ is minimal. We may assume that we are working with a generating set $\{s_1, \ldots, s_n\}$ with $s_1 = s$.

By the inductive hypothesis we have that $G_1 = \langle\{s_1\}\rangle$ and $G_2 = \langle\{s_2, \ldots, s_n\}\rangle$ are both isomorphic to a direct sum of cyclic groups. We wish then to show that G is isomorphic to $G_1 \oplus G_2$. Suppose not. Then there is a sequence $m_1 s_1 + m_2 s_2 + \ldots + m_n s_n = 0$ with $m_1 s_1 \neq 0$ or $m_2 s_2 + \ldots + m_n s_n \neq 0$. Since the second case implies the first we assume that $m_1 s_1 \neq 0$. By changing the sign of some of the terms, we may assume that $m_1, \ldots, m_n$ are all nonnegative. Since G_1 is cyclic, we may assume that $m_1 < \#G_1$. From Lemma 2.60 we obtain a generating set $\{r_1, \ldots, r_n\}$ such that

$$0 = m_1 s_1 + \ldots m_n s_n = r_1 \mathrm{gcd}(m_1, \ldots, m_n)$$

But this implies that the new generating set has n elements and $\#\langle\{r_1\}\rangle < \#\langle\{s_1\}\rangle$, a contradiction. ∎

Corollary 2.62. *Let G be a infinite finitely generated abelian group. There is a surjective homomorphism $f : G \to \mathbb{Z}$.*

Proof. Since G is isomorphic to the direct sum of a finite family of cyclic groups, there is a surjective homomorphism from G to each of these cyclic factors (the composition of the isomorphism with one of the projection maps). Since G is infinite, one of the factors must be infinite. But every infinite cyclic group $G = \langle\{g\}\rangle$ is isomorphic to $\mathbb{Z}$, though the isomorphism $G \to \mathbb{Z}$ given by $g^n \mapsto n$. ∎

Proposition 2.63. *Let G be a finitely generated group and $k \in \{1, 2, \ldots\}$. There are finitely many subgroups of G with index k.*

Proof. Let $S = \{1, 2, \ldots, k\}$. For all $H \subset G$ subgroups of index k, fix a bijection $f_H : {}^{G}/_{H} \to S$ that takes He to 1. Consider the action $\Psi_H : G \to \mathrm{Sym}({}^{G}/_{H})$ defined as $g(Hg') = H(g'g)$ (here ${}^{G}/_{H}$ is just a set, no group structure is needed). From f_H we can induce an action $\Phi_H : G \to \mathrm{Sym}(S)$.

In general there are only finitely many homomorphisms from a finitely generated group to a finite group: the behaviour of the homomorphism is completely determined by its value on the elements of some finite generating set. Thus Φ_H is one among finitely many actions from G on S.

But note that the stabilizer subgroup G_1, of elements $g \in G$ satisfying $g \cdot 1 = 1$, (with respect to the action Φ_H) is equal to H, regardless the specific bijection f_H we choose. Thus different subgroups induce different actions on S and, since there are finitely many actions, there must be finitely many subgroups. ∎

Proposition 2.64. *Let G be a finitely generated group. Every subgroup $H \subset G$ of finite index k is also finitely generated. If G has a generating set of n elements, then H must have a generating set with at most nk elements.*

Proof. Let $S = \{s_1, \ldots, s_n\}$ be a generating set for G. As in the proof of Proposition 2.63, consider the action $\Psi_H : G \to \mathrm{Sym}({}^{G}/_{H})$ defined as $g(Hg') = H(g'g)$. For each $g \in G$, a permutation is induced on ${}^{G}/_{H} = \{Hg_1, \ldots, Hg_k\}$, where we can assume that $g_1 = e$. If an element of the generating set s_j takes some Hg_i to some $Hg_{k_{i,j}}$, we may write $g_i s_j = h_{i,j} g_{k_{i,j}}$ for some $h_{i,j} \in H$.

An element $h \in H$ is an element of G and thus it can be written as $h = s_{j_1} \ldots s_{j_N}$. But each $s_j = g_1 s_j$ can be written as $h_{i,j} g_{k_{i,j}}$. Therefore we can make the following repeated substitutions:

$$\begin{aligned} h &= (h_{i_1,1} g_{k_{i_1,1}}) s_{i_2} s_{i_3} \ldots s_{i_n} \\ &= h_{i_1,1} (g_{k_{i_1,1}} s_{i_2}) s_{i_3} \ldots s_{i_n} \\ &= h_{i_1,1} (h_{k_{i_1,1},i_2} g_{k_{k_{i_1,1},i_2}}) s_{i_3} \ldots s_{i_n} \\ &= h_{i_1,1} h_{k_{i_1,1},i_2} (g_{k_{k_{i_1,1},i_2}} s_{i_3}) \ldots s_{i_n} \\ &= \cdots \end{aligned}$$

After a finite number of steps we will have expressed h as a combination of the elements of the set $\{h_{i,j}\}$, with i ranging in $\{1, 2, \ldots, k\}$ an j ranging in $\{1, 2, \ldots, n\}$. ∎

The following is known both as the subgroup lattice theorem and the correspondence theorem for groups (see Proposition 8.9, Chapter II on [Alu09]):

Proposition 2.65. *Let G be a group and $H \subset G$ be a normal subgroup, with $\pi : G \to {}^{G}/_{H}$ being the quotient map. Let $\mathcal{G}$ be the set of all subgroups of G that contain H and let $\mathcal{H}$ the set of all subgroups of ${}^{G}/_{H}$. Then the function $\Pi : \mathcal{G} \to \mathcal{H}$ given by $\Pi(A) = \pi(A)$ (usually written ${}^{A}/_{H}$) is well defined and it is a bijection.*

Moreover, $A \in \mathcal{G}$ is a normal subgroup of G if and only of $^{A}/_{H}$ is a normal subgroup of $^{G}/_{H}$. In this case, the groups $^{G}/_{A}$ and $\frac{G/H}{A/H}$ are isomorphic.

A subgroup $B \in \mathcal{H}$ is abelian if and only if H contains the commutator of $\Pi^{-1}(B)$

2.3.1 Metric actions

We have commented on the importance of homomorphisms from a group to the group of automorphisms of an object. We are going to focus on automorphisms of metric spaces, i.e., homomorphisms $f : G \to \mathrm{Iso}(X)$, where $\mathrm{Iso}(X) \subset \mathrm{Sym}(X)$ is the group of bijective isometries. We call a homomorphism of this sort a *metric action.* We wish to define an appropriate distance between points on $^{X}/_{G}$, the orbit space.

Definition 2.66. Let X be a metric space and let G be a group acting on X. Let $a, b \in X$. The induced distance between $O_G(a)$ and $O_G(b)$ is defined as

$$d(O_G(a), O_G(b)) = \inf\{d(a', b') \ : \ a' \in O_G(a),\ b' \in O_G(b)\}.$$

Proposition 2.67. *The distance above defines a pseudometric on $^{X}/_{G}$.*

Proof. Most easily, $d(O_G(a), O_G(a)) \le d(a, a) = 0$. For symmetry, take $a, b \in X$ and apply the definition of inf. Let $\varepsilon > 0$. There are $a' \in O_G(a)$ and $b' \in O_G(b)$ such that

$$d(b', a') < d(O_G(b), O_G(a)) + \varepsilon.$$

Thus we only need to choose small enough values for ε:

$$\begin{aligned} d(O_G(a), O_G(b)) &\le d(a', b') = d(b', a') \\ &< d(O_G(b), O_G(a)) + \varepsilon. \end{aligned}$$

For the triangle inequality we take $a, b, c \in X$ and apply the definition of the infimum. Let $\varepsilon > 0$. There are $a' \in O_G(a)$ and $b' \in O_G(b)$ such that

$$d(a', b') < d(O_G(a), O_G(b)) + \varepsilon.$$

Similarly, there are $b'' \in O_G(b)$ and $c' \in O_G(c)$ such that

$$d(b'', c') < d(O_G(b), O_G(c)) + \varepsilon.$$

Let $g \in G$ such that $gb' = b''$. We have

$$\begin{aligned} d(O_G(a), O_G(c)) &\le d(a', g^{-1}c') \\ &\le d(a', b') + d(b', g^{-1}c') \\ &\overset{!}{=} d(a', b') + d(gb', gg^{-1}c') \\ &= d(a', b') + d(b'', c') \\ &< d(O_G(a), O_G(b)) + d(O_G(b), O_G(c)) + 2\varepsilon. \end{aligned}$$

Since ε is chosen arbitrarily we have what was wanted. Notice that the equality marked by ! could fail if G wasn't required to act *isometrically* on X. ■

2.3.2 Topological groups

Once again we reflect on how nice the spaces $\mathbb{R}^n$ are. Not only these are metric spaces that have the topological and smooth structures that make topological and smooth manifolds possible, but they are also groups under vector addition. More importantly, the group operation as a function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and the inversion function $\mathbb{R}^n \to \mathbb{R}^n$ are continuous and smooth, i.e., the Euclidean spaces are groups in a way that is *compatible* with its geometrical structures. Again, this is something to generalize:

Definition 2.68. Let G be a set, τ a topology on G and $\cdot : G \times G \to G$ a group operation. Then $(G, \tau, \cdot)$ is called a topological group if the group operation and the inversion function $G \to G$ defined as $a \mapsto a^{-1}$ are continuous (where $G \times G$ is given the product topology).

If, additionally, $\overline{\mathfrak{A}}$ is a smooth structure of G such that the group operation and inversion function are smooth, then $(G, \tau, \cdot, \overline{\mathfrak{A}})$ is called a Lie group.

Proposition 2.69. *Let X be a metric space. The group of bijective isometries $Iso(X)$ is a topological group.*

Proof. We can see $\text{Iso}(X)$ as a topological space simply as a subset of the Cartesian product $X^X = \prod_{a\in X} X$ (this is known as the topology of pointwise convergence). By Proposition 2.45, it suffices to show that for all $a \in X$ the inversion function $I_a : \text{Iso}(X) \to X$, defined as $I_a(f) = f^{-1}(a)$, and the composition function $P_a : \text{Iso}(X) \times \text{Iso}(X) \to X$, defined as $P_a(f, g) = f(g(a))$, are both continuous.

Let $\varepsilon > 0$ and $f \in \text{Iso}(X)$ and consider the set

$$A = I_a^{-1}(B_\varepsilon(f(a))) = \{f' \in \text{Iso}(X) \ : \ d(f(a), f'(a)) < \varepsilon\}.$$

This is the inverse image of a basis element of X under I_a, so showing that these sets are open is enough to show that the function is continuous. Note that A is a basis element for $\text{Iso}(X)$: it equals the Cartesian product $\prod_{a'\in X} A_{a'}$, where $A_{a'} = X$ for all $a' \in X \setminus \{a\}$ and $A_a = B_\varepsilon(f(a))$. Thus the set is open, which is what was needed.

Now let $\varepsilon > 0$ and $(f, g) \in \text{Iso}(X) \times \text{Iso}(X)$ and consider

$$A' = P_a^{-1}(B_\varepsilon(f(g(a)))) = \{(f', g') \in \text{Iso}(X) \times \text{Iso}(X) \ : \ d(f(g(a)), f'(g'(a))) < \varepsilon\}.$$

Given $(f', g') \in A'$, let $\varepsilon' = \varepsilon - d(f(g(a)), f'(g'(a)))$. Define $A'' = \prod_{(a_1,a_2)\in X\times X} A_{(a_1,a_2)}$, where $A_{(g'(a),a)} = B_{\varepsilon'/2}(f'(g'(a))) \times B_{\varepsilon'/2}(g'(a))$ and $A_{(a_1,a_2)} = X \times X$ if $a_1 \neq a$ or $a_2 \neq g(a)$ (note that this is a basis element for $\text{Iso}(X) \times \text{Iso}(X)$ containing (f', g')). For all $(f'', g'') \in A''$ we have

$$\begin{aligned} d(f(g(a)), f''(g''(a))) &\leq d(f(g(a)), f'(g'(a))) + d(f'(g'(a)), f''(g'(a))) + d(f''(g'(a)), f''(g''(a))) \\ &= \varepsilon - \varepsilon' + d(f'(g'(a)), f''(g'(a))) + d(g'(a), g''(a)) \\ &< \varepsilon - \varepsilon' + \varepsilon'/2 + \varepsilon'/2 = \varepsilon \end{aligned}$$

That is, $(f', g') \in A'' \subset A$, which is what was needed. ■

In the year 1900, David Hilbert published a list of 23 unsolved problems that proved to be very influential in the development of mathematics in the 20th-century. The fifth of these problems asked whether the assumption of smoothness in the definition of a Lie group would exclude any topological group that is also a topological manifold. The solution was eventually presented in the 1950s by Andrew Gleason, Deane Montgomery and Leo Zippin [MZ55] as the following Theorem states:

Theorem 2.70. *Let $(G, \tau, \cdot)$ be a topological group. Then there exists a smooth structure $\overline{\mathfrak{A}}$ on (G, τ) that makes G into a Lie group if, and only if, (G, τ) is a topological manifold. In such case, this smooth structure is unique.*

A version of this Theorem (which can be found in [Gro81]) that is tailored to the topology on the group of bijective isometries given in Proposition 2.69 is the following:

Lemma 2.71. *Let X be a finite dimensional, locally compact, connected, locally connected homogeneous metric space. Its topological group of bijective isometries $Iso(X)$ is a topological manifold (and thus a Lie group) and has only finitely many connected components.*

For those that wish to know how this Lemma came to be: it can be shown (the last Corollary of Section 8 in [Are46], for instance) that the group of bijective isometries of a locally compact metric space is also locally compact under the topology of pointwise convergence. We have shown in Example 2.44 that it must also be a Hausdorff space. A result deeply tied to Hilbert's Fifth Problem, known as the Gleason-Yamabe Theorem (Theorem 1 on [Tao11]) then implies that the group of bijective isometries arises as an *inverse limit* of Lie groups (the inverse limit is an important categorical concept that unfortunately found no place in this text). The other hypotheses then match the first Corollary of section 6.3 in [MZ55], yielding the result.

These results are very useful, seeing that Lie groups have nice properties that are not shared in general by topological groups. Let us comment on one nice property that we are going to need later: associated to every n-dimensional Lie group G we have an n-dimensional vector space $\text{Lie}(G)$, known as its *Lie algebra*. The Lie algebra is itself a Lie group (over a metric topology inherited from some inner product, essentially the same as $\mathbb{R}^n$) and one can define a map $\exp : \text{Lie}(G) \to G$, known as the *exponential map*, satisfying the following properties:

1. $\exp : \mathrm{Lie}(G) \to G$ is a smooth map;
2. $\exp((\lambda_1 + \lambda_2)u) = \exp(\lambda_1 u) \cdot \exp(\lambda_2 u)$, for all $\lambda_1, \lambda_2 \in \mathbb{R}$ and $u \in \mathrm{Lie}(G)$;
3. $\exp(u)^{-1} = \exp(-u)$, for all $u \in \mathrm{Lie}(G)$ (thus $\exp(0) = e$);
4. There is an open set $A \subset \mathrm{Lie}(G)$ containing 0 and some open set $B \subset G$ containing e such that the restriction $\exp : A \to B$ is a diffeomorphism.

Definition 2.72. A topological group is said to satisfy the "No Small Subgroups" (NSS) property if it has some open set containing the identity that does not contain any subgroup except the trivial one.

Lemma 2.73. *Lie groups satisfy the NSS property.*

Proof. Using Property 4, let $A \subset \mathrm{Lie}(G)$ and $B \subset G$ such that $\exp : A \to B$ is a diffeomorphism. Since $\mathrm{Lie}(G)$ is a metric space, we may assume that A is some open ball centered at 0 with a radius $\varepsilon > 0$. Let A' be the concentric ball of radius $\varepsilon/2$. Assume that there is some $H \subset B' = \exp(A')$ that is a subgroup of G. Let $h \in H$ and $u \in A'$ such that $\exp(u) = h$.

Since H is a subgroup, we have $h^2 \in H \subset B'$, and so there is some $v \in A'$ such that $\exp(v) = h^2$. By Property 2 above we have $\exp(2u) = h^2$, which means $\exp(2u) = \exp(v)$. Since the metric on $\mathrm{Lie}(G)$ comes from some vector norm $| - |$, we have $d(e, 2u) = |2u| < \varepsilon$, thus $2u \in A$. Therefore $\exp(2u) = \exp(v) \implies 2u = v$, which means $2u \in A'$. Interactively, we have $2^n u \in A'$ for all $n \in \{1, 2, \ldots\}$. Since $|2^n u| = 2^n |u| < \varepsilon$, we must have $|u| = 0$ and thus $u = 0$. Then $h = \exp(u) = \exp(0) = e$. ■

3 The Gromov-Hausdorff distance

In the Section 2 we defined the Hausdorff distance between subsets of a fixed metric space. This idea can be generalized by considering all possible ways to embed two spaces into a common space. Similarly as with the Hausdorff distance, we define a pseudometric in a few equivalent ways and then show that it can be restricted into an actual metric. We show that the space of all compact metric spaces is a metric space in its own right, with respect to this metric.

Proposition 3.25 offers a condition for a family of compact metric spaces to have an accumulation point in the space. This result must be adapted to non-compact spaces, which we do in Theorem 3.38. First, we define the pointed Gromov-Hausdorff distance, which applies to metric spaces with a distinguished point. We define the notion of convergence of functions between pointed metric spaces and in Proposition 3.35 we give a condition for a sequence of functions to have a convergent subsequence.

Definition 3.1. Given two metric spaces (X, d_X), (Y, d_Y), let $\mathfrak{S}$ be the class of all pairs (f, g), where $f : X \to Z$ and $g : Y \to Z$ are isometries from X and Y to a common metric space Z (here Z is not fixed, but it ranges over all metric spaces that allow such isometries to exist!). We define the Gromov-Hausdorff distance between X and Y as

$$d_{GH}(X, Y) = \inf_{(f,g)\in\mathfrak{S}} \{d_H(f(X), f(Y))\}.$$

The class $\mathfrak{S}$ above is, of course, quite large. We can make things simpler by considering not all metric spaces that X and Y can be isometrically embedded into, but all pseudometrics that can be given to the disjoint union $X \sqcup Y$ that agree with d_X and d_Y in their domains.[3]

Definition 3.2. Let (X, d_X) and (Y, d_Y) be metric spaces, we say that a pseudometric d on $X \sqcup Y$ is *admissible* w.r.t. d_X and d_Y if $d(x_0, x_1) = d_X(x_0, x_1)$ for all $x_0, x_1 \in X$ and $d(y_0, y_1) = d_Y(y_0, y_1)$ for all $y_0, y_1 \in Y$.

Proposition 3.3. *Let (X, d_X) and (Y, d_Y) be two metric spaces. Define*

$$d'_{GH}(X, Y) = \inf\{d_H(X, Y)\}.$$

Where the $d_H(X, Y)$ ranges over the Hausdorff distances between X and Y relative to all pseudometrics on $X \sqcup Y$ that are admissible. Then $d'_{GH}(X, Y) = d_{GH}(X, Y)$.

