Mathematical
Sciences

University of
Southampton

MATH6138 Geometric Group Theory
Lecture notes

Autumn 2025-2026

Ashot Minasyan

Contents

CHAPTER 0

Notation and background

In this chapter we will cover the notation and basic notions from Group Theory, most of which you have already seen in MATH2003.

0.1. Notation

Let us start by establishing some notation that will be used throughout these notes:

- $\mathbb{N} = \{1, 2, 3, \dots, \}$ will denote the set of all natural numbers;
- $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$;
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ will be used to denote the set of all integers and also the abelian group $(\mathbb{Z}, +)$ of integers under addition;
- $\mathbb{Q} = \{m/n \mid m \in \mathbb{Z}, n \in \mathbb{N}\}$ will denote the set of all rational numbers (and also the corresponding group under addition);
- $\mathbb{R}$ will denote the set of all real numbers (and also the corresponding group under addition);
- $\mathbb{Z}_n = \{\overline{0}, \overline{1}, \dots, \overline{n-1}\}$, where $n \in \mathbb{N}$, will denote the group of residues modulo (under addition modulo n);
- C_n, where $n \in \mathbb{N}$, will denote the cyclic group of order n under multiplication; if we denote by $c \in C_n$ a generator of this group then $C_n = \{1, c, c^2, \dots, c^{n-1}\}$;
- C_∞ will denote the infinite cyclic group; if we let $c \in C_n$ be a generator of this group then $C_\infty = \{c^k \mid k \in \mathbb{Z}\}$;
- if X is a set then $|X|$ will denote its cardinality (the number of elements in the case when X is finite);
- if x and y are elements of a group G, we will write $[x, y] = xyx^{-1}y^{-1} \in G$ for the *commutator of x and y* in G.

In these notes, given a group G we will assume that the operation of G is multiplication (not necessarily commutative), unless specified otherwise. We will denote by 1 the identity element of the group and by $g^{-1} \in G$ the inverse of an element $g \in G$. Sometimes, to distinguish the identity elements of different groups, we will write 1_G for the identity element of a group G.

If H is a subgroup of a group G, we will write $H \leqslant G$. Recall that a group G is said to be *abelian* if the product on it is commutative, i.e., $xy = yx$, for all $x, y \in G$. On some abelian groups (such as $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{Z}_n$) we will use the additive notation $a + b$ to denote the group operation.

If G, H are groups then their *direct product* $G \times H$ is also a group and the multiplication in this group is performed coordinate-wise:

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2), \quad \text{for all } (g_1, h_1), (g_2, h_2) \in G \times H.$$

Given any $n \in \mathbb{N}$, the *free abelian group of rank n*, $\mathbb{Z}^n = \mathbb{Z} \times \cdots \times \mathbb{Z}$, is defined as the n-fold direct product of $\mathbb{Z}$ with itself. Thus $\mathbb{Z}^n$ consists of all n-tuples of integers with coordinate-wise addition. We can similarly define the groups $\mathbb{Q}^n$ and $\mathbb{R}^n$.

0.2. Group morphisms

Definition 0.2.1. Let G, H be groups. A function $\phi : G \to H$ is called a *homomorphism* if $\phi(xy) = \phi(x)\phi(y)$, for all $x, y \in G$.

Exercise 0.2.2. Suppose that $\phi : G \to H$ is a group homomorphism. Prove that

(i) $\phi(1_G) = 1_H$;
(ii) $\phi(g^{-1}) = \phi(g)^{-1}$ and $\phi(g^n) = \phi(g)^n$, for all $g \in G$ and all $n \in \mathbb{Z}$;

(iii) if g has finite order $k \in \mathbb{N}$ in G then $\phi(g)$ has finite order $l \in \mathbb{N}$ in H and l divides k.

Definition 0.2.3. A group homomorphism $\phi : G \to H$ is called

- an *epimorphism* if ϕ is surjective;
- a *monomorphism* if ϕ is injective;
- an *isomorphism* if ϕ is bijective.

In the case when there is an isomorphism between the groups G and H, we will say that these groups are *isomorphic*, writing $G \cong H$.

A homomorphism $\alpha : G \to G$ is called an *endomorphism*. An *automorphism* of G is an isomorphism $G \to G$. We will use $\mathrm{Aut}(G)$ to denote the set of all automorphisms of G.

Example 0.2.4. Let G and H be some groups.

(a) The *trivial homomorphism* $\tau : G \to H$ sends every element of G to 1_H.
(b) The *identity automorphism* $\mathrm{Id} : G \to G$ is the identity map, defined by $\mathrm{Id}(g) = g$, for all $g \in G$.
(c) Any element $g \in G$ gives rise to the *inner automorphism* $\psi_g \in \mathrm{Aut}(G)$, defined by $\psi_g(x) = gxg^{-1}$, for all $x \in G$. (Note that $(\psi_g)^{-1} = \psi_{g^{-1}}$.)

Exercise 0.2.5. Show that $C_\infty \cong \mathbb{Z}$ and $C_n \cong \mathbb{Z}_n$, for all $n \in \mathbb{N}$.

Exercise 0.2.6. Let $n \in \mathbb{N}$ and $\phi : \mathbb{Z} \to \mathbb{Z}_n$ be the function sending every integer $m \in \mathbb{Z}$ to the remainder after dividing m by n. Prove that ϕ is an epimorphism. Is ϕ a monomorphism?

Exercise 0.2.7. Suppose that $\phi : G \to H$ is a homomorphism between groups G and H. Prove that

(i) for every subgroup $K \leqslant G$ its *image* $\phi(K) = \{\phi(x) \mid x \in K\}$ is a subgroup of H;
(ii) for every subgroup $L \leqslant H$ its *full preimage* $\phi^{-1}(L) = \{g \in G \mid \phi(g) \in L\}$ is a subgroup of G.

Exercise 0.2.8. Let $\phi : \mathbb{Z} \to \mathbb{Z}_9$ be the epimorphism from Exercise 0.2.6. Find the image of the subgroup $2\mathbb{Z}$, of even integers in $\mathbb{Z}$, under ϕ. What is the full preimage of the subgroup $\{\overline{0}, \overline{3}, \overline{6}\} \leqslant \mathbb{Z}_9$ under ϕ?

0.3. Normal subgroups and quotients

Definition 0.3.1. Let H be a subgroup of a group G. For every $g \in G$ the

- the subset $gH = \{gh \mid h \in H\} \subseteq G$ is a *left coset modulo H*;
- the subset $Hg = \{hg \mid h \in H\} \subseteq G$ is a *right coset modulo H*.

Remark 0.3.2. If G is a group and $g \in G$ is a fixed element then the map $x \mapsto gx$, for all $x \in G$, defines a bijection $G \to G$ (indeed, its inverse is the map $x \mapsto g^{-1}x$). For any subgroup $H \leqslant G$ the image of H under this bijection is gH, which implies that $|gH| = |H|$, for all $g \in G$. Similarly, the map $x \mapsto xg$ is also a bijection $G \to G$, whence $|Hg| = |H|$, for all $g \in G$.

Exercise 0.3.3. If $H \leqslant G$ is a subgroup and $f, g \in G$ are elements then either $fH = gH$ or $fH \cap gH = \emptyset$. A similar statement is also true for the right cosets Hf and Hg (either they coincide as subsets of G are they are disjoint).

This exercise implies that every subgroup H of a group G defines a partition of G into the disjoint union of left (right) cosets modulo H. We denote by G/H the set of all left cosets modulo H in G, and by $H\backslash G$ – the set of all right cosets. If $gH \in G/H$, any element $g' \in gH$ is called a *coset representative* of gH. Exercise 0.3.3 easily implies that if g and g' are representative of the same coset modulo H then $gH = g'H$.

Definition 0.3.4. Given a group G and a subgroup $H \leqslant G$, the cardinality of the set of left cosets G/H is called the *index of H in G*, denoted $|G : H|$.

Exercise 0.3.5. Show that if $H \leqslant G$ then the sets G/H and $H\backslash G$ have the same cardinality. Thus the index $|G:H|$ is also the number of right cosets modulo H in G. [*Hint:* check that the map $G \to G$, $g \mapsto g^{-1}$ induces a bijection between G/H and $H\backslash G$.]

Example 0.3.6. Take any $n \in \mathbb{N}$ and let $n\mathbb{Z}$ denote the subgroup of $\mathbb{Z}$ consisting of integers divisible by n. Then $\mathbb{Z}$ contains exactly n cosets modulo $n\mathbb{Z}$, i.e., $|\mathbb{Z} : n\mathbb{Z}| = n$. Each of these cosets has the form $m + n\mathbb{Z} = \{m + nk \mid k \in \mathbb{Z}\}$, for some $m \in \{0, 1, \dots, n-1\}$. Thus, the coset $m + n\mathbb{Z}$ consists of all integers with remainder m when divided by n (in other words, all integers congruent to m modulo n).

After combining Remark 0.3.2 with Exercise 0.3.3 we obtain the following well-known statement.

Theorem 0.3.7 (Lagrange's theorem). *Let G be a finite group with a subgroup H. Then we have $|G:H| = |G|/|H|$, thus both $|H|$ and $|G:H|$ divide $|G|$.*

Definition 0.3.8. A subgroup N of a group G is *normal* if $gN = Ng$ (equivalently, $gNg^{-1} = N$), for all $g \in G$. We will write $N \lhd G$ to indicate that N is a normal subgroup of G.

Example 0.3.9. (a) Any subgroup of an abelian group is normal.

(b) If $\mathbb{K}$ is a field and $n \in \mathbb{N}$, the group of all invertible $n \times n$ matrices over $\mathbb{K}$ is the *general linear group* $\mathrm{GL}(n,\mathbb{K})$. Check that the *special linear group*

$$\mathrm{SL}(n,\mathbb{K}) = \{A \in \mathrm{GL}(n,\mathbb{K}) \mid \det(A) = 1\}$$

is a normal subgroup of $\mathrm{GL}(n,\mathbb{K})$.

Definition 0.3.10. If G is a group and $N \lhd G$ we can define the *quotient group* G/H as the set of left cosets modulo H under the operation

$$fH \cdot gH = (fg)H, \qquad \text{for all } fH, gH \in G/H.$$

Exercise 0.3.11. Check that the above product on G/H is well-defined (i.e., it does not depend on the choice of cosets representatives $f \in fH$ and $g \in gH$) because N is normal in G. Show that G/H, equipped with this product, is a group. The identity element of this quotient group is the left coset $H = 1_G H$, and for every $gH \in G/H$ its inverse is $g^{-1}H$.

Example 0.3.12. If you check the definition of the group $\mathbb{Z}_n$ (for some $n \in \mathbb{N}$) then you will see that it is defined precisely as the quotient of the group of integers $\mathbb{Z}$ by the normal subgroup $n\mathbb{Z}$.

0.4. Isomorphism theorems

Definition 0.4.1. Let $\phi : G \to H$ be a group homomorphism. The *kernel* of ϕ is defined as

$$\ker\phi = \{g \in G \mid \phi(g) = 1_H\} \subseteq G.$$

The *image* of ϕ is the subset of H defined by

$$\operatorname{im}\phi = \phi(G) = \{\phi(g) \mid g \in G\} \subseteq H.$$

Exercise 0.4.2. If $\phi : G \to H$ is a group homomorphism then $\ker\phi$ is a normal subgroup of G and $\operatorname{im}\phi$ is a subgroup of H.

Exercise 0.4.3. Show that a group homomorphism $\phi : G \to H$ is a monomorphism (i.e., it is injective) if and only if $\ker\phi = \{1_G\}$. And ϕ is an epimorphism (it is surjective) if and only if $\operatorname{im}\phi = H$.

Exercise 0.4.4. Given a group G and a normal subgroup $N \lhd G$, check that the map $\phi : G \to G/N$, defined by $\phi(g) = gN$, for all $g \in G$, is a homomorphism such that $\ker\phi = N$ and $\operatorname{im}\phi = G/N$.

Definition 0.4.5. The homomorphism $\phi : G \to G/N$ defined in Exercise 0.4.4 will be called the *natural homomorphism* from G to its quotient group G/N.

Theorem 0.4.6 (First Isomorphism Theorem). *Let $\phi : G \to H$ be a homomorphism between groups G and H, and let $N = \ker\phi \lhd G$. Then $G/N \cong \operatorname{im}\phi$. More precisely, this isomorphism is defined by the map*

$$\psi : G/N \to \operatorname{im}\phi, \quad \psi(gN) = \phi(g), \text{ for all } g \in G.$$

Proof. Exercise (or see a textbook). □

The First Isomorphism Theorem is very useful for identifying quotients of groups. For example, we can apply it in the following.

Exercise 0.4.7. If $\mathbb{K}$ is a field then the quotient $\mathrm{GL}(n,\mathbb{K})/\mathrm{SL}(n,\mathbb{K})$ is naturally isomorphic to the multiplicative group of this field $\mathbb{K}^* = \mathbb{K} \setminus \{0\}$.

The following statement is a direct consequence of Theorem 0.4.6 and Theorem 0.3.7:

Corollary 0.4.8. *If G is a finite group and $\phi : G \to H$ is a group homomorphism then $|\operatorname{im}\phi|$ divides $|G|$.*

Example 0.4.9. Let $m, n \in \mathbb{N}$ be coprime integers and let G, H be groups such that $|G| = m$ and $|H| = n$. Suppose that $\phi : G \to H$ is a homomorphism. By Corollary 0.4.8, the order $|\operatorname{im}\phi|$ must divide m. On the other hand, the order of any subgroup of H divides n, by Lagrange's Theorem. Therefore the order of $\operatorname{im}\phi$ must divide $\gcd(m,n) = 1$, hence $|\operatorname{im}\phi| = 1$, whence ϕ is the trivial homomorphism. Thus the only homomorphism from G to H is the trivial one.

The First Isomorphism Theorem is also helpful for finding images of subgroups under homomorphisms.

Corollary 0.4.10. *Suppose that $\phi : G \to H$ is a group homomorphism and $F \leqslant G$. Then the image $\phi(F)$, of F under ϕ, is a subgroup of H and $\phi(F) \cong F/(F \cap \ker\phi)$.*

Proof. Denote by $\psi : F \to H$ the restriction of ϕ to F. Then ψ is obviously a group homomorphism, $\ker\psi = F \cap \ker\phi$ and $\operatorname{im}\psi = \phi(F)$. Therefore, $\phi(F)$ is a subgroup of H and $\phi(F) = \operatorname{im}\psi \cong F/(F \cap \ker\phi)$ by Theorem 0.4.6. □

Theorem 0.4.11 (Second Isomorphism Theorem). *Let N be a normal subgroup of a group G and let F be any subgroup of G. Then $FN = \{fx \mid f \in F,\ x \in N\}$ is a subgroup of G, $N \lhd FN$, $F \cap N \lhd F$ and we have an isomorphism $\psi : F/(F\cap N) \to FN/N$, defined by*

$$\psi : f(F \cap N) \mapsto fN, \quad \text{for all } f \in F.$$

Proof. Exercise (or see a textbook). □

The Third Isomorphism Theorem will be used quite frequently in these notes, so we will state three slightly different versions of it.

Theorem 0.4.12 (Third Isomorphism Theorem, version 1). *Let M and N be normal subgroups of a group G, such that $N \subseteq M$. Then $K = M/N$ is a normal subgroup of $H = G/N$ and $H/K \cong G/M$.*

Proof. Exercise (or see a textbook). □

Theorem 0.4.13 (Third Isomorphism Theorem, version 2). *Let $\phi : G \to H$ be a group epimorphism and let $L \lhd H$ be a normal subgroup. Then $M = \phi^{-1}(L)$ is a normal subgroup of G containing $\ker\phi$ and $H/L \cong G/M$.*

Proof. This follows almost immediately from Theorem 0.4.12, once we recall that $G/\ker\phi \cong H$ by Theorem 0.4.6. □

Theorem 0.4.14 (Third Isomorphism Theorem, version 3). *Suppose that G, H, K are groups, $\phi : G \to H$ is an epimorphism and $\psi : G \to K$ is a homomorphism. If $\ker\phi \subseteq \ker\psi$ then ψ factors through ϕ, that is there is a homomorphism $\eta : H \to K$ such that $\psi = \eta \circ \phi$. In other words, the following diagram commutes:*

$$\begin{tikzcd} & G \arrow[dl, "\phi"'] \arrow[dr, "\psi"] & \\ H \arrow[rr, dashed, "\eta"] & & K \end{tikzcd}$$

Moreover, if ψ is surjective then so is η.

Proof. First we define $\eta : H \to K$ as follows. For every $h \in H$ there exists $g \in G$ such that $h = \phi(g)$ because ϕ is surjective, so we set

$$\eta(h) = \psi(g). \tag{0.1}$$

It is important to check that this formula for $\eta(h)$ is well-defined, i.e., it does not depend on the choice of $g \in G$ satisfying $\phi(g) = h$. So, suppose that $g' \in G$ is another element such that $\phi(g) = h$, then $\phi(g) = \phi(g')$, hence $\phi(g^{-1}g') = 1_H$, so $g^{-1}g' \in \ker\phi$. But since $\ker\phi \subseteq \ker\psi$, by the assumptions, we obtain $g^{-1}g' \in \ker\psi$, which yields that $\psi(g) = \psi(g')$, as required.

If $h, h' \in H$ are two elements, then there are $g, g' \in G$ such that $\phi(g) = h$ and $\phi(g') = h'$, hence $hh' = \phi(gg')$, and, by the definition (0.1), we have $\eta(h) = \psi(g)$, $\eta(h') = \psi(g')$, so

$$\eta(hh') = \psi(gg') = \psi(g)\psi(g') = \eta(h)\eta(h').$$

Thus $\eta : H \to K$ is a homomorphism.

The definition of η immediately implies that $\eta(\phi(g)) = \psi(g)$, for all $g \in G$, hence $\psi = \eta \circ \phi$. Finally, if ψ is surjective then for every $x \in K$ there is $g \in G$ such that $x = \psi(g) = \eta(\phi(g))$, hence η is also surjective and the theorem is proved. □

0.5. Group generators

Definition 0.5.1. Let G be a group and let $X \subseteq G$ be a subset. The *subgroup generated by X*, denoted $\langle X\rangle$, is the smallest subgroup of G containing X. In other words,

$$\langle X\rangle = \bigcap\{H \mid H \leqslant G \text{ and } X \subseteq H\}. \tag{0.2}$$

Example 0.5.2. Let G be a group.

(a) Formula (0.2) implies that $\langle\emptyset\rangle = \{1\}$ is the trivial subgroup of G.
(b) Evidently, $\langle G\rangle = G$.
(c) If $x \in G$ then $\langle x\rangle$ is called the *cyclic subgroup of G generated by x*.

Exercise 0.5.3. Suppose that X is a non-empty subset of a group G. Show that $\langle X\rangle$ consists of all products of the form

$$x_1^{\varepsilon_1}x_2^{\varepsilon_2}\cdots x_k^{\varepsilon_k}, \quad \text{for arbitrary } k \in \mathbb{N}_0,\ x_1,\dots,x_k \in X \text{ and } \varepsilon_1,\dots,\varepsilon_k \in \{\pm 1\}.$$

In particular, if $x \in G$ then $\langle x\rangle = \{x^n \mid n \in \mathbb{Z}\}$ consists of all powers of x.

Definition 0.5.4. A subset X of a group G is called a *generating set* of G if $G = \langle X\rangle$. This means that every element of G can be written as a product of elements from $X^{\pm 1}$. In this case we say that G is *generated by X*.

Given any $n \in \mathbb{N}_0$, a group is said to be *n-generated* if it can be generated by a subset with at most n elements. A group is *finitely generated* if it has a finite generating set, i.e., if it is n-generated, for some $n \in \mathbb{N}_0$.

Example 0.5.5. (a) The only 0-generated group is the trivial group.
(b) A group is cyclic if and only if it is 1-generated.
(c) If p is a prime and $n \in \mathbb{N}$ then the group $(\mathbb{Z}_p)^n$ is n-generated but not $(n-1)$-generated. Indeed, we can view $(\mathbb{Z}_p)^n$ as an n-dimensional vector space over the field $\mathbb{Z}_p$. One can then observe that a subset $X \subseteq (\mathbb{Z}_p)^n$ generates this group if and only if it spans the vector space. And from Linear Algebra we know that any minimal spanning set of an n-dimensional vector space is a basis and has exactly n vectors.

(d) For any $n \geq 2$, the symmetric group S_n, on the set $\{1, \ldots, n\}$, is generated by the set of transpositions $\{(i\ i+1) \mid i = 1, \ldots, n-1\}$. This can be proved by induction on n, using the fact that S_n is generated by S_{n-1} (where we view S_{n-1} as the subgroup of S_n fixing n) and $(n-1\ n)$.

In fact, $S_n = \langle \sigma, \tau \rangle$ is 2-generated, where $\sigma = (1\ 2)$ and $\tau = (1\ 2\ \ldots\ n)$ (because $\tau^{i-1}\sigma\tau^{1-i} = (i\ i+1)$).

(e) The group $\mathbb{R}$, of real numbers under addition, is not finitely generated. Indeed, $\mathbb{R}$ is uncountable, but any finitely generated group is countable by Exercise 0.5.3.