Proof. Let $f : X \to Z$ and $g : Y \to Z$ be two isometries, where Z caries the metric d_Z. Let us define a pseudometric d on $X \sqcup Y$ such that $d_H(X, Y) \leq d_H(f(X), g(Y)) = r$. Set $d(x, y) = d_Z(f(x), g(y))$ for all $x \in X$ and $y \in Z$, as in Remark 2.4.

By the definition of the Hausdorff distance between $f(X)$ and $g(X)$ (in fact, due to Proposition 2.19), for all $\varepsilon > 0$ and all $f(x) \in f(X)$, $g(y) \in g(Y)$, we have $g(y') \in g(Y)$ and $f(x') \in f(X)$ such that $d_Z(f(x), g(y')) < r + \varepsilon$ and $d_Z(g(y), f(x')) < r + \varepsilon$. This equivalent to $d(x, y') < r + \varepsilon$ and $d(y, x') < r + \varepsilon$, which is enough to say that $d_H(X, Y) \leq r$.

Conversely, let d be a pseudometric on $X \sqcup Y$ that is admissible w.r.t. d_X and d_Y. Let (Z, d_Z) be the metric space associated with d, where $\pi : X \sqcup Y \to Z$ is the quotient map (see Remark 2.2). Note that admissibility is equivalent to the inclusions $i_X : X \to X \sqcup Y$ and $i_Y : Y \to X \sqcup Y$ being isometries. Also note that, since d acts as a true metric when restricted to X and Y, the restrictions of π to X and Y are isometries. Let $f = \pi \circ i_X$ and $g = \pi \circ i_Y$.

Let $r = d_H(X, Y)$ (with respect to d_Z). For all $\varepsilon > 0$ and $x \in X$, $y \in Y$, we have $y' \in Y$ and $x' \in X$ such that $d_Z(x, y') < r + \varepsilon$ and $d_Z(y, x') > r + \varepsilon$. This is precisely the same as saying $d(f(x), g(y')) < r + \varepsilon$ and $d(g(y), f(x')) < r + \varepsilon$, which in turn implies $d(f(X), g(Y)) \leq r$. ■

Corollary 3.4. *Let X and Y be two bounded metric spaces. We have:*

$$d_{GH}(X, Y) \leq 1/2(diam(X) + diam(Y)).$$

Proof. Let $K = 1/2(\mathrm{diam}(X) + \mathrm{diam}(Y))$. Define a pseudometric d on $X \sqcup Y$ by setting $d(x, y) = K$ for all $x \in X$ and $y \in Y$ and letting d agree d_X and d_Y in their domains. It is clear that d will be

[3] Strictly, the definition of the Hausdorff distance given previously was only meant for subsets of true metric spaces, but we extend it verbatim to pseudometrics.

symmetric and that $d(x,x) = d(y,y) = 0$ for all $x \in X$ and $y \in Y$. To show the triangle inequality, let $x, x' \in X$ and $y \in Y$ and let us show that triangle inequality is satisfied:

$$d(x,y) = K \leq d(x,x') + K = d(x,x') + d(x',y),$$

$$d(x,x') \leq 2K = d(x,y) + d(y,x').$$

Thus, d is a pseudometric on the disjoint union. Of course, $X \subset \overline{B}_K(Y)$ and $Y \subset \overline{B}_K(X)$, which implies the needed inequality. ■

A third equivalent definition will prove itself useful in the same way r-neighbourhoods were useful in studying the Hausdorff distance: by making it easy to establish a triangle inequality.

Definition 3.5. A correspondence between two sets X and Y is a relation $R \subset X \times Y$ such that

- $\forall x \in X$, there exists $y \in Y$ such that $(x,y) \in R$;
- $\forall y \in Y$, there exists $x \in X$ such that $(x,y) \in R$.

Example 3.6. Let $f : X \to Y$ be a surjective function. Then its graph $\{(x, f(x)) \ : \ x \in X\}$ is a correspondence between X and Y.

Definition 3.7. Let $R \subset X \times Y$ be any relation between two metric spaces X and Y. The distortion of R is defined as.

$$\mathfrak{D}(R) = \sup_{(x,y),(x',y') \in R} \{|d_X(x,x') - d_Y(y,y')|\}$$

If $f : X \to Y$ is a function between metric spaces, we define $\mathfrak{D}(f)$ as the distortion of its graph.

We are measuring the difference between the distance of pairs elements in X and the distance between correspondent pairs of elements in Y. If $f : X \to Y$ is a function, the distortion serves a way to measure how far it is from being an isometry:

Remark 3.8. Let $f : X \to Y$ be a surjective function. Then $\mathfrak{D}(f) = 0$ if and only if f is an isometry. Indeed, $\sup_{x,x' \in X}\{|d(x,x') - d(f(x), f(x'))|\} = 0$ is equivalent to $d(x,x') = d(f(x), f(x'))$ for all $x, x' \in X$.

Proposition 3.9. *For any two metric spaces X, Y we have*

$$d_{GH}(X,Y) = 1/2 \inf\{\mathfrak{D}(R)\}$$

where the infimum is taken from all the correspondences between X and Y

Proof. We begin by assuming that $d_{GH}(X,Y) \leq r$. There is a metric space Z and isometries $f : X \to Z$, $g : Y \to Z$ such that $d_H(f(X), g(Y)) < r$. Define $R \subset X \times Y$ as

$$R = \{(x,y) \in X \times Y \ : \ d(f(x), g(y)) \leq r\}.$$

Note that $d(f(X), \overline{B}(g(Y))) \leq r$ is equivalent to $\forall x \in X, \ \exists y \in Y : (x,y) \in R$ and similarly for $d(g(Y), \overline{B}(f(X))) \leq r$, which means R is indeed a correspondence. Let $(x,y), (x',y') \in R$. We have an upper bound on the distortion of R:

$$d(x,x') - d(y,y') \leq d(x,y) + d(y,y') + d(y',x') - d(y,y') = d(x,y) + d(x',y') < 2r;$$

$$d(y,y') - d(x,x') \leq d(y,x) + d(x,x') + d(x',y') - d(x,x') = d(y,x) + d(y',x') < 2r.$$

To show the other inequality, we assume $\mathfrak{D}(R) \leq 2r$ and let us construct an appropriate pseudometric d on $X \sqcup Y$. For pairs of points within X or Y, we simply define d to agree with d_X and d_Y. If $x \in X$ and $y \in Y$ we set

$$d(x,y) = \inf\{d_X(x,x') + r + d_Y(y,y') \ : \ (x',y') \in R\}$$

All axioms except the triangle inequality are clearly satisfied. we must then establish the inequality. Without loss of generality, it suffices to show $d(x,y) \leq d(x,x') + d(x',y)$ and $d(x,x') \leq d(x,y) + d(y,x')$ for all $x, x' \in X$ and $y \in Y$.

Let $\varepsilon > 0$ and choose $(x'', y'') \in R$ such that $d(x', y) + \varepsilon > d_X(x', x'') + r + d_Y(y, y'')$. Then:

$$\begin{aligned} d(x, y) &\leq d_X(x, x'') + r + d_Y(y, y'') \\ &\leq d_X(x, x') + d_X(x', x'') + r + d_Y(y, y'') \\ &< d_X(x, x') + d(x', y) + \varepsilon. \end{aligned}$$

Hence $d(x, y) \leq d_X(x, x') + d(x', y)$. For the other inequality, let $(x'', y''), (x''', y''') \in R$ such that

$$\begin{cases} d(x, y) + \varepsilon > d_X(x, x'') + r + d_Y(y'', y) \\ d(x', y) + \varepsilon > d_X(x', x''') + r + d_Y(y''', y) \end{cases}$$

Since $\mathfrak{D}(R) \leq 2r$ we have $d_X(x'', x''') - d_Y(y'', y''') \leq 2r$. Therefore:

$$\begin{aligned} d_X(x, x') &\leq d_X(x, x'') + d_X(x'', x''') + d_X(x''', x') \\ &\leq d_X(x, x'') + d_Y(y'', y''') + 2r + d_X(x''', x') \\ &\leq d_X(x, x'') + r + d_Y(y'', y) + d_X(x''', x') + r + d_Y(y, y''') \\ &< d(x, y) + d(x', y) + 2\varepsilon. \end{aligned}$$

Thus, $d_X(x, x') \leq d(x, y) + d(x', y)$. After having proved that d is a pseudometric on $X \sqcup Y$, it is easy to find an upper bound $d_H(X, Y)$: if $(x, y) \in R$, we have $d(x, y) \leq d_X(x, x) + r + d_Y(y, y) = r$. Thus R being a correspondence is again equivalent to $X \subset \overline{B}_r(Y)$ and $Y \subset \overline{B}_r(X)$, which in turn implies $d_H(X, Y) \leq r$. ∎

Corollary 3.10. *Let X and Y be two bounded metric spaces. We have*

$$d_{GH}(X, Y) \geq 1/2 |diam(X) - diam(Y)|.$$

Proof. For some $\varepsilon > 0$, let $x, x' \in X$ such that $d_X(x, x') > \operatorname{diam}(X) - \varepsilon$. Then, for every correspondence R between X and Y and all $y, y' \in Y$ such that $(x, y), (x', y') \in R$,

$$\mathfrak{D}(R) \geq d_X(x, x') - d_Y(y, y') > \operatorname{diam}(X) - \operatorname{diam}(Y) - \varepsilon.$$

Therefore, $\mathfrak{D}(R) \geq \operatorname{diam}(X) - \operatorname{diam}(Y)$ and, by symmetry, $\mathfrak{D}(R) \geq \operatorname{diam}(Y) - \operatorname{diam}(X)$ which, due to Proposition 3.9, implies the needed inequality. ∎

Example 3.11. It follows from Corollaries 3.10 and 3.4 that if X is bounded and Y contains a single point, then $d_{GH}(X, Y) = 1/2 \operatorname{diam}(X)$. This one of the few cases where we have an exact number for the Gromov-Hausdorff distance of two spaces.

Definition 3.12. Let X, Y and Z be sets and let $R \subset X \times Y$, $S \subset Y \times Z$ be relations. We define the composition of R and S as

$$S \circ R = \{(x, z) \in X \times Z \ : \ \exists y \in Y \text{ such that } (x, y) \in R \text{ and } (y, z) \in S\}$$

The composition operation satisfies a sort of triangle inequality that will translate directly into an analogous inequality for the Gromov-Hausdorff distance:

Lemma 3.13. *The composition $S \circ R$ of two correspondences $R \subset X \times Y$ and $S \subset Y \times Z$ is a correspondence. Furthermore, $\mathfrak{D}(S \circ R) \leq \mathfrak{D}(R) + \mathfrak{D}(S)$*

Proof. Let $x \in X$. There is some $y \in Y$ such that $(x, y) \in R$. But there is $z \in Z$ such that $(y, z) \in S$. Then $(x, z) \in S \circ R$. Symmetrically, the other requirement for $S \circ R$ to be a correspondence also holds.

As for the distortion, let $\varepsilon > 0$ and let $(x, z), (x', z') \in S \circ R$ such that $\mathfrak{D}(S \circ R) < |d_X(x, x') - d_Y(z, z')| + \varepsilon$. We have, for all $y, y' \in Y$:

$$\begin{aligned} |d_X(x, x') - d_Z(z, z')| &= |d_X(x, x') + d_Y(y, y') - d_Y(y, y') + d_Z(z, z')| \\ &\leq |d_X(x, x') + d_Y(y, y')| + |d_Y(y, y') - d_Z(z, z')| \\ &\leq \mathfrak{D}(R) + \mathfrak{D}(S). \end{aligned}$$

Since ε is arbitrary, $\mathfrak{D}(S \circ R) \leq \mathfrak{D}(R) + \mathfrak{D}(S)$. ∎

Proposition 3.14. *Let X, Y, Z be any three metric spaces. We have*

$$d_{GH}(X,Z) \leq d_{GH}(X,Y) + d_{GH}(Y,Z)$$

Proof. We use Proposition 3.9 and Lemma 3.13:

$$\begin{aligned} 2d_{GH}(X,Z) &= \inf\{\mathfrak{D}(T)\} \\ &\leq \inf\{\mathfrak{D}(S \circ R)\} \\ &\leq \inf\{\mathfrak{D}(R) + \mathfrak{D}(S)\} \\ &\leq \inf\{\mathfrak{D}(R)\} + \inf\{\mathfrak{D}(S)\} \\ &\leq 2d_{GH}(X,Y) + 2d_{GH}(Y,Z) \end{aligned}$$

■

This shows that the Gromov–Hausdorff distance is a pseudometric on the class of all metric spaces. Issues may again arise in the case of non-compact spaces, and we shall show that compact spaces will behave quite nicely. Again we resort to an equivalent definition

Definition 3.15. A function $f : X \to Y$ between metric spaces X and Y is called an ε-isometry if

- $\mathfrak{D}(f) \leq \varepsilon$;
- $f(X)$ is a ε-net of Y.

Remark 3.16. A 0-isometry is the same as a bijective isometry.

Lemma 3.17. *For any two metric spaces X, Y we have*

- *If there exists an ε-isometry between X and Y, then there exists a correspondence R between X and Y with $\mathfrak{D}(R) \leq 3\varepsilon$*
- *If there exists a correspondence R between X and Y with $\mathfrak{D}(R) \leq \varepsilon$, then there exists an ε-isometry between X and Y.*

Proof. For the first task, define $R \subset X \times Y$ as

$$R = \{(x,y) \in X \times Y \ : \ d_Y(y, f(x)) \leq \varepsilon\}$$

This is a correspondence: For all $x \in X$, we have $(x, f(x)) \in R$ and for all $y \in Y$ we have some $f(x) \in \overline{B}_\varepsilon(y)$, since $f(X)$ is a ε-net. Let us estimate its distortion by taking arbitrary $(x,y),(x',y') \in R$:

$$\begin{aligned} d_X(x,x') - d_Y(y,y') &\leq d_X(x,x') - d_Y(y,f(x)) - d_Y(f(x),f(x')) - d_Y(f(x'),y') \\ &\leq |d_X(x,x') - d_Y(y,f(x)) - d_Y(f(x),f(x')) - d_Y(f(x'),y')| \\ &\leq |d_X(x,x') - d_Y(f(x),f(x'))| + d_Y(y,f(x)) + d_Y(f(x'),y'). \end{aligned}$$

$$\begin{aligned} d_Y(y,y') - d_X(x,x') &\leq d_Y(y,f(x)) + d_Y(f(x),f(x')) + d_Y(f(x'),y') - d_X(x,x') \\ &\leq |d_Y(y,f(x)) + d_Y(f(x),f(x')) + d_Y(f(x'),y') - d_X(x,x')| \\ &\leq |d_X(x,x') - d_Y(f(x),f(x'))| + d_Y(y,f(x)) + d_Y(f(x'),y'). \end{aligned}$$

That is,

$$|d_X(x,x') - d_Y(y,y')| \leq |d_X(x,x') - d_Y(f(x),f(x'))| + d_Y(y,f(x)) + d_Y(f(x'),y') \qquad \leq 3\varepsilon$$

As for the existence of a ε-isometry from the assumption $\mathfrak{D}(R) \leq \varepsilon$, define $f : X \to Y$ by taking $x \in X$ to some $y \in Y$ such that $(x,y) \in R$. Since the graph of f will be contained in R, we have $\mathfrak{D}(f) \leq \mathfrak{D}(R)$. To see that $f(X)$ is a ε-net of Y, let $y \in Y$. Choose $x \in X$ with $(x,y) \in R$. Since (x,y) and $(x,f(x)) \in R$, we have

$$|d_X(x,x) - d_Y(y,f(x)| \leq \varepsilon$$

$$\implies d_Y(y,f(x)) \leq \varepsilon$$

■

Note that, in particular, we obtain the following characterization:

Corollary 3.18. *Let X, Y be two metric spaces. Then the following are equivalent:*

- $d_{GH}(X,Y) = 0$.
- *For all $\varepsilon > 0$, there exists a ε-isometry between X and Y.*

A much stronger characterization is possible for compact spaces:

Proposition 3.19. *Let X, Y be two compact metric spaces. Then $d_{GH}(X,Y) = 0$ if and only of X and Y are isometric.*

Proof. If X and Y are isometric, then we have a 0-isometry between them. Thus Corollary 3.18 implies $d_{GH}(X,Y) = 0$.

If $d_{GH}(X,Y) = 0$, Corollary 3.18 implies that there is a sequence $\{f_n\}$ of functions between X and Y such that each $f_n : X \to Y$ is a $1/n$-isometry. All compact metric spaces are separable, so let $S = \{s_1, s_2, \ldots\} \subset X$ be countable and dense.

Let us find a sequence of indices $\{k_n\}_{n=1}^{\infty}$ such that $f_{k_n}(s)$ converges as a sequence of points in Y for all $s \in S$: there is a subsequence $\{f_n^1\}$ of $\{f_n\}$ such that $\{f_n^1(s_1)\}$ converges, due to the compactness of Y. Inductively, after having obtained a nested sequence of subsequences $\{f_n^1\} \supset \{f_n^2\} \supset \ldots \supset \{f_n^i\}$ such that $f_n^j(s_j)$ converges for all $j \in \{1, \ldots, j\}$, we find a subsequence f_n^{i+1} of f_n^i such that $f_n^{i+1}(s_{i+1})$ converges. Thus, by taking the diagonal subsequence $\{g_n\}_{n=1}^{\infty}$ defined as $g_n = f_n^n$, we guarantee pointwise convergence in S.

Let us show that $f : S \to Y$, defined as $f(s) = \lim_{n\to\infty} g_n(s)$, is an isometry. Let $s, s' \in S$. For all $\varepsilon > 0$, we have $d_Y(g_n(s), f(s)) < \varepsilon$ and $d_Y(g_n(s'), f(s')) < \varepsilon$ for all but finitely many $n \in \{1, 2, \ldots\}$.

$$\begin{aligned} d_Y(f(s), f(s')) &\le d_Y(f(s), g_n(s)) + d_Y(g_n(s), g_n(s')) + d_Y(g_n(s'), f(s')) \\ &< d_Y(g_n(s), g_n(s')) + 2\varepsilon \end{aligned}$$

$$\begin{aligned} d_Y(g_n(s), g_n(s')) &\le d_Y(g_n(s), f(s)) + d_Y(f(s), f(s')) + d_Y(f(s'), g_n(s')) \\ &< d_Y(f(s), f(s')) + 2\varepsilon \end{aligned}$$

The distortion of g_n is bounded above by $1/n$. That is:

$$d_Y(g_n(s), g_n(s')) \le d_X(s, s') + 1/n$$

$$d_X(s, s') \le d_Y(g_n(s), g_n(s')) + 1/n$$

Therefore:

$$\begin{aligned} d_Y(f(s), f(s')) &< d_X(s, s') + 1/n + 2\varepsilon \\ d_X(s, s') &< d_Y(f(s), f(s')) + 1/n + 2\varepsilon \\ \implies |d_Y(f(s), f(s')) &- d_X(s, s')| < 1/n + 2\varepsilon \end{aligned}$$

Where above n can be all but finitely many positive integers. Since s, s', ε are all arbitrary, we have that f is an isometry. By Corollary 2.14 we can extend f into an isometry on X. Symmetrically, there is a isometry between Y and X. By Proposition 2.17, X and Y must be isometric. ∎

Propositions 3.14 and 3.19 give us something remarkable: An analogue of Theorem 2.21 for the space of isometry classes of compact metric spaces, which we call the *Gromov space*.

Theorem 3.20. *The class of all isometry classes of compact metric spaces forms a metric space with respect to the Gromov-Hausdorff metric.*

3.1 Convergence of compact metric spaces

We wish to obtain a better understanding of the topology of the Gromov space. Immediately, we have the following:

Remark 3.21. The set of all finite metric spaces is dense in the Gromov space. Indeed, for any given compact space X and all $\varepsilon > 0$, we have a finite ε-net S. Clearly, if $i : S \to X$ is the inclusion and $\mathrm{id} : X \to X$ is the identity, we have $d_H(i(S), \mathrm{id}(X)) \leq \varepsilon$.

This simple observation motivates us to find a characterization of convergence within this space in terms of convergence of finite metric spaces. The following will suffice:

Proposition 3.22. *A sequence of compact metric spaces $\{X_n\}_{n=1}^{\infty}$ converges to a space X if, and only if, for every $\varepsilon > 0$ there are ε-nets S_n^ε of each X_n and a ε-net S^ε of X such that $\{S_n^\varepsilon\}_{n=1}^{\infty}$ converges to S^ε.*

Proof. Since every space is a ε-net of itself the first implication is immediate.

Assuming $\{S_n\}_{n=1}^{\infty}$ is a sequence of ε-nets of each X_n that converges to S, we may also assume that there are a sequence of correspondences $R_n \subset S_n \times S$ such that $\mathfrak{D}(R_n)$ becomes arbitrarily small as $n \to \infty$. Let us extend each R_n into a correspondence $\tilde{R}_n$ between X_n and X. Define $c_n : X_n \to S_n$ such that $d(c_n(x_n), x_n) \leq \varepsilon$ for all $x_n \in X$ (which is possible, since S_n is a ε-net). Similarly, let $c : X \to S$ with $d(c(x), x) \leq \varepsilon$ for all $x \in X$.

For all $x_n \in X_n$, there is some $f_n(x_n) \in S$ such that $(c_n(x_n), f_n(x_n)) \in R_n$, since this is a correspondence. Similarly, for all $x \in X$, there is some $f(x) \in S_n$ such that $(f(x), c(x)) \in R_n$. Define the functions $f_n : X_n \to S$ and $f : X \to S_n$ as such. Now, we create a correspondence from the union of the graphs $\tilde{R}_n = \{(x_n, f_n(x_n)) : x_n \in X_n\} \cup \{(f(x), x) : x \in X\}$. This correspondence is close to R_n in terms of distortion, as we can check the three possible cases:

- $x_n, x_n' \in X_n$:

$$\begin{aligned}
&|d(x_n, x_n') - d(f_n(x_n), f_n(x_n'))| \\
\leq\ &|d(x_n, x_n') - d(c_n(x_n), c_n(x_n'))| + |d(c_n(x_n), c_n(x_n')) - d(f_n(x_n), f_n(x_n'))| \\
\leq\ &2\varepsilon + \mathfrak{D}(R_n).
\end{aligned}$$

- $x, x' \in X$:

$$\begin{aligned}
&|d(f(x), f(x')) - d(x, x')| \\
\leq\ &|d(f(x), f(x')) - d(c(x), c(x'))| + |d(c(x), c(x')) - d(x, x')| \\
\leq\ &\mathfrak{D}(R_n) + 2\varepsilon.
\end{aligned}$$

- $x_n \in X_n$ and $x \in X$

$$\begin{aligned}
&|d(x_n, f(x)) - d(f_n(x_n), x)| \\
\leq\ &|d(x_n, f(x)) - d(c_n(x_n), f(x))| + |d(c_n(x_n), f(x)) - d(f_n(x_n), c(x))| \\
&+ |d(f_n(x_n), c(x)) - d(f_n(x_n), x)| \\
\leq\ &\varepsilon + \mathfrak{D}(R_n) + \varepsilon.
\end{aligned}$$

Hence $\mathfrak{D}(\tilde{R}_n)$ also becomes arbitrarily small as $n \to \infty$, which implies $\lim_{n\to\infty}(X_n) = X$, due to Proposition 3.9. ■

Definition 3.23. A subset $S \subset X$ of a metric space X is *precompact* if its closure is compact.

We wish to establish a condition that ensures that a family of compact metric spaces is precompact with respect to the Gromov-Hausdorff distance.

Definition 3.24. A family $\{X\}_{\lambda \in L}$ of compact metric spaces is *uniformly totally bounded* if

- every space in the family is bounded by the same constant: There exists $K > 0$ such that $\mathrm{diam}(X_\lambda) < K$ for all $\lambda \in L$
- for every $\varepsilon > 0$, there is some $N(\varepsilon) \in \{1, 2, \ldots\}$ such that all elements of the family admit a ε-net with at most $N(\varepsilon)$ points.

This is the precompactness condition we hoped for:

Proposition 3.25. *Let $\{X_n\}_{n=1}^{\infty}$ be a uniformly totally bounded sequence of compact metric spaces. Then it admits a convergent subsequence that converges to some compact metric space X.*

Proof. We must construct a compact space. We may assume that for every $k \in \mathbb{N}$, each X_n admits a $1/k$-net with exactly $N(1/k)$ elements, allowing repetition in case of finite spaces. Let $S_n^1 = \{x_n^1, \cdots, x_n^{N(1)}\}$ be a 1-net of X_n, $S_n^2 = \{x_n^{N(1)+1}, \cdots, x_n^{N(1)+N(1/2)}\}$ be a $1/2$-net, and so on. In general, we have a family $\{x_n^i\}_{i=1}^{\infty}$, containing the $N(1/k)$ points of some $1/k$-net of X_n, for every $k \in \mathbb{N}$ (note that this is a dense set on X_n).

Consider the sequence $\{d(x_n^i, x_n^j)\}_{n=1}^{\infty}$ for fixed i, j. It takes values on the compact set $[0, K]$, where K is the constant in Definition 3.24. Thus there is a convergent subsequence. Since the set $\{(i,j) \in \mathbb{N}^2\}$ is countable we may construct a sequence of nested subsequences, each allowing the limit of $d(x_n^i, x_n^j)$ to exist for some (i,j). By taking the diagonal subsequence, we may assume that $\{d(x_n^i, x_n^j)\}_{n=1}^{\infty}$ converges for all (i,j) (here we are discarding a subset of $\{X_n\}$. Only the remaining subsequence will be considered from now on).

Thus we may define X to be the set of all sequences $x_i = \{x_n^i\}_{n=1}^{\infty}$. The following distance is well defined:

$$d(x_i, x_j) = \lim_{n\to\infty} d(x_n^i, x_n^j).$$

Since the triangle inequality is valid for each n, it will also be satisfied by the limit. Symmetry and non-negativity are trivial. The problem of positivity can be solved by replacing X with the quotient construction described in Remark 2.2.

We will show that X is totally bounded: given $i \in \mathbb{N}$, each x_n^i is an element of X_n. Since there is a $1/k$-net of X_n contained in $\{x_n^i\}_{i=1}^{\infty}$, select $x_n^{j_n}$ such that $d(x_n^i, x_n^{j_n}) < 1/k$ for each n. Since the $1/k$-nets of each X_n contains exactly $N(1/k)$ elements, the sequence $\{j_n\}_{n=1}^{\infty}$ must have a constant subsequence. Let j be that constant. Then the sequence $d(x_n^i, x_n^j)$ must have a subsequence bounded by $1/k$. Therefore, the limit $d(x_i, x_j)$ must also be bounded by $1/k$. Note that we have just shown that $S_1 = \{x_1, \cdots, x_{N(1)}\}$ is a 1-net of X, as $S_2 = \{x_{N(1)+1}, \cdots, x_{N(1)+N(1/2)}\}$ is a $1/2$-net, and so on.

Since X is totally bounded its completion must also be totally bounded and thus compact. Therefore, by Proposition 3.22, it only remains to show that $S_n^k \to S_k$ for each k. We use the correspondence definition of d_{GH} from Proposition 3.9 and note that the distortion of $R_n^k = \{(x_n^i, x_i) : i \in \mathbb{N}\}$ can be estimated using the limit that defined the metric of X:

$$\lim_{n\to\infty} \sup_{i,j\in\mathbb{N}} |d(x_n^i, x_n^j) - d(x_i, x_j)| = 0 \iff [\lim_{n\to\infty} d(x_n^i, d_n^j) = d(x_i, x_j)\ \forall i, j \in \mathbb{N}].$$

∎

3.2 The pointed Gromov-Hausdorff distance

In [Gro81], Gromov employs a variation of the Gromov-Hausdorff distance that has nice convergence properties even for unbounded spaces. It would be nice, for instance, if there was a sense in which the sequence of open intervals intervals $I_n = (-n, n)$, $n \in \mathbb{N}$ converged to the real line, even though the Gromov-Hausdorff distance between I_n and $\mathbb{R}$ is always infinite. Note, though, that for arbitrary but fixed $r > 0$, the sequence $I_n \cap (-r, r)$ does converge to $\mathbb{R} \cap (-r, r)$, which might be used as a justification to expect $I_n \to \mathbb{R}$.

The idea of spaces converging for every bounded neighbourhood of the origin can be formalized as the modified Gromov-Hausdorff convergence of pointed spaces, in which general metric spaces receive a sort of origin.

Definition 3.26. A *pointed metric space* is a metric space X with a distinguished point $x \in X$, which we denote (X, x) (a slightly confusing notation, since we already use the ordered pair (X, d) to refer to the set X endowed with the metric d).

We can define a notion of convergence of pointed spaces that is less restrictive than how it was previously defined, being well behaved even for some unbounded spaces.

Definition 3.27. Given two pointed metric spaces (X, x) and (Y, y), we define their pointed Gromov-Hausdorff distance $d_{pGH}((X, x), (Y, y))$ as the infimum of all $r \geq 0$ such that there exist a pseudometric d on the disjoint union $X \sqcup Y$ with the property that

$$d(x, y) \leq r \text{ and } \overline{B}_{1/r}(x) \subset \overline{B}_r(Y) \text{ and } \overline{B}_{1/r}(y) \subset \overline{B}_r(X).$$

Indeed, this definition satisfies our intuition:

Example 3.28. The sequence of intervals $I_n = (-n, n)$ (seen as the pointed spaces $(I_n, 0)$) converges to $(\mathbb{R}, 0)$ with respect to d_{GH}. We can show that $d_{GH}((I_n, 0), (\mathbb{R}, 0)) < 2/n$ by first embedding I_n and $\mathbb{R}$ disjointly into $\mathbb{R}^2$ as

$$Z = \{(x, 0) : x \in \mathbb{R}\} \cup \{(x, 1/n) : x \in I_n\}.$$

Then $d((0, 0), (0, 1/n) = 1/n < 2/n$. Also (assuming $n > 1$),

$$\begin{aligned} B_{n/2}((0,0)) &= [(-n/2, n/2) \times (-n/2, n/2)] \cap Z \\ &= \{(x, 0) : x \in (-n/2, n/2)\} \cup \{(x, 1/n) : x \in (-n/2, n/2)\} \\ &\subset B_{2/n}(\{(x, 1/n) : x \in I_n\}) \cap Z \\ &= \{(x, 0) : x \in (-n, n)\} \cup \{(x, 1/n) : x \in I_n\}; \end{aligned}$$

$$\begin{aligned} B_{n/2}((0, 1/n)) &= [(-n/2, n/2) \times (1/n - n/2, 1/n + n/2)] \cap Z \\ &= \{(x, 0) : x \in (-n/2, n/2)\} \cup \{(x, 1/n) : x \in (-n/2, n/2)\} \\ &\subset B_{2/n}(\{(x, 0) : x \in \mathbb{R}\}) \\ &= Z. \end{aligned}$$

3.3 Convergence of functions

We will use the following alternative definition of pointed convergence:

Definition 3.29. Let $\{(X_n, x_n)\}$ be a sequence of pointed metric spaces and (Y, y) be a pointed metric space. Let $\{d_n\}_{n=1}^{\infty}$ be a sequence of pseudometrics on $X_n \sqcup Y$. We say that $\{(X_n, x_n)\}$ converges to (Y, y) with respect to $\{d_n\}$ if for every $\varepsilon > 0$ and $r \geq 0$ there is some $n_0 \in \mathbb{N}$ such that for every $n > n_0$ we have

$$d_n(x_n, y) \leq \varepsilon,$$

$$\overline{B}_r(x_n) \subset \overline{B}_\varepsilon(Y) \text{ and } \overline{B}_r(y) \subset \overline{B}_\varepsilon(X_n).$$

where the closed balls above are constructed with respect to d_n. We call this *definite* convergence w.r.t. $\{d_n\}$.

Proposition 3.30. *The following are equivalent:*

- *The sequence $\{(X_n, x_n)\}$ converges to (Y, y) with respect to some $\{d_n\}$*
- *The sequence $d_{pGH}(X_n, Y)$ converges to zero.*

Proof. Let $\rho > 0$. If (X_n, x_n) converges definitely to (Y, y) w.r.t. d_n, then there is some $n_0 \in \{1, 2, \ldots\}$ such that, for all $n > n_0$, we have

$$d_n(x_n, y) \leq \rho,$$

$$\overline{B}_{1/\rho}(x_n) \subset \overline{B}_\rho(Y) \text{ and } \overline{B}_{1/\rho}(y) \subset \overline{B}_\rho(X_n).$$

which implies that $d_{pGH}((X_n, x_n), (Y, y)) \leq \rho$ for all $n > n_0$. Since ρ is arbitrary, we have shown that $\lim_{n\to\infty} d_{pGH}((X_n, x_n), (Y, y)) = 0$.

Conversely, assume $\lim_{n\to\infty} d_{pGH}((X_n, x_n), (Y, y)) = 0$. Let $\varepsilon > 0$ and $r \geq 0$. If we set $\rho < \min\{\varepsilon, 1/r\}$, there exists $n_0 \in \{1, 2, \ldots\}$ such that, for all $n > n_0$,

$$d_n(x_n, y) \leq \rho$$

$$\overline{B}_{1/\rho}(x_n) \subset \overline{B}_\rho(Y) \text{ and } \overline{B}_{1/\rho}(y) \subset \overline{B}_\rho(X_n)$$

Which implies

$$d_n(x_n, y) \leq \varepsilon$$
$$\overline{B}_r(x_n) \subset \overline{B}_\varepsilon(Y) \text{ and } \overline{B}_r(y) \subset \overline{B}_\varepsilon(X_n)$$

Which in turn implies $\{d_n\}$ induces definite convergence of (X_n, x_n) to (Y, y). ∎

Corollary 3.31. *Let $\{(X_n, x_n)\}_{n=1}^\infty$ be a convergent sequence of proper pointed spaces. Then, for every $r > 0$, the sequence of compact metric spaces $\{\overline{B}_r(x_n)\}$ converges.*

Proof. There exists a pointed metric space (Y, y) and a sequence of metrics $\{d_n\}_{n=1}^\infty$, each on one of $X_n \sqcup Y$, such that for all $r \geq 0$ and $\varepsilon > 0$ we have, for all large enough n:

$$d_n(x_n, y) < \varepsilon;$$
$$\overline{B}_r(x_n) \subset \overline{B}_\varepsilon(Y);$$
$$\overline{B}_r(y) \subset \overline{B}_\varepsilon(X).$$

Let $x \in \overline{B}_r(x_n)$ and $y' \in Y$ such that $d(x, y') \leq \varepsilon$. We have

$$d(y', y) \leq d(y', x) + d(x, x_n) + d(x_n, y) \leq r + 2\varepsilon.$$

Similarly, given $y' \in \overline{B}_r(y)$ and $x \in X$ with $d(x, y') < \varepsilon$:

$$d(x, x_n) \leq d(x, y') + d(y', y) + d(y, x_n) \leq r + 2\varepsilon.$$

Thus $\overline{B}_r(x_n) \subset \overline{B}_{2\varepsilon}(\overline{B}_r(y))$ and $\overline{B}_r(y) \subset \overline{B}_{2\varepsilon}(B_r(x_n))$, which gives an upper bound of 2ε on the Gromov distance of between $\overline{B}_r(x_n)$ and $\overline{B}_r(y)$. ∎

The reason why we differentiate definite convergence w.r.t. a fixed sequence of metrics and usual convergence w.r.t pGH is to be able to measure distance between points in X_n and points in Y in exactly one way. For instance, with definite convergence it makes sense to say that a sequence of points $p_n \in X_n$ converges to some $p \in Y$, while for usual convergence there could be uncountably many different metrics on each $X_n \sqcup Y$. The following definition also only makes sense for definite convergence.

Definition 3.32. Let $\{(X_n, x_n)\}$ and $\{(X'_n, x'_n)\}$ be two sequences of pointed metric spaces that converge to (Y, y) and (Y', y'), respectively, w.r.t the family of metrics $\{d_n\}_{n=1}^\infty$ and $\{d'_n\}_{n=1}^\infty$.

We say that a sequence of functions $f_n : X_n \to X'_n$ converges to a function $f : Y \to Y'$ when for all $\varepsilon > 0$ and $r \geq 0$ there is some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that, for all $n > n_0$, we have

$$p \in B_r(x_n), \text{ and } q \in B_r(y) \text{ and } d_n(p, q) < \delta$$
$$\implies d'_n(f_n(p), f(q)) < \varepsilon.$$

A good definition of convergence is of course one that allows nice properties shared by each element of the sequence to be preserved by the limit. The following properties will perhaps justify Definition 3.32 as a good one:

Proposition 3.33. *Let f_n and f be as in Definition 3.32. If each f_n is an isometry, then f is an isometry.*

Proof. Let $a, b \in Y$, $\varepsilon > 0$ and δ as in the definition above (we may assume $\delta \leq \varepsilon$). Take $a' \in X_n$ and $b' \in X_n$ (with n large enough) such that $d_n(a, a') < \delta$ and $d_n(b, b') < \delta$. Then $d_n(f_n(a'), f(a)) < \varepsilon$ and $d_n(f_n(b'), f(b)) < \varepsilon$. Thus

$$\begin{aligned} d_n(f(a), f(b)) &\leq d_n(f(a), f_n(a')) + d_n(f_n(a'), f_n(b')) + d_n(f_n(b'), f(b)) \\ &< d_n(f_n(a'), f_n(b')) + 2\varepsilon \\ &= d_n(a', b') + 2\varepsilon \\ &< d_n(a', a) + d_n(a, b) + d_n(b, b') + 2\varepsilon \\ &< d_n(a, b) + 4\varepsilon. \end{aligned}$$

Similarly:

$$\begin{aligned} d_n(a,b) &\le d_n(a,a') + d_n(a',b') + d_n(b',b) \\ &< d_n(a',b') + 2\varepsilon \\ &= d_n(f_n(a'), f_n(b')) + 2\varepsilon \\ &\le d_n(f_n(a'), f(a)) + d_n(f(a), f(b)) + d_n(f(b), f_n(b')) + 2\varepsilon \\ &< d_n(f(a), f(b)) + 4\varepsilon. \end{aligned}$$

Thus $|d_n(f(a), f(b)) - d_n(a,b)| < \varepsilon$ for arbitrary ε. ∎

For the sake of simplicity from now on every convergence of pointed metric spaces is assumed to be definite and the fixed underlying metrics shall remain implicit.