Remark 0.5.6. Suppose that G is a group and X is a set, equipped with a set map $f : X \to G$. If G is generated by $f(X)$ then we will sometimes abuse the notation and simply write that G is generated by X (even though X need not be a subset of G in general).

Exercise 0.5.7. For each $n \geq 2$ give an example of a 2-generated group G containing a subgroup that is not n-generated.

In contrast to the above, 1-generated groups behave better.

Proposition 0.5.8. *Every subgroup of a cyclic group is cyclic.*

PROOF. Suppose that $G = \langle g \rangle$ is a cyclic group, generated by an element $g \in G$, and $H \leqslant G$ is a non-trivial subgroup. Let $n \in \mathbb{N}$ be the smallest natural number such that $g^n \in H$. For any $h \in H$ there exists $m \in \mathbb{Z}$ such that $h = g^m$ because G is generated by g. We can then divide m by n:

$$m = qn + r, \text{ where } q \in \mathbb{Z},\ r \in \mathbb{N}_0 \text{ and } r < n.$$

Since H is a subgroup, we have $g^r = g^m(g^n)^{-q} \in H$, which implies that $r = 0$ by the choice of n. Therefore $m = qn$, so $h = (g^n)^q \in \langle g^n \rangle$. The latter shows that $H \subseteq \langle g^n \rangle$. But the opposite inclusion also holds because $g^n \in H$ and H is a subgroup containing g^n. Hence $H = \langle g^n \rangle$ is cyclic, as required. □

Exercise 0.5.9. Consider the group of all rational numbers $\mathbb{Q}$ under addition.

(i) Prove that every finitely generated subgroup of $\mathbb{Q}$ is cyclic. [*Hint:* given rational numbers $m_1/n_1, \ldots, m_k/n_k \in \mathbb{Q}$ find $n \in \mathbb{N}$ such that $m_i/n_i \in \langle 1/n \rangle$, for all $i = 1, \ldots, k$.]

(ii) Show that $\mathbb{Q}$ is not finitely generated.

In fact, as we will later show (see Proposition 1.3.3), a 2-generated group can contain a subgroup that is not finitely generated.

Exercise 0.5.10. Suppose that G is a group with a finite index subgroup $H \leqslant G$. Show that if H is finitely generated then so is G.

The converse of Exercise 0.5.10 is also true: see Corollary 3.8.12.

The next statement lists important properties of generating sets that will be used throughout this course.

Lemma 0.5.11. *Let G be a group generated by a subset X.*

(i) *For any group H and any homomorphism $\phi : G \to H$, ϕ is uniquely determined by its restriction to X. In other words, for every set map $f : X \to H$ there exists at most one group homomorphism $\phi : G \to H$ extending f (so that $\phi(x) = f(x)$, for all $x \in X$).*

(ii) *If $\psi : F \to G$ is a group homomorphism such that $X \subseteq \operatorname{im}\psi$ then $\operatorname{im}\psi = G$, i.e., ψ is surjective.*

(iii) *An element $g \in G$ is* central in G *(i.e., $ga = ag$, for all $a \in G$) if and only if g commutes with each element $x \in X$ in G.*

(iv) *A subgroup $N \leqslant G$ is normal in G if and only if $xNx^{-1} = N$ for all $x \in X$.*

(v) *If $\phi : G \to H$ is a group epimorphism then $\phi(X)$ is a generating set of H.*

PROOF. (i) Suppose that $\xi : G \to H$ is another homomorphism with the same restriction to X as ϕ, i.e., $\xi(x) = \phi(x)$, for all $x \in X$. By Exercise 0.5.3, for any $g \in G$ there exist $x_1, \ldots, x_k \in X$ and $\varepsilon_1, \ldots, \varepsilon_k \in \{\pm 1\}$ such that $g = x_1^{\varepsilon_1} \cdots x_k^{\varepsilon_k}$. Therefore,

$$\xi(g) = \xi(x_1^{\varepsilon_1} \cdots x_k^{\varepsilon_k}) = \xi(x_1)^{\varepsilon_1} \cdots \xi(x_k)^{\varepsilon_k} = \phi(x_1)^{\varepsilon_1} \cdots \phi(x_k)^{\varepsilon_k} = \phi(g).$$

Thus $\xi(g) = \phi(g)$, for all $g \in G$.

The proofs of claims (ii) and (iii) are left as exercises.

(iv) Recall that N is normal in G if $gNg^{-1} = N$, for each $g \in G$. So, assume that $xNx^{-1} = N$ for all $x \in X$. After conjugating both sides by x^{-1}, we see that $x^{-1}Nx = N$, for all $x \in X$. If $g \in G$ then $g = x_1^{\varepsilon_1} \cdots x_k^{\varepsilon_k}$, for some $x_1, \dots, x_k \in X$ and $\varepsilon_1, \dots, \varepsilon_k \in \{\pm 1\}$, which gives

$$gNg^{-1} = x_1^{\varepsilon_1} \cdots x_k^{\varepsilon_k} N x_k^{-\varepsilon_k} \cdots x_1^{-\varepsilon_1} = x_1^{\varepsilon_1} \cdots x_{k-1}^{\varepsilon_{k-1}} N x_{k-1}^{-\varepsilon_{k-1}} \cdots x_1^{-\varepsilon_1} = \cdots = x_1^{\varepsilon_1} N x_1^{-\varepsilon_1} = N.$$

Hence $N \lhd G$.

(v) Since ϕ is onto, for any $h \in H$ there exists $g \in G$ such that $h = \phi(g)$. Now, $g = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$, for some $x_1, \dots, x_n \in X$ and $\varepsilon_1, \dots, \varepsilon_n \in \{\pm 1\}$, because $G = \langle X \rangle$. It follows that

$$h = \phi(g) = \phi(x_1)^{\varepsilon_1} \dots \phi(x_n)^{\varepsilon_n} \quad \text{in } H.$$

Thus $H = \langle \phi(X) \rangle$, by Exercise 0.5.3. ☐

Fact 0.5.12. The converse of claim (i) in Lemma 0.5.11 is also true: if G is a group and $X \subseteq G$ is a subset such that for every group H, each homomorphism $\phi : G \to H$ is uniquely determined by its restriction to X then $G = \langle X \rangle$. However, proving this requires some knowledge of free constructions over groups that we do not have yet.

Exercise 0.5.13. Suppose that G is generated by a subset X such that $|X| = k \in \mathbb{N}_0$. Prove that for any finite group M, G admits at most $|M|^k$ distinct homomorphisms to M.

Example 0.5.14. Consider the group $\mathrm{SL}(2, \mathbb{Z}) = \{A \in \mathrm{M}(2, \mathbb{Z}) \mid \det A = 1\}$. Let us show that this group is generated by the matrices

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Note that $B^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I_2$, so $B^4 = I_2$. On the other hand, one can easily check that $A^m = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$, for all $m \in \mathbb{Z}$.

Set $H = \langle A, B \rangle \leqslant \mathrm{SL}(2, \mathbb{Z})$. To show that $\mathrm{SL}(2, \mathbb{Z}) = H$ it is enough to find for any matrix $M \in \mathrm{SL}(2, \mathbb{Z})$ an element $h \in H$ such that $hM \in H$ (then $M \in h^{-1}H = H$).

Consider any matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$, where $a, b, c, d \in \mathbb{Z}$ and $\det M = ad - bc = 1$. Let us prove that $M \in H$ by induction on $\min\{|a|, |c|\}$. To establish the base of induction, assume that $\min\{|a|, |c|\} = 0$.

Case 1.1. Suppose that $c = 0$ then $ad = 1$, so $a = d = \pm 1$. If $a = d = 1$ then $M = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = A^b \in H$, and if $a = d = -1$ then $M = \begin{pmatrix} -1 & b \\ 0 & -1 \end{pmatrix}$, so that $M = -A^{-b} = B^2 A^{-b} \in H$.

Case 1.2. Suppose that $a = 0$. Then

$$B\,M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & b \\ c & d \end{pmatrix} = \begin{pmatrix} -c & -d \\ 0 & b \end{pmatrix},$$

whence $B\,M \in H$ by Case 1.1.

Thus the base of induction has been established, so we can further assume that $|a| > 0$ and $|c| > 0$.

Case 2.1. Suppose that $|a| \geq |c|$, then we can divide a by c:

$$a = qc + r, \quad \text{where } q, r \in \mathbb{Z} \text{ and } 0 \leq r < |c|.$$

It follows that $a - qc = r$, whence

$$A^{-q}\,M = \begin{pmatrix} 1 & -q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} r & b - qd \\ c & d \end{pmatrix},$$

and $\min\{|r|,|c|\} = |r| < \min\{|a|,|c|\}$, so $A^{-q}\, M \in H$ by induction.

Case 2.2. Suppose that $|c| < |a|$. Then $B\, M = \begin{pmatrix} -c & -d \\ a & b \end{pmatrix}$, so $B\, M \in H$ by Case 2.1.

Therefore, we have shown that $M \in H$, for every $M \in \mathrm{SL}(2,\mathbb{Z})$, which means that $\mathrm{SL}(2,\mathbb{Z}) = H$ is generated by A and B.

CHAPTER 1

Free groups

This chapter introduces and studies basic properties of free groups that will play a central role in this course.

1.1. The Universal Property

There are many ways to define free groups, and different definitions are more convenient for different purposes. We will start with the abstract definition via the *Universal Property*, which is essentially a strengthening of claim (i) of Lemma 0.5.11.

Definition 1.1.1 (Free groups via the Universal Property)**.** Let F be a group and let X be a subset of F. Then F is said to be a *free group on* X if for any group H and any set map $\phi : X \to H$ there exists a unique group homomorphism $\hat{\phi} : F \to H$ extending ϕ (so that $\hat{\phi}(x) = \phi(x)$, for all $x \in X$). In particular, we have the following commutative diagram, where $i : X \to F$ denotes the inclusion of X in F:

$$\begin{tikzcd} X \arrow[r, hook, "i"] \arrow[dr, "\phi"'] & F \arrow[d, dashed, "\hat{\phi}"] \\ & H \end{tikzcd}$$

In this case we say that the group F is *free* and X is a *free generating set* (or a *free basis*) of F.

Example 1.1.2. (a) The trivial group $\{1\}$ is free on the empty subset $\emptyset$.
(b) The infinite cyclic group C_∞, generated by an element $x \in C$, is free on $\{x\}$. This is less obvious, so let's prove it.

Suppose that H is an arbitrary group and $\phi : \{x\} \to H$ is any set map. Denote $h = \phi(x) \in H$ and define $\hat{\phi} : C_\infty \to H$ by the formula

$$\hat{\phi}(x^n) = h^n, \text{ for all } n \in \mathbb{Z}.$$

Here we used the assumption that $C_\infty = \langle x \rangle$, so every element is a power of x. If $x^n = x^m$ in C_∞ then $m = n$ (by definition of the infinite cyclic group different powers of x represent different elements), hence $\hat{\phi}(x^n) = \hat{\phi}(x^m)$, so the map $\hat{\phi}$ is well-defined. Moreover, $\hat{\phi}$ is a group homomorphism because for any $x^n, x^m \in C_\infty$ we have

$$\hat{\phi}(x^n\, x^m) = \hat{\phi}(x^{n+m}) = h^{n+m} = h^n\, h^m = \hat{\phi}(x^n)\, \hat{\phi}(x^m) \text{ in } H.$$

Finally, $\hat{\phi}$ is the unique group homomorphism from C_∞ to H extending ϕ because C_∞ is generated by $\{x\}$ (see Lemma 0.5.11.(i)). Thus we have checked that $C_\infty = \langle x \rangle$ is a free group on $\{x\}$.

Lemma 1.1.3. *If F is a free group on a subset $X \subseteq F$ then X generates F.*

Proof. Let $H = \langle X \rangle \leqslant F$. Clearly X is a subset of H, so we can let $\phi : X \to H$ denote the inclusion map (i.e., $\phi(x) = x$, for all $x \in X$). By the Universal Property of F (see Definition 1.1.1) there exists a unique group homomorphism $\hat{\phi} : F \to H$ such that $\hat{\phi}(x) = \phi(x) = x$, for all $x \in X$. Since $H \leqslant F$ we can also treat $\hat{\phi}$ as a homomorphism $F \to F$, extending the inclusion map $X \hookrightarrow F$. But the identity automorphism $\mathrm{Id} : F \to F$ is evidently also a homomorphism extending this inclusion $X \hookrightarrow F$, therefore $\hat{\phi} = \mathrm{Id}$, by the Universal Property of F. Thus $\hat{\phi}(f) = f$, for all $f \in F$, in particular $F = \mathrm{im}(\hat{\phi}) \subseteq H \subseteq F$, so $F = H = \langle X \rangle$. Hence X generates F. ☐

Example 1.1.4. If $k \in \mathbb{N}$ and $k \geq 2$ then the cyclic group C_k is not free (on any subset).

Indeed, arguing by contradiction, suppose that C_k is free on a subset X. By Lemma 1.1.3, X must generate C_k, hence $X \neq \emptyset$ as C_k is not the trivial group. Choose any element $x \in X$ and define the map $\phi : X \to C_\infty$, where $C_\infty = \langle c \rangle$, by $\phi(x) = c$ and $\phi(y) = 1$, for all $y \in X \setminus \{x\}$. If C_k is free on X then there must exist a group homomorphism $\hat{\phi} : C_k \to C_\infty$ such that $\hat{\phi}(x) = c$. But then the order of c in C_∞ must be finite, because the order of x in C_k divides k (see Exercise 0.2.2.(iii)). This contradicts the definition of C_∞, finishing our argument.

Lemma 1.1.5. *Let F be a free group on an infinite subset X. Then F is not finitely generated.*

PROOF. Since X is infinite, there are infinitely many pairwise distinct functions $\phi : X \to \mathbb{Z}_2$, and every such function extends to a group homomorphism $\hat{\phi} : F \to \mathbb{Z}_2$, so there are infinitely many such homomorphisms. However, a finitely generated group admits at most finitely many homomorphisms to $\mathbb{Z}_2$, by Exercise 0.5.13. Therefore F cannot be finitely generated. □

Exercise 1.1.6. Let F be a free group on a subset X, with $|X| = k \in \mathbb{N}_0$. Prove that F admits exactly 2^k pairwise distinct homomorphisms to $\mathbb{Z}_2$.

THEOREM 1.1.7. *Let F_i be a free group on a subset $X_i \subseteq F_i$, for $i = 1, 2$. Then the following are equivalent:*

(i) *there exists a bijection $X_1 \to X_2$ (i.e., $|X_2| = |X_2|$);*
(ii) *$F_1 \cong F_2$ (F_1 and F_2 are isomorphic as groups).*

PROOF. First let's show that (i) implies (ii). Suppose that there is a bijection $\phi_1 : X_1 \to X_2 \subseteq F_2$, and let $\phi_2 : X_2 \to X_1 \subseteq F_1$ be the inverse of ϕ_1. By the Universal Property of F_1 and F_2, there are group homomorphisms $\hat{\phi}_1 : F_1 \to F_2$ and $\hat{\phi}_2 : F_2 \to F_1$, extending ϕ_1 and ϕ_2 respectively. Then the composition $\hat{\phi}_2 \circ \hat{\phi}_1 : F_1 \to F_1$ is a homomorphism satisfying

$$(\hat{\phi}_2 \circ \hat{\phi}_1)(x) = \hat{\phi}_2(\hat{\phi}_1(x)) = \hat{\phi}_2(\phi_1(x)) = \phi_2(\phi_1(x)) = x, \text{ for all } x \in X_1,$$

as $\phi_2 = \phi_1^{-1}$. Thus the restriction of $\hat{\phi}_2 \circ \hat{\phi}_1$ to X_1 is the identity map, so this homomorphism extends the inclusion map $X_1 \hookrightarrow F_1$. By the Universal Property of F_1, such an extension is unique and given by the identity automorphism $\mathrm{Id}_{F_1} : F_1 \to F_1$, hence we have $\hat{\phi}_2 \circ \hat{\phi}_1 = \mathrm{Id}_{F_1}$. Similarly, we can show that $\hat{\phi}_1 \circ \hat{\phi}_2 = \mathrm{Id}_{F_2}$, which implies that $\hat{\phi}_2 = \hat{\phi}_1^{-1}$, so $\hat{\phi}_1 : F_1 \to F_2$ is an isomorphism.

Now, let us show that (ii) implies (i). We will only do it under the assumption that X_1 (or X_2) is finite. The general case can be proved similarly, but it requires some background in Set Theory. So, suppose that $|X_1| \in \mathbb{N}_0$ and $F_1 \cong F_2$. Then F_1 is finitely generated, hence so is F_2 (because it is isomorphic to F_1), which implies that $|X_2| < \infty$, by Lemma 1.1.5. Now, by Exercise 1.1.6, F_i admits exactly $2^{|X_i|}$ homomorphisms to $\mathbb{Z}_2$, for $i = 1, 2$. Therefore, since $F_1 \cong F_2$, we must have $2^{|X_1|} = 2^{|X_2|}$, whence $|X_1| = |X_2|$, as required. □

Theorem 1.1.7 and its proof have the following important corollary.

Corollary 1.1.8. *If F is a free group then any two free generating sets X and Y of F have the same cardinality. Moreover, any bijection $X \to Y$ extends to an automorphism $\psi : F \to F$.*

A non-trivial free group will generally have multiple free generating sets.

Exercise 1.1.9. Prove that $C_\infty = \langle c \rangle$ (see Example 1.1.2.(b)) has precisely two free generating sets $\{c\}$ and $\{c^{-1}\}$.

In fact, if $|X| \geq 2$ then a free group F on X will have infinitely many free generating sets (see Exercise 1.5.3 below).

Corollary 1.1.8 allows us to give the following definition.

Definition 1.1.10. If F is a free group on a subset $X \subseteq F$ then the cardinality $|X|$ is called the *rank* of F, and is denoted $\mathrm{rank}(F)$.

Remark 1.1.11. By Theorem 1.1.7, two free groups are isomorphic if and only if they have the same rank. Therefore, for every $k \in \mathbb{N}_0$ it makes sense to talk about *the free group of rank k*, F_k.

Example 1.1.12. In view of Example 1.1.2, we see that F_0, the free group of rank 0, is the trivial group, and F_1, the free group of rank 1, is the infinite cyclic group.

Exercise 1.1.13. Show that F_k cannot be generated by less than k elements.

Thus the rank of F_k is also the smallest number of elements required to generate this group (in fact, any generating set of F_k with exactly k elements is a free basis, see Theorem 1.3.7 below).

1.2. Construction of free groups

In Section 1.1 we have defined free groups as groups satisfying a certain Universal Property. However, we have not shown that such groups actually exist. In this section we will remedy this by providing an explicit construction of a free group on any set X.

Let X be a non-empty set, and let $\overline{X}$ be a copy of X, with a fixed bijection $\sigma : X \to \overline{X}$. We should think of X and $\overline{X}$ as disjoint sets of symbols.

Notation 1.2.1. For each $x \in X$ we denote $\sigma(x) \in \overline{X}$ by x^{-1}, and for each $y \in \overline{X}$ we let y^{-1} denote $\sigma^{-1}(y) \in X$. Thus $\overline{X} = \{x^{-1} \mid x \in X\}$ and $(x^{-1})^{-1} = x$, for all $x \in X$. This justifies the notation that we will further use: $X^{-1} = \overline{X}$ and $X^{\pm 1} = X \cup X^{-1}$. For each $x \in X$, it is also convenient to set $x^1 = x$.

Definition 1.2.2. A *word* over $X^{\pm 1} = X \cup X^{-1}$ is a finite string

$$w = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \dots x_k^{\varepsilon_k}, \text{ where } k \in \mathbb{N}_0,\ x_i \in X \text{ and } \varepsilon_i \in \{\pm 1\}, \text{ for } i = 1, \dots, k. \tag{1.1}$$

The number k is called the *length* of the word w and is denoted $\|w\|$. A *subword* of w is any substring $x_i^{\varepsilon_i} \dots x_j^{\varepsilon_j}$, where $1 \le i \le j \le k$. The *empty word* e is given by the empty string, thus $\|e\| = 0$.

Example 1.2.3. If $X = \{x, y, z\}$, then $x^{-1}zyy^{-1}y^{-1}x^{-1}zzzy$ is a word of length 10 over $X^{\pm 1}$, and z, $yy^{-1}y^{-1}$ and zzy are some of its subwords.

Notation 1.2.4. Given any $x \in X$ and $n \in \mathbb{N}$ we will write x^n to denote the word $xx \dots x$ (of length n), and x^{-n} to denote the word $x^{-1}x^{-1} \dots x^{-1}$ (of length n). We will also use x^0 as another way to write the empty word.

Example 1.2.5. The word $x^{-1}zyy^{-1}y^{-1}x^{-1}zzzy$ over $\{x, y, z\}^{\pm 1}$ can also be written as $x^{-1}zyy^{-2}x^{-1}z^3y$.

Definition 1.2.6. Suppose that w is a word over $X^{\pm 1}$, given in (1.1). Then w is said to be *reduced* if it does not contain any subwords of the form xx^{-1} or $x^{-1}x$, where $x \in X$.

Example 1.2.7. If $X = \{x, y, z\}$, then the words $x^{-1}zyy^{-2}x^{-1}z^3y$ and $z^{-7}x^{-3}y^6y^{-4}x^{-5}x$ are not reduced. On the other hand, the words z^{10}, $y^{-9}z^8y^4x^5$, $xyxyxyx$, $xy^{-1}x^{-1}y$ are reduced.