Lemma 3.34. *All proper spaces are complete.*

Proof. Take a Cauchy sequence in a metric space X. Since Cauchy sequences are bounded every term in a Cauchy sequence is contained within a closed and bounded set $K \subset X$. Since K is compact, the sequence must converge. ∎

We now prove a beautiful analogue of the precompactness criterion for sequences of isometries.

Proposition 3.35 (Gromov's Isometry Lemma)**.** *Let $\{(X_n, x_n)\}_{n=1}^{\infty}$ and $\{(X'_n, x'_n)\}_{n=1}^{\infty}$ be two sequences of pointed metric spaces that converge to proper pointed spaces (Y, y) and (Y', y'), respectively. Let $f_n : X_n \to X'_n$ be a sequence of isometries such that $d(x'_n, f_n(x_n))$ is bounded above by a fixed constant C. Then there is a subsequence of $\{f_n\}$ that converges to an isometry $f : Y \to Y'$.*

Proof. Let $\varepsilon_n = \frac{1}{4^n}$ and $r_n = nC + 1$. Due to the hypothesis of convergence we may make the following assumptions by passing to subsequences:

$$d(x_n, y) < \varepsilon_n,\ d(x'_n, y') < \varepsilon_n,$$
$$B_{r_n}(y) \subset B_{\varepsilon_n}(X_n),\ B_{r_n}(y') \subset B_{\varepsilon_n}(X'_n),$$
$$B_{1/2 + r_n + C}(x'_n) \subset B_{\varepsilon_n}(Y').$$

Since Y and Y' are proper, each ball is totally bounded. Thus for each $m \in \mathbb{N}$ there is a sequence S_m of ε_m-nets of $B_{r_m}(y)$ and a sequence S'_m of ε_m-nets of $B_{r_m}(y')$ (it is convenient to assume that $S_m \subset S_{m+1}$ and $S'_m \subset S'_{m+1}$). We wish to initially define a family of functions $g_{m,n} : S_m \to S'_{m+1}$ for $n \ge m$.

Let $s \in S_m$. Since each $s \in B_{r_m}(y) \subset B_{r_n}(y) \subset B_{\varepsilon_n}(X_n)$ we can choose some $p \in X_n$ such that $d(p, s) < \varepsilon_n$. Note that

$$\begin{aligned} d(f_n(p), x'_n) &\le d(f_n(p), f_n(x_n)) + d(f_n(x_n), x'_n) \\ &= d(p, x_n) + d(f_n(x_n), x'_n) \\ &< d(p, x_n) + C \\ &\le d(p, s) + d(s, y) + d(y, x_n) + C \\ &< \varepsilon_n + r_m + \varepsilon_n + C \\ &\le 1/2 + r_m + C. \end{aligned}$$

Thus $f_n(p) \in B_{1/2 + r_m + C}(x'_n)$, and so is some $q \in Y'$ such that $d(f_n(p), q) < \varepsilon_m$. We can show that q is close enough to y' after applying the triangle inequality a few more times:

$$\begin{aligned} d(q, y') &\le d(q, f_n(p)) + d(f_n(p), x'_n) + d(x'_n, y') \\ &\le \varepsilon_n + 1/2 + r_m + C + \varepsilon_n \\ &\le 1 + r_m + C \\ &= 1 + (m+1)C \\ &= r_{m+1}. \end{aligned}$$

Therefore $q \in B_{r_{m+1}}(y')$. We can choose any $s' \in S'_{m+1}$ such that $d(q, s') < \varepsilon_{m+1}$. Define $g_{m,n}(s) = s'$.

Each S'_{m+1} is finite. Thus for some fixed $m \in \mathbb{N}$ and $s \in S_m$, there is a sequence of indices $\{n_k\}_{k=1}^{\infty}$ such that $g_{m,n_k}(s)$ is constant with respect to k. Since S_m is also finite there is subsequence $\{n_{k_{k'}}\}_{k'=1}^{\infty}$ and we have that $g_{m,n_{k'}}$ depends only on m. It remains to show that $\{f_{n_{k_{k'}}}\}$ is a convergent subsequence of $\{f_n\}$. From now on we pass to the subsequences in order to write f_n instead of the unwieldy $f_{n_{k_{k'}}}$. Also, we write g_m instead of $g_{m,n}$.

Suppose that we had started with some other $\tilde{s} \in B_{r_m}(y)$ and that we found some $\tilde{p} \in X_n$ with $d(\tilde{p}, \tilde{s}) < \varepsilon_n$. We would then have

$$\begin{aligned} d(p, \tilde{p}) &\leq d(p, s) + d(s, \tilde{s}) + d(\tilde{s}, \tilde{p}) \\ &< d(s, \tilde{s}) + 2\varepsilon_n. \end{aligned}$$

Also, if we had chosen some $\tilde{q} \in Y'$ such that $d(f(\tilde{p}), \tilde{q}) < \varepsilon_m$ we would have

$$\begin{aligned} d(q, \tilde{q}) &\leq d(q, f(p)) + d(f(p), f(\tilde{p})) + d(f(\tilde{p}), \tilde{q}) \\ &< d(f(p), f(\tilde{p})) + 2\varepsilon_m \\ &= d(p, \tilde{p}) + 2\varepsilon_m \\ &< d(s, \tilde{s}) + 2\varepsilon_m + 2\varepsilon_n. \end{aligned}$$

Finally, if we had chosen some $\tilde{s'} \in S'_{m+1}$ with $d(\tilde{q}, \tilde{s'}) < \varepsilon_{m+1}$ we would have

$$\begin{aligned} d(s', \tilde{s'}) &\leq d(s', q) + d(q, \tilde{q}) + d(\tilde{q}, \tilde{s'}) \\ &< d(q, \tilde{q}) + 2\varepsilon_{m+1} \\ &< d(s, \tilde{s}) + 2\varepsilon_m + 2\varepsilon_n + 2\varepsilon_{m+1} \\ &\leq d(s, \tilde{s}) + 6\varepsilon_{m+1}. \end{aligned}$$

Thus if we set $S = \bigcup_{m=1}^{\infty} S_m$ and $S' = \bigcup_{m=1}^{\infty} S'_m$ the function $g : S \to S'$, defined to coincide with each g_m on S_m, takes Cauchy sequences $\{s_m\}_{m=1}^{\infty}$ with each $s_m \in S_m$ to Cauchy sequences $\{s'_m\}_{m=1}^{\infty}$ with $s'_m \in S'_m$. Thus by Proposition 2.15 g is uniformly continuous. Since S is dense on a complete metric space, due to Proposition 3.34, we extend $g : S \to Y'$ to $g : Y \to Y'$ as in Proposition 2.13.

Now given $x \in B_{r_m}(x_n)$, $y \in B_{r_m}(y)$ we write $y = \lim_{m\to\infty} s_m$ and we apply continuity of the metric to obtain $d(f_n(x), g(y)) = \lim_{m\to\infty} d(f_n(x), g_m(s_m))$. We wish to obtain an estimate of $d(f_n(x), g(y))$ through this limit. Let us recall the notation from above, with $g_m(s) = s'$

$$\begin{aligned} d(f_n(x), s') &\leq d(f_n(x), f_n(p)) + d(f_n(p), q) + d(q, s') \\ &< d(f_n(x), f_n(p)) + \varepsilon_m + \varepsilon_{m+1} \\ &= d(x, p) + \varepsilon_m + \varepsilon_{m+1} \\ &\leq d(x, s) + d(s, p) + \varepsilon_m + \varepsilon_{m+1} \\ &< d(x, s) + \varepsilon_n + \varepsilon_m + \varepsilon_{m+1} \\ &< d(x, s) + 3\varepsilon_{m+1}. \end{aligned}$$

That is, $d(f_n(x), g_m(s_m)) < d(x, s_m) + 3\varepsilon_{m+1}$. Therefore $d(f_n(x), g(y)) \leq d(x, y)$, immediately satisfying the convergence requirement. By Proposition 3.33, f must be an isometry. ■

3.4 Consequences of the Isometry Lemma

Proposition 3.36. *Let $\{(X_n, x_n)\}_{n=1}^{\infty}$ be a sequence of proper pointed spaces such that for every $r \in \{0, 1, 2, \ldots\}$ the sequence of compact metric spaces $\{\overline{B}_r(x_n)\}$ converges to some Y_r. Then $\{(X_n, x_n)\}_{n=1}^{\infty}$ has a subsequence that converges to some proper pointed space (Y, y).*

Proof. We see each $\overline{B}_r(x_n)$ as a pointed space in the obvious way. To see each Y_r as a pointed space, we endow $\overline{B}_r(x_n) \sqcup Y_r$ with a sequence of metrics $\{d_n^r\}_{n=1}^\infty$ such that for all ε we have $\overline{B}_r(x_n) \subset \overline{B}_\varepsilon(Y_r)$ and $Y_r \subset \overline{B}_\varepsilon(\overline{B}_r(x_n))$ for all large enough n. That is, there is some point $y(\varepsilon) \in Y_r$ such that $d_n^r(x_n, y(\varepsilon)) < \varepsilon$. We pass to a subsequence and assume that $\{y(1/n)\}_{n=1}^\infty$ converges to some $y_r \in Y_r$, which is our distinguished point.

We will need to construct a limit space Y by joining each limit space Y_r. We begin by using Proposition 3.35 to find a subsequence of $\{(X_n, x_n)\}_{n=1}^\infty$ for which the inclusions $\overline{B}_0(x_n) \to \overline{B}_1(x_n)$ converge to an isometry $\varphi_0 : Y_0 \to Y_1$ with respect to d_n^0 and d_n^1. Then we inductively find nested subsequences of this subsequence such that each family of inclusions $\overline{B}_r(x_n) \to \overline{B}_{r+1}(x_n)$ converges to an isometry $\varphi_r : Y_r \to Y_{r+1}$ with respect to d_n^r and d_n^{r+1}. From the definition of convergence of functions, the fact that $d_n^r(x_n, y_r)$ becomes arbitrarily small as $n \to \infty$ implies that $d_n^{r+1}(x_n, \varphi_r(y_r))$ also approaches zero. Therefore, $\varphi_r(y_r) = y_{r+1}$.

Thus for each $r < s$ we can define $\varphi_{r,s} = \varphi_{s-1} \circ \ldots \circ \varphi_r$, which is an isometry of Y_r into Y_s that takes y_r to y_s. Proceed by setting

$$Y' = \bigsqcup_{r=0}^{\infty} Y_r.$$

Given $a \in Y_r$ and $b \in Y_s$ with $r < s$ define $d(a,b) = d(\varphi_{r,s}(a), b)$. This will be a pseudometric because for any $a, b, c \in Y'$ the values of $d(a,b)$, $d(b,c)$, $d(a,c)$ will agree with the values given by the metric of some Y_R, $R \gg 1$.

The positivity axiom, though, will be not be satisfied in general. If $b = \varphi_{r,s}(a)$, then $d(a,b) = 0$ even if $a \neq b$. Thus we can once again apply the construction that induces a metric on the quotient space $Y := Y'/\sim$ under the equivalence relation $a \sim b \iff d(a,b) = 0$. Note that positivity was already satisfied within each $Y_r \subset Y'$, which means under the quotient it remains unchanged: We can still identify Y_r with its image in Y. Also, all the distinguished points $y_0, y_1, \ldots$ become identified as a single point $y \in Y$. This is the distinguished point we give to Y.

We now wish to characterize $Y_r \subset Y$ precisely as the ball $\overline{B}_r(y)$. Take some $y' \in Y_r$. With respect to d_n^r (for arbitrary ε and large n), we have $d_n^r(y, x_n) < \varepsilon$ and some $x_n' \in \overline{B}_r(x_n)$ with $d_n^r(y', x_n') < \varepsilon$. Then

$$d(y, y') \leq d_n^r(y, x_n) + d_n^r(x_n, x_n') + d_n^r(x_n', y') < r + 2\varepsilon.$$

Thus $d(y, y') \leq r$ and, therefore, $Y_r \subset \overline{B}_r(y)$. Conversely, choose $y' \in Y_s \setminus Y_r$ for some $s > r$ (if possible. Otherwise, put $Y_r = Y$ and we are done) and let us show that $y' \notin \underline{B_r(y)}$. Let $K = d(y', Y_r) > 0$. For all $\varepsilon > 0$ and all but finitely many $n \in \{1, 2, \ldots\}$, there is some $x_n' \in \overline{B}_s(x_n)$ such that $d_n^s(y', x_n') < \varepsilon$. Consider the distance between x_n' and some arbitrary $y'' \in Y_r$:

$$K \leq d(y', y'') \leq d_n^r(y', x_n') + d_n^r(x_n', y'') < d_n^r(x_n', y'') + \varepsilon$$

$$\implies d_n^r(x_n', y'') \geq K - \varepsilon$$

Hence we may assume that $d_n^r(x_n', y'') > K/2$ for all but finitely n, as ε is arbitrary. Since, by hypothesis, $\overline{B}_r(x_n) \subset \overline{B}_{K/2}(Y_r)$ for all but finitely many n, we see that $x_n' \notin \overline{B}_r(x_n)$, as long as n is large enough. But then we have

$$r < d_n^s(x_n', x_0) \leq d_n^s(x_n', y') + d(y', y) + d_n^s(y, x_0) \leq d(y', y) + 2\varepsilon$$

$$\implies d(y', y) > r - 2\varepsilon$$

$$\implies d(y', y) \geq r$$

That is, $y' \notin B_r(y)$. We have shown that $B_r(y) \subset Y_r$. Since $Y_r \subset Y$ is closed, this implies $\overline{B}_r(y) \subset Y_r$, as was needed.

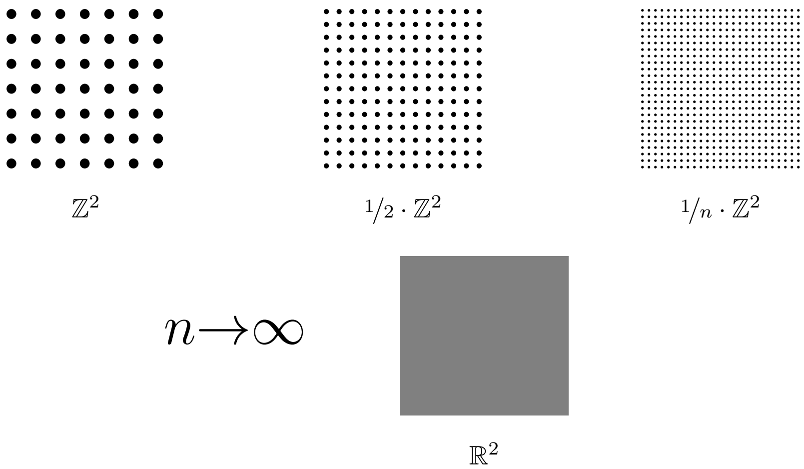
What we have amounts to this: for arbitrary $r, n \in \{1, 2, \ldots\}$ there is a metric d_n^r on $\overline{B}_r(x_n) \sqcup Y_r$ which, when extended do $X_n \sqcup Y$ (see Remark 2.5), satisfies

$$d_n^r(x_n, y) < 1/n$$

$$\overline{B}_r(x_n) \subset \overline{B}_{1/n}(Y_r) \subset \overline{B}_{1/n}(Y)$$

$$\overline{B}_r(y) = Y_r \subset \overline{B}_{1/n}(\overline{B}_r(x_n)) \subset \overline{B}_{1/n}(X_n)$$

If we then define $d_n = d_n^n$, the sequence of metrics $\{d_n\}_{n=1}^\infty$ clearly induces the needed definite convergence of $\{(X_n, x_n)\}_{n=1}^\infty$ to (Y, y). ■



The sequence of metric spaces $\mathbb{Z}_n^2$ as it converges to $\mathbb{R}^2$

Example 3.37. Consider sequence of pointed metric spaces $\{(X_n, x)\}_{n=1}^{\infty}$, each over the same set X but with the metric obtained from a fixed metric d on X by setting $d_n(x, x') = \frac{1}{n} d(x, x')$. One might picture X_n as the space X after being *shrunk* by a factor of n.

Interesting spaces can be obtained by shrinking other spaces. Take, for instance, $(\mathbb{Z}^k, 0)$ with the metric inherited from $\mathbb{R}^k$. Then $\{(\mathbb{Z}_n^k, 0)\}_{n=1}^{\infty}$ can be shown to have a subsequence that converges to $\mathbb{R}^k$ as $n \to \infty$ (in fact, the entire sequence converges, but we wish to use the statement from Theorem 3.36). Note that the ball of radius r in $\mathbb{Z}_n^k$ is embedded in the ball of the same radius in $\mathbb{R}^k$, consisting of the points with all coordinates being integer multiples of $\frac{1}{n}$. The Hausdorff distance between the two balls is then bounded by $\frac{1}{n}$, which also bounds the Gromov-Hausdorff distance between them, in turn satisfying the hypothesis of the Theorem. The idea of shrinking a discrete space into a non-discrete one will come up again later on.

We finally know enough about the pointed Gromov convergence to assert a precompactness condition. By relating the pointed convergence to the well behaved convergence of compact spaces we have made our job quite easy:

Theorem 3.38. *Let $\{(X_n, x_n)\}_{n=1}^{\infty}$ be a sequence of proper pointed metric spaces such that for each $r \in \{0, 1, 2 \ldots\}$ the sequence $\{B_r(x_n)\}_{n=1}^{\infty}$ is uniformly totally bounded. Then it admits a subsequence that converges to some proper pointed space.*

Proof. Since each sequence of balls is a sequence of uniformly totally bounded compact spaces we may apply the former precompactness condition as in Proposition 3.25. Each $\{B_r(x_n)\}$ has a convergent subsequence. By Cantor's diagonal argument there is a sequence of indices $\{n_k\}_{k=1}^{\infty}$ such that each $\{B_r(x_{n_k})\}_{k=1}^{\infty}$ converges. Then by Proposition 3.36, $\{(X_{n_k}, x_{n_k})\}_{k=1}^{\infty}$ has a subsequence that converges to a proper pointed space. ∎

Another consequence of the isometry lemma concerns homogeneous spaces:

Proposition 3.39. *Let $\{(X_n, x_n)\}_{n=1}^{\infty}$ be a sequence of pointed metric spaces such that each X_n is homogeneous and such that the sequence converges to a proper pointed space (Y, y). Then Y is homogeneous.*

Proof. Let $y', y'' \in Y$. There are sequences $\{x_n'\}_{n=1}^{\infty}$ and $\{x_n''\}_{n=1}^{\infty}$ that converge to y', y'', respectively (w.r.t. a fixed family of metrics on the disjoint unions). Let $\{f_n\}_{n=1}^{\infty}$ be a family of isometries such that each $f_n(x_n') = x''$. By passing to a subsequence we may assume that the sequence $\{f_n\}_{n=1}^{\infty}$ converges to an isometry $f : Y \to Y$.

Let us consider the value of $d(f(y'), y''))$. Let $\varepsilon > 0$. There is some $\delta > 0$ such that $d(f_n(x_n'), f(y')) < \varepsilon/2$, as long as $d(x_n', y') < \delta$ (which is true for all but finitely many n). Assuming n is large enough, we also have $d(y'', x_n'') < \varepsilon/2$. Of course we also have $d(f_n(x_n'), x_n'') = 0$. By combining everything, we get

$$d(f(y'), y'')) \leq d(f(y'), f_n(x_n')) + d(f_n(x_n'), x_n'') + d(x_n'', y'')) < \varepsilon/2 + 0 + \varepsilon/2 = \varepsilon.$$

∎

4 Groups of polynomial growth

To finish our task, we return to finitely generated groups and their growth. We will properly define the growth rate of a group, showing that the definition does not depend on the choice of a generator set. We proceed to explore properties of the growth rate of a finitely generated group, particularly how it is related to the growth rate of its subgroups and how it behaves in relation to exact sequences. We prove in Corollary 4.14 the result that Gromov's Theorem is the converse of: all finitely generated almost nilpotent groups have polynomial growth.

In Proposition 2.71 we tie many results from the previous two sections by showing that a group of polynomial growth, when endowed with a certain sequence of left-invariant metrics, converges with respect to the Gromov-Hausdorff distance to a metric space with finite Hausdorff dimension whose group of isometries is a Lie group. Finally, we conclude the section with a proof of Gromov's Theorem.

Definition 4.1. Let G be a finitely generated group. Let $S \subset G$ be a generating set that is symmetric (i.e., $x \in S \implies x^{-1} \in S$). Each element of G can be written (usually in multiple ways) as a finite string of elements of S. We denote the norm of g with respect to S as the minimal possible length of such strings, denoted $|g|_S$.

Proposition 4.2. *The norm defined above is indeed a group norm as a function $G \to \mathbb{R}^{\geq 0}$, i.e., for a finitely generated group G and a fixed generator set S we have the following (where $g, h \in G$ are arbitrary elements and $e \in G$ is the group identity):*

1. $|g|_S = 0 \iff g = e$;
2. $|g^{-1}|_S = |g|_S$;
3. $|gh|_S \leq |g|_S + |h|_S$.

Note that an for abelian group a group norm looks exactly like a usual norm for vector spaces, where (2.) could be reinterpreted as a sort of absolute homogeneity for all invertible scalars of the ring.

A norm $|\cdot|$ defined on a group G induces a metric on G as defined by

$$d(g, h) = |g^{-1}h|.$$

This metric is known as a *left-invariant* metric. The name becomes justified when we consider the value of $d(tg, th)$ for $t, g, h \in G$:

$$d(tg, th) = |g^{-1}t^{-1}th| = |g^{-1}h| = d(g, h).$$

That is, left-multiplication by an element of G induces an isometry on G. Metric spaces constructed from group metrics are always homogeneous: Given $g, h \in G$, there is a isometry taking g to h, namely left-multiplication by hg^{-1}. Also, these isometries are a way for subgroups of G to act on G:

Proposition 4.3. *Let $H \subset G$ be a subgroup, where G carries the left-invariant metric defined by a group norm. For each $h \in H$, the function $g \mapsto hg$ is an isometry on G. Moreover, the corresponding function $H \to Iso(G)$ is an action.*

Proof. As we have seen above, left multiplication defines an isometry for all elements of G, elements of H in particular. To show that it is a homomorphism, take $h, h' \in H$ and $g \in G$. Let f_h be defined as $g \mapsto hg$ and define $f_{h'}$ and $f_{hh'}$ similarly. We have

$$f_{hh'}(g) = (hh')g = h(h'g) = f_h(h'g) = f_h(f_{h'}(g)) = (f_h \circ f_{h'})(g).$$

■

By the construction above, each finitely generated group can be seen as a pointed metric space (G, e). We are interested on the asymptotic behaviour of the function $b_S : \mathbb{N} \to \mathbb{N}$ defined by sending n to the cardinality of the closed the ball centered at e with radius n with respect to the metric induced by S. We first remove the arbitrariness regarding S.

Proposition 4.4. *Let $S \subset G$ and $S' \subset G$ be two finite generating sets. There is some constant $C > 0$ such that, for all $g \in G$*

$$C^{-1}|g|_{S'} \leq |g|_S \leq C|g|_{S'}.$$

Proof. Let $S' = \{s'_1, \ldots, s'_n\}$. Write an element of G as $g = s'_{j_1} \ldots s'_{j_k}$, where $k = |g|_{S'}$. Write each s'_j as $s'_j = s_{1,j} \ldots s_{m_j,j}$ for $s_{i,j} \in S$. Choose C such that $C > m_j$ for every $j \in \{1, \ldots, n\}$. Then g can be written as a string of $k \cdot C$ elements of S. Thus $|g|_S \leq k \cdot C = C|g|_{S'}$, where C does not depend on g.

Symmetrically we have $|g|_{S'} \leq D|g|_S$, and therefore $D^{-1}|g|_{S'} \leq |g|_S$. By choosing C larger than D we obtain the other inequality. ∎

Thus if we have $g \in B_n^S(e)$ then $|g|_S \leq n$, which means $|g|_{S'} \leq Cn$ and $g \in B_{Cn}^{S'}(e)$. Therefore, $b_S(n) \leq b_{S'}(Cn)$. Similarly, if $|g|_{S'} \leq C^{-1}n$, we have $|g|_S \leq n$ and thus $b_{S'}(C^{-1}n) \leq b_S(n)$.

Definition 4.5. Let $f, g : \mathbb{N} \to \mathbb{N}$. We say that f has a *polynomial growth rate* of degree $d \geq 0$ when there are positive constants K_1, K_2 such that, for all $n \in \mathbb{N}$ we have

$$K_1 n^d \leq f(n) \leq K_2 n^d.$$

Similarly, we say that f has *exponential growth rate* when there are constants $K_1, K_2 > 1$ such that, for all $n \in \mathbb{N}$

$$K_1^n \geq f(n) \geq K_2^n.$$

Remark 4.6. If a function $f : \mathbb{N} \to \mathbb{N}$ has polynomial growth rate, then its degree of growth is unique. Otherwise, we would have positive K_1, K'_2 and $d > d'$ such that

$$K_1 n^d \leq f(n) \leq K'_2 n'^d$$

$$\implies K_1 n^d - K'_2 n'^d \leq 0.$$

This is a contradiction, since the polynomial $K_1 n^d - K'_2 n'^d$ has a positive leading coefficient and must satisfy $\lim_{n\to\infty} K_1 n^d - K'_2 n'^d = \infty$.

Remark 4.7. If G is a finitely generated group and S and S' are two finite generating sets, and $b_{S'}$ has polynomial growth of degree d, then the same can be said of b_S. Indeed, from Proposition 4.4, if $K_1 n^d \leq b_{S'}(n) \leq K_2 n^d$ we have $K_1 C^{-d} n^d \leq b_S(n) \leq K_2 C^d n^d$.

Definition 4.8. Let G be a group generated by a finite $S \subset G$. We say that G is a group of *polynomial* (resp., *exponential*) *growth* if b_S is a function of polynomial (resp., exponential) growth.

We define the *growth rate* $\Gamma(G)$ of G as the degree the polynomial growth of b_S, if it is of polynomial growth, and set $\Gamma(G) = \infty$ otherwise.

Example 4.9. Finitely generated abelian groups have polynomial growth. If a group is abelian and generated by a finite subset $S = \{s_1, \ldots, s_k\}$, then all strings with letters in S can be rearranged into the simple form $s_1^{m_1} \ldots s_n^{m_k}$, with $m_1, \ldots, m_k \in \mathbb{Z}$. The value of $b_S(n)$ is precisely the number of elements in the group that can written as one of such strings with $|m_1| + \ldots + |m_k| = n$. An exercise in combinatorics shows that this is bounded by $2^k \binom{n+k}{k} \leq 2^k n^k$.

Example 4.10. Free groups (see Example 2.52) with finitely many but at least two generators have exponential growth. If a generating set S has k elements and for each $s \in S$ we have $s^{-1} \in S$, then each string with i letters is uniquely determined by a choice of the leftmost element among k choices, and each of the subsequent $i - 1$ letters can be any one of $k - 1$ elements (since we must avoid the inverse of the previous letter to ensure uniqueness of the word). Thus there are $1 + \sum_{i=1}^{n} k(k-1)^{i-1} = 1 + \frac{1-(k-1)^n}{1-k} \geq (4/3)^k$ strings with at most k letters.

4.1 Some motivation: Volume growth in compact Riemannian manifolds

Now that we have defined the growth rate properly, let us try and put the problem of group growth in context and justify (somewhat informally) our interest in it. We refer to [Lee11] and [Lee12] for the background on the fundamental group and on Riemannian manifolds that will be needed in the rest of the section. See [Mil68] for the original proof of Proposition 4.11 below and more information on the application of the theory of group growth to differential geometry.

Let M be a compact smooth Riemannian manifold with a basepoint $x_0 \in M$ and let $G = \pi_1(M, x_0)$ be its fundamental group (which is finitely generated due to the compactness of M). Let

$$p : (\tilde{M}, \tilde{x}_0) \to (M, x_0)$$

be its universal covering map. If we endow $\tilde{M}$ with a Riemannian metric, then a geodesic ball $B_r(\tilde{x}_0)$ with radius r centered at $\tilde{x}_0$ has a well-defined, positive and finite volume $v_r(\tilde{x}_0)$, given by integral of the Riemannian volume form over the ball.

If the Riemannian metric on $\tilde{M}$ is the pullback by p of the metric given to M, then G acts isometrically on $\tilde{M}$ by the identification of the fundamental group with the group of deck transformations of the universal cover. This action is properly discontinuous, i.e., every point in $\tilde{M}$ has a neighbourhood U such that $gU \cap U = \varnothing$ for all $g \in G \setminus \{e\}$.

Let us assume that $v_r(\tilde{x}_0)$ has polynomial growth with respect to r, i.e., there are positive constants K, d such that

$$v_r(\tilde{x}_0) \le Kr^d.$$

Let $\varepsilon > 0$ be small enough such that $gB_\varepsilon(\tilde{x}_0) \cap B_\varepsilon(\tilde{x}_0) = \varnothing$ for $g \in G \setminus \{e\}$. Let $S = \{g_1, \ldots, g_k\}$ be a finite generating set for G with a corresponding growth function b_S and let r_0 be large enough such that $g_1\tilde{x}_0, \ldots, g_k\tilde{x}_0 \in B_{r_0}(\tilde{x}_0)$. Then if $g \in G$ can be written as a string of n elements of S, successive application of the triangle inequality implies $g\tilde{x}_0 \in B_{r_0 n}(\tilde{x}_0)$. We have $gB_\varepsilon(\tilde{x}_0) \subset B_{r_0 n+\varepsilon}(\tilde{x}_0)$, which in terms of volume implies

$$b_S(n) \le \frac{v_{r_0 n+\varepsilon}(\tilde{x}_0)}{v_\varepsilon(\tilde{x}_0)} \le \frac{K(r_0 n + \varepsilon)^d}{v_\varepsilon(\tilde{x}_0)}.$$

That is, b_S is of polynomial growth. Thus we expose a connection between the algebraic growth of the fundamental group of a space and the geometric growth of its universal cover. Let us seek an inequality in the other direction, assuming $b_S(n) \le Kn^d$.

Let δ be the diameter of M as a metric space. We have that the restriction of $p : \tilde{X} \to X$ to $\overline{B}_\delta(\tilde{x}_0)$, the closed ball centered at $\tilde{x}_0$, is surjective. Then $\tilde{X}$ has the cover by compact sets

$$\tilde{X} = \bigcup_{g \in G} g\overline{B}_\delta(\tilde{x}_0).$$

Note that for every $r > 0$, the ball $\overline{B}_r(\tilde{x}_0)$ intersects finitely many of these translates. For, otherwise, there would be infinitely many points in the form $g\tilde{x}_0$ at a distance less than $r + \delta$ from $\tilde{x}_0$, which would mean there are infinitely many balls with positive volume $B_\varepsilon(\tilde{x}_0)$ contained in the ball with radius $r + \delta + \varepsilon$ centered at $\tilde{x}_0$, a contradiction due to the finite volume of those balls.

Let S' be the finite subset of G consisting of all $g' \in G$ that satisfy

$$g'\overline{B}_\delta(\tilde{x}_0) \cap \overline{B}_\delta(\tilde{x}_0) \ne \varnothing.$$

We will show that this set generates G as a group. Let $\eta > 0$ be the infimum of all the (positive) distances between $\overline{B}_\delta(\tilde{x}_0)$ and $g\overline{B}_\delta(\tilde{x}_0)$, where $g \in G \setminus S'$. Fix some $g \in G \setminus S'$ and some $t \in \{1, 2, \ldots\}$ such that $d(\tilde{x}_0, g\overline{B}_\delta(\tilde{x}_0)) < \eta t + \delta$.

We have some $y_0 \in g\overline{B}_\delta(\tilde{x}_0)$ with $d(\tilde{x}_0, y_0) < \eta t + \delta$. Choose a minimal geodesic from y_0 to $\tilde{x}_0$ (the existence of a minimal geodesic relies on the completeness of $\tilde{X}$, which is guaranteed by the fact that X is itself complete) and take points $y_1, \ldots, y_t$ along its image such that $d(x_0, y_t) \le \delta$ and $d(y_i, y_{i+1}) < \eta$ for all $i \in \{0, \ldots, t-1\}$. Each y_i is contained in some ball $h_i\overline{B}_\delta(\tilde{x}_0)$, where in particular we choose $h_t = e$ and $h_0 = g$. If we write $g'_i = h_i^{-1}y_{i-1}$ we have

$$g'_t \ldots g'_1 = h_t^{-1}h_{t-1}h_{t-1}^{-1} \ldots h_1h_1^{-1}h_0 = h_t^{-1}h_0 = g.$$

Note that $h_i^{-1}y_i \in \overline{B}_\delta(\tilde{x}_0)$ and $h_i^{-1}y_{i-1} = h_i^{-1}h_{i-1}h_{i-1}^{-1}y_{i-1} \in g'_i\overline{B}_\delta(\tilde{x}_0)$. But since G acts *isometrically* on $\tilde{X}$, we have

$$d(h_i^{-1}y_i, h_i^{-1}y_{i-1}) = d(y_i, y_{i-1}) < \eta.$$

Thus we have a group element, g'_i, such that $g'_i\overline{B}_\delta(\tilde{x}_0)$ is closer to $\overline{B}_\delta(\tilde{x}_0)$ than the minimum reached by any element in $G \setminus S'$. It follows that $g'_i \in S$ for all $i \in \{1, 2, \ldots, t\}$. Since $g'_t \ldots g'_1 = g$ and g was taken arbitrarily from $G \setminus S'$, we see that S' generates G.

As we will see later, the asymptotic behaviour of the growth function does not change when we choose a different finite generating set, so we might as well assume that $S' = S$ and that $b_S(n)$ is still the growth function associated with this generating set. Every y_0 is in some $g\overline{B}_\delta(\tilde{x}_0)$ for $g \in G$ and above we shown that even if $g \notin S$, if additionally we have $d(\tilde{x}_0, y_0) < \eta t + \delta$, then g can be written as a string of no more than t elements of S. Then the ball $B_{\eta t+\delta}(\tilde{x}_0)$ must be completely covered by translates of $g\overline{B}_\delta(\tilde{x}_0)$, where g is one of such strings. Since there are no more than $b_S(t)$ distinct elements of G that can be expressed in this form, we have

$$v_{\eta t+\delta}(\tilde{x}_0) \leq v_\delta(\tilde{x}_0) b_S(t) \leq v_\delta(\tilde{x}_0) K t^d.$$

Having acquired this second inequality, we have the following very nice result:

Proposition 4.11. *Let M be a compact Riemannian manifold and $\tilde{M}$ be its universal cover, endowed with the Riemannian metric pulled back from the covering map. Then $\pi_1(M)$ has polynomial group growth if and only if $\tilde{M}$ has polynomial volume growth.*

In the terminology of Geometric Group Theory, what we have really accomplished in the proof of Proposition 4.11 is showing that the fundamental group of a compact Riemannian manifold is *quasi-isometric* to its universal covering space. This is a very particular case of what is now known as the Švarc–Milnor Lemma, or the Fundamental Observation of Geometric Group Theory (see Chapter IV of [dlH00] for a definition of quasi-isometries and a full statement of the Lemma).

Under the language of quasi-isometries, Remark 4.7 could be re-stated as "The metrics induced on a finitely generated group by different finite generating sets are quasi-isometric to each other", while the proof of Proposition 4.12 below implies that "A finitely generated group is quasi-isometric to its subgroup of finite index".

4.2 Growth rate of subgroups

We begin our delve into groups of polynomial growth by inquiring about the growth rate of subgroups of a group of polynomial growth. This the the first significant result:

Proposition 4.12. *Let G be a finitely generated group and $H \subset G$ a subgroup of finite index. Then G has polynomial growth if and only if H does, in which case we have $\Gamma(G) = \Gamma(H)$.*

Proof. If G is generated by a finite set $S \subset G$, then H is generated by a finite set $R \subset G$, due to Proposition 2.64. Then G is also generated by $R \cup S$ and of course we have $b_R(n) \leq b_{R\cup S}(n)$, which implies that $\Gamma(H) \leq \Gamma(G)$.

Conversely, let $H \subset G$ be a subgroup of finite index and of polynomial growth, generated by some finite $R \subset H$. Let us assume temporarily that H is normal in G. Then we have a finite quotient group $G/H = \{Hg_1, Hg_2, \ldots, Hg_n\}$. We may assume that $g_1 = e$ and let $U = \{g_1, \ldots, g_n\}$. Normality implies that each $g_i^{-1} r g_i \in H$ for $r \in R$ and $g_i \in U$. Then H is also generated by the finite set

$$R' = \{g_i^{-1} r g_i \ : \ g_i \in U, \ r \in R\}.$$

Since all elements of G can be written as $g = hg_i$ for some $h \in H$ and $g_i \in U$, we see that $S = R' \cup U$ is a finite generating set for G. In particular for all $\varepsilon_1, \varepsilon_2 \in \{-1, 1\}$ and $i, j \in \{1, \ldots, n\}$ we can write $g_i^{\varepsilon_1} g_j^{\varepsilon_2} = g_{k(\varepsilon_1 i, \varepsilon_2 j)} h(\varepsilon_1 i, \varepsilon_2 j)$ for some $h(\varepsilon_1 i, \varepsilon_2 j) \in H$ and $k(\varepsilon_1 i, \varepsilon_2 j) \in \{1, \ldots, n\}$. There are finitely many combinations of $\varepsilon_1, \varepsilon_2, i, j$, so let $N \in \{1, 2 \ldots\}$ be such that all $h(\varepsilon_1 i, \varepsilon_2 j)$ can be written as a string of, at most, N elements of R'.

Let $r' = g_i^{-1} r g_i \in R'$ and $g_j \in U$. We have

$$\begin{aligned} r' g_j &= g_i^{-1} r (g_i g_j) \\ &= g_i^{-1} r g_{k(i,j)} h(i,j) \\ &= (g_i^{-1} g_{k(i,j)})(g_{k(i,j)}^{-1} r g_{k(i,j)}) h(i,j) \\ &= g_{k(-i,k(i,j))} h(-i, k(i,j))(g_{k(i,j)}^{-1} r g_{k(i,j)}) h(i,j) \end{aligned}$$

That is, we can write $r' g_j$ as an element of U followed by a string of, at most, $2N + 1$ elements of R'. Now write some $g \in G$ as a string of m elements of S, that is, with letters in the form $g_j \in U$ or $r' = g_i^{-1} r g_i \in R'$.