Definition 1.2.8. An *elementary reduction* of a word w over $X^{\pm 1}$ is the deletion of a subword of the form xx^{-1} or $x^{-1}x$, where $x \in X$. We will write $w \mapsto w'$ to signify that a word w' has been obtained from w by applying an elementary reduction.

Note that every elementary reduction decreases the length of the word by 2, so any word can be brought to a reduced word by applying a finite sequence of elementary reductions.

Example 1.2.9. Suppose that $X = \{x, y, z\}$ and $w = x^{-1}zyy^{-2}x^{-1}z^{-2}z^3y$. Clearly w is not reduced but we can apply a sequence of three elementary reductions to bring it to a reduced word:

$$w \mapsto x^{-1}zy^{-1}x^{-1}z^{-2}z^3y \mapsto x^{-1}zy^{-1}x^{-1}z^{-1}z^2y \mapsto x^{-1}zy^{-1}x^{-1}zy.$$

Note that in the above example the word w admits two different elementary reductions (cancellation of y with y^{-1} and of z^{-1} with z).

Lemma 1.2.10. *Let w be a word over $X^{\pm 1}$ admitting two elementary reductions $w \mapsto w_1$ and $w \mapsto w_2$, with $w_1 \neq w_2$. Then there exist a word w_3 and elementary reductions $w_1 \mapsto w_3$, $w_2 \mapsto w_3$, such that we have the following diagram:*

$$\begin{tikzcd}
 & w_1 \arrow[dr, mapsto] & \\
w \arrow[ur, mapsto] \arrow[dr, mapsto] & & w_3 . \\
 & w_2 \arrow[ur, mapsto] &
\end{tikzcd}$$

Proof. We need to consider two cases, depending on the overlaps between the two reductions of w.

Case 1: the elementary reductions affect disjoint subwords of w. In other terms,

$$w = u_1 x_1 x_1^{-1} u_2 x_2 x_2^{-1} u_3,$$

where u_1, u_2, u_3 are (possibly empty) words over $X^{\pm 1}$, $x_1, x_2 \in X^{\pm 1}$, and the elementary reductions under consideration are

$$w \mapsto u_1 u_2 x_2 x_2^{-1} u_3 = w_1 \quad \text{and} \quad w \mapsto u_1 x_1 x_1^{-1} u_2 u_3 = w_2.$$

Then we clearly have the elementary reductions

$$w_1 \mapsto u_1 u_2 u_3 \quad \text{and} \quad w_2 \mapsto u_1 u_2 u_3,$$

so we can set $w_3 = u_1 u_2 u_3$.

Case 2: the subwords affected by the two elementary reductions of w overlap. In this case, without loss of generality, we can assume that $w = u_1 x x^{-1} x u_2$, where u_1, u_2 are some words over $X^{\pm 1}$, $x \in X^{\pm 1}$, w_1 is obtained from w by deleting the subword $x x^{-1}$ and w_2 is obtained from w by deleting the subword $x^{-1} x$. But then $w_1 = u_1 x u_2 = w_2$, contradicting the assumption that w_1 and w_2 are distinct. Thus Case 2 is impossible. ☐

Theorem 1.2.11. *If w is a word over $X^{\pm 1}$ then it can be brought to a reduced word by a sequence of elementary reductions and this reduced word is unique.*

Proof. The fact that w can be brought to a reduced word by a sequence of elementary reductions is obvious because every such reduction decreases the length $\|w\|$. To prove that such a reduced word is unique we will employ induction on $\|w\|$.

If $\|w\| = 0$ then w is the empty word, so it is already reduced. Thus we can assume that $\|w\| \geq 1$ and w is not reduced. Let

$$w \mapsto w_1' \mapsto w_2' \mapsto \cdots \mapsto w_k' \text{ and } w \mapsto w_1'' \mapsto w_2'' \mapsto \cdots \mapsto w_l'' \tag{1.2}$$

be two sequences of elementary reductions of w resulting in reduced words w_k' and w_l'' over $X^{\pm 1}$, where $k, l \in \mathbb{N}$. Suppose that $w_k' \neq w_l''$, then there exists $i \in \{1, \ldots, \min\{k, l\}\}$ such that $w_i' \neq w_i''$ and i is the minimal index with such property (so that $w_j' = w_j''$, for all $j = 1, \ldots, i-1$). Thus the two sequences of reductions in (1.2) can be re-written as

$$\begin{tikzcd}
 & & & & w_i' \arrow[r, mapsto] & \ldots \arrow[r, mapsto] & w_k' \\
w \arrow[r, mapsto] & w_1' = w_1'' \arrow[r, mapsto] & \ldots \arrow[r, mapsto] & w_{i-1}' = w_{i-1}'' \arrow[ur, mapsto] \arrow[dr, mapsto] & & & . \\
 & & & & w_i'' \arrow[r, mapsto] & \ldots \arrow[r, mapsto] & w_l''
\end{tikzcd}$$

By Lemma 1.2.10, there is a word u over $X^{\pm 1}$ and elementary reductions $w_i' \mapsto u$, $w_i'' \mapsto u$. Of course, u also admits a sequence of elementary reductions bringing it to some reduced word v over $X^{\pm 1}$:

$$\begin{tikzcd}
 & & & & w'_i \arrow[r, mapsto] \arrow[dr, mapsto] & \dots \arrow[r, mapsto] & w'_k & \\
w \arrow[r, mapsto] & w'_1 = w''_1 \arrow[r, mapsto] & \dots \arrow[r, mapsto] & w'_{i-1} = w''_{i-1} \arrow[ur, mapsto] \arrow[dr, mapsto] & & u \arrow[r, mapsto] & \dots \arrow[r, mapsto] & v \cdot \\
 & & & & w''_i \arrow[r, mapsto] \arrow[ur, mapsto] & \dots \arrow[r, mapsto] & w''_l &
\end{tikzcd}$$

By construction, we have $\|w'_i\| < \|w\|$, and since w'_i can be brought to two reduced words w'_k and v, we can use the induction hypothesis to conclude that $w'_k = v$. Similarly, we see that $v = w''_l$, hence $w'_k = w''_l$, which contradicts our assumptions that these words are distinct. This contradiction completes the proof of the theorem. ☐

Theorem 1.2.11 allows us to introduce the following.

Definition 1.2.12. If w is a word over $X^{\pm 1}$, we denote by $\overline{w}$ the unique reduced word over $X^{\pm 1}$ to which w can be brought by a sequence of elementary reductions.

Recall that, given two word u, v over $X^{\pm 1}$, the *concatenation* uv, of u with v, is the word obtained by attaching the string for v at the end of the string for u. For example, if $u = zxy^{-1}z^3$ and $v = z^{-3}yx^{-4}y$ then $uv = zxy^{-1}z^3z^{-3}yx^{-4}y$.

Definition 1.2.13. Let X be a non-empty set. We define $F(X)$ as the set of all reduced words over $X^{\pm 1}$. We also define the product $u \cdot v$, of two reduced words $u, v \in F(X)$, as $\overline{uv}$ (the reduced word for the concatenation of u and v).

Example 1.2.14. Let $X = \{x, y, z\}$, then $u = zxy^{-1}z^3$, $v = z^{-3}yx^{-4}y$ are elements of $F(X)$, and

$$u \cdot v = \overline{(zxy^{-1}z^3)\,(z^{-3}yx^{-4}y)} = \overline{(zxy^{-1})\,(yx^{-4}y)} = \overline{(zx)\,(x^{-4}y)} = zx^{-3}y \in F(X).$$

Theorem 1.2.15. *For any non-empty set X, $F(X)$ is a group with respect to the multiplication given in Definition 1.2.13.*

Proof. Firstly, the empty word e evidently works as the identity element of $F(X)$ because for any reduced word u over $X^{\pm 1}$ we have

$$u \cdot e = \overline{ue} = \overline{u} = u \text{ and } e \cdot u = \overline{eu} = \overline{u} = u.$$

Secondly, if $w = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$ is a reduced word from $F(X)$, where $x_1, \dots, x_n \in X$ and $\varepsilon_1, \dots, \varepsilon_n \in \{\pm 1\}$, then the word $w^{-1} = x_n^{-\varepsilon_n} \dots x_1^{-\varepsilon_1}$ is also reduced, hence $w^{-1} \in F(X)$. Moreover,

$$w^{-1} \cdot w = \overline{(x_n^{-\varepsilon_n} \dots x_1^{-\varepsilon_1})(x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n})} = \overline{(x_n^{-\varepsilon_n} \dots x_2^{-\varepsilon_2})(x_2^{\varepsilon_2} \dots x_n^{\varepsilon_n})} = \dots = \overline{x_n^{-\varepsilon_n} x_n^{\varepsilon_n}} = e.$$

And, similarly, $w \cdot w^{-1} = e$, thus w^{-1} is the inverse of w in $F(X)$.

It remains to check that the multiplication on $F(X)$ is associative. Let $u, v, w \in F(X)$ be arbitrary. Observe that the reduced words $\overline{u\overline{vw}}$ and $\overline{\overline{uv}w}$ can both be obtained by applying reductions to the word uvw, hence $\overline{u\overline{vw}} = \overline{\overline{uv}w}$, by Theorem 1.2.11. It follows that

$$u \cdot (v \cdot w) = u \cdot \overline{vw} = \overline{u\overline{vw}} = \overline{\overline{uv}w} = \overline{uv} \cdot w = (u \cdot v) \cdot w.$$

Thus we have verified the associativity of the product, so $F(X)$ is indeed a group. ☐

Remark 1.2.16. Since $F(X)$ is, by definition, the set of all reduced words over $X^{\pm 1}$, we see that distinct reduced words over $X^{\pm 1}$ give rise to distinct elements of $F(X)$. In particular, we can treat X as a subset of $F(X)$, after identifying it with the set of 1-letter words $\{x \mid x \in X\}$ in $F(X)$. Moreover, X is clearly a generating set of $F(X)$.

The next theorem fulfills our promise at the beginning of the section by showing that the group $F(X)$ that we constructed is indeed a free group on X.

Theorem 1.2.17. *Let X be a non-empty set. Then the group $F(X)$ constructed above is free on X in the sense of Definition 1.1.1.*

Proof. Suppose that H is a group and $\phi : X \to H$ is a set map. We start by defining a map $\hat{\phi} : F(X) \to H$ as follows. We set $\hat{\phi}(e) = 1_H$ and for any non-empty reduced word $w = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n} \in F(X)$, where $x_1, \dots, x_n \in X$ and $\varepsilon_1, \dots, \varepsilon_n \in \{\pm 1\}$, we set

$$\hat{\phi}(w) = \phi(x_1)^{\varepsilon_1} \dots \phi(x_n)^{\varepsilon_n} \in H. \tag{1.3}$$

Let us show that the map $\hat{\phi}$ is actually a group homomorphism.

Let $u, v \in F(X)$ be two reduced words. Then we can write

$$u = x_1^{\alpha_1} \dots x_k^{\alpha_k} y_1^{\beta_1} \dots y_l^{\beta_l} \quad \text{and} \quad v = y_l^{-\beta_l} \dots y_1^{-\beta_1} z_1^{\gamma_1} \dots z_m^{\gamma_m},$$

where $x_1, \dots, x_k, y_1, \dots, y_l, z_1 \dots, z_m \in X$, $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_l, \gamma_1, \dots, \gamma_m \in \{\pm 1\}$ and $z_1^{\gamma_1} \neq x_k^{-\alpha_k}$, so that

$$u \cdot v = \overline{uv} = x_1^{\alpha_1} \dots x_k^{\alpha_k} z_1^{\gamma_1} \dots z_m^{\gamma_m} \quad \text{in } F(X).$$

Then, according to (1.3), we have

$$\begin{aligned} \hat{\phi}(u)\hat{\phi}(v) &= \left(\phi(x_1)^{\alpha_1} \dots \phi(x_k)^{\alpha_k} \phi(y_1)^{\beta_1} \dots \phi(y_l)^{\beta_l}\right) \left(\phi(y_l)^{-\beta_l} \dots \phi(y_1)^{-\beta_1} \phi(z_1)^{\gamma_1} \dots \phi(z_m)^{\gamma_m}\right) \\ &= \phi(x_1)^{\alpha_1} \dots \phi(x_k)^{\alpha_k} \phi(z_1)^{\gamma_1} \dots \phi(z_m)^{\gamma_m} = \hat{\phi}(u \cdot v). \end{aligned}$$

Thus $\hat{\phi} : F(X) \to H$ is indeed a group homomorphism.

In view of (1.3), $\hat{\phi}(x) = \phi(x)$, for every $x \in X$, thus $\hat{\phi}$ extends the map ϕ. Finally, since $F(X) = \langle X \rangle$ (see Remark 1.2.16), we know that $\hat{\phi}$ is the unique extension of ϕ by Lemma 0.5.11.(i). ☐

Corollary 1.2.18. *Let F be a free group on a subset $X \subseteq F$. Then $F \cong F(X)$ and every element of F can be written uniquely as a reduced word over $X^{\pm 1}$.*

Proof. The isomorphism $F \cong F(X)$ follows from Theorem 1.2.17 and the uniqueness of a free group on X, given by Theorem 1.1.7. In fact, this isomorphism sends very element of X to itself, so the second claim follows from Remark 1.2.16. ☐

Exercise 1.2.19. Let F be a free group on a subset X. Prove that if $|X| \geq 2$ then F is not abelian.

If F is any group and X is a generating set of F then any element of F can be written as a reduced word over $X^{\pm 1}$, however this reduced word is not always unique. In fact, the unique representation of elements by reduced words characterizes the freeness of F.

Corollary 1.2.20. *Suppose that F is a group with a generating set X. Then the following statements are equivalent:*

(i) *F is free on X;*
(ii) *every element of F has a unique expression as a reduced word over $X^{\pm 1}$ in F;*
(iii) *every non-empty reduced word over $X^{\pm 1}$ represents a non-trivial element of F.*

Proof. The implication (i)⇒(ii) is given by Corollary 1.2.18. Statement (iii) is a special case of statement (ii), as the identity element 1 is represented by the empty word, so (ii) implies (iii).

To prove (iii)⇒(i), recall that since $F(X)$ is free on X (by Theorem 1.2.17), there is a homomorphism $\psi : F(X) \to F$ extending the identity map on X. This homomorphism is surjective because $F = \langle X \rangle$ (see Lemma 0.5.11.(ii)). The injectivity of this homomorphism follows the fact that $\ker \psi = \{e\}$ by (iii). Therefore ψ is an isomorphism between $F(X)$ and F, hence (i) holds. ☐

Notation 1.2.21. From now on we will often write uv instead of $u \cdot v$, for the product of elements $u, v \in F(X)$. We will also sometimes write 1 for the empty word $e \in F(X)$.

1.3. Subgroups of free groups

One of the fundamental theorems in the theory of free groups is that any subgroup of a free group is itself free. This is a non-trivial fact, which will be proved much later in these notes. However, in this section we will investigate particular instances of it.

Exercise 1.3.1. Let F be a free group on a subset X. Show that for any subset $Y \subseteq X$ the subgroup $H = \langle Y \rangle \leqslant F$ is free, freely generated by Y.

This exercise has the following immediate consequence.

Corollary 1.3.2. *The free group of rank n, where $n \in \mathbb{N}_0$ or $n = \infty$, contains a free subgroup of rank m, for any $0 \leq m \leq n$.*

Proposition 1.3.3. *The free group of rank 2 contains free subgroups of any countable rank.*

Proof. Let $F = F(\{x,y\})$ be the free group with the free generating set $\{x,y\}$. By Corollary 1.3.2 it is enough to show that F contains a subgroup isomorphic to the free group on an infinite countable set. To this end, define elements $z_i \in F$ by $z_i = x^i y x^{-i}$, for $i \in \mathbb{N}_0$, and consider the subset $Z = \{z_i \mid i \in \mathbb{N}_0\} \subseteq F$. We will show that Z is a free generating set of the subgroup $H = \langle Z \rangle \leqslant F$, thus H is a free group of infinite countable rank.

Consider any non-empty reduced word w over $Z^{\pm 1}$. Then it can be written as

$$w = z_{i_1}^{\alpha_1} \dots z_{i_k}^{\alpha_k}, \quad \text{where } k \in \mathbb{N},\ i_1, \dots, i_k \in \mathbb{N}_0,\ \alpha_1, \dots, \alpha_k \in \mathbb{Z} \setminus \{0\}, \text{ and} \tag{1.4}$$

$$i_j \neq i_{j+1}, \quad \text{for every } j = 1, \dots, k-1. \tag{1.5}$$

Note that $z_i^\alpha = x^i y^\alpha x^{-i}$ in F, for all $i \in \mathbb{N}_0$ and all $\alpha \in \mathbb{Z}$. Therefore, in the free group F, w represents the element

$$f = x^{i_1} y^{\alpha_1} x^{-i_1} x^{i_2} y^{\alpha_2} x^{-i_2} \dots x^{i_k} y^{\alpha_k} x^{-i_k} = x^{i_1} y^{\alpha_1} x^{i_2 - i_1} y^{\alpha_2} x^{i_3 - i_2} \dots x^{i_k - i_{k-1}} y^{\alpha_k} x^{-i_k}.$$

By (1.4) and (1.5), the latter word for f is non-empty and reduced over $\{x,y\}^{\pm 1}$, therefore $f \neq 1$ in F. Thus w represents a non-trivial element of the subgroup $H = \langle Z \rangle$. Since the latter is true for every non-empty reduced word over $Z^{\pm 1}$, we can conclude that H is free on Z, by Corollary 1.2.20. ☐

Exercise 1.3.4. Prove that the free group of rank 1 cannot contain a subgroup isomorphic to the free group of rank 2.

We now state two key theorems about subgroups of free groups, but we defer the proofs until later chapters. The first one was proved by Jakob Nielsen and Otto Schreier, who are widely regarded as forefathers of Combinatorial Group Theory.

Theorem 1.3.5 (Nielsen-Schreier Theorem). *Every subgroup of a free group is itself free.*

Example 1.3.6. Let F be a group on a subset X. Consider arbitrary elements $h_1, \dots, h_k \in F$. Then, according to Theorem 1.3.5, the subgroup $H = \langle h_1, \dots, h_k \rangle$ is free. However, the set $\{h_1, \dots, h_k\}$ may not necessarily be a free generating set of H. For example, suppose that $x, y \in X$ and set $a = x$, $b = xy^2$ and $c = xy^4$ in F. Then the subgroup $H = \langle a, b, c \rangle \leqslant F$ is free, but $\{a, b, c\}$ is not a free basis of H by Corollary 1.2.20, because the distinct reduced words $(a^{-1}b)^2 = a^{-1}ba^{-1}b$ and $a^{-1}c$ represent the same element $y^4 \in H$.

Theorem 1.3.7. *Let F be the free group of rank $n \in \mathbb{N}_0$. If $Y \subseteq F$ is a generating set such that $|Y| = n$ then Y is a free basis of F.*

Example 1.3.8. Consider the free group $F = F(\{x,y\})$, of rank 2. Evidently, $\{x,y\}$ is a free generating set of F. Consider the elements $u = xyx^{-1}$ and $v = xy^{-5}$ in F. One easily checks that $u^5 v = x$, whence $(u^5 v)^{-1} u (u^5 v) = y$, thus $\{x,y\} \subseteq \langle u, v \rangle$ in F. It follows that $\langle u, v \rangle = F$ (by properties of generating sets, see Definition 0.5.1). Therefore, $\{u, v\}$ is also a free generating set of F, by Theorem 1.3.7.

Theorems 1.3.5 and 1.3.7 will be proved in Section 3.8 and Section 5.3 respectively.

1.4. Centralizers in free groups

Let us first recall the following definition.

Definition 1.4.1. If G is a group and $g \in G$ is any element then the *centralizer of g in G*, $\mathrm{C}_G(g)$, is defined by

$$\mathrm{C}_G(g) = \{h \in G \mid hg = gh\} \subseteq G.$$

The *center of G*, $\mathrm{Z}(G)$, is defined by

$$\mathrm{Z}(G) = \{h \in G \mid hg = gh, \quad \text{for all } g \in G\} = \bigcap_{g \in G} \mathrm{C}_G(g) \subseteq G.$$

Exercise 1.4.2. Suppose that G is a group and $g \in G$. Prove that $\mathrm{C}_G(g)$ is a subgroup of G, containing $\langle g \rangle$, and $\mathrm{Z}(G)$ is a normal subgroup of G.