There is a rightmost occurrence of a g_j in the string that that is not part if a conjugate $g_i^{-1} r g_i$. We can translate this g_j right to left through the string, replacing each $r' g_j$ with some $g_{j'} w_0$, where

w_0 is a string of length not greater than $2N+1$ elements of R'. If another lone g_k is encountered, replace $g_k g_{j'}$ with $g_{j''} w_1$, where w_1 is a string of not more than N elements in R. After a finite number of steps, we are left with an equivalent string in the form $g_J w_2$, where w_2 is a product of no more than $m(3N+1)$ elements of R'.

Thus, every $g \in B^S_m(e)$ can be written as a product of one of the $[G:H]$ elements of U by one of the elements of $B^{R'}_{m(3N+1)}(e)$. It follows that $b_S(m) \leq [G:H] b_{R'}(m(3N+1))$, which is bounded above by a polynomial of degree $\Gamma(H)$. Therefore, $\Gamma(G) \leq \Gamma(H)$.

Even if H is not normal in G, the intersection $N = \bigcap_{g \in G} g^{-1} H g$ is normal in G. Note that if $g' = hg$ for some $g \in G$ and $h \in H$, we have

$$g'^{-1} H g' = (hg)^{-1} H (hg) = g^{-1} h^{-1} H h g = g^{-1} H g$$

Hence if H has finite index in G, it has finitely many conjugate subgroups. Then N is an intersection of finitely many subgroups of finite index and thus has finite index in both G and H, which means $\Gamma(N) \leq \Gamma(H)$ and $\Gamma(G) \leq \Gamma(N)$, as we have just proved. ∎

We have shown in Example 4.9 that finitely generated abelian groups have polynomial growth. Let us do better and prove something much stronger:

Proposition 4.13. *Finitely generated nilpotent groups have polynomial growth.*

Proof. We prove by induction on the length of the lower central series, where the first step is complete as it is precisely Example 4.9.

For the inductive step, let $G = \langle \{s_1, \ldots, s_p\} \rangle$ be a finitely nilpotent group with lower central series of length N. Let $g = s_{i_1} \ldots s_{i_\ell}$ be some string of elements of $S = \{s_1, \ldots, s_p\}$. If G were abelian, one could put any such string in the simple form $s_1^{m_1} \ldots s_p^{m_p}$ with $|m_1| + \ldots + |m_p| \leq \ell$ by repeatedly swapping elements with their neighbour to the left to ensure all occurrences of s_1 come first, then all occurrences of s_2, and so on. Since G is merely nilpotent, any swap made along the way can generate non-trivial commutators, as $g_1 g_2 = g_2 g_1 [g_1, g_2]$

Every element of S present in the string will have to swap places with, at most, ℓ other elements of S in order to be placed in its adequate position. This will generate no more than ℓ^2 commutators in the form $[s_i, s_j]$ (elements of G_1) that have to be inserted into an appropriate position so that the string remains equal to g. This means that every element of S present in the string will also have to swap places with possibly each one of these commutators, adding ℓ^3 entries to the string that have the form $[[s_i, s_j], s_k]$, which are elements if G_2. We proceed as such, swapping elements of S with commutators as needed.

The result will be an equivalent string $g = s_1^{m_1} \ldots s_p^{m_p} w$, where w is a product of, at most ℓ^2 elements of G_1, ℓ^3 elements of G_2, and so on. Of course, no elements of G_N will be needed because we can commute with elements of G_{N-1} as needed.

Applying the inductive hypothesis, we know that there is some $K > 0$ such that w is one of no more than $K \left(\sum_{i=2}^{N-1} \ell^i \right)^{\Gamma(G_1)} \leq K \ell^{N \Gamma(G_1)}$ possible elements of G_1. Since we know that there are, at most, $2^p \binom{\ell+p}{p} \leq 2^p \ell^p$ elements of G in the form $s_1^{m_1} \ldots s_p^{m_p}$ with $|m_1| + \ldots + |m_p| \leq \ell$, we are able to bound the growth function of G with respect to S by the expression $K \ell^{p + N\Gamma(G_1)}$, which implies $\Gamma(G) \leq p + N\Gamma(G_1)$. ∎

Propositions 4.12 and 4.13 are combined in an important result. We say that a group is *almost* nilpotent if it has a nilpotent subgroup of finite index (similarly for almost abelian or almost solvable groups). Thus:

Corollary 4.14. *Finitely generated almost nilpotent groups have polynomial growth.*

Gromov's Theorem is the much more interesting converse of this Corollary: Groups of polynomial growth must always be almost nilpotent. Gromov's proof relied on the following previous results, respectively by Tits (Corollary 1 on [Tit72]) and Wolf (Theorem 4.8 on [Wol68])

Theorem 4.15. *A finitely generated subgroup of a Lie group with finitely many components either contains a free group with at least two generators or a solvable group of finite index*

Theorem 4.16. *A finitely generated solvable group either has exponential growth or is almost nilpotent.*

By combining both results we obtain a partial converse to Corollary 4.14:

Lemma 4.17. *If a finitely generated subgroup of a Lie group with finitely many components has polynomial growth, then it is almost nilpotent.*

Our objective henceforth will be to extend this result to all groups of polynomial growth, acquiring a better understanding of the growth rate along the way. An obvious next step will be to seek a version of Proposition 4.12 for subgroups of infinite index.

To do so we return our attention to metric aspect of the growth function. Propositions 2.67 and 4.3 tell us that there is a pseudometric on G/H built from the metric on G (here we do not need to see G/H as a group, merely as the set of left cosets). Looking back at the definition we have

$$d(Hg, Hh) = \inf\{d(g', h') \ : \ g' \in Hg, \ h' \in Hh\}.$$

Since the metric on G only takes integer values, we notice that $d(Hg, Hh) = 0$ implies that $d(g', h') = 0$ for some $g' \in Hg$ and $h' \in Hh$. Thus $Hg = Hh$. Therefore, the pseudometric induced on G/H is an actual metric in this case. This space has the following useful property:

Definition 4.18. A metric space is said to have the *integral connectivity property* if for all $a, b \in \{1, 2 \ldots\}$ and $x \in X$ we have $\overline{B}_a(\overline{B}_b(x)) = \overline{B}_{a+b}(x)$.

Proposition 4.19. *Let X be a metric space such that $d(x, x') \in \mathbb{Z}$ for all $x, x' \in X$. Then the following are equivalent:*

1. *X has the integral connectivity property.*
2. *For all points $x_0, x_p \in X$ with $d(x_0, x_p) = p$ there are $x_1, \ldots, x_{p-1} \in X$ with $d(x_{i-1}, x_i) = 1$ for all $i \in \{1 \ldots, p\}$.*

Proof. We begin by assuming the first property and proving the second. Let $x_0, x_p \in X$ with $d(x_0, x_p) = p$. Then $x_p \in \overline{B}_p(x_0) \subset \overline{B}_1(\overline{B}_{p-1}(x_0))$. We see that x_p must be contained in $\overline{B}_1(\overline{B}_{p-1}(x_0)) \backslash \overline{B}_{p-1}(x_0)$ (Otherwise, $d(x_p, x_0) \leq p-1$). Thus there must be some $x_{p-1} \in \overline{B}_{p-1}(x_0)$ with $d(x_p, x_{p-1}) = 1$. Also, we have

$$p = d(x_0, x_p) \leq d(x_0, x_{p-1}) + d(x_{p-1}, x_p) = d(x_0, x_{p-1}) + 1.$$

Thus $d(x_0, x_{p-1}) = p - 1$. We may then proceed inductively with x_{p-1}.

For the other implication, note that the inclusion $\overline{B}_a(\overline{B}_b(x)) \subset \overline{B}_{a+b}(x)$ is true for every metric space. For the converse inclusion, let $x' \in X$ with $d(x, x') = p \leq a + b$ and $p > b$ (if $p \leq b$ the inclusion $x' \in \overline{B}_a(\overline{B}_b(x))$ is immediate). Then there are $x = x_0, x_1, \ldots, x_p = x'$ such that $d(x_{i-1}, x_i) = 1$. If $i > j$ we have

$$d(x_i, x_j) \leq d(x_i, x_{i-1}) + \ldots + d(x_{j+1}, x_j) = i - j.$$

In particular, $d(x_b, x_0) \leq b$ and $d(x_p, x_b) \leq p - b \leq a$. Thus $x_p \in \overline{B}_a(\overline{B}_b(x_0))$. ∎

Proposition 4.20. *Let G be a finitely generated group endowed with the metric induced by the finite generating set $S = \{s_1, \ldots, s_n\}$. Then G has the integral connectivity property*

Proof. We apply Proposition 4.19 by using the alternative definition. Given $x_0, x_p \in X$, we write $x_p = x_0(x_0^{-1}x_p)$, where $x_0^{-1}x_p = s_{j_1} \ldots s_{j_p}$ is some minimal string of elements of S. Then we can make $x_i = x_0 s_{j_1} \ldots s_{j_i}$. Indeed we have

$$0 < d(x_{i-1}, x_i) \leq |(x_0 s_{j_1} \ldots s_{j_{i-1}})^{-1}(x_0 s_{j_1} \ldots s_{j_i})|_S = |s_{j_i}|_S = 1.$$

∎

Corollary 4.21. *Let G be a finitely generated group and $S \subset G$ be a finite generating set. Then $b_S(n) \leq b_S(1)^n$.*

Proof. Due to Proposition 4.20 we have, with respect to the metric induced on G by S, the inclusion $\overline{B}_{a+b}(e) \subset \overline{B}_a(\overline{B}_b(e))$ for all positive integers a, b. Then $b_S(a + b) \leq b_S(a)b_S(b)$. In particular, $b_S(n) \leq b_S(1)b_S(n - 1) \leq \ldots \leq b_S(1)^n$. ∎

It turns out that this property is inherited by ${}^{G}/_{H}$:

Proposition 4.22. *Let X be a metric space such that $d(x, x') \in \mathbb{Z}$. Let $f : G \to Iso(X)$ be a metric action. If X has the integral connectivity property, then so does ${}^{X}/_{G}$ (here, ${}^{X}/_{G}$ is an actual metric space due to the fact that the metric on X takes only integer values).*

Proof. Again we use the alternative definition from Proposition 4.19. We take $O_G(x_0), O_G(x_p) \in {}^{X}/_{G}$ with $d(O_G(x_0), O_G(x_p)) = p$. Since the metric of X takes only integer values, we may assume that $d(x_0, x_p) = p$ (that is, we assume that x_0 and x_p are the elements of their respective classes that are as close as possible to each other). There must be a a sequence $x_0, x_1, \ldots, x_p$ with $d(x_{i-1}, x_i) = 1$. We claim that $d(O_G(x_{i-1}), O_G(x_i)) = 1$

$$\begin{aligned} p = d(O_G(x_0), O_G(x_p)) &\leq d(O_G(x_0), O_G(x_1)) + \ldots + d(O_G(x_{p-1}), O_G(x_p)) \\ &\leq d(x_0, x_1) + \ldots + d(O_G(x_{i-1}), O_G(x_i)) + \ldots d(x_{p-1}, x_p) \\ &= d(O_G(x_{i-1}), O_G(x_i)) + p - 1. \end{aligned}$$

Thus $d(O_G(x_{i-1}), O_G(x_i)) \geq 1$. Since we know that $d(O_G(x_{i-1}), O_G(x_i)) \leq d(x_{i-1}, x_i) = 1$, we have the needed equality. ∎

Lemma 4.23. *Let X be an infinite metric space with the integral connectivity property. Each ball $B_r(x)$, $r \in \{0, 1, 2, \ldots\}$, $x \in X$ has at least $r + 1$ elements.*

Proof. Fix some r and x. If $B_r(x)$ is infinite we are done. Otherwise there is some $x' \in X$ with $d(x, x') > r$. Then we have a sequence $x = x_0, x_1, \ldots, x_p = x'$ with $p = d(x, x')$ and $d(x_{i-1}, x_i) = 1$. Note that each $x_i \in B_i(x)$, which means $\#B_r(x) \geq p + 1 \geq r + 1$. ∎

Proposition 4.24. *Let G be a finitely generated group of polynomial growth and let $H \subset G$ be a finitely generated subgroup of infinite index. Then H has polynomial growth and $\Gamma(H) < \Gamma(G)$.*

Proof. We may assume that G is finitely generated by some S such that H is generated by some $S' = S \cap H$. Due to the previous lemma, we know that the ball in ${}^{G}/_{H}$ centered at He with radius r has at least $r + 1$ distinct elements $Hx_0, \ldots, Hx_r$, where we may assume $d(Hx_i, He) = d(x_i, e)$ because the distance taking integer values implies that the infimum of the distances does achieve a minimum in this case.

For all $h \in H$, we have $|h|_S \leq |h|_{S'}$ which implies that the ball B' of radius r in H centered at the identity w.r.t. S' in contained in the concentric ball of radius r in H w.r.t. S. Hence we have $B' \subset \overline{B}_r(e) \cap H$.

Consider the sets $B'x_i$ of all elements xx_i, for all $x \in B'$. Since $B' \subset H$, we know that each pair $B'x_i$, $B'x_j$ is disjoint, as x_i and x_j belong to different classes of G modulo H. Since $d(Hx_i, He) = d(x_i, e)$ we have that $x_0, \ldots, x_r \in \overline{B}_r(e)$. Thus $B'x_i \subset \overline{B}_{2r}(e)$. We have shown that $(r + 1)$ disjoint copies of B' are contained within the ball of radius $2r$ in G. Thus

$$\#B' \leq \frac{\#\overline{B}_{2r}(e)}{(r+1)} \leq \frac{K(2r)^{\Gamma(G)}}{(r+1)} < K(2r)^{\Gamma(G)-1}$$

Where K is the constant in the definition of growth and r is taken arbitrarily. Thus $\Gamma(H) \leq \Gamma(G) - 1$. ∎

4.3 Group growth and exact sequences

In order to make use Proposition 4.24 we are going to need a condition that ensures that a subgroup with infinite index of a group of polynomial growth is finitely generated. This will suffice:

Proposition 4.25. *Let G a group of polynomial growth. Let $f : G \to H$ be a surjective group homomorphism such that imf contains an infinite, cyclic group. Then* $\ker f$ *is finitely generated.*

Proof. Choose some $s_0 \in G$ such that $f(s_0)$ is a generator of the infinite cyclic subgroup of H. If $S = \{s_0, s_1, \ldots, s_n\} \subset G$ is a finite generating set containing s_0, note that $\{s_0, s_1 s_0^{k_1}, \ldots, s_n s_0^{k_n}\}$ (with each $k_i \in \mathbb{Z}$) is still a generating set. By setting $k_i = -f(s_i)$ for all $i \in \{1, 2, \ldots, n\}$, we get that $\{s_1 s_0^{k_1}, \ldots, s_n s_0^{k_n}\} \subset \ker f$. Thus by replacing S we may assume that $S \setminus \{s_0\} \subset \ker f$.

Consider the nested sequence of subgroups $\{G_i\}_{i=0}^{\infty}$ defined as such

$$G_i = \langle \{s_0^j s_k s_0^{-j} \ : \ k \in \{1, \dots, n\}, \ j \in \{-i, \dots, i\}\} \rangle.$$

Note that each $G_i \subset \ker f$, since $s_k \in \ker f \implies s_0^j s_k s_0^{-j} \in \ker f$, due to the normality of the kernel. Also, if $g \in \ker f$, then g can be written as a string of elements of S such that s_0 and s_0^{-1} appear the same amount of times, which we can call N. A bit of combinatorics will convince one that then we must have $g \in G_N$. Thus

$$\ker f = \bigcup_{i=0}^{\infty} G_i.$$

Note that, since each G_i is finitely generated, no more work is needed if the sequence of subgroups halts, i.e., if $G_i = \ker f$ for some $i \in \{0, 1, \dots\}$. Assuming the contrary, we have a sequence $\{g_i\}_{i=1}^{\infty}$ with each $g_i = s_0^{j_i} s_{k_i} s_0^{-j_i}$ with $k_i \in \{1, \dots, n\}$ and $j \in \{-i, i\}$ such that $g_i \notin G_{i-1}$.

Let $S_i = \{0, 1\}^i$ (The i-times Cartesian product of the set $\{0,1\}$). Define the following function

$$F_i : S_i \to \ker f$$

$$(e_1, \dots, e_i) \mapsto g_1^{e_1} \dots g_i^{e_i}$$

This is an injection: Assume that there are $E = (e_1, \dots, e_i)$ and $E' = (e_1', \dots, e_i')$, two different elements of S_i such that $F_i(E) = F_i(E')$. Let $I \in \{1, 2, \dots\}$ be the maximum index such that $e_I \neq e_I'$ (We may assume that $e_I = 1$ and $e_I' = 0$). Thus

$$\begin{aligned} g_1^{e_1} \dots g_i^{e_i} = g_1^{e_1'} \dots g_i^{e_i'} &\implies g_1^{e_1} \dots g_I^{e_I} = g_1^{e_1'} \dots g_I^{e_I'} \\ &\implies g_1^{e_1} \dots g_I = g_1^{e_1'} \dots g_{I-1}^{e_{I-1}'} \\ &\implies g_I = g_{I-1}^{e_{I-1}} \dots g_1^{e_1} g_1^{e_1'} \dots g_{I-1}^{e_{I-1}'}, \end{aligned}$$

a contradiction, since all terms on the right are elements of G_{I-1}. Thus $\mathrm{im} F_i \subset \ker f$ contains at least $\# S_i = 2^i$ elements. Applying the group norm $| - |$ induced by the generating set S we have

$$g \in \mathrm{im} F_i \implies |g| \leq \sum_{j=1}^{i} |g_j| = \sum_{j=1}^{i} |s_0^{\pm i} s_{k_j} s_0^{\mp i}| \leq \sum_{j=1}^{i} (2i+1) = (2i+1)i$$

That is, if the sequence $\{G_i\}_{i=0}^{\infty}$ doesn't halt, then for all $i \in \{1, 2, \dots\}$ the ball of radius $(2i+1)i$ has cardinality bounded below by 2^i, while by hypothesis we know that it must be bounded above by a polynomial with degree $\Gamma(G) + 2$ (in reality we have only used the fact that the growth is less-than-exponential). ■

Corollary 4.26. *The commutator subgroup of a group of polynomial growth is finitely generated.*

Proof. If the commutator is of finite index, then by Proposition 2.64 it must be finitely generated. Otherwise, by Proposition 2.61, the commutator is the kernel of a surjective homomorphism onto a direct sum of finitely many cyclic groups, one of which being infinite. ■

Proposition 4.25 is sometimes put in a different language. We use the term *exact sequence* of groups to refer to a sequence of groups $\{G_i\}_{i=0}^{\infty}$ and groups homomorphisms $\{f_i : G_i \to G_{i+1}\}_{i=0}^{\infty}$ such that

$$\ker f_{i+1} = \mathrm{im} f_i.$$

The most ubiquitous exacts sequences are *short*, meaning they have the form

$$\{e\} \to A \xrightarrow{f} B \xrightarrow{g} C \to \{e\},$$

where $\{e\}$ is the trivial group and the rightmost and leftmost arrows are the only possible group homomorphisms in this case. Note that we have f injective and g surjective. We often identify A with $\mathrm{im} f = \ker g \subset B$ and C with $\frac{B}{\ker g} = {}^{B}/_{A}$. Reciprocally, if B is any group and $A \subset B$ is any normal subgroup, then the inclusion $A \to B$ and the quotient $B \to {}^{B}/_{A}$ define a short exact sequence.