Lemma 1.4.3. *Let F be a free group on a subset $X \subseteq F$. Then for every $x \in X$, we have $\mathrm{C}_F(x) = \langle x \rangle$.*

Proof. By Exercise 1.4.2, $\langle x \rangle \subseteq \mathrm{C}_F(x)$, so we only need to prove the opposite inclusion. Throughout this proof we will identify F with $F(X)$, the group of all reduced words over $X^{\pm 1}$ (see Corollary 1.2.18). Arguing by contradiction, suppose that there exists a reduced word $w \in \mathrm{C}_F(x)$ such that w contains letters from $X \setminus \{x\}$. Then w must have the form $x^\alpha y x^\beta$ or $x^\alpha yuzx^\beta$, where $\alpha, \beta \in \mathbb{Z}$, $y, z \in X^{\pm 1} \setminus \{x^{\pm 1}\}$ and u is a reduced word over $X^{\pm 1}$. In the former case we have

$$xw = x^{\alpha+1} y x^\beta \quad \text{and} \quad wx = x^\alpha y x^{\beta+1}.$$

And in the latter case,

$$xw = x^{\alpha+1} yuzx^\beta \quad \text{and} \quad wx = x^\alpha yuzx^{\beta+1}.$$

Since $y, z \neq x^{\pm 1}$, in both cases the right-hand sides are distinct reduced words over $X^{\pm 1}$, so, by Corollary 1.2.20, $xw \neq wx$ in F, contradicting the assumption that $w \in \mathrm{C}_F(x)$. This shows that $\mathrm{C}_F(x) \subseteq \langle x \rangle$, hence $\mathrm{C}_F(x) = \langle x \rangle$ and the lemma is proved. □

Corollary 1.4.4. *If F is a free group on a subset X, with $|X| \geq 2$, then $\mathrm{Z}(F) = \{1\}$.*

Proof. By the assumptions, there exist distinct elements $x, y \in X$. Then, according to Lemma 1.4.3, we have $\mathrm{C}_F(x) = \langle x \rangle$ and $\mathrm{C}_F(y) = \langle y \rangle$. Note that $x^m \neq y^n$ in F, for any $m, n \in \mathbb{Z} \setminus \{0\}$, by Corollary 1.2.20, which implies that $\langle x \rangle \cap \langle y \rangle = \{1\}$ in F. Therefore,

$$\mathrm{Z}(F) \subseteq \mathrm{C}_F(x) \cap \mathrm{C}_F(y) = \langle x \rangle \cap \langle y \rangle = \{1\},$$

i.e., $\mathrm{Z}(F) = \{1\}$, as required. □

Given a group G with an element $g \in G$, we will say that *g is a proper power* in G if there exists an element $h \in G$ and an integer $k \geq 2$ such that $g = h^k$ in G. Thus g is not a proper power in G if and only if whenever $g = h^k$, for some $h \in G$ and $k \in \mathbb{Z}$, we must have $k = \pm 1$.

Theorem 1.4.5. *Let F be a free group and let $f \in F \setminus \{1\}$ be a non-trivial element. Then $\mathrm{C}_F(f)$ is a cyclic subgroup of F. In fact, if f is not a proper power in F then $\mathrm{C}_F(f) = \langle f \rangle$.*

Proof. By the Nielsen-Schreier Theorem (Theorem 1.3.5), the subgroup $H = \mathrm{C}_F(f)$ is a free group. Clearly $f \in \mathrm{Z}(H)$, so $\mathrm{Z}(H) \neq \{1\}$ as $f \neq 1$. Therefore $\mathrm{rank}(H) \leq 1$ by Corollary 1.4.4, and so H must be cyclic. The latter means that there exists $h \in H$ such that $H = \langle h \rangle$, whence $f = h^k$, for some $k \in \mathbb{Z} \setminus \{0\}$. If f is not a proper power in F then we must have $k = \pm 1$, so that $h = f^{\pm 1}$, which implies that $H = \langle f \rangle$. □

Example 1.4.6. Let $F = F(\{x, y\})$ be the free group of reduced words over $\{x, y\}^{\pm 1}$. Let us show that $\mathrm{C}_F\left((x^2y^{-3})^7\right) = \langle x^2y^{-3} \rangle$.

By Theorem 1.4.5, $\mathrm{C}_F\left((x^2y^{-3})^7\right)$ is cyclic, hence it is abelian. In particular, every element of this centralizer commutes with $x^2y^{-3} \in \mathrm{C}_F\left((x^2y^{-3})^7\right)$, hence $\mathrm{C}_F\left((x^2y^{-3})^7\right) \subseteq \mathrm{C}_F(x^2y^{-3})$. The opposite inclusion is obvious, so

$$\mathrm{C}_F\left((x^2y^{-3})^7\right) = \mathrm{C}_F(x^2y^{-3}). \tag{1.6}$$

Now, we need to show that the element x^2y^{-3} is not a proper power in F. This can be done by considering reduced words and *cyclically reduced* words, but we will use an alternative method (which does not always work), to demonstrate the general idea how the Universal Property may be used. Suppose that $x^2y^{-3} = h^k$, for some $k \in \mathbb{Z}$. By the Universal Property of free groups, there exists a homomorphism $\psi : F \to \mathbb{Z}$ such that $\psi(x) = \psi(y) = -1$ and $\psi(h) = l$, for some $l \in \mathbb{Z}$. Here the group operation on $\mathbb{Z}$ is addition, so

$$\psi(h^k) = k\psi(h) = kl \quad \text{and} \quad \psi(x^2y^{-3}) = 2\psi(x) - 3\psi(y) = 1,$$

thus $kl = 1$, so $k = \pm 1$. This shows that x^2y^{-3} is not a proper power of another element in F. By Theorem 1.4.5 the latter implies that $\mathrm{C}_F(x^2y^{-3}) = \langle x^2y^{-3}\rangle$. We can combine this equality with (1.6) to conclude that $\mathrm{C}_F\left((x^2y^{-3})^7\right) = \langle x^2y^{-3}\rangle$.

1.5. Additional exercises

Exercise 1.5.1. Let F be a free group on a subset X. Prove that there is a unique automorphism $\psi \in \mathrm{Aut}(F)$ such that $\psi(x) = x^{-1}$, for every $x \in X$.
[*Hint:* Use the Universal Property together with the observation that $\psi \circ \psi = \mathrm{Id}_F$.]

Exercise 1.5.2. Suppose that F is free on a subset X. Show that for any automorphism $\alpha \in \mathrm{Aut}(F)$, $\alpha(X)$ is also a free generating set of F. Use this fact to prove that the following subsets of F are free generating sets:

(a) $fXf^{-1} = \{fxf^{-1} \mid x \in X\}$, for any $f \in F$;
(b) $X^{-1} = \{x^{-1} \mid x \in X\}$.

Recall that two elements x, y of a group G are said to be *conjugate* if there exists $g \in G$ such that $y = gxg^{-1}$. Note that conjugacy is an equivalence relation on the group G: it is reflexive, symmetric and transitive.

Exercise 1.5.3. Let F be a free group on a set X, with $|X| \geq 2$.

(i) Show that no two distinct elements of X are conjugate in F.
(ii) Prove that F has infinitely many free generating sets. [*Hint:* Claim (i) implies that if $fXf^{-1} = gXg^{-1}$ then $gxg^{-1} = fxf^{-1}$ for each $x \in X$ and all $f, g \in F$. Combine this observation with Lemma 1.4.3 to show that $fXf^{-1} \neq gXg^{-1}$ if $f \neq g$ in F.]

Corollary 1.1.8 essentially gives the following statement, which can be regarded as a converse to Exercise 1.5.2.(a).

Exercise 1.5.4. Suppose that X and Y are two free generating sets of the same free group F. Prove that there is an automorphism $\psi \in \mathrm{Aut}(F)$ such that $\psi(X) = Y$. Is this automorphism always unique?

Exercise 1.5.5. Let F be a free group.

(a) Show that every abelian subgroup of F is cyclic.
(b) Prove that every non-trivial abelian subgroup $C \leqslant F$ is contained in a unique maximal abelian subgroup $A \leqslant F$ (that is, $C \subseteq A$ and if B is another abelian subgroup of F containing C then $B \subseteq A$).

CHAPTER 2

Group presentations

The subject of this chapter are group presentations, which give a convenient way to define infinite groups in terms of their generators and relators. The study of groups via presentations is the central theme of Combinatorial Group Theory.

2.1. Presenting groups via generators and relators

Free groups are "universal objects" in the realm of groups and can be used to define other groups as their quotients. Let us start with the following observation.

Proposition 2.1.1. *Every group is a quotient of a free group.*

Proof. Let G be an arbitrary group. Choose a generating set $X \subseteq G$ (for example, we can take $X = G$) and let $\phi : X \to G$ denote the inclusion map. Consider the free group $F = F(X)$ on X. By the Universal Property of F (see Definition 1.1.1), there is a homomorphism $\hat{\phi} : F \to G$ extending ϕ. Since $X = \phi(X) = \hat{\phi}(X) \subseteq \operatorname{im}(\hat{\phi})$, we can use Lemma 0.5.11.(ii) to conclude that $\hat{\phi}$ is surjective. Therefore $G \cong F/N$, where $N = \ker \hat{\phi} \lhd G$, by the First Isomorphism Theorem (Theorem 0.4.6). Thus G is a quotient of F. □

The above proposition shows that every group can be represented (up to isomorphism) as the quotient $F(X)/N$, where X is any generating set of G and $N \lhd F(X)$ is a suitable normal subgroup. Generally N will be infinite, but in many cases it will be "normally generated" by finitely many elements.

Definition 2.1.2. Let R be a subset of a group F. The *normal closure of R in F*, denoted $\langle\langle R \rangle\rangle^F$, is the smallest normal subgroup of F containing R. In other words,

$$\langle\langle R \rangle\rangle^F = \bigcap_{N \lhd F,\ R \subseteq N} N.$$

The normal closure of the empty subset is thus defined to be the trivial subgroup $\{1\} \leqslant G$.

Exercise 2.1.3. Suppose that F is a group and $R \subseteq F$ is a non-empty subset. Prove that $\langle\langle R \rangle\rangle^F$ consists of all products of the form

$$f_1 r_1^{\varepsilon_1} f_1^{-1} f_2 r_2^{\varepsilon_2} f_2^{-1} \dots f_k r_k^{\varepsilon_k} f_k^{-1}, \tag{2.1}$$

for arbitrary $f_1, \dots, f_k \in F$, $r_1, \dots, r_k \in R$ and $\varepsilon_1, \dots, \varepsilon_k \in \{\pm 1\}$.

The above exercise tells us that the normal closure of R consists of all the elements that can be written as products of conjugates of elements from R.

Example 2.1.4. Let F be any group and let $x, y \in F$ be two elements.

(a) The relation $[x, y] = 1$ is a *consequence* of the relation $xy^{-m} = 1$ (which can be re-written as $x = y^m$), for any $m \in \mathbb{Z}$. By this we mean that $[x, y] \in \langle\langle xy^{-m} \rangle\rangle^F$ in F. This can be easily checked explicitly:

$$[x, y] = xyx^{-1}y^{-1} = xy^{-m}\, y(xy^{-m})^{-1}y^{-1} \in \langle\langle xy^{-m} \rangle\rangle^F.$$

An alternative way to see that $[x, y] = 1$ follows from $x = y^m$ is simply to observe that in any group a power of an element always commutes with that element.

(b) Take $R = \{[x, y]\}$, where $[x, y] = xyx^{-1}y^{-1}$ is the commutator of x and y in F. By Exercise 2.1.3, the normal closure $N = \langle\langle R \rangle\rangle^F$ consists of the product of all conjugates of $[x, y]$ and

$[x,y]^{-1}=[y,x]$. Let us show that $[x^m,y^n]\in N$, for any $m,n\in\mathbb{Z}$. To this end, it suffices to check that the image of $[x^m,y^n]$ is trivial in the quotient $Q=F/N$. But since $[x,y]\in N$, the images of x and y commute in Q, which evidently implies that the images of x^m and y^n commute in Q. Therefore, $[x^m,y^n]\in N$.

This is easy to check explicitly for small values of m and n. For instance, if $m=-1$ and $n=2$, we have

$$\begin{aligned}[x^{-1},y^2]&=x^{-1}y^2xy^{-2}=x^{-1}y(yxy^{-1}x^{-1})xy^{-1}=(x^{-1}y)[y,x](x^{-1}y)^{-1}(x^{-1}yxy^{-1})\\&=(x^{-1}y)[y,x](x^{-1}y)^{-1}\,x^{-1}[y,x]x=f_1r^{-1}f_1^{-1}\,f_2r^{-1}f_2^{-1}\in N,\end{aligned}$$

where $f_1=x^{-1}y$, $f_2=x^{-1}$ and $r=[x,y]$.

Exercise 2.1.5. Let $\phi:F\to G$ be a surjective group homomorphism and $R\subseteq F$. Prove that $\phi(\langle\!\langle R\rangle\!\rangle^F)=\langle\!\langle\phi(R)\rangle\!\rangle^G$.

Definition 2.1.6. Let G be a group isomorphic to a quotient of a free group $F=F(X)$ by a normal subgroup $N\lhd F$. If $N=\langle\!\langle R\rangle\!\rangle^F$, for some subset $R\subseteq F$, then we say that $\langle X\mid R\rangle$ is a *presentation of* G. In this case elements of X are called generators and elements of R are called *defining relators* of (this presentation of) G. Any element of $N=\langle\!\langle R\rangle\!\rangle^F$ is a *relator* of G, and is a *consequence* of the defining relators from R.

Remark 2.1.7. In view of Proposition 2.1.1, every group G has a presentation in the sense of Definition 2.1.6.

Remark 2.1.8. If we combine Definition 2.1.6 with the First Isomorphism Theorem, the we see that a group G has presentation $\langle X\mid R\rangle$ if and only if there is an epimorphism $\psi:F(X)\to G$ such that $\ker\psi=\langle\!\langle R\rangle\!\rangle^{F(X)}$. Since $F(X)$ is generated by X follows that $\psi(X)$ generates G (see Lemma 0.5.11.(v)), which justifies calling elements of X generators of G. This also shows that a group G is finitely generated if and only if it admits a presentation $\langle X\mid R\rangle$ with $|X|<\infty$.

Convention 2.1.9. Suppose that G has a presentation $\langle X\mid R\rangle$, so that $G\cong F(X)/\langle\!\langle R\rangle\!\rangle^F$. Formally speaking, the set of generators X is not a subset of G. However, we will often abuse the notation by identifying every element $x\in X$ with its image in G.

Example 2.1.10. Let $F=F(X)$ be the free group on X, for some non-empty set X.

(a) If we take $R=\emptyset$ then $N=\langle\!\langle R\rangle\!\rangle^F$ is the trivial subgroup of F, whence $F/N\cong F$. Thus the presentation $\langle X\mid\ \rangle$, with the empty set of defining relators, defines the free group F on X. In particular, $\langle a\mid\ \rangle$ defines the infinite cyclic group generated by a, and $\langle x,y\mid\ \rangle$ defines the free group of rank 2 on the set $\{x,y\}$.

(b) Suppose that X consists of a single element x and consider the group presentation $\langle x\mid x^n\rangle$, where $n\in\mathbb{N}$ is some integer. We claim that this presentation defines the cyclic group C_n, of order n. Indeed, as we know from Example 1.1.2.(b), the group $F=F(\{x\})$ is simply the infinite cyclic group $C_\infty=\langle x\rangle$. If $R=\{x^n\}\subseteq F$ then the normal closure $\langle\!\langle R\rangle\!\rangle^F$ is equal to the cyclic subgroup $\langle x^n\rangle$, because $\langle x^n\rangle$ is clearly the smallest normal subgroup of $\langle x\rangle$ containing x^n. Therefore $F/\langle\!\langle R\rangle\!\rangle^F\cong\langle x\rangle/\langle x^n\rangle\cong C_n$ (cf. Example 0.3.12), as claimed.

Notation 2.1.11. We will often write $G=\langle X\mid R\rangle$ to indicate that G is the group defined by the presentation $\langle X\mid R\rangle$ (in other words, $G=F(X)/\langle\!\langle R\rangle\!\rangle^F$).

We will also sometimes represent elements of R as equations, writing $u=v$ instead of the element $uv^{-1}\in R$ in $F(X)$. We will then say that $u=v$ is a *defining relation* of $G=\langle X\mid R\rangle$. For example, the presentation $\langle a,b\mid[a,b]\rangle$ can also be written as $\langle a,b\mid[a,b]=1\rangle$ or $\langle a,b\mid ab=ba\rangle$.

Exercise 2.1.12. Suppose that X is a set and R,S are subsets of the free group $F(X)$. If $R\subseteq\langle\!\langle S\rangle\!\rangle^{F(X)}$ then the presentations $\langle X\mid R\cup S\rangle$ and $\langle X\mid S\rangle$ define the same group.

2.2. Finding presentations of specific groups

The following theorem is a analogue of the Universal Property of free groups, and will be very useful for showing that some groups are quotients of others. This statement is sometimes attributed to the German mathematician Walther von Dyck.

Theorem 2.2.1 (von Dyck's Theorem). *Let G be a group with presentation $\langle X \mid R\rangle$, where $R \subseteq F(X)$. Suppose that we have a set map $\phi : X \to H$, where H is another group, such that ϕ sends every defining relator of G to the identity element of H, which means that for every $r \in R$, given by the reduced word $x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$ in $F(X)$ (where $x_j \in X$ and $\varepsilon_j \in \{\pm 1\}$, $j = 1, \dots, n$), we have*

$$\phi(x_1)^{\varepsilon_1} \dots \phi(x_n)^{\varepsilon_n} = 1 \quad \text{in } H. \tag{2.2}$$

Then there is a unique group homomorphism $\tilde{\phi} : G \to H$ extending ϕ (i.e., $\tilde{\phi}(x) = \phi(x)$, for each $x \in X$, where X is identified with its image in G, as in Convention 2.1.9).

Proof. By Definition 2.1.6, there is an epimorphism $\psi : F(X) \to G$, whose kernel is the normal closure of R in $F(X)$. Since ψ is surjective and X generates $F(X)$, we see that $\psi(X)$ generates G (see Lemma 0.5.11.(v)). By the Universal Property of $F(X)$, there is a group homomorphism $\hat{\phi} : F(X) \to H$ extending the given set map $\phi : X \to H$, and we need to find a homomorphism $\tilde{\phi} : G \to H$ completing the commutative diagram below:

$$\begin{tikzcd} X \arrow[r, hook] \arrow[dr, "\phi"'] & F(X) \arrow[r, two heads, "\psi"] \arrow[d, "\hat{\phi}"] & G \arrow[dl, dashed, "\tilde{\phi}"] \\ & H & \end{tikzcd} \tag{2.3}$$

Recall that the homomorphism $\hat{\phi}$ is defined by formula (1.3) in the proof of Theorem 1.2.17, thus (2.2) simply says that $\hat{\phi}(r) = 1$ in H, for all $r \in R$. Therefore $R \subseteq \ker \hat{\phi}$, hence $\langle\langle R\rangle\rangle^F \subseteq \ker \hat{\phi}$, by Definition 2.1.2. Since $\langle\langle R\rangle\rangle^F = \ker \psi$, we can apply Theorem 0.4.14 to find a homomorphism $\tilde{\phi} : G \to H$ such that $\hat{\phi} = \tilde{\phi} \circ \psi$, which completes the commutative diagram (2.3).

Thus $\tilde{\phi}$ extends ϕ, in the sense that $\tilde{\phi}(\psi(x)) = \phi(x)$, for all $x \in X$. Finally, $\tilde{\phi}$ is the unique group homomorphism with this property by Lemma 0.5.11.(i) because $\psi(X)$ generates G. □

We will now use von Dyck's Theorem to find presentations for some groups. The following observation is useful when looking at finite groups.

Remark 2.2.2. Suppose that G and H are finite sets and $\phi : G \to H$ is a surjective map. If $|G| \le |H|$ then ϕ is a bijection.

Example 2.2.3. Take $n \in \mathbb{N}$, $n \ge 3$, and let D_n denote the dihedral group of order $2n$. Recall that D_n is the group of all symmetries of a regular n-gon P_n. It consists of n rotations about the center of P_n (including the trivial rotation) and n reflections in symmetry lines of P_n.

Let σ_0 and τ_0 denote reflections in two axes of symmetries of P_n that have the angle π/n between them. Observe that $\sigma_0^2 = \tau_0^2 = 1$ and $\rho_0 = \sigma_0\tau_0$ is a rotation by angle $2\pi/n$, whence $\rho_0^n = (\sigma_0\tau_0)^n = 1$ in D_n. Moreover, D_n is generated by σ_0 and τ_0: the cyclic subgroup $\langle \rho_0\rangle \subseteq \langle \sigma_0, \tau_0\rangle$ contains n rotations and its coset $\sigma_0\langle\rho_0\rangle \subseteq \langle\sigma_0, \tau_0\rangle$ contains n reflections.

We claim that

$$\langle \sigma, \tau \mid \sigma^2,\ \tau^2,\ (\sigma\tau)^n\rangle \tag{2.4}$$

is a presentation of D_n. Let G be the group defined by presentation (2.4). By Theorem 2.2.1, there is a group homomorphism $\xi : G \to D_n$ such that $\xi(\sigma) = \sigma_0$ and $\xi(\tau) = \tau_0$ (because $\sigma_0^2 = \tau_0^2 = (\sigma_0\tau_0)^n = 1$ in D_n as we have observed above, thus ξ sends defining relators of G to relators of D_n). Note that ξ is surjective because D_n is generated by σ_0 and τ_0, both of which are contained in the image of ξ (see Lemma 0.5.11.(ii)).

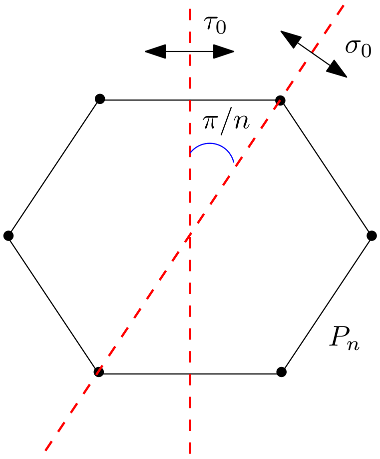


FIGURE 2.1. Symmetries of a regular hexagon

The elements σ and τ generate G and satisfy $\sigma^{2k} = \tau^{2k} = 1$ and $\sigma^{2k+1} = \sigma$, $\tau^{2k+1} = \tau$, for all $k \in \mathbb{Z}$. Moreover, the element $\rho = \sigma\tau$ satisfies $\rho^n = 1$ and $\rho^{-1} = \tau\sigma$ in G. It follows that every element $g \in G$ can be written in one of the following four forms:

(1) $g = \underbrace{\sigma\tau\ldots\sigma\tau}_{2m} = \rho^m$;

(2) $g = \underbrace{\tau\sigma\ldots\tau\sigma}_{2m} = \rho^{-m}$;

(3) $g = \underbrace{\sigma\tau\ldots\sigma\tau\sigma}_{2m+1} = \sigma\rho^{-m}$;

(4) $g = \underbrace{\tau\sigma\ldots\tau\sigma\tau}_{2m+1} = \sigma\underbrace{\sigma\tau\ldots\sigma\tau}_{2m+2} = \sigma\rho^{m+1}$,

for some $m \in \mathbb{N}_0$. This shows that $G \subseteq \langle\rho\rangle \cup \sigma\langle\rho\rangle$, and since $\langle\rho\rangle$ contains at most n elements (as $\rho^n = 1$), we can conclude that $|G| \leq 2n$. In view of Remark 2.2.2, the existence of the surjective map $\xi : G \to D_n$ shows that ξ must be a bijection. Therefore, ξ is a group isomorphism between G and D_n, so (2.4) is indeed a presentation of D_n.

Example 2.2.4. In general every group G will have infinitely many different presentations because there are infinitely many epimorphisms from free groups to G. In this example we will show that the dihedral group D_n from Example 2.2.3 also has the presentation

$$\langle \sigma, \rho \mid \sigma^2 = 1,\ \rho^n = 1,\ \sigma\rho\sigma^{-1} = \rho^{-1}\rangle. \tag{2.5}$$

Rather than proving this directly, as we did in Example 2.2.3, we will use a different method, relying on the fact that we already have the presentation (2.4) for D_n.

Let

$$G_1 = \langle \sigma_1, \tau_1 \mid \sigma_1^2,\ \tau_1^2,\ (\sigma_1\tau_1)^n\rangle \text{ and } G_2 = \langle \sigma_2,\ \rho_2 \mid \sigma_2^2,\ \rho_2^n,\ \sigma_2\rho_2\sigma_2^{-1}\rho_2\rangle.$$

Thus $G_1 \cong D_n$ by Example 2.2.3, and the given presentation for G_2 is a simple re-writing of (2.5).

Our goal is to show that $G_1 \cong G_2$. To this end, we define the set map $\phi : \{\sigma_1, \tau_1\} \to G_2$ by $\phi(\sigma_1) = \sigma_2$ and $\phi(\tau_1) = \sigma_2\rho_2$. Then

$$\phi(\sigma_1)^2 = \sigma_2^2 = 1,\ \phi(\tau_1)^2 = (\sigma_2\rho_2)^2 = \sigma_2\rho_2\sigma_2^{-1}\rho_2 = 1 \text{ and } (\phi(\sigma_1)\phi(\tau_1))^n = (\sigma_2^2\rho_2)^n = 1 \text{ in } G_2.$$

Thus we have checked that ϕ sends the defining relators of G_1 to relators of G_2. Therefore, by Theorem 2.2.1, we have a group homomorphism $\tilde{\phi} : G_1 \to G_2$ which extends ϕ. To show that $\tilde{\phi}$ is actually an isomorphism, we will prove that it has an inverse using a similar method.

Consider the set map $\psi : \{\sigma_2, \rho_2\} \to G_1$, defined by $\psi(\sigma_2) = \sigma_1$ and $\psi(\rho_2) = \sigma_1\tau_1$. As before, one can easily check that this map extends to a group homomorphism $\tilde{\psi} : G_2 \to G_1$. Moreover,

$$(\tilde{\psi} \circ \tilde{\phi})(\sigma_1) = \sigma_1 \text{ and } (\tilde{\psi} \circ \tilde{\phi})(\tau_1) = \tilde{\psi}(\sigma_2\rho_2) = \sigma_1\sigma_1\tau_1 = \tau_1.$$

Thus the homomorphism $\tilde{\psi} \circ \tilde{\phi} : G_1 \to G_1$ sends each generator of G_1 to itself, hence it must be the identity automorphism (for example, by Lemma 0.5.11.(i)), so $\tilde{\psi} \circ \tilde{\phi} = \mathrm{Id}_{G_1}$. Similarly, we can check that $\tilde{\phi} \circ \tilde{\psi} = \mathrm{Id}_{G_2}$, which shows that $\tilde{\psi} = \tilde{\phi}^{-1}$, so $\tilde{\phi} : G_1 \to G_2$ is an isomorphism, as required.

The following observation should be intuitively clear.

Lemma 2.2.5. *Let $\langle X \mid R \rangle$ be a presentation of a group G, let w be a reduced word over $X^{\pm 1}$, and let $h \in G$ be the element represented by w. If $N = \langle\!\langle g \rangle\!\rangle^G \lhd G$ then the quotient G/N has the presentation $\langle X \mid R \cup \{w\} \rangle$.*

Proof. Since $G = \langle X \mid R \rangle$, we have an epimorphism $\phi : F(X) \to G$, where $F(X)$ is the free group of X and $\ker \phi = \langle\!\langle R \rangle\!\rangle^{F(X)} \lhd F(X)$. By Theorem 0.4.13, $F(X)/\phi^{-1}(N) \cong G/N$, so it's sufficient to show that $\phi^{-1}(N) = \langle\!\langle R \cup \{w\} \rangle\!\rangle^{F(X)}$ in $F(X)$.

Since $R \subseteq \ker \phi$ and $\phi(w) = g \in N$, we see that $R \cup \{w\} \subseteq \phi^{-1}(N)$, hence $\langle\!\langle R \cup \{w\} \rangle\!\rangle^{F(X)} \subseteq \phi^{-1}(N)$, by the definition of normal closure. To establish the opposite inclusion, note that since $\phi(w) = g$, we can apply Exercise 2.1.5 to deduce that $N = \langle\!\langle g \rangle\!\rangle^G = \phi(\langle\!\langle w \rangle\!\rangle^{F(X)})$. Since for any element $v \in \phi^{-1}(N)$ we have $\phi(v) \in N$, there must exist $u \in \langle\!\langle w \rangle\!\rangle^{F(X)}$ such that $\phi(v) = \phi(u)$ in G, i.e., $u^{-1}v \in \ker \phi = \langle\!\langle R \rangle\!\rangle^{F(X)}$. It follows that

$$v \in u \ker \phi \subseteq \langle\!\langle w \rangle\!\rangle^{F(X)} \langle\!\langle R \rangle\!\rangle^{F(X)} \subseteq \langle\!\langle R \cup \{w\} \rangle\!\rangle^{F(X)}.$$

Hence, we have shown that $\phi^{-1}(N) \subseteq \langle\!\langle R \cup \{w\} \rangle\!\rangle^{F(X)}$, so $\phi^{-1}(N) = \langle\!\langle R \cup \{w\} \rangle\!\rangle^{F(X)}$. Consequently, $F(X)/\langle\!\langle R \cup \{w\} \rangle\!\rangle^{F(X)} \cong G/N$ and the lemma is proved. □

Example 2.2.6. According to Example 2.2.4, for each $n \geq 3$, the dihedral group D_n has the presentation (2.5). Suppose that $m \geq 3$ is a natural number dividing n. Then, in view of Lemma 2.2.5, the quotient $Q = D_n/\langle\!\langle \rho^m \rangle\!\rangle^{D_n}$ has the presentation

$$Q = \langle \sigma, \rho \mid \sigma^2 = 1,\ \rho^n = 1,\ \sigma\rho\sigma^{-1} = \rho^{-1},\ \rho^m = 1 \rangle.$$

Observe that the relation $\rho^n = 1$ is a consequence of the relation $\rho^m = 1$, because m divides n, so, by Exercise 2.1.12,

$$Q \cong \langle \sigma, \rho \mid \sigma^2 = 1,\ \rho^m = 1,\ \sigma\rho\sigma^{-1} = \rho^{-1} \rangle \cong D_m.$$

Thus the dihedral group D_m is a quotient of D_n, whenever m divides n.

2.3. Finite group presentations

Definition 2.3.1. A presentation $\langle X \mid R \rangle$ of a group G is called *finite* if $|X| < \infty$ and $|R| < \infty$. A group G is *finitely presented* if it has a finite presentation.

Examples 2.2.3 and 2.2.3 give two different finite presentations (2.4) and (2.5) for the dihedral group D_n.

The easiest example of a group that is not finitely presented is the free group of infinite rank, which is not even finitely generated (see Lemma 1.1.5). There are also finitely generated groups which do not have finite presentations. Explicit examples of such groups are harder to produce, however, a theorem of Bernhard Neumann states that there exists a continuum of 2-generated groups, while it is easy to see that there can be at most countably many of finitely presented groups (because there are only countably many finite group presentations!). Thus set-theoretic considerations show that most finitely generated groups are not finitely presented.

Proposition 2.3.2. *Every finite group is finitely presented.*

Proof. Let G be a finite group. We can enumerate its elements $G = \{g_1, \dots, g_n\}$, so that $g_1 = 1$, where $n = |G| \in \mathbb{N}$. Let $X = \{x_1, \dots, x_n\}$ be a finite set of cardinality n, and let $F = F(X)$ be the free group on X. Since G is finite, its multiplication table can be encoded in an $n \times n$ array, giving us n^2 relations in G. More precisely, let $\theta : \{1, \dots, n\} \times \{1, \dots, n\} \to \{1, \dots, n\}$ be the function such that

$$g_i\, g_j = g_{\theta(i,j)}, \quad \text{for all } i, j \in \{1, \dots, n\}.$$

Consider the group P given by the presentation

$$P = \langle X \mid x_i x_j = x_{\theta(i,j)}, \quad \text{for all } i, j \in \{1, \dots, n\}\rangle \tag{2.6}$$

Evidently P is finitely presented, so it remains to show that $P \cong G$. By von Dyck's Theorem, there is a homomorphism $\phi : P \to G$ such that $\phi(x_i) = g_i$, for all $i = 1, \dots, n$. In particular, ϕ is surjective. Note that X is a generating set of P and the defining relators from (2.6) can be used to completely describe the multiplication table in P (alternatively, see Exercise 2.3.3), which implies that every element of P is equal to some $x_i \in X$, thus $|P| \le n$. We can therefore use Remark 2.2.2 to conclude that $\phi : P \to G$ is a group isomorphism. Thus G is finitely presented. ☐

Exercise 2.3.3. Suppose that X is a finite subset of a group P such that $X\,X \subseteq X$ in P. Prove that X is a subgroup of P. Give an example showing that this may fail to be true if X is infinite.

The presentation (2.6) of a finite group G, provided by Proposition 2.3.2, is far from efficient. Indeed, for a cyclic group C_n, of order $n \in \mathbb{N}$, it gives a presentation with n generators and n^2 defining relators, while the easiest presentation for C_n has 1 generator and 1 defining relator (see Example 2.1.10.(b)).

In principle, a presentation of a group completely determines it. However, extracting information from such a presentation may not be an easy task. Some questions are even undecidable in general. For example, by the famous Adian-Rabin theorem, there exists no algorithm taking on input a finite group presentation $\langle X \mid R\rangle$ and deciding whether the group that it represents is trivial or non-trivial.

Example 2.3.4. Given any non-zero integers $k, l \in \mathbb{Z}$ we can consider the Baumslag-Solitar group

$$BS(k, l) = \langle a, t \mid ta^k t^{-1} = a^l\rangle.$$

Let's show that this group is always infinite.

Indeed, consider the infinite cyclic group $C_\infty = \langle x\rangle$ and the set map $\phi : \{a, t\} \to C_\infty$ defined by $\phi(a) = 1$ and $\phi(t) = x$. Since

$$\phi(t)\phi(a)^k\phi(t)^{-1} = x\,x^{-1} = 1 = \phi(a)^l \;\text{ in } C_\infty,$$

by Theorem 2.2.1 there is a group homomorphism $\tilde{\phi} : G \to C_\infty$ extending ϕ. In particular, $x = \tilde{\phi}(t) \in \operatorname{im}\tilde{\phi}$, hence $\tilde{\phi}$ is surjective by Lemma 0.5.11.(ii). Since G admits a surjective homomorphism onto the infinite group C_∞, we can conclude that $|G| = \infty$, as claimed.

Example 2.3.5. The third Fibonacci group $F(2, 3)$ is defined by the presentation

$$F(2, 3) = \langle a, b, c \mid ab = c,\; bc = a,\; ca = b\rangle.$$

Let us show that $a^4 = b^4 = (ab)^4 = 1$ in this group. Indeed, using the first defining relation $c = ab$ we can re-write the two other defining relations as

$$\left\{\begin{array}{l} bab = a \\ aba = b \end{array}\right. \implies \left\{\begin{array}{l} ba = ab^{-1} \\ a^2 = b^2 \end{array}\right. \implies \left\{\begin{array}{l} a^2 = b^2 \\ a^{-1}ba = b^{-1} \end{array}\right. \implies \left\{\begin{array}{l} a^2 = b^2 \\ a^{-1}b^2a = b^{-2} \end{array}\right. \implies \left\{\begin{array}{l} a^2 = b^2 \\ a^2 = a^{-2} \end{array}\right. .$$

The latter equation implies that $a^4 = 1$. Since $b^2 = a^2$, we deduce that $b^4 = 1$. Finally, since $c = ab$ and $aba = b$, we obtain $c^2 = abab = b^2$, so $c^4 = b^4 = 1$ in $F(2, 3)$.

Exercise 2.3.6. Prove that the group $F(2, 3)$ from Example 2.3.5 is isomorphic to the quaternion group Q_8.

2.4. Free products of groups

One important use of groups presentations stems from the fact that they can be used to construct new groups from the existing ones. Free products introduced in this section are one example of such a construction.

Definition 2.4.1. Let I be an index set, and let $\{A_i\}_{i\in I}$ be a family of groups with presentations $A_i = \langle X_i \mid R_i\rangle$. The *free product* $*_{i\in I} A_i$, of this family, is the group defined by the presentation $\langle X \mid R\rangle$, where $X = \bigsqcup_{i\in I} X_i$ and $R = \bigcup_{i\in I} R_i$. The groups A_i, $i \in I$, are called *(free) factors* of the free product.

Note that in the above definition we think of each R_i, $i \in I$, as a subset of the free group $F(X_i)$, consisting of reduced words over $X_i^{\pm 1}$. Since $X = \bigsqcup_{i\in I} X_i$, every reduced word from R_i is also a reduced word over $X^{\pm 1}$, so that the union $R = \bigcup_{i\in I} R_i$ is a collection of reduced words from $F(X)$.

Example 2.4.2. (a) If $I = \{1, \dots, n\}$, for some $n \in \mathbb{N}$, and $A_i = \langle X_i \mid R_i\rangle$ then the free product $A_1 * \cdots * A_n$ has the presentation

$$\langle X_1 \sqcup \cdots \sqcup X_n \mid R_1 \sqcup \cdots \sqcup R_n\rangle.$$

In particular, if $X_i = \{x_i\}$ and $R_i = \emptyset$ (so that $A_i = \langle x_i \mid \, \rangle$ is the infinite cyclic group), for each i, then $A_1 * \cdots * A_n = \langle x_1, \dots, x_n \mid \, \rangle$ is the free group of rank n.

(b) Let $A = \langle x \mid x^2 = 1\rangle$ and $B = \langle y \mid y^3 = 1\rangle$ be cyclic groups of orders 2 and 3 respectively. Then the free product $A * B$ has the presentation

$$A * B = \langle x, y \mid x^2 = 1,\ y^3 = 1\rangle.$$

Exercise 2.4.3. Show that the free product of groups is associative, that is, if A, B and C are some groups then $(A * B) * C \cong A * (B * C)$.

The reason why we call this construction a free *product* is the fact that every free factor naturally embeds in the free product.

Proposition 2.4.4. *Suppose that A_i, $i \in I$, are groups with presentations $A_i = \langle X_i \mid R_i\rangle$, $i \in I$, $X = \bigsqcup_{i\in I} X_i$ and $R = \bigsqcup_{i\in I} R_i$. Then for every $i \in I$ the inclusion $X_i \hookrightarrow X$ gives rise to an injective homomorphism (an* embedding*) of A_i in the free product $*_{i\in I} A_i$.*

Proof. Take any $i \in I$. By Theorem 2.2.1, there is a group homomorphism $\tilde\phi : A_i \to G = *_{i\in I} A_i$, extending the natural inclusion ϕ of X_i in G (here we identify X with its image in G). To show that ϕ is injective, we will find a left inverse for it. To this end, we define the map $\psi : X \to A_i$ by $\psi(x) = x$, for all $x \in X_i$, and $\psi(y) = 1$, for all $y \in X \setminus X_i$. It is easy to see that all the assumptions of Theorem 2.2.1 are satisfied, so ψ extends to a group homomorphism $\tilde\psi : G \to A_i$. Clearly the composition $\tilde\psi \circ \tilde\phi$ defines the identity map on X_i:

$$\tilde\psi(\tilde\phi(x)) = \tilde\psi(x) = x, \quad \text{for all } x \in X_i.$$

Since the identity automorphism $\mathrm{Id}_{A_i} : A_i \to A_i$ has the same restriction to X_i as this composition and $A_i = \langle X_i\rangle$, Lemma 0.5.11.(a) tells us that $\tilde\psi \circ \tilde\phi = \mathrm{Id}_{A_i}$. The later implies that $\tilde\phi$ is injective, as required. □

Convention 2.4.5. Proposition 2.4.4 tells us that each free factor naturally embeds in a free product. We will often use this fact to identify every free factor with its image under this embedding.

This convention, together with Definition 2.4.1, immediately imply the following observation.

Remark 2.4.6. If $G = *_{i\in I} A_i$ then $\bigcup_{i\in I} A_i$ is a generating set of G.

Theorem 2.4.7 (Universal Property of free products). *Let I be an index set, let $\{A_i\}_{i\in I}$ be a family of groups and let $G = *_{i\in I} A_i$. Suppose that for some group H we are given homomorphisms $\phi_i : A_i \to H$, $i \in I$. Then there is a unique homomorphism $\psi : G \to H$, whose restriction to A_i is ϕ_i (i.e., $\psi(a) = \phi_i(a)$, for all $a \in A_i$), for each $i \in I$.*

Proof. The argument is an easy consequence of von Dyck's Theorem and is left as an exercise for the reader. □

An important issue arising from Definition 2.4.1 that we have not yet addressed is the question whether a free product $G = *_{i\in I} A_i$ depends on the choice of the presentations for the factors A_i, $i \in I$.

Proposition 2.4.8. *Let* $\{A_i\}_{i\in I}$ *and* $\{B_i\}_{i\in I}$ *be two families of groups indexed by the same set* I. *If* $A_i \cong B_i$, *for each* $i \in I$, *then the free products* $G = *_{i\in I}A_i$ *and* $H = *_{i\in I}B_i$ *are isomorphic.*

Proof. Let $\phi_i : A_i \to B_i$ be an isomorphism, $i \in I$. By Proposition 2.4.4 we can think of each B_i as a subgroup of H, so that ϕ_i becomes a homomorphism from A_i to H, $i \in I$. Therefore, we can apply Theorem 2.4.7 to get a homomorphism $\psi_1 : G \to H$, whose restriction to A_i is ϕ_i, for all $i \in I$. We can similarly find a homomorphism $\psi_2 : H \to G$, whose restriction to B_i is ϕ_i^{-1}, for each $i \in I$.

Note that the composition $\psi_2 \circ \psi_1 : G \to G$ restricts to the identity map on each A_i, $i \in I$. The "uniqueness" claim from Theorem 2.4.7, then tells us that $\psi_2 \circ \psi_1 = \mathrm{Id}_G$. Similarly, $\psi_1 \circ \psi_2 = \mathrm{Id}_H$, hence $\psi_2 = \psi_1^{-1}$, so $\psi_1 : G \to H$ is a group isomorphism. □

We can use Proposition 2.4.8 to immediately deduce that the free product is a group-theoretic construction, i.e., the resulting group only depends on the isomorphism types of the factors:

Corollary 2.4.9. *Given a family of groups* $\{A_i\}_{i\in I}$, *the free product* $G = *_{i\in I}A_i$ *does not depend on the choices of presentations for the groups* A_i, $i \in I$, *up to isomorphism.*

2.5. Normal forms in free products

In this section we will discuss a more explicit realization of free products, using normal forms that, which is a sense generalizes the construction of free groups using reduced words, in Section 1.2.

Throughout this section we assume that $\{A_i\}_{i\in I}$ is a family of groups indexed by a set I. We can then consider the disjoint union $\bigsqcup_{i\in I} A_i$ as an alphabet.