Thus Proposition 4.25 says that if B is of polynomial growth and C has $\mathbb{Z}$ as a subgroup, then A is finitely generated. Often, we can say something about a group that is part of a short exact sequence based on properties of the two other groups. The following is a classic example:

Proposition 4.27. *Let $\{e\} \to A \xrightarrow{f} B \xrightarrow{g} C \to \{e\}$ be a short exact sequence. Then A and C are solvable if and only if B is solvable.*

Proof. Suppose B is solvable. If $A^{(m)}$ and $B^{(m)}$ are the m-th terms of the derived series of A and B, respectively, we can show that $A^{(m)} \subset B^{(m)}$ inductively on m (where we are using the appropriate identification of A with a subgroup of B): For $m = 1$, simply note that the commutator of a subgroup is contained in the commutator of the entire group. But if $A^{(m)} \subset B^{(m)}$, then $B^{(m+1)}$ is the commutator of $B^{(m)}$, and thus $A^{(m+1)} \subset B^{(m+1)}$ (we have proved, more generally, that subgroups of solvable groups are solvable).

Similarly, we prove by induction that if $C^{(n)}$ is the n-th term of the derived series of C, then $C^{(n)} = \frac{B^{(n)}A}{A}$ (where $B^{(n)}A = \{ba \ : \ b \in B^{(n)}, \ a \in A\}$). First we note that for all $Ab_1, Ab_2 \in {}^{B}/_{A}$ we have $[Ab_1, Ab_2] = A[b_1, b_2]$. Thus $[C, C] = \frac{[B,B]A}{A}$. Inductively, since $C^{(n+1)}$ is just the commutator of $C^{(n)}$, we have $C^{(n+1)} = [C^{(n)}, C^{(n)}] = \frac{[B^{(n)}, B^{(n)}]A}{A} = \frac{B^{(n+1)}A}{A}$.

Conversely, suppose and A and C are both solvable, with derived series of lengths m and n, respectively. We know that $\{e\} = C^{(n)} = \frac{B^{(n)}A}{A}$. Thus $B^{(n)} \subset A$. But then $B^{(n+i)} \subset A^{(i)}$ for all $i \in \{1, 2, \ldots\}$. In particular, $B^{(m+n)} \subset A^{(n)} = \{e\}$, and we conclude that B is solvable. ■

More appropriate to our discussion we have:

Proposition 4.28. *Let $\{e\} \to A \xrightarrow{f} B \xrightarrow{g} \mathbb{Z} \to \{e\}$ be a short exact sequence, where B is of polynomial growth. If A has a solvable subgroup of finite index, then so does B.*

Proof. We know from Proposition 4.25 that A is finitely generated. Again, we see A as a subgroup of B through the identification provided by f. There is some $M \subset A$ solvable with finite index. From Proposition 2.63 we know that A has finitely many subgroups with index $q = [A : M]$. Let $A' \subset A$ be the intersection of all subgroups of A with index q (A' is itself solvable and has finite index on A) and $s \in B$ such that $g(s) = 1 \in \mathbb{Z}$. Let $B' \subset B$ be the subgroup generated by $A' \cup \{s\}$.

The subgroup A' is normal in B': If $b \in B'$, we must show that $b^{-1}A'b \subset A'$. This is certainly true if $b \in A'$. Thus it is enough to show that $s^{-1}A's \subset A'$. Every $a \in A'$ is an element of every subgroup of A with index q. But a conjugate of such a group by s must also be contained in A, since A is normal in B and $s \in B$. Thus $s^{-1}as$ is an element of every subgroup of A with index q and thus we have $s^{-1}A's \subset A'$.

Consider the kernel of the restriction $g : B' \to \mathbb{Z}$. It consists of the words

$$s^{k_1}a_1s^{k_2}a_2 \ldots a_N s^{k_{N+1}},$$

$$k_1 + k_2 + \ldots + k_{N+1} = 0.$$

We can always rearrange such words to be a combination of conjugates of elements of A' by powers of s, replacing $s^{k_1}a_1s^{k_2}a_2s^{k_3}$ with $s^{k_1}a_1s^{-k_1}s^{k_1+k_2}a_2s^{-k_1-k_2}s^{k_1+k_2+k_3}$, and so on. Due to A' being normal, all of such words must then be in A', which implies that A' is the kernel of g, which is the same as $A \cap B'$.

As in the beginning of the proof of Proposition 4.25, we may assume that B is generated by a set $\{s, s_1, \ldots, s_n\}$, with $s_i \in A$ for all $i \in \{1, 2, \ldots\}$. Thus we have $B = AB'$ and

$$[B : B'] = [AB' : B'] = [A : A \cap B'] = [A : A'] < \infty;$$

$$[B : B'][B' : A'][A' : M] = [B : M] = [B : A][A : M] = \infty;$$

$$\implies [B' : A'] = \infty.$$

The image of the restriction $g : B' \to \mathbb{Z}$ must then be infinite and thus isomorphic to $\mathbb{Z}$ itself. Therefore, the restricted sequence $\{e\} \to A' \xrightarrow{f} B' \xrightarrow{g} g(B') \to \{e\}$ remains exact. Since both A' and $g(B') \subset \mathbb{Z}$ are solvable, we apply Proposition 4.27 and obtain that B' is solvable. ■

4.4 Growth rate and dimension

An important aspect of the growth rate we will investigate is how large a ball of a given radius is when compared to balls of fixed lesser radius. More specifically, we wish to cover the ball $\overline{B}_n(e)$ with translates of the ball of smaller radius $\overline{B}_\varepsilon(e)$ and obtain an estimate of the number of balls that is needed (note how this can remind one of the Hausdorff dimension).

Theorem 4.29. *Let G be a finitely generated group with a fixed set of generators. Let $r, \varepsilon \in \{1, 2, \ldots\}$, with $\varepsilon \le r$. The ball $\overline{B}_r(e)$ can be covered by $\frac{\#\overline{B}_r(e)}{\#\overline{B}_\varepsilon(e)}$ balls of radius 3ε. Moreover, if the group is of polynomial growth rate d, there is some $K > 0$ (independent of r, ε) such that a cover with $K\left(\frac{r}{\varepsilon}\right)^d$ balls is possible.*

Proof. Let $S \subset G$ be a maximal 2ε-separated net of $\overline{B}_{r-\varepsilon}(e)$ as per Remark 2.10. The balls $\overline{B}_\varepsilon(s)$ for $s \in S$ are pairwise disjoint and contained in $\overline{B}_r(e)$. Thus

$$\#S \le \frac{\#\overline{B}_r(e)}{\#\overline{B}_\varepsilon(e)}.$$

The balls $\overline{B}_{2\varepsilon}(s)$ for all $s \in S$ cover $\overline{B}_{r-\varepsilon}(e)$, since S is a 2ε-net. The balls $\overline{B}_{3\varepsilon}(s) = \overline{B}_\varepsilon(\overline{B}_{2\varepsilon}(s))$ cover $\overline{B}_r(e) = \overline{B}_\varepsilon(\overline{B}_{r-\varepsilon}(e))$, due to the integral connectivity property.

Further assuming that the group is of polynomial growth, let K_1, K_2 positive constants such that $K_1 r'^d \le \#\overline{B}_{r'}(e) \le K_2 r'^d$ for all positive integers r'. By combining this with inequality above we see that the ball $\overline{B}_r(e)$ can be covered by $\frac{\#\overline{B}_r(e)}{\#\overline{B}_\varepsilon(e)} \le \frac{K_2}{K_1}\left(\frac{r}{\varepsilon}\right)^d$ balls of radius 3ε. ■

We will now combine Theorems 3.38 and 4.29 into a generalization of Example 3.37. As before, we take a finitely generated group G of polynomial growth rate d (in the example we took the group $\mathbb{Z}^k$) and denote as d the metric induced on G by some fixed set of generators. Then we define a metric space G_n over the same set G but with the scaled metric $d_n = \frac{1}{n}d$.

The ball $\overline{B}_r(e_n)$ of radius r centered at the identity in G_n is just the ball $\overline{B}_{nr}(e)$ of radius nr centered at the identity in G but with the distances shrunk by a factor of n. Since $\overline{B}_{nr}(e)$ can be covered by $K(\frac{nr}{n\varepsilon})^d = K(\frac{r}{\varepsilon})^d$ isometric copies of $\overline{B}_{n\varepsilon}(e)$, we know that $\overline{B}_r(e_n)$ can be covered by $K(\frac{r}{\varepsilon})^d$ copies of $\overline{B}_\varepsilon(e_n)$. This implies that the family of balls $\{\overline{B}_r(e_n)\}_{n=0}^\infty$ is uniformly totally bounded. Each space G_n is of course proper since all balls are finite in cardinality and thus compact. Thus the sequence $\{(G_n, e_n)\}_{n=1}^\infty$ satisfies the hypothesis of Theorem 3.38 and we have a limit space (Y_G, y_0). This space happens to have particularly nice properties.

Proposition 4.30. *Let G be a group of polynomial growth rate d and let Y_G be as above. We have*

1. *Y_G is connected and locally path connected.*
2. *Y_G is a locally compact and homogeneous space.*
3. *$\dim(Y_G) \le d$.*

Proof.

1. Denote by δ the metric on Y_G. Let $a, b \in Y_G$. By applying Proposition 3.30 we know that there is a sequence of metrics $\{\delta_n\}_{n=1}^\infty$, each agreeing with both δ and d_n on the disjoint union $G_n \sqcup Y_G$, such that $\{G_n\}$ converges to $\{Y_G\}$ w.r.t. $\{\delta_n\}$. Each space G_n is just G with the metric scaled by a positive factor $\lambda_n \le 1/n$ (the sequence $\{\lambda_n\}_{n=1}^\infty$ is by construction a subsequence of $\{1/n\}_{n=1}^\infty$). By passing to a subsequence, we may assume that there are $a_n, b_n \in G_n$ such that $\delta_n(a_n, a) < \lambda_n$ and $\delta_n(b_n, b) < \lambda_n$.

   By applying Proposition 4.19 to G_n we know that there is a sequence $a_n = x_0, \ldots, x_i, \ldots, x_p = b_n$, each belonging to G_n such that $d_n(x_{i-1}, x_i) = \lambda_n$ and $p = \frac{d_n(a_n, b_n)}{\lambda_n}$. Setting j to $\frac{p}{2}$ rounded down to the nearest integer and applying the triangle inequality j times we see that

$$\begin{aligned} d_n(a_n, x_j) &\le j\lambda_n \\ &\le \frac{d_n(a_n, b_n)}{2} + \lambda_n \\ &\le \frac{\delta(a, b)}{2} + 2\lambda_n. \end{aligned}$$

Very similarly:

$$\begin{aligned} d_n(x_j, b_n) &\leq (j+1)\lambda_n \\ &\leq \frac{d_n(a_n, b_n)}{2} + 2\lambda_n \\ &\leq \frac{\delta(a,b)}{2} + 3\lambda_n. \end{aligned}$$

By choosing some $y_n \in Y_G$ such that $\delta_n(y_n, x_j) < \lambda_n$ we get

$$\delta(a, y_n) \leq \frac{\delta(a,b)}{2} + 3\lambda_n,$$

$$\delta(y_n, b) \leq \frac{\delta(a,b)}{2} + 4\lambda_n.$$

Then, the resulting sequence $\{y_n\}_{n=1}^{\infty}$ can be seen to be a Cauchy sequence that by completeness converges to some $y \in Y_G$ satisfying the inequalities $\delta(a, y) \leq \frac{\delta(a,b)}{2}$ and $\delta(y, b) \leq \frac{\delta(a,b)}{2}$. Note that by applying the triangle inequality once more, we have $\delta(a, b) \leq \delta(a, y) + \delta(y, b)$, which would imply the contradiction $\delta(a, b) < \delta(a, b)$ if either of the inequalities above were strict. Thus we have the interesting property that for all $a, b \in Y_G$ there is some $y \in Y_G$ that is precisely *halfway* between a and b:

$$\delta(a, y) = \frac{\delta(a,b)}{2} = \delta(y, b).$$

By applying induction, the above implies that for all points $t \in [0, 1]$ that can be written as a integer multiple of 2^{-k} for some $k \in \{1, 2, \ldots\}$ there is some $y_t \in Y_G$ such that

$$\delta(a, y_t) = t\delta(a, b),$$

$$\delta(y_t, b) = (1-t)\delta(a, b).$$

All of the points y_t are contained in the ball of radius $\delta(a, b)$ centered at either a or b. By applying Proposition 2.13 we obtain a path $[0, 1] \to Y_G$ with $0 \mapsto a$ and $1 \mapsto b$, not only showing that Y_G is path connected but all of the balls are path connected, which in turn implies that Y_G is locally path connected.

2. The limit space being proper is one of the things guaranteed by Theorem 3.38, which implies that the space is locally compact. Homogeneity is proved by Proposition 3.39, since a metric induced by a group norm is always homogeneous and this property is directly inherited by each G_n.

3. The ball of radius 1 centered at the origin of Y_G is covered by $\frac{1}{\varepsilon^d}$ balls of radius ε. Thus we can bound its d-dimensional Hausdorff measure, limiting also its dimension:

$$H^d(Y_G) \leq \frac{1}{\varepsilon^d}(2\varepsilon)^d \leq 2^d.$$

By Corollary 2.33 we only need to show that there is a countable, dense subset of Y_G. This is possible because Y_G is a proper space: each closed ball has a countable dense subset because it is compact. The union of the countable subsets of Y_G that are dense in each of the closed balls with rational radii centered at some arbitrary point will also be countable and will be dense in the entire space.

■

From Lemma 2.71 we obtain

Proposition 4.31. *The group of bijective isometries* Iso(Y_G) *is a Lie group with finitely many connected components.*

4.5 Some necessary algebra

We now invoke an important result from the theory of Lie groups that is going to assist us ([Rag72], Theorem 8.29)

Theorem 4.32 (Jordan–Schur Theorem for Lie groups)**.** *Let L be a Lie group with finitely many connected components. There is a positive number $k(L)$ such that every finite subgroup of $H \subset L$ has an abelian subgroup $H' \subset H$ such that $[H : H'] \leq k(L)$.*

Corollary 4.33. *Let L be a Lie group with finitely many connected components. Let G be a finitely generated group such that, for all $n \in \{1, 2, \ldots\}$ there is a group homomorphism $f_n : G \to L$ such that $n \leq \#\mathrm{im} f_n < \infty$. Then there is a subgroup of $G' \subset G$ such that $[G : G'] < \infty$ and $[G' : [G', G']] = \infty$. In particular, G' admits a surjective homomorphism $f : G' \to \mathbb{Z}$.*

Proof. Let $f_n : G \to L$ be the required homomorphism. Since its image is finite for each n, we apply Theorem 4.32 and obtain an abelian subgroup $H_n \subset \mathrm{im} f_n$ with $[\mathrm{im} f_n : H_n] \leq k(L)$. By Proposition 2.65, there is some $G_n \subset G$ such that $[G_n, G_n] \subset \ker f_n \subset G_n$ and $H_n = \frac{G_n}{\ker f_n}$. We also have

$$\# \frac{G}{G_n} = \# \frac{G/\ker f}{G_n/\ker f_n} = \frac{\#\mathrm{im} f_n}{\# H_n} \leq k(L),$$

$$\#[G_n : [G_n, G_n]] \geq [G_n, \ker f_n] = \# H_n = \frac{\#\mathrm{im} f_n}{[\mathrm{im} f_n : H_n]} \geq \frac{n}{k(L)}.$$

Consider the subgroup $G' = \bigcap_{n=1}^{\infty} G_n$. From Proposition 2.63 we see that this is really an intersection of finitely many subgroups of finite index and thus the index of G' in G is also finite. Note that

$$[G : [G', G']] \geq [G_n : [G_n : G_n]] \geq \frac{n}{k(L)}.$$

Since $\frac{n}{k(L)}$ can be made arbitrarily large, we see that the commutator of G' has infinite index in G. But G' has finite index in G and thus $[G : G'][G' : [G', G']] = \infty$, which implies that $[G' : [G', G']] = \infty$.

In conclusion, note that the index of the commutator of G' being infinite is equivalent to its abelianization being infinite in cardinality. From Proposition 2.64 we know that G' is finitely generated, which in turn implies its abelianization also is finitely generated. Thus from Corollary 2.62 we see that there is a surjective homomorphism from the abelianization to $\mathbb{Z}$. ■

Some of the previous results will be combined in this Lemma:

Lemma 4.34. *Let G be a finitely generated group of polynomial growth and L be a Lie group with finitely many connected components. Assume that G has a subgroup of finite index $H \subset G$ satisfying one of the following:*

- *H is abelian.*
- *For all $n \in \{1, 2, \ldots\}$ there is a group homomorphism $f_n : H \to L$ with $\#\mathrm{im} f \geq n$.*

Then G is almost nilpotent.

Proof. We prove by induction on $d = \Gamma(G)$. If $d = 0$, the group is finite and even the trivial group $\{e\} \subset G$ is nilpotent with finite index.

Let $H \subset G$ be the subgroup of finite index satisfying the hypothesis (note that H must also be infinite). One of three cases apply:

- If H is abelian, then due to Proposition 2.61 we know that H is direct sum of finitely many cyclic groups, one of them being isomorphic to $\mathbb{Z}$. The projection $H \to \mathbb{Z}$ is a surjective homomorphism.
- If there are homomorphisms $f_n : H \to L$ with finite but arbitrarily large images, from Corollary 4.33 we obtain $H' \subset H$ with $[H : H'] < \infty$ and a surjective homomorphism $H' \to \mathbb{Z}$. We may replace H with H' for simplicity, since it will also have finite index in G

- If there is a homomorphism $f : H \to L$ with infinite image, then from Lemma 4.17, $\mathrm{im} f$ is an infinite group which has a solvable subgroup of finite index $K \subset \mathrm{im} f$. That is, K has a finite derived series $K^{(0)} \supset K^{(1)} \supset \ldots \supset \{e\}$. Since $K^{(0)} = K$ is infinite, one of the terms of the sequence must have infinite index on K. Replacing $K^{(0)}$ with the last term with finite index (and H with the inverse image under f of that term), we have that $K^{(1)} = [K^{(0)}, K^{(0)}]$ has infinite index. The composition of f with the abelianization map $K^{(0)} \to {}^{K^{(0)}}/_{K^{(1)}}$, followed by the projection map ${}^{K^{(0)}}/_{K^{(1)}} \to \mathbb{Z}$ obtained in Proposition 2.61 is a surjective homomorphism $H \to \mathbb{Z}$.