Definition 2.5.1. A *word* over the alphabet $\bigsqcup_{i\in I} A_i$ is a finite sequence

$$w = a_1a_2\dots a_n, \quad \text{where for each } j \in \{1,\dots,n\} \text{ there is } i_j \in I \text{ such that } a_j \in A_{i_j}. \tag{2.7}$$

The elements $a_1,\dots,a_n$ are called *syllables* of w and n is called the *length* of w. If we treat each A_i as a subgroup of the free product $G = *_{i\in I}A_i$ then the word w *represents* the element $g = a_1 \cdot a_2 \cdots a_n \in G$.

Definition 2.5.2. A non-empty word w from (2.7) is said to be *reduced* if the following two conditions are satisfied:

- $a_j \neq 1$ in A_{i_j}, for every $j = 1,\dots,n$;
- $i_j \neq i_{j+1}$, for $j = 1,\dots,n-1$.

The empty word over $\bigsqcup_{i\in I} A_i$ is reduced, by definition.

A reduced word w representing some element of $g \in *_{i\in I}A_i$ is called a *normal form* (*reduced form*) of g.

In other words, the word w is reduced if all of its syllables are non-trivial elements of the corresponding groups A_i and no two consecutive syllables are in the same free factor. By Remark 2.4.6, every element of the free product $G = *_{i\in I}A_i$ can be represented as a word $w = a_1a_2\dots a_n$ over the alphabet $\bigsqcup_{i\in I} A_i$. If this word is not reduced then it can obviously be shortened, while still representing the same element of the free product. Therefore, every $g \in G$ has at least one normal form. The following statement is an analogue of Theorem 1.2.11.

Theorem 2.5.3 (Normal Form theorem for free products). *Let* $G = *_{i\in I}A_i$, *where* I *is an index set and* $\{A_i\}_{i\in I}$ *is a family of groups. Then every element of* G *has a unique normal form.*

Proof. The existence of a normal forms has already been explained, so we only need to show the uniqueness. To this end we will employ an argument invented by Bartel Leendert van der Waerden. Let $\mathcal{W}$ denote the set of all reduced words over $\bigsqcup_{i\in I} A_i$, including the empty word e. We define an action of the free product $G = *_{i\in I}A_i$ on $\mathcal{W}$ as follows. Given any $i \in I$ and $a \in A_i$, a will act on $\mathcal{W}$ by a permutation $\overline{a}$. If $a = 1$ in A_i then $\overline{a}$ is the identity permutation, otherwise for every reduced word $a_1a_2\dots a_n \in \mathcal{W}$ we set

$$\overline{a}(a_1a_2\dots a_n) = \begin{cases} aa_1a_2\dots a_n, & \text{if } a_1 \in A_j, \text{ for some } j \in I \setminus \{i\} \\ (aa_1)a_2\dots a_n, & \text{if } a_1 \in A_i, \text{ and } aa_1 \neq 1 \text{ in } A_i \\ a_2\dots a_n, & \text{if } a_1 = a^{-1} \text{ in } A_i \end{cases} . \tag{2.8}$$

We can easily check that $\overline{a}$ is a permutation of $\mathcal{W}$ by verifying that $\overline{a^{-1}}$ induces the inverse permutation. Thus, $\overline{a}$ belongs to the group $\mathrm{Sym}(\mathcal{W})$, of all permutations of the set $\mathcal{W}$. Moreover, the map $a \mapsto \overline{a}$ gives rise to a group homomorphism from A_i to $\mathrm{Sym}(\mathcal{W})$, for each $i \in I$. Indeed, if $a, b \in A_i$ then, by examining the formula (2.8), one sees that $\overline{ab} = \overline{a}\,\overline{b}$ in $\mathrm{Sym}(\mathcal{W})$.

Thus for each $i \in I$ we have a homomorphism $A_i \to \mathrm{Sym}(\mathcal{W})$, which, by Theorem 2.4.7, can be extended to a homomorphism $\psi : G \to \mathrm{Sym}(\mathcal{W})$. Now, suppose that $u = a_1 \dots a_n$ and $v = b_1 \dots b_m$ are two distinct reduced words over $\bigsqcup_{i\in I} A_i$. Let g and h be the two elements of the free product G represented by these words. Then, according to (2.8), we have

$$\psi(g)(e) = a_1 \dots a_n = u \quad\text{and}\quad \psi(h)(e) = b_1 \dots b_m = v.$$

Since $u \neq v$ in $\mathcal{W}$ we can conclude that $\psi(g) \neq \psi(h)$ in $\mathrm{Sym}(\mathcal{W})$, hence $g \neq h$ in G. This shows that distinct reduced words represent distinct elements of G, so the proof is complete. □

Example 2.5.4. Let $G = \langle x, y \mid x^3 = 1,\ y^5 = 1\rangle$. Then $G = A * B$, where $A = \langle x\rangle_3 \cong C_3$ and $B = \langle y\rangle_5 \cong C_5$. We can write $A = \{1, x, x^2\}$ and $B = \{1, y, y^2, y^3, y^4\}$, so the elements of G will be represented by words over $A \sqcup B$.

For example, let us consider the elements $f, g, h \in G$ represented by the reduced words $x^2y^4xy^2$, $y^3x^2yx^2$ and $xyx^2y^3x^2$ respectively. Then f, g and h are pairwise distinct in G by Theorem 2.5.3 because their reduced words are different. Keeping in mind that $x^3 = y^5 = 1$ in G, we can find the reduced words for fg and h^{-1} in G as follows:

$$fg = (x^2y^4xy^2)\,(y^3x^2yx^2) = x^2y^4xy^5x^2yx^2 = x^2y^4x^3yx^2 = x^2y^5x^2 = x^4 = x, \quad\text{and}$$

$$h^{-1} = (xyx^2y^3x^2)^{-1} = x^{-2}y^{-3}x^{-2}y^{-1}x^{-1} = xy^2xy^4x^2.$$

Example 2.5.5. The *infinite dihedral group* D_∞ can be defined as the group of isometries of the real line $\mathbb{R}$ preserving the integers. It can also be viewed as the group of all symmetries of the regular simplicial line, see Figure 2.2.

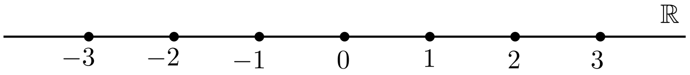


FIGURE 2.2. The simplicial line

Every element of D_∞ is either a reflection in $n/2$, for some $n \in \mathbb{Z}$, or a translation by an integer $m \in \mathbb{Z}$. It is easy to see that D_∞ is generated by the reflection σ_0 about the origin ($\sigma_0 : \mathbb{R} \to \mathbb{R}$, $x \overset{\sigma_0}{\mapsto} -x$) and by the reflection τ_0 about $1/2$ ($\tau_0 : \mathbb{R} \to \mathbb{R}$, $x \overset{\tau_0}{\mapsto} 1 - x$). Note that the composition $\rho_0 = \sigma_0 \circ \tau_0$ is the translation by -1: $\rho_0 : \mathbb{R} \to \mathbb{R}$, $\rho_0(x) = x - 1$. In particular, $\sigma_0^2 = 1 = \tau_0^2$ and ρ_0 has infinite order in D_∞.

Let G be the group with presentation

$$\langle \sigma, \tau \mid \sigma^2 = 1,\ \tau^2 = 1\rangle. \tag{2.9}$$

Clearly, $G \cong C_2 * C_2$ is the free product of two cyclic groups of order 2, generated by σ and τ respectively. By the Universal Property of free products Theorem 2.4.7 (or by von Dyck's Theorem 2.2.1), there is a group homomorphism $\psi : G \to D_\infty$ such that $\psi(\sigma) = \sigma_0$ and $\psi(\tau) = \tau_0$.

Note that ψ is surjective because $D_\infty = \langle \sigma_0, \tau_0\rangle$ (see Lemma 0.5.11.(ii)). To show that ψ is injective, denote $\rho = \sigma\tau \in G$ and suppose that $g \in G$ is a non-trivial element. Then, according to Theorem 2.5.3, g can be written as a non-empty word $\sigma\tau\sigma\tau\dots$ or $\tau\sigma\tau\sigma\dots$. Just like in Example 2.2.3, we can see that $g = \rho^m$, for some $m \in \mathbb{Z} \setminus \{0\}$ (if the length of the word is even), or $g = \sigma\rho^n$, for some $n \in \mathbb{Z}$ (if this length is odd). In the former case $\psi(g) = \rho^m \neq 1$ in D_∞ because it is the translation of $\mathbb{R}$ by $-m$, and in the latter case $\psi(g) = \sigma_0\rho_0^n \neq 1$ because it has the formula

$\psi(g)(x) = n - x$, for all $x \in \mathbb{R}$ (it is the reflection of $\mathbb{R}$ in $n/2$). Thus we have shown that $\psi(g) \neq 1$ for every $g \in G \setminus \{1\}$, so ψ is injective.

Therefore $\psi : G \to D_\infty$ is a group isomorphism, which shows that $D_\infty \cong C_2 * C_2$ and (2.9) is a presentation of D_∞.

Exercise 2.5.6. Suppose that A and B are non-trivial groups. Prove that the free product $A * B$ is infinite.

2.6. Presentations of direct products

Recall that for groups A and B their *direct product* $A \times B$ is the group whose elements are pairs (a, b), where $a \in A$ and $b \in B$, and the product is defined coordinate-wise:

$$(a_1, b_1)\,(a_2, b_2) = (a_1 a_2, b_1, b_2), \quad \text{for all } (a_1, b_1), (a_2, b_2) \in A \times B.$$

Note that A and B embed as subgroups into $A \times B$ under the homomorphisms $a \mapsto (a, 1)$, for all $a \in A$, and $b \mapsto (1, b)$, for all $b \in B$. Moreover these copies of A and B in $A \times B$ element-wise commute with each other:

$$(a, 1)(1, b) = (a, b) = (1, b)(a, 1), \quad \text{for all } a \in A,\ b \in B.$$

We can similarly define direct products for arbitrary (possibly infinite) families of groups.

Definition 2.6.1. Let I be an index set, and let $\{A_i\}_{i\in I}$ be a family of groups indexed by I. The direct product $\times_{i\in I} A_i$ is the group whose elements are functions of $f : I \to \bigsqcup_{i\in I} A_i$ such that $f(i) \in A_i$, for all $i \in I$, and $f(i) \neq 1_{A_i}$ only for finitely many $i \in I$ (i.e., each function has *finite support*). The multiplication in this group is defined coordinate-wise: given $f, g \in \times_{i\in I} A_i$, we set

$$(fg)(i) = f(i)\, g(i), \quad \text{for all } i \in I.$$

The groups A_i, $i \in I$, are called *direct factors* of $\times_{i\in I} A_i$. When $|I| = n$ is finite, elements of $\times_{i\in I} A_i$ can be regarded as n-tuples, where for each $i \in I$ the i-th coordinate belongs to the i-th group A_i.

Let $\operatorname{supp}(f)$ denote the support of any function $f \in \times_{i\in I} A_i$, i.e.,

$$\operatorname{supp}(f) = \{i \in I \mid f(i) \neq 1_{A_i}\}.$$

The above definition implies that if $f, g \in \times_{i\in I} A_i$, then $\operatorname{supp}(fg) \subseteq \operatorname{supp}(f) \cup \operatorname{supp}(g)$, hence fg is again a function of finite support. Thus the multiplication in Definition 2.6.1 is well-defined. It is also not difficult to check that $\times_{i\in I} A_i$ is a group under this multiplication: the identity element will be the function $e : I \to \sqcup_{i\in I} A_i$ such that $e(i) = 1_{A_i}$, for all $i \in I$, and for a function $f \in \times_{i\in I} A_i$ we define its inverse $f^{-1} \in \times_{i\in I} A_i$ by $f^{-1}(i) = f(i)^{-1} \in A_i$, for all $i \in I$.

Example 2.6.2. (a) Suppose that $m, n \in \mathbb{N}$. As we know from the basic Group Theory course, if m and n are coprime then $C_m \times C_n \cong C_{mn}$, otherwise the direct product $C_m \times C_n$ is not cyclic.

(b) For any set I the direct product $\mathbb{Z}^I = \times_{i\in I} \mathbb{Z}$ is called the *free abelian group of rank* $|I|$.

Just like in the case of two factors, every A_i can be considered as a subgroup of the direct product $\times_{i\in I} A_i$ and these subgroups commute element-wise.

Lemma 2.6.3. *Consider a family of groups $\{A_i\}_{i\in I}$, where I is an index set, and let $G = \times_{i\in I} A_i$.*

(a) Given any $j \in I$ we have a natural monomorphism $\eta_j : A_j \to G$ defined as follows: for every $a \in A_j$, $\eta(a)$ is the function from I to $\bigsqcup_{i\in I} A_i$ such that $\eta_j(a)(j) = a$ and $\eta_j(a)(i) = 1_{A_i}$, provided $i \in I \setminus \{j\}$.

(b) The direct product G is generated by the subset $\bigcup_{j\in I} \eta_j(A_j)$.

(c) If $j, k \in I$ are distinct indices and $f \in \eta_j(A_j)$, $g \in \eta_k(A_k)$ then $fg = gf$ in G.

Proof. The proof of (a) is an exercise. Claim (b) is also easy to verify. Indeed, choose any element $g \in G$. Then g is a function $g : I \to \bigsqcup_{i\in I} A_i$ that has finite support $J \subseteq I$, with $|J| < \infty$. For each $j \in J$ consider the function $f_j \in \eta_j(A_j)$ defined by $f_j(j) = g(j)$ and $f_j(i) = 1_{A_i}$, for $i \in I \setminus \{j\}$. Clearly, $f_j = \eta_j(g(j)) \in \eta_j(A_j)$ and $g = \prod_{j\in J} f_j$ in G, which implies that $G = \langle \eta_j(A_j) \mid j \in I \rangle$.

To establish claim (c), we only need to observe that both fg and gf result in the same function $h : I \to \bigsqcup_{i\in I} A_i$ such that $h(j) = g(j) \in A_j$, $h(k) = g(k) \in A_k$ and $h(i) = 1_{A_i}$ for all $i \in I \setminus \{j,k\}$. □

Lemma 2.6.3 allows us to treat direct factors as commuting subgroups of a direct product; moreover, the direct product is generated by these subgroups.

Direct products possess a useful Universal Property.

Proposition 2.6.4 (Universal Property of direct products)**.** *Let $\{A_i\}_{i\in I}$ be a family of groups indexed by a set I, let $G = \times_{i\in I} A_i$ and let H be any group. Suppose that we have a family of homomorphisms $\phi_i : A_i \to H$, $i \in I$, such that elements from the images of A_i and A_j pairwise commute in H, whenever $i, j \in I$ are distinct indices (in other words, $\phi_i(a)\phi_j(b) = \phi_j(b)\phi_i(a)$ in H, for all $i, j \in I$, $i \neq j$, and all $a \in A_i$, $b \in A_j$). Then there exists a unique group homomorphism $\phi : G \to H$ whose restriction to (the copy of) A_j is ϕ_j, for every $j \in I$. More formally, the latter means that for each $j \in I$ we have the following commutative diagram (where $\eta_j : A_j \to G$ is defined in Lemma 2.6.3):*

$$\begin{tikzcd}
A_j \arrow[r, "\eta_j"] \arrow[dr, "\phi_j"'] & G \arrow[d, dashed, "\phi"] \\
& H
\end{tikzcd}$$

Proof. Let us start by defining the map $\phi : G \to H$. For any function $g : I \to \bigsqcup_{i\in I} A_i$ in G, we set

$$\phi(g) = \prod_{i\in I} \phi_i(g(i)) \in H.$$

Note that in the latter product only finitely many elements $\phi_i(g(i))$ are non-trivial in H because g has finite support. To show that ϕ is a homomorphism, suppose that $f : I \to \bigsqcup_{i\in I} A_i$ is another element of G. Since every element of $\phi_i(A_i)$ commutes with every element of $\phi_j(A_j)$ in H, for $i \neq j$, we have

$$\phi(fg) = \prod_{i\in I} \phi_i((fg)(i)) = \prod_{i\in I} \phi_i(f(i)g(i)) = \prod_{i\in I} \phi_i(f(i)) \prod_{i\in I} \phi_i(g(i)) = \phi(f)\,\phi(g) \text{ in } H.$$

Thus ϕ is a group homomorphism. By construction, its restriction to the natural copy of A_j in G is ϕ_j, for all $j \in I$. The uniqueness of ϕ follows from Lemma 2.6.3.(b) and Lemma 0.5.11.(i). □

Example 2.6.5. Let $m, n \in \mathbb{N}$ be two distinct primes, and let A be an abelian group of order mn. By Cauchy's theorem there exist elements $a, b \in A$ of orders m and n respectively. The cyclic subgroups $\langle a\rangle$ and $\langle b\rangle$ of A are obviously isomorphic to C_m and C_n respectively, thus we have group monomorphisms $\phi_1 : C_m \to A$ and $\phi_2 : C_n \to A$, where $C_m = \langle x\rangle$ and $C_n = \langle y\rangle$, such that $\phi_1(x) = a$ and $\phi_2(y) = b$. By Proposition 2.6.4, there exists a homomorphism $\phi : C_m \times C_n \to H$ such that $\phi((x,1)) = \phi_1(x) = a$ and $\phi((1,y)) = \phi_2(y) = b$. It follows that

$$\phi((x^k, y^l)) = \phi((x,1)^k (1,y)^l) = \phi((x,1))^k\, \phi((1,y))^l = a^k\, b^l \text{ in } A, \quad \text{for all } k, l \in \mathbb{Z}.$$

Suppose that $a^k b^l = 1$ in A, for some $k, l \in \mathbb{Z}$. Then $a^{-k} = b^l$ is an element whose order divides both m and n in A, hence $a^{-k} = b^l = 1$ in A because m and n are distinct primes. It follows that k is divisible by m and l is divisible by n, which implies $(x^k, y^l) = (1,1)$ in $C_m \times C_n$. Thus we have demonstrated that $\ker\phi = \{(1,1)\}$, so $\phi : C_m \times C_n \to A$ is injective. Since $|C_m \times C_n| = mn = |A|$, we can conclude that ϕ is bijective, so it is an isomorphism between $C_m \times C_n$ and A. Since C_{mn} is another abelian group of order mn, we can also deduce that $C_{mn} \cong C_m \times C_n \cong A$, i.e., every abelian group of order mn is isomorphic to the cyclic group of that order.

Exercise 2.6.6. Use Proposition 2.6.4 to show that every finitely generated abelian group is a quotient of the free abelian group $\mathbb{Z}^n$, for some $n \in \mathbb{N}$.

We are now able to describe presentations of direct products.

Theorem 2.6.7. *Let I be an index set and let $\{A_i\}_{i\in I}$ be a family of groups with presentations $A_i = \langle X_i \mid R_i\rangle$, $i \in I$. Then the direct product $G = \times_{i\in I} A_i$ has the presentation $\langle X \mid R\rangle$, where*

$$X = \bigsqcup_{i\in I} X_i \quad \textit{and} \quad R = \bigcup_{i\in I} R_i \cup \bigcup_{i,j\in I,\ i\neq j} [X_i, X_j],$$

and $[X_i, X_j] = \{[x, y] \mid x \in X_i,\ y \in X_j\}$ is the set of commutators of elements from X_i with elements from X_j in $F(X)$.

Proof. For simplicity, we will only prove this statement in the case of two factors, i.e., when $G = A \times B$, $A = \langle X \mid R\rangle$ and $B = \langle Y \mid S\rangle$. Let P be the group defined by the presentation

$$P = \langle X \sqcup Y \mid R \cup S \cup [X, Y]\rangle.$$

By von Dyck's theorem (Theorem 2.2.1), there exist group homomorphisms $\phi_1 : A \to P$ and $\phi_2 : B \to P$ such that $\phi_1(x) = x$ and $\phi_2(y) = y$, for all $x \in X$ and $y \in Y$. The defining relators $[X, Y] = \{[x, y] \mid x \in X,\ y \in Y\}$ clearly imply that each element of the subgroup $\langle X\rangle$ commutes with each element of $\langle Y\rangle$ in P, hence we can apply Proposition 2.6.4 to find a homomorphism $\phi : A \times B \to P$ such that $\phi((x, 1)) = \phi_1(x) = x$, for all $x \in X$, and $\phi((1, y)) = \phi_2(y) = y$, for all $y \in Y$.

On the other hand, another application of von Dyck's theorem gives us a homomorphism $\psi : P \to A\times B$ such that $\psi(x) = (x, 1)$ and $\psi(y) = (1, y)$, for all $x \in X$ and $y \in Y$. We now observe that the homomorphism $\phi \circ \psi : P \to P$ satisfies

$$(\phi \circ \psi)(x) = \phi((x, 1)) = x \quad \text{and} \quad (\phi \circ \psi)(y) = \phi((1, y)) = y, \quad \text{for all } x \in X,\ y \in Y.$$

Thus $\phi \circ \psi = \mathrm{Id}_P$ because P is generated by $X \cup Y$ (see Lemma 0.5.11.(i)). We can similarly check that $\psi \circ \phi = \mathrm{Id}_{A\times B}$, whence $\psi = \phi^{-1}$, so $\phi : A \times B \to P$ is an isomorphism. This finishes the proof of the theorem. □

Example 2.6.8. (a) Given two cyclic groups C_m and C_n, of orders $m, n \in \mathbb{N}$, Example 2.1.10.(b) together with Theorem 2.6.7 tell us that the direct product $C_m \times C_n$ has the presentation

$$C_m \times C_n = \langle x, y \mid x^m = 1,\ y^n = 1,\ [x, y] = 1\rangle.$$

(b) The free abelian group $\mathbb{Z}^n$, of rank $n \in \mathbb{N}_0$, has the presentation

$$\mathbb{Z}^n = \langle x_1, \dots, x_n \mid [x_i, x_j],\ \text{for all } 1 \le i < j \le n\rangle. \tag{2.10}$$

Note that, Theorem 2.6.7 also tells us to include the commutators $[x_j, x_i]$, for $i < j$, as defining relators in the above presentation (2.10), however these relators are redundant by Exercise 2.1.12, since $[x_j, x_i] = [x_i, x_j]^{-1}$.

Theorem 2.6.7 immediately implies the following.

Corollary 2.6.9. *If A and B are finitely presented groups then so is the direct product $A \times B$.*

2.7. Additional examples

Example 2.7.1. Consider the group G given by the presentation

$$G = \langle x, y \mid x^4 = 1,\ x^2 = y^3\rangle. \tag{2.11}$$

(In fact, $G \cong \mathrm{SL}(2, \mathbb{Z})$, see Corollary 5.2.3, but we'll pretend that we don't know this.)

Let us show that

(a) x has order 4 in G;
(b) y has order 6 in G.

(a) The given presentation of G implies that the order of x in G divides 4. We will show that it is exactly 4, by finding a quotient of G where the image of x has order 4.

By von Dyck's theorem (Theorem 2.2.1), there exists a homomorphism $\phi : G \to C_4 = \langle a\rangle$ such that $\phi(x) = a$ and $\phi(y) = a^2$. Indeed, it is easy to see that

$$\phi(x)^2 = a^2 = a^6 = \phi(y)^3 \ \text{ in } C_4.$$

Since $a = \phi(x)$ has order 4 in C_4, we can conclude that the order of x in G must be divisible by 4 (see Exercise 0.2.2.(iii)). Therefore this order is exactly 4.

(b) The presentation (2.11) implies that $y^6 = (y^3)^2 = x^4 = 1$, so the order of y in G divides 6. By Exercise 2.7.2 below, the image of y has order less than 6 in every homomorphism $G \to C_6$, which is why we will map G to $C_{12} = \langle b \rangle$. Note that in C_{12} we have $(b^3)^4 = 1$ and $(b^3)^2 = (b^2)^3$, hence, by Theorem 2.2.1, there is a group homomorphism $\psi : G \to C_{12}$ such that $\psi(x) = b^3$ and $\psi(y) = b^2$. Since the order of $\psi(y)$ is 6 in C_{12}, we can use Exercise 0.2.2.(iii) to conclude that 6 divides the order of y in G. Hence this order is exactly 6.

Alternative argument for (b): we can also use the homomorphism $\xi : G \to C_3 = \langle c \rangle$ such that $\xi(x) = 1$ and $\xi(y) = c$ to deduce that the order of y in G must be divisible by 3. Using the homomorphism $\phi : G \to C_4$ from part (a), we see that this order is also divisible by the order of $\phi(y)$ in C_4, which is 2. Thus the order of y in G must be divisible by 6.

Exercise 2.7.2. Let G be given by (2.11). Prove that there is no homomorphism $\psi : G \to C_6$ such that $\psi(y)$ has order 6 in C_6.

Example 2.7.3. Let P be the group given by the presentation

$$P = \langle a, b \mid ba^2b^{-1} = a^{-2},\ ab^2a^{-1} = b^{-2} \rangle. \tag{2.12}$$

This group P is sometimes called the Promislow group, but it is also isomorphic to the sixth Fibonacci group $F(2,6)$. Set $N = \langle a^2, b^2, (ab)^2 \rangle \leqslant P$. Let us show that N is an abelian normal subgroup of index 4 in P.

First of all, the defining relations from (2.12) easily imply that the three generators of N commute with each other in P, hence N is abelian. Denote $M = \langle a^2, b^2 \rangle \leqslant N$. The presentation of P shows that $aMa^{-1} = M$ and $bMb^{-1} = M$, which implies that $M \lhd P$ by Lemma 0.5.11.(iv). According to Lemma 2.2.5, the quotient group P/M has the presentation

$$P/M \cong \langle a, b \mid ba^2b^{-1} = a^{-2},\ ab^2a^{-1} = b^{-2},\ a^2 = 1,\ b^2 = 1 \rangle.$$

Note that the first pair of defining relations in this presentation are consequences of the second pair. Therefore, the first two relations are redundant (see Exercise 2.1.12), whence

$$P/M \cong \langle a, b \mid a^2 = 1,\ b^2 = 1 \rangle$$

is the infinite dihedral group (as seen in Example 2.5.5). Thus we have a homomorphism $\xi : P \to D_\infty$ such that $\ker \xi = M$, $\xi(a) = \sigma_0$ and $\xi(b) = \tau_0$ (here we use the same notation as in Example 2.5.5).

Observe that for $\rho_0 = \sigma_0\theta_0$ and any $n \in \mathbb{N}$, we have

$$\sigma_0\rho_0^n\sigma_0^{-1} = \rho_0^{-n} \text{ and } \tau_0\rho_0^n\tau_0^{-1} = \rho_0^{-n},$$

hence the subgroup $\langle \rho_0^2 \rangle$ is normal in D_∞. By Lemma 2.2.5, the quotient $D_\infty/\langle \rho_0^2 \rangle$ has the presentation

$$D_\infty/\langle \rho_0^2 \rangle \cong \langle \sigma, \tau \mid \sigma^2,\ \tau^2,\ (\sigma\tau)^2 \rangle,$$

which defines the Klein 4-group $C_2 \times C_2$ (check this!). It follows that $|D_\infty/\langle \rho_0^2 \rangle| = 4$. By Theorem 0.4.13 (version 2 of the Third Isomorphism Theorem), $N' = \xi^{-1}(\langle \rho_0^2 \rangle) \lhd P$ and

$$P/N' \cong D_\infty/\langle \rho_0^2 \rangle \cong C_2 \times C_2,$$

so N' has index 4 in P.

It remains to show that $N' = N$. This is true because $\xi(N) = \xi(\langle a^2, b^2, (ab)^2 \rangle) = \langle \rho_0^2 \rangle$ in D_∞ and $N = \xi^{-1}(\xi(N))$, as $\ker \xi = M \leqslant N$.

CHAPTER 3

Groups and graphs

In this chapter we will discuss graphs and group actions on them. The main result will be a characterization of free groups as groups admitting free actions on trees. We will then use it to give a proof of the Nielsen-Schreier theorem from Chapter 2 (Theorem 1.3.5).

3.1. Definition of a graph

The reader may be familiar with the notion of a graph, as an object that consists of vertices and edges joining these vertices. The definition that we give below follows this pattern, however, we will require that all edges are directed and every edge has an inverse. This version of the definition was invented by Jean-Pierre Serre, who, together with Hyman Bass, created the theory of groups acting on trees.

Definition 3.1.1. A *graph* Γ is a 5-tuple $(V, E, \alpha, \omega, ^{-})$, where

- V and E are sets;
- $\alpha : E \to V$ and $\omega : E \to V$ are functions;
- $^{-} : E \to E$ is a bijection.

Moreover, we require that the following hold for all $e \in E$:

- $\overline{e} \neq e$ and $\overline{\overline{e}} = e$ (thus $^{-}$ is an *involution*);
- $\alpha(\overline{e}) = \omega(e)$;
- $\omega(\overline{e}) = \alpha(e)$.

Let us also recall some standard terminology.

Definition 3.1.2. Suppose that $\Gamma = (V, E, \alpha, \omega, ^{-})$ is a graph.

- Elements of V are the *vertices* of Γ, elements of E are its *edges*, the functions α and ω are the *incidence maps*.
- For each $e \in E$, $\overline{e}$ will be called the *inverse* of e, $\alpha(e) \in V$ will be called the *initial vertex* of e, and $\omega(e)$ will be called the *terminal vertex* of e.
- An edge $e \in E$ is said to be *incident* to a vertex $v \in V$ if $v \in \{\alpha(e), \omega(e)\}$.
- Two vertices $u, v \in V$ are said to be *adjacent* if either $u = v$ or there exists an edge $e \in E$ such that $\alpha(e) = u$ and $\omega(e) = v$.
- An edge $e \in E$ is a *loop* if $\alpha(e) = \omega(e)$.
- The graph Γ is *finite* if $|V| < \infty$ and $|E| < \infty$.

Notation 3.1.3. If $\Gamma = (V, E, \alpha, \omega, ^{-})$ is a graph, we will write $V\Gamma$ and $E\Gamma$ for the sets of vertices V and edges E respectively.

Definition 3.1.4. Let $\Gamma = (V, E, \alpha, \omega, ^{-})$ be a graph. Choose an edge from each mutually inverse pair of edges $\{e, \overline{e}\}$ in E, and denote the resulting set of edges E^+, so that $E = E^+ \sqcup E^-$, where $E^- = \{\overline{e} \mid e \in E^+\}$. Once such a choice has been made, the graph Γ will be called *oriented*, the edges in E^+ will be called *positively oriented* and the edges in E^- – *negatively oriented*.

Clearly, when describing an oriented graph it is sufficient to specify only its positively oriented edges, which is why we adopt the following convention.

Convention 3.1.5. When depicting a graph $\Gamma = (V, E, \alpha, \omega, ^{-})$ we will choose an orientation and will only draw positively oriented edges $e \in E^+$. Every such edge will be depicted as an arrow from $\alpha(e)$ to $\omega(e)$, labelled by e.

Example 3.1.6. Figure 3.1 depicts three graphs Γ_1, Γ_2 and Γ_3.

The graph Γ_1 has 4 vertices, $V\Gamma_1 = \{u_1, u_2, u_3, u_4\}$, but no edges. Such a graph (with $E\Gamma = \emptyset$) is said to be *totally disconnected*.

The graph Γ_2 has 4 vertices, $V\Gamma_2 = \{v_1, v_2, v_3, v_4\}$, and 6 edges, $E\Gamma_2 = \{e_1, \overline{e}_1, e_2, \overline{e}_2, e_3, \overline{e}_3\}$. This graph is *connected*, that is, starting from any vertex it is possible to reach any other vertex by following along the edges. The incidence maps in Γ_2 can be described by $\alpha(e_1) = v_1$, $\omega(e_1) = v_2$, $\alpha(e_2) = v_2$, $\omega(e_2) = v_3$, $\alpha(e_3) = v_3$, $\omega(e_3) = v_4$.

The graph Γ_3 has 3 vertices and 8 edges:

$$V\Gamma_3 = \{w_1, w_2, w_3\} \text{ and } E\Gamma_3 = \{f_1, \overline{f}_1, f_2, \overline{f}_2, f_3, \overline{f}_3, f_4, \overline{f}_4\}.$$

Note that the pair of edges f_1 and f_2 have the same initial and terminal vertices in Γ_3, thus Γ_3 has *multiple edges*. Moreover, both the edge f_3 and its inverse $\overline{f}_3$ are *loops*, because they start and end at the same vertex w_3.

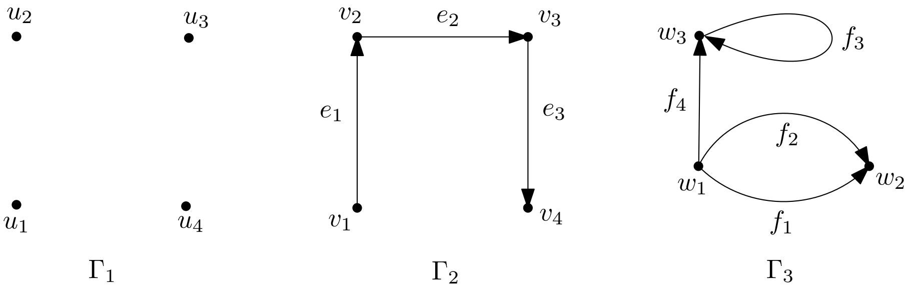


FIGURE 3.1. Graphs Γ_1, Γ_2 and Γ_3

For the remainder of the section assume that $\Gamma = (V, E, \alpha, \omega, ^{-})$ is a graph.

Definition 3.1.7. A *path* p in Γ is a sequence of edges $e_1e_2 \ldots e_n$ (with their vertices) such that $\alpha(e_{j+1}) = \omega(e_j)$, for all $j = 1, \ldots, n-1$. In this case the number n is called the *length* of p, denoted $\|p\|$. We also say that p starts at $\alpha(p) = \alpha(e_1)$, terminates at $\omega(p) = \omega(e_n)$ and *joins* $\alpha(p)$ with $\omega(p)$. The *inverse* path $\overline{p}$ is defined as the path $\overline{e}_n \ldots \overline{e}_2\overline{e}_1$, so that $\|\overline{p}\| = \|p\|$, $\alpha(\overline{p}) = \omega(p)$ and $\omega(\overline{p}) = \alpha(p)$.

A path p is said to be *closed* if $\alpha(p) = \omega(p)$.

The *trivial* (*degenerate*) path p_0 has no edges, and consists of a single vertex $u = \alpha(p_0) = \omega(p_0)$. Thus a path is trivial if and only if it has length 0.

A path $p = e_1e_2 \ldots e_n$ is called a *cycle* if it is non-trivial, closed and $e_{j+1} \neq \overline{e}_j$, for each $j = 1, \ldots, n-1$.

Example 3.1.8. Consider the graph Γ depicted on Figure 3.2. Then $p_0 = \{v_3\}$, $p_1 = e_4e_3\overline{e}_2\overline{e}_1$, $p_2 = e_2\overline{e}_3\overline{e}_4$, $p_3 = e_1e_2\overline{e}_3\overline{e}_4\overline{e}_1$ and $p_4 = \overline{e}_3e_3$ are paths in Γ. Note that p_0 is a trivial path, p_2, p_3 are cycles and p_4 is a closed path that is not a cycle.

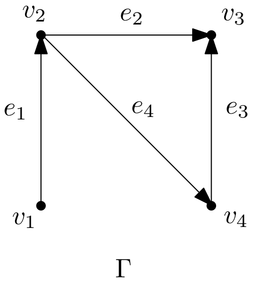


FIGURE 3.2. Paths in a graph

Definition 3.1.9. A *subgraph* Δ of the graph Γ is a 5-tuple $(V_1, E_1, \alpha_1, \omega_1, ^{-1})$, where V_1 is a subset of V, E_1 is a subset of E that is closed under inversion (i.e., $\overline{E_1} = E_1$) and $\alpha_1 : E_1 \to V_1$, $\omega_1 : E_1 \to V_1$, $^{-1} : E_1 \to E_1$ are simply the restrictions of the maps α, ω, $^-$ to E_1 (i.e., $\alpha_1(e) = \alpha(e)$, $\omega_1(e) = \omega(e)$ and $\overline{e}^1 = \overline{e}$, for all $e \in E_1$).

Definition 3.1.10. The graph Γ is said to be *connected* if for any two vertices $u, v \in V$ there is a path p in Γ joining these vertices, i.e., $\alpha(p) = u$ and $\omega(p) = v$. A *connected component* of a graph is a maximal connected subgraph.

Definition 3.1.11. A graph that does not contain any cycles is called a *forest.* A connected forest is called a *tree.*

Example 3.1.12. The graph Γ_1 on Figure 3.3 is a forest with 3 connected components. The graph Γ_2 is a tree.

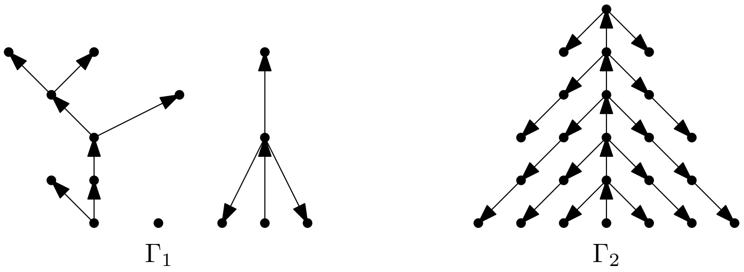


FIGURE 3.3. Forests and trees

FIGURE 3.4. An Autumn forest, as imagined and drawn by Nina Minasyan.

Connected graphs are important for us because they come with a natural notion of distance.

Definition 3.1.13. Let Γ be a connected graph. The *path metric* on Γ is the distance function d on the set of vertices of Γ defined as follows. For any two vertices $u, v \in V\Gamma$, $\mathrm{d}(u, v)$ is the length of a shortest path joining u with v in Γ.

One can easily check that the above distance function satisfies the standard properties of a metric on $V\Gamma$ (i.e., $\mathrm{d}(u,u)=0$, it is symmetric and the triangle inequality holds).

Example 3.1.14. Considering the connected graph Γ on Figure 3.2, we have $\mathrm{d}(v_1,v_2)=1$, $\mathrm{d}(v_3,v_2)=1$ and $\mathrm{d}(v_4,v_1)=2$.

3.2. Cayley graphs

Let G be a group with a subset $X \subseteq G$. Denote by $\overline{X}$ a formal copy of X disjoint from X, so that we have a bijection $x \mapsto \overline{x}$ from X to $\overline{X}$.

Definition 3.2.1. The *Cayley graph* of G with respect to X, denoted $\Gamma(G,X)$, is the oriented graph $(V,E,\alpha,\omega,{}^{-})$, where $V=G$, $E=E^+ \sqcup E^-$, the positively oriented edges are $E^+=G\times X$ and the negatively oriented edges are $E^-=G\times\overline{X}$. For any $e=(g,x)\in E^+$ (where $g\in G$ and $x\in X$) we have $\alpha(e)=g\in G$, $\omega(e)=gx\in G$ and $\overline{e}=(gx,\overline{x})\in E^-$.

Thus each $e=(g,x)\in E^+$ starts at g and terminates at gx, while its inverse $\overline{e}\in E^-$ starts at gx and terminates at g. Further on we will identify the set $\overline{X}$ with $X^{-1}\subseteq G$, via the bijection $\overline{x}\mapsto x^{-1}$. This way we can say that the set of edges of $\Gamma(G,X)$ is

$$E=E^+\sqcup E^-=G\times X\sqcup G\times X^{-1}=\{(g,x)\mid g\in G,\ x\in X\}\sqcup\{(g,x^{-1})\mid g\in G,\ x^{-1}\in X^{-1}\}.$$

Therefore, any edge $e\in E$ has the form (g,x^{ε}), where $g\in G$, $x\in X$ and $\varepsilon\in\{\pm1\}$, and satisfies $\alpha(e)=g\in G$ and $\omega(e)=gx^{\varepsilon}\in G$.

Definition 3.2.2. The Cayley graph $\Gamma(G,X)$ is a *labelled graph*, where each edge is labelled by a letter $\mathrm{Lab}(e)$ from the alphabet $X^{\pm1}=X\sqcup X^{-1}$. More precisely, if $e=(g,x)\in E^+$ then we set $\mathrm{Lab}(e)=x\in X$ and $\mathrm{Lab}(\overline{e})=x^{-1}\in X^{-1}$.

Note that Definition 3.2.1 does not require X to generate G. For example, X could be empty, in which case the Cayley graph would consist of vertices only and would have no edges.

Definition 3.2.3. Let G be a group with a subset $X\subseteq G$ and let $p=e_1e_2\dots e_n$ be a path in the Cayley graph $\Gamma(G,X)$. Suppose that $\mathrm{Lab}(e_i)=x_i^{\varepsilon_i}$, where $x_i\in X$ and $\varepsilon_i\in\{\pm1\}$, $i=1\dots,n$. Then the *label of* p is the word

$$\mathrm{Lab}(p)=x_1^{\varepsilon_1}x_2^{\varepsilon_2}\dots x_n^{\varepsilon_n}\ \text{ over } X^{\pm1}.$$

The label of a trivial path is the empty word.

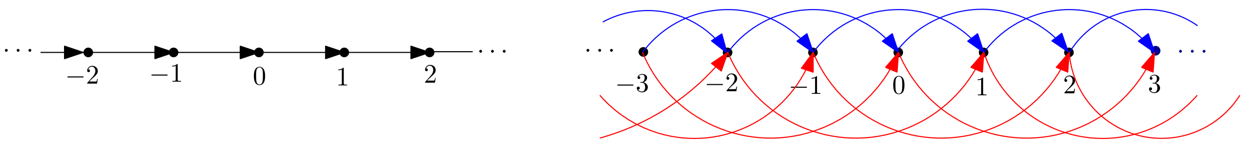
Remark 3.2.4. Given any word w over $X^{\pm1}$, where X is a subset of a group G, and any vertex $g\in G$, there is a unique path p starting at g and labelled by w in the Cayley graph $\Gamma(G,X)$. Moreover, $\omega(p)=gh$, where $h\in G$ is the element represented by the word w in G.

Example 3.2.5. (a) Figure 3.5 shows Cayley graphs of $\mathbb{Z}$ with respect to the generating subsets $\{1\}$ and $\{2,3\}$ (as usual, we only depict the positively oriented edges). In the Cayley graph $\Gamma(\mathbb{Z},\{1\})$ all positive edges are labelled by 1. In $\Gamma(\mathbb{Z},\{2,3\})$ edges labelled by 2 are blue and edges labelled by 3 are red. Note that these graphs are infinite, so we can only draw finite pieces of them.