Thus we have a surjective group homomorphism $g : H \to \mathbb{Z}$, regardless of which case. Since $\mathbb{Z}$ is abelian, we know that $\ker f$ contains the commutator $[H, H]$. Thus

$$[H : [H : H]] \geq [H, \ker f] = \#\mathbb{Z} = \infty.$$

We know from Corollary 4.25 that $\ker g$ is finitely generated. Therefore, Proposition 4.24 tells us that the kernel has polynomial growth rate bounded above by $d-1$. We then apply the inductive hypothesis: $\ker g$ is almost nilpotent and, a fortiori, almost solvable. Thus we have the short exact sequence

$$\{e\} \to \ker g \xrightarrow{f} H \xrightarrow{g} \mathbb{Z} \to \{e\}$$

Proposition 4.28 implies that H is almost solvable. Finally, Theorem 4.16 implies that it must be almost nilpotent. ■

4.6 A Proof of Gromov's Theorem

Let us rehash the ideas from Proposition 4.30. A certain group of polynomial growth G gives rise to a sequence of metric spaces $\{G_n\}_{n=1}^{\infty}$, each being identical to G excepts for having its metric scaled be a factor λ_n, with the sequence $\{\lambda_n\}_{n=1}^{\infty}$ converging to zero. Our understanding of Gromov's convergence confirms that the sequence of spaces converges to (or has a subsequence that converges to) a very nice pointed space (Y_G, y_0), endowed with a metric δ. The theory of Gleason, Montgomery and Zippin then implies that the group of bijective isometries of this space is a Lie group with finitely many connected components $L = \mathrm{Iso}(Y_G)$.

Let us now see this from the perspective of Proposition 3.35, Gromov's Isometry Lemma. Each group G_n acts isometrically on itself, with each $g' \in G_n$ defining an isometry $f_{g',n} \in \mathrm{Iso}(G_n)$, taking some $g \in G_n$ to $g'g$, and we have

$$d_n(g, g'g) = \lambda_n |g^{-1} g' g|.$$

In particular, $d_n(f_{g',n}(e), e) = \lambda_n |g'| \leq |g'|$. Since the limit space Y_G is proper, this matches exactly with the hypothesis of the Isometry Lemma and thus we obtain a subsequence of $\{f_{g',n}\}_{n=1}^{\infty}$ that converges to some $f_{g'} \in L$.

Now let $g_1, g_2 \in G_n$, defining $f_{g_1}, f_{g_2} \in L$. Reminding ourselves of the definition of convergence of functions between pointed metric spaces, we know that for all $\varepsilon > 0$ and $r \geq 0$ there are $\eta > 0$ and $n_0 \in \{1, 2, \ldots\}$ such that, for all $n > n_0$:

$$\delta_n(g, y) < \eta \implies \delta_n(f_{g_1,n}(g), f_{g_1}(y)) < \varepsilon \text{ and } \delta_n(f_{g_2,n}(g), f_{g_2}(y)) < \varepsilon.$$

Where δ_n is a metric on $G_n \sqcup Y_G$ that coincides with both d_n and δ, the points $g \in G_n$ and $y \in Y_G$ are in the balls of radius r centered at the distinguished points of their respective spaces, and $n \geq n_0$. Let $g' \in G_n$ such that $\delta_n(g', f_{g_1}(y)) < \min(\varepsilon, \eta)$. We have the following inequalities:

$$\begin{aligned}
\delta_n(f_{g_2,n}(f_{g_1,n}(g)), f_{g_2}(f_{g_1}(y))) &\leq d_n(f_{g_2,n}(f_{g_1,n}(g)), f_{g_2,n}(g')) + \delta_n(f_{g_2,n}(g'), f_{g_2}(f_{g_1}(y))) \\
&\leq d_n(f_{g_1,n}(g), g') + \delta_n(f_{g_2,n}(g'), f_{g_2}(f_{g_1}(y))) \\
&< d_n(f_{g_1,n}(g), g') + \varepsilon \\
&< \delta_n(f_{g_1,n}(g), f_{g_1}(y)) + \delta_n(f_{g_1}(y), g') + \varepsilon \\
&< 3\varepsilon.
\end{aligned}$$

Thus we see that the sequence of isometries $\{f_{g_2,n} \circ f_{g_1,n}\}_{n=1}^{\infty}$ converges to $f_{g_2} \circ f_{g_1}$. Since G is countable (due to the fact of it being finitely generated), we apply a diagonalization argument and obtain a group homomorphism $F : G \to L$ given by $g' \mapsto f_{g'}$, which is a metric action of G on Y_G. We may write $g'y$ instead of $F(g')(y)$. Let us determine $\ker F$.

Let $g' \in G$ and $r \geq 0$. We define the following value

$$\Delta_r(g') = \sup\{|g^{-1}g'g| \ : \ g \in \overline{B}_r(e)\}.$$

Lemma 4.35. *Let $F : G \to L$ be as above. We have*

$$\ker F = \{g' \in G \ : \ \lim_{n\to\infty} \lambda_n \Delta_{\lambda_n^{-1}}(g') = 0\}.$$

Proof. Let $g' \in G$ and $y \in Y_G$ such that $\delta(y, g'y) > \varepsilon$ for some $\varepsilon > 0$. Let $r > \delta(y, y_0)$. There is some $n_0 \in \{1, 2, \ldots\}$, such that, for all $n \geq n_0$ we may choose $g \in G_n$ with $d_n(g, e) \leq r$, $\delta_n(g, y) \leq \varepsilon/4$ and $\delta_n(g'g, g'y) \leq \varepsilon/4$. Therefore

$$\delta(y, g'y) \leq \delta_n(y, g) + d_n(g, g'g) + \delta_n(g'g, g'y) \leq d_n(g, g'g) + \varepsilon/2$$

$$\begin{aligned} &\implies d_n(g, g'g) = \lambda_n|g^{-1}g'g| \geq \varepsilon/2 \\ &\implies \lambda_n\Delta_r(g') \geq \varepsilon/2 \\ &\implies \lim_{n\to\infty} \lambda_n \Delta_{\lambda_n^{-1}}(g') \neq 0. \end{aligned}$$

Conversely, suppose there is some $\varepsilon > 0$ such that for all $n_0 \in \{1, 2, \ldots\}$ we have $n \geq n_0$ such that $\lambda_n \Delta_{\lambda_n^{-1}}(g') \geq \varepsilon$. Then there is some $g \in G$ with $d_n(g, e) \leq \lambda_n^{-1}$ and $\lambda_n|g^{-1}g'g| \geq \varepsilon$. Take some $y \in Y$ with $\delta(y, y_0) \leq \lambda_n^{-1}$ and $\delta_n(g, y) \leq \varepsilon/4$. We may also assume that $\delta_n(g'y, g'g) < \varepsilon/4$. We obtain

$$d_n(g, g'g) \leq \delta_n(g, y) + \delta(y, g'y) + \delta_n(g'y, g'g) \leq \delta(y, g'y) + \varepsilon/2$$

$$\begin{aligned} &\implies d(y, g'y) \geq \varepsilon/2 \\ &\implies g' \notin \ker F. \end{aligned}$$

■

Consider what it means for the function $\Delta_r(g')$ to be bounded as $r \to \infty$. Let $C : G \to \mathrm{Aut}(G)$ be the conjugation action given by $C(g)(g') = g^{-1}g'g$. The orbit $O_{g'}(G)$ w.r.t. the action C is finite if and only if $\Delta_r(g')$ is bounded. As we have seen in Remark 2.55, this is equivalent to the stabilizer of g' having finite index as a subgroup of G.

Theorem 4.36 (Gromov's Theorem)**.** *All groups of polynomial growth have a nilpotent subgroup of finite index.*

Proof. Let G be a group of polynomial growth. If $L = \mathrm{Iso}(Y_G)$, Lemma 4.34 tells us that we must find some $G' \subset G$, a subgroup of finite index, that is either abelian or endowed with homomorphisms $G' \to L$ with arbitrarily large images. This is quite easy if the image of the group action $F : G \to L$ described above is infinite.

If the image is not infinite (i.e., the kernel of the action has finite index), there are two possibilities: First, suppose that there is some fixed constant $C > 0$ such that $\Delta_r(s) < C$ for all $r > 0$ and with s ranging in some finite generating set $S \subset \ker F$ (the kernel is finitely generated for it has finite index). Then the stabilizer G_s of each s under the conjugation action has finite index, and so does the intersection $H = \bigcap_{s\in S} G_s$. Every element of H commutes with every element of $\ker F$ (see Example 2.57), and so $G' = H \cap \ker F$ is abelian (while having finite index).

The second possibility is more interesting, for we are going to use a different artifice to create arbitrarily large homomorphic images of $\ker F$ on L (that is, we set $G' = \ker F$, which has finite index). Lemma 2.73 says that L as a topological group has the No Small Subgroups property, i.e., there is a open set $U \subset L$ containing the identity that contains no subgroups. We may assume that U is a basic open set: there is some $\varepsilon' > 0$ and a finite collection of points $y_1, \ldots, y_k \in Y$ such that all isometries $\ell \in L$ that satisfy $\delta(y_i, \ell(y_i)) < \varepsilon'$ for all $i \in \{1, \ldots, k\}$ are in U. Let $R > 0$ be the maximum distance of all y_i to the origin y_0.

Let $\ell \subset U$ be an isometry that generates in L a subgroup of order N. Consider what happens if we assume that $\delta(y, \ell(y)) < \varepsilon'/N$: for all $y \in \overline{B}_{R+\varepsilon'}(y_0)$:

$$\delta(y, \ell^p(y)) \leq \delta(y, \ell(y)) + \delta(\ell(y), \ell^2(y)) + \ldots, \delta(\ell^{p-1}(y), \ell^p(y)) < \frac{p\varepsilon'}{N} \leq \varepsilon'$$

We see that each of the isometries in $\{\ell, \ell^2, \dots, \ell^N\} = \langle \ell \rangle$ is still in U, a contradiction since $\langle \ell \rangle$ is a subgroup. Thus, if we are somehow able to construct an isometry $\ell \in L$ with the property that $\delta(y, \ell(y)) < \delta(y_0, y)\frac{\varepsilon'}{(R+\varepsilon')N}$ for all $y \in Y_G$, we would know that $\#\langle \ell \rangle > N$. Let us fix $\varepsilon < \frac{\varepsilon'}{(R+\varepsilon')N}$ and proceed on this direction.

The function $r \mapsto \Delta_r(s)$ is unbounded for $r > 0$ and s ranging in any fixed generating set $S \subset \ker F$. Proposition 4.20 tells us that for every $\rho > 0$ and $g_r \in \overline{B}_{r+\rho}(e)$ there is some $g_\rho \in \overline{B}_\rho(e)$ with $d(g_r, g_\rho) \leq r$. Thus, for all $g' \in G$ we have

$$d(g_r, g'g_r) \leq d(g_r, g_\rho) + d(g_\rho, g'g_\rho) + d(g'g_\rho, g'g_r) \leq \Delta_\rho(g') + 2r$$

$$\implies \Delta_{r+\rho}(g') \leq \Delta_\rho(g') + 2r.$$

Note also that, for arbitrary $\alpha \in G$ and $g \in \overline{B}_r(e)$,

$$d(g, \alpha^{-1}g'\alpha g) = d(\alpha g, g'\alpha g) \leq \Delta_{|\alpha g|}(g')$$

$$\implies \Delta_r(\alpha^{-1}g'\alpha) \leq \Delta_{r+|\alpha|}(g') \leq \Delta_r(g') + 2|\alpha|.$$

Let $\varepsilon > 0$. We are assuming that $\Delta_r(s)$ is unbounded for $r > 0$ and $s \in S$. Lemma 4.35 tells us that, for all $s_0 \in S$ and large enough $n \in \{1, 2, \dots\}$, we have

$$\Delta_{\lambda_n^{-1}}(s_0) < \lambda_n^{-1}\varepsilon.$$

But, from unboundedness, there is some $\tilde{g} \in \overline{B}_R(e)$ for some $R > 0$ and $s_1 \in S$ such that $|\tilde{g}^{-1}s_1\tilde{g}| > \lambda_n^{-1}\varepsilon$. Let $\alpha = \tilde{g}g^{-1}$ for some $g \in \overline{B}_{\lambda_n^{-1}}(e)$ and we obtain

$$\Delta_{\lambda_n^{-1}}(\alpha^{-1}s_1\alpha) \geq |g^{-1}\alpha^{-1}s_1\alpha g| = |(\alpha g)^{-1}s_1(\alpha g)| = |\tilde{g}^{-1}s_1\tilde{g}| > \lambda_n^{-1}\varepsilon.$$

Since $\ker F \subset G$ is normal we have that s_1 and s_0 are conjugate. Thus by replacing α we may assume that $\Delta_{\lambda_n^{-1}}(\alpha^{-1}s_0\alpha) > \lambda_n^{-1}\varepsilon$ Apply the integral connectivity property again, using the alternative definition from Proposition 4.19. There is a sequence

$$e = \alpha_0, \alpha_1, \dots, \alpha_{k-1}, \alpha_k = \alpha,$$

such that $d(\alpha_i, \alpha_j) \leq |i - j|$. Thus

$$|\Delta_{\lambda_n^{-1}}(\alpha_i^{-1}s_0\alpha_i) - \Delta_{\lambda_n^{-1}}(\alpha_j^{-1}s_0\alpha_j)| \leq 2|\alpha_i^{-1}\alpha_j| \leq 2|i - j|.$$

Since $\Delta_{\lambda_n^{-1}}(\alpha_0^{-1}s_0\alpha_0) < \lambda_n^{-1}\varepsilon < \Delta_{\lambda_n^{-1}}(\alpha_k^{-1}s_0\alpha_k)$, there must be some intermediate $\alpha_i = \beta$ such that

$$\Delta_{\lambda_n^{-1}}(\alpha_i^{-1}s_0\alpha_i) < \lambda_n^{-1}\varepsilon < \Delta_{\lambda_n^{-1}}(\alpha_{i+1}^{-1}s_0\alpha_{i+1})$$

$$\implies |\Delta_{\lambda_n^{-1}}(\beta^{-1}s_0\beta) - \lambda_n^{-1}\varepsilon| < 2.$$

Since n is only required to be large enough, for all large enough n we have β_n such that

$$|\Delta_{\lambda_n^{-1}}(\beta_n^{-1}s_0\beta_n) - \lambda_n^{-1}\varepsilon| < 2$$

$$\implies \lim_{n\to\infty} |\lambda_n\Delta_{\lambda_n^{-1}}(\beta_n^{-1}s_0\beta_n) - \varepsilon| = \lim_{n\to\infty} 2\lambda_n = 0$$

$$\implies \lim_{n\to\infty} \lambda_n\Delta_{\lambda_n^{-1}}(\beta_n^{-1}s_0\beta_n) = \varepsilon.$$

Define a family of actions $F_n : \ker F \to \mathrm{Iso}(G_n)$ as

$$F_n(g')(g) = \beta_n^{-1}g'\beta_n g.$$

Note that the distance $d_n(F_n(g')(e), e) = \lambda_n d(e, \beta_n^{-1}g'\beta_n) \leq \lambda_n\Delta_{\lambda_n^{-1}}(\beta_n^{-1}g'\beta_n)$ cannot be much greater than ε as $n \to \infty$, which again satisfies the hypothesis for Proposition 3.35. We obtain a new action $F' : \ker F \to \mathrm{Iso}(Y_G)$ as its limit (passing to subsequences a countable amount of times as needed).

Consider the distance between some $y \in Y_G$ and its image under the action of s_0. We must take a family $g_n \in G_n$ such that $\delta_n(g_n, y) < 1/n$, where δ_n is a metric on $G_n \sqcup Y_G$ that coincides with the metrics of both spaces on their domain. Therefore

$$\begin{aligned}
\delta(y, F'(s_0)(y)) &= \lim_{n\to\infty} d_n(g_n, F_n(s_0)(g_n)) \\
&= \lim_{n\to\infty} \lambda_n d(g_n, \beta_n^{-1} s_0 \beta_n g_n) \\
&\leq \lim_{n\to\infty} \lambda_n \Delta_{|g_n|}(\beta_n^{-1} s_0 \beta_n) \\
&\leq \lim_{n\to\infty} \frac{\lambda_n}{\lambda_m} \lambda_m \Delta_{\lambda_m^{-1}}(\beta_n^{-1} s_0 \beta_n) \\
&= \delta(y_0, y)\varepsilon,
\end{aligned}$$

where above we chose a minimal $m \in \{1, 2, \ldots\}$ with $\lambda_m^{-1} \geq |g_n|$. This is precisely what was needed: the isometry $F'(s_0)$ generates a subgroup of order greater than N on L, where $N \in \{1, 2, \ldots\}$ was taken arbitrarily. Thus, $\#F'(\ker F) \geq \#F'(\langle s_0 \rangle) = \#\langle F'(s_0) \rangle > N$. ■

References

[Alu09] Paolo Aluffi. *Algebra: Chapter 0.* American Mathematical Society, 2009.

[Are46] Richard Arens. Topologies for homeomorphism groups. *American Journal of Mathematics*, 68(4):593, October 1946.

[BBI22] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry.* American Mathematical Society, Providence, RI, January 2022.

[BO08] Nicholas Bingham and A. Ostaszewski. Normed versus topological groups: Dichotomy and duality. *Dissertationes Mathematicae*, 472, 01 2008.

[CFL15] Francesca Cagliari, Barbara Di Fabio, and Claudia Landi. The natural pseudo-distance as a quotient pseudo-metric, and applications. *Forum Mathematicum*, 27(3):1729–1742, 2015.

[dlH00] Pierre de la Harpe. *Topics in geometric group theory.* University Of Chicago Press, 2000.

[dlH21] Pierre de la Harpe. On the prehistory of growth of groups, 2021. Available on `arxiv.org`.

[Edw75] David A. Edwards. *The Structure of Superspace*, page 121–133. Elsevier, 1975.

[Gri85] R I Grigorchuk. Degrees of growth of finitely generated groups, and the theory of invariant means. *Mathematics of the USSR-Izvestiya*, 25(2):259–300, April 1985.

[Gro81] Mikhael Gromov. Groups of polynomial growth and expanding maps. *Publications mathématiques de l'IHÉS*, 53(1):53–78, December 1981.

[Lee11] John M. Lee. *Introduction to Topological Manifolds.* Springer New York, 2011.

[Lee12] John M. Lee. *Introduction to Smooth Manifolds.* Springer New York, 2012.

[Mil68] J. Milnor. A note on curvature and fundamental group. *Journal of Differential Geometry*, 2(1), January 1968.

[MZ55] Deane Montgomery and Leo Zippin. *Topological Transformation Groups.* Interscience Publishers, University of California, 1955.

[Rag72] M S Raghunathan. *Discrete Subgroups of Lie Groups.* Springer, 1972.

[Sch55] Albert S. Schwarz. A volume invariant of coverings. *Proceedings of the USSR Academy of Sciences*, 105:32–34, 1955.

[Tao11] Terence Tao. 254a, notes 5: The structure of locally compact groups, and hilbert's fifth problem, Oct 2011. Available on `terrytao.wordpress.com`.

[Tit72] J Tits. Free subgroups in linear groups. *Journal of Algebra*, 20(2):250–270, February 1972.

[vdDW84] L van den Dries and A.J Wilkie. Gromov's theorem on groups of polynomial growth and elementary logic. *Journal of Algebra*, 89(2):349–374, August 1984.

[Wol68] Joseph A. Wolf. Growth of finitely generated solvable groups and curvature of riemannian manifolds. *Journal of Differential Geometry*, 2(4), January 1968.