(b) Given $n\in\mathbb{N}$, the Cayley graph $\Gamma(C_n,\{x\})$ of the cyclic group $C_n=\langle x\rangle$ is a cycle of length n: see Figure 3.6 for the Cayley graphs in the cases $n=2$ and $n=6$ (all edges are labelled by x).

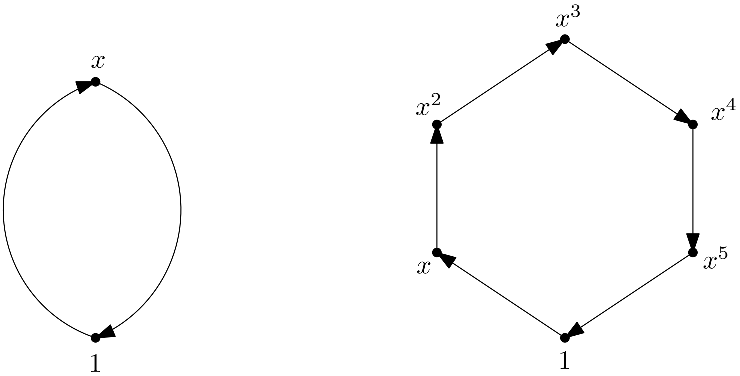
(c) Figure 3.7 depicts the Cayley graphs of the cyclic group $\mathbb{Z}_6$, with respect to the generating subset $\{\overline{2},\overline{3}\}$, and of the symmetric group S_3, with respect to the generating subset $\{(1\ 2),(1\ 2\ 3)\}$. In $\Gamma(\mathbb{Z}_6,\{\overline{2},\overline{3}\})$ edges labelled by $\overline{2}$ are blue and edges labelled by $\overline{3}$ are red; in $\Gamma(S_3,\{(1\ 2),(1\ 2\ 3)\})$ edges labelled by $(1\ 2\ 3)$ are blue and edges labelled by $(1\ 2)$ are red. One can note that although these two groups are rather different (for example, one is abelian while the other is not), their Cayley graphs are remarkably similar. In fact, since with every edge a graph also contains its inverse, one can show that these two Cayley graphs are isomorphic as graphs (but they will not be isomorphic as *labelled graphs*).

Exercise 3.2.6. Sketch the Cayley graphs of the dihedral group D_n, $n\geq 3$, with respect to the subsets $\{\sigma_0,\tau_0\}$ and $\{\sigma_0,\rho_0\}$ (see Example 2.2.3).



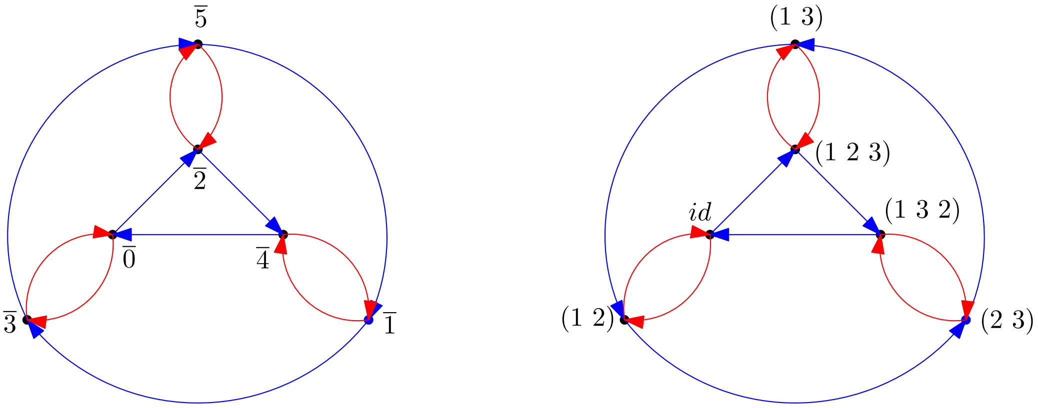
$\Gamma(\mathbb{Z}, \{1\})$ $\Gamma(\mathbb{Z}, \{2, 3\})$

FIGURE 3.5. Parts of Cayley graphs of $\mathbb{Z}$ with respect to two different generating sets



$\Gamma(C_2, \{x\})$ $\Gamma(C_6, \{x\})$

FIGURE 3.6. Cayley graphs of C_2 and C_6 with respect to a single generator x.



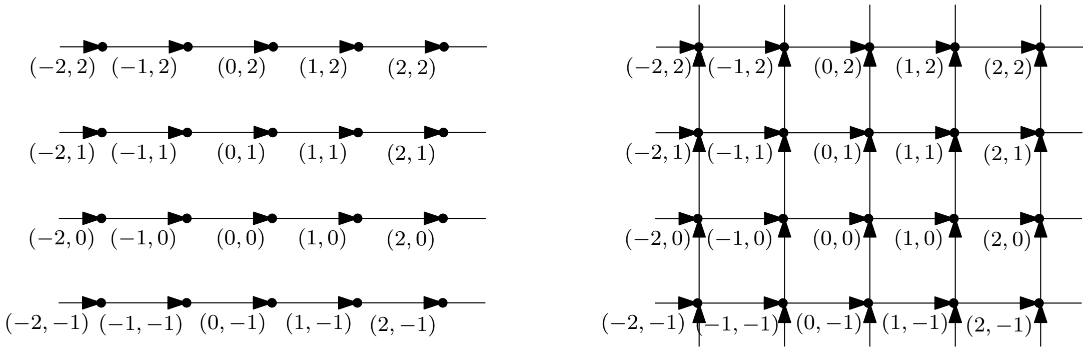
$\Gamma(\mathbb{Z}_6, \{\overline{2}, \overline{3}\})$ $\Gamma(S_3, \{(1\ 2), (1\ 2\ 3)\})$

FIGURE 3.7. Cayley graphs of $\mathbb{Z}_2$ and S_3 with respect to two generators.

Example 3.2.7. The reader may note that for all of the groups in Example 3.2.5 we have chosen X as generating sets. This is because the Cayley graph relative to a non-generating subset will be disconnected, hence it carries less information: see Lemma 3.2.8 below. Figure 3.8 demonstrates this when $G = \mathbb{Z}^2$ (all horizontal edges are labelled by $(1, 0)$ and all vertical edges are labelled by $(0, 1)$).

Lemma 3.2.8. *Given a group G and a subset $X \subseteq G$, the Cayley graph $\Gamma(G, X)$ is connected if and only if $G = \langle X \rangle$.*

PROOF. To prove the sufficiency, assume that X generates G. Then for every element $g \in G$ there exist $x_1 \ldots, x_n \in X$ and $\varepsilon_1, \ldots, \varepsilon_n \in \{\pm 1\}$ such that $g = x_1^{\varepsilon_1} \ldots x_n^{\varepsilon_n}$ in G. Let w be the word $x_1^{\varepsilon_1} \ldots x_n^{\varepsilon_n}$ over the alphabet $X^{\pm 1}$. According to Remark 3.2.4, the path p, starting at 1 and



$\Gamma(\mathbb{Z}^2, \{(1,0)\})$ $\Gamma(\mathbb{Z}^2, \{(1,0),(0,1)\})$

FIGURE 3.8. Parts of Cayley graphs of $\mathbb{Z}^2$ with respect to a non-generating subset $\{(1,0)\}$ and a generating subset $\{(1,0),(0,1)\}$.

labelled by w, terminates at the vertex g of $\Gamma(G,X)$. Thus, we have found a path joining the vertex 1 with the vertex g, for every vertex g of $\Gamma(G,X)$. It follows that $\Gamma(G,X)$ is connected.

Now, let us assume that $\Gamma(G,X)$ is a connected graph. Take any $g \in G$ and let p be a path in $\Gamma(G,X)$ joining 1 with g. Then $\mathrm{Lab}(p) = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$ is some word over $X^{\pm 1}$, so by Remark 3.2.4, this word represents the element $\omega(p) = g$, i.e., $g = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$ in G. It follows that $G = \langle X \rangle$, as claimed. □

3.3. Group actions on graphs

Much of this chapter is devoted to looking at actions of groups on graphs. In order to define them, we first need to recall the concepts of isomorphisms and automorphisms of graphs. Informally speaking, two graphs are isomorphic if there exist bijections between their sets of vertices and their sets of edges which respect the incidence maps and the edge inversion.

Definition 3.3.1. Let $\Gamma = (V, E, \alpha, \omega, {}^{-})$ and $\Gamma' = (V', E', \alpha', \omega', {}^{-\prime})$ be two graphs. An *isomorphism* between these graphs is a pair $\phi = (\phi^0, \phi^1)$, where $\phi^0 : V \to V'$ and $\phi^1 : E \to E'$ are bijections satisfying the following conditions for all $e \in E$:

(i) $\alpha'(\phi^1(e)) = \phi^0(\alpha(e))$ and $\omega'(\phi^1(e)) = \phi^0(\omega(e))$;
(ii) $\phi^1(\overline{e}) = \overline{\phi^1(e)}'$.

In this case the graphs Γ and Γ' are said to be *isomorphic*. An *automorphism* of graph Γ is an isomorphism $\phi : \Gamma \to \Gamma$ and we will use $\mathrm{Aut}(\Gamma)$ to denote the set of all such automorphisms.

Exercise 3.3.2. Show that for any non-empty graph $\Gamma = (V, E, \alpha, \omega, {}^{-})$ the set $\mathrm{Aut}(\Gamma)$ is a group under the composition of pairs of maps. More precisely, if $\phi = (\phi^0, \phi^1)$ and $\psi = (\psi^0, \psi^1)$ are in $\mathrm{Aut}(\Gamma)$, we define $\psi \circ \phi = (\psi^0 \circ \phi^0, \psi^1 \circ \phi^1)$, where $\psi^0 \circ \phi^0 : V \to V$ and $\psi^1 \circ \phi^1 : E \to E$ are the usual compositions of maps.

Example 3.3.3. Consider the graph Γ depicted on Figure 3.9. The map $\phi = (\phi^0, \phi^1) : \Gamma \to \Gamma$ is defined by

$$\phi^0(v_1) = v_4,\ \phi^0(v_2) = v_3,\ \phi^0(v_3) = v_2,\ \phi^0(v_4) = v_1$$
$$\text{and}\ \ \phi^1(e_1) = \overline{e}_3,\ \phi^1(e_2) = \overline{e}_2,\ \phi^1(e_3) = \overline{e}_1.$$

To satisfy condition (ii) from Definition 3.3.1 we must also set $\phi^1(\overline{e}) = \overline{\phi^1(e)}$, for all $e \in E\Gamma^-$. Then it is not difficult to check that ϕ is an automorphism of Γ.

Let us now recall the definition of a group action on a set.

Definition 3.3.4. Let G be a group and let S be a set. An *action* of G on S is a homomorphism $\xi : G \to \mathrm{Sym}(S)$, where $\mathrm{Sym}(S)$ denotes the group of all self-bijections (permutations) of S. Equivalently, we have a map $G \times S \to S$, $(g, s) \mapsto g.s$, obeying the following properties:

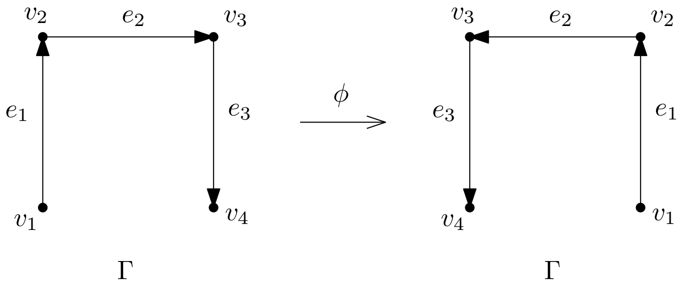


FIGURE 3.9. An automorphism ϕ of a graph Γ.

- $1_G.s = s$, for all $s \in S$;
- $(gh).s = g.(h.s)$, for all $g, h \in G$ and all $s \in S$.

Exercise 3.3.5. Show that the two definitions of a group action given in Definition 3.3.4 are indeed equivalent to each other.

Example 3.3.6. (a) For each $n \in \mathbb{N}$ the symmetric group S_n acts naturally on the set $\{1, 2, \dots, n\}$ by permutations.

(b) The group $\mathbb{Z}$ acts on the real line $\mathbb{R}$ by translation: $n.x = x + n$, for all $n \in \mathbb{Z}$ and $x \in \mathbb{R}$.

(c) Any group G acts on itself by left multiplication: $g.x = gx$, for all $g, x \in G$.

Definition 3.3.7. Let G be a group acting on a set S.

(i) If $s \in S$ then the *stabilizer* of s in G is defined as $\mathrm{Stab}_G(s) = \{g \in G \mid g.s = s\} \subseteq G$.

(ii)The *orbit* of an element $s \in S$ under the G-action is the subset $G.s = \{g.s \mid g \in G\} \subseteq S$.

Exercise 3.3.8. Let G be a group acting on a set S and let $s, t \in S$.

(a) Show that $\mathrm{Stab}_G(s)$ is a subgroup of G.

(b) Prove that either $G.s \cap G.t = \emptyset$ or $G.s = G.t$. Thus S decomposes into the disjoint union of orbits of G.

The following basic result was discussed in the Group Theory module.

Exercise 3.3.9 (Orbit-stabilizer theorem)**.** Suppose that G is a finite group acting on a set S. Then for each $s \in S$ we have $|G.s| = |G : \mathrm{Stab}_G(s)| = \dfrac{|G|}{|\mathrm{Stab}_G(s)|}$.

Group actions on graphs are defined similarly, except the maps need to be graph automorphisms in the sense of Definition 3.3.1.

Definition 3.3.10. Let G be a group and let $\Gamma = (V, E, \alpha, \omega, ^-)$ be a graph. An *action* of G on Γ is a homomorphism $\xi : G \to \mathrm{Aut}(\Gamma)$, where $\mathrm{Aut}(\Gamma)$ is the automorphism group of the graph Γ. Equivalently, there are actions of G on the vertex set V and on the edge set E that are compatible in the following sense: for all $g \in G$ and all $e \in E$ we have

- $g.\alpha(e) = \alpha(g.e)$ and $g.\omega(e) = \omega(g.e)$;
- $g.\overline{e} = \overline{g.e}$.

Let us make the following important observation, which stems from the fact that graph automorphisms send paths to paths of the same length.

Remark 3.3.11. Suppose that Γ is a connected graph and let d be the path metric on its vertex set, as given in Definition 3.1.13. If G is a group acting on Γ then the induced action of G on the vertex set $V\Gamma$ is by isometries. This means that for arbitrary vertices $u, v \in V\Gamma$ and any $g \in G$ we have $\mathrm{d}(g.u, g.v) = \mathrm{d}(u, v)$.

Definition 3.3.12. An action of a group G on a graph Γ is said to be *without edge inversions* if $g.e \neq \overline{e}$, for every $e \in E\Gamma$.

Example 3.3.13. Let $n \in \mathbb{N}$, $n \geq 3$, and let P_n be an n-gon, considered as an oriented graph with n vertices and n positively oriented edges. The dihedral group D_n acts on P_n naturally by automorphisms that can be thought of as reflections and rotations of the graph P_n.

For example, consider the hexagon P_6 and the reflections σ_0 and τ_0 depicted on Figure 3.10. Thus $VP_6 = \{v_1, \ldots, v_6\}$, $EP_6^+ = \{e_1, \ldots, e_6\}$,

$$\sigma_0(v_1) = v_1,\ \sigma_0(v_2) = v_6,\ \sigma_0(v_3) = v_5,\ \sigma_0(v_4) = v_4,\ \sigma_0(v_5) = v_3,\ \sigma_0(v_6) = v_2,\ \text{and}$$

$$\sigma_0(e_1) = \overline{e}_6,\ \sigma_0(e_2) = \overline{e}_5,\ \sigma_0(e_3) = \overline{e}_4,\ \sigma_0(e_4) = \overline{e}_3,\ \sigma_0(e_5) = \overline{e}_2,\ \sigma_0(e_6) = \overline{e}_1.$$

Note that we only need to define the images of the positively oriented edges under a graph automorphism, as their inverses have to go to inverses of the images, by Definition 3.3.1.

In particular, we see that the action of the cyclic group $\langle \sigma_0 \rangle$ on P_6 is without edge inversions. On the other hand, the reflection $\tau_0 \in D_6$ does invert some edges of P_6, because $\tau_0(e_6) = \overline{e}_6$ and $\tau_0(e_3) = \overline{e}_3$. Thus, the action of D_6 on P_6 has edge inversions.

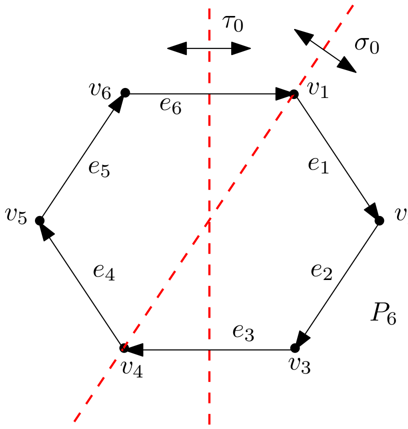


FIGURE 3.10. The action of D_6 on the hexagon graph P_6.

Definition 3.3.14. An action of a group G on an oriented graph $\Gamma = (V, E^+ \sqcup E^-, \alpha, \omega, {}^-)$ is said to be *orientation-preserving* if $g.E^+ = E^+$, for all $g \in G$. It follows that $g.E^- = E^-$, for all $g \in G$, so elements of G send positively oriented edges to positively oriented edges and negatively oriented edges to negatively oriented edges.

Remark 3.3.15. Note that every orientation-preserving action of a group on a graph is without edge inversions.

Exercise 3.3.16. Let G act on an (unoriented) graph Γ without edge inversions. Prove that there exists an orientation on Γ such that the given G-action on Γ is orientation-preserving.

One way to solve Exercise 3.3.16 is to represent $E\Gamma$ as a disjoint union of G-orbits of edges, and then choose one orbit from every pair of "mutually inverse" orbits $(G.e, G.\overline{e})$. Afterwards, we can proclaim all chosen orbits to consist of positively oriented edges and the remaining orbits to consist of negatively oriented edges. For infinite graphs such a proof would have to invoke the Axiom of choice or Zorn's Lemma.

Example 3.3.17. The action of the cyclic group $\langle \sigma_0 \rangle$ does not preserve the given orientation on P_6 on Figure 3.10. But since σ_0 does not invert any edges of P_6, it is possible to define a different orientation on the edges of P_6 in such a way that this group acts by preserving this new orientation: see Figure 3.11.

Lemma 3.3.18. *Let G be a group with a subset X. Then G acts on the Cayley graph $\Gamma(G, X)$ naturally, preserving the orientation and labels of the edges.*

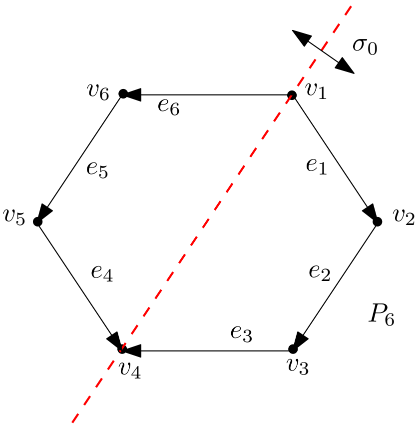


FIGURE 3.11. A new orientation on the edges of P_6 that is preserved by the action of $\langle\sigma_0\rangle$.

PROOF. Recall that, by Definition 3.2.1, $\Gamma(G,X)$ is the oriented graph $(V,E,\alpha,\omega,{}^{-})$, where $V=G$, $E=E^+\sqcup E^-$, $E^+=G\times X$ and $E^-=G\times X^{-1}$. Moreover, for each edge $e=(g,x^\varepsilon)\in E$, where $g\in G$, $x\in X$ and $\varepsilon\in\{\pm1\}$, we have $\alpha(e)=g$, $\omega(e)=gx^\varepsilon\in G$, $\mathrm{Lab}(e)=x^\varepsilon\in X^{\pm1}$ and $\overline{e}=(gx^\varepsilon,x^{-\varepsilon})$.

The natural action of G on $\Gamma(G,X)$ is defined as follows: given any $h\in G$ we set

$$h.g=hg,\ \text{for all } g\in V=G,\quad \text{and}\quad h.(g,x^\varepsilon)=(hg,x^\varepsilon),\ \text{for all } (g,x^\varepsilon)\in E.$$

Evidently the above defines actions of G on the set of vertices V and the set of edges E of $\Gamma(G,X)$, which preserve the edge labels and orientation. It is also easy to check that the compatibility conditions from Definition 3.3.10 are satisfied in this case. ☐

Example 3.3.19. (a) The group of integers $\mathbb{Z}$ acts on its two Cayley graphs, depicted on Figure 3.5, by translations.

(b) The Cayley graph of the cyclic group $C_6=\langle x\rangle$, with respect to the generating set $\{x\}$, is an oriented hexagon (see Figure 3.6). The group C_6 acts on $\Gamma(C_6,\{x\})$ by "rotations".

3.4. Characterizing Cayley graphs in terms of group action

In this section we will give a criterion for a graph, equipped with an action of a group, to be isomorphic to a Cayley graph of this group. Let us start by recalling the following two standard definitions from the theory of group actions.

Definition 3.4.1. Suppose that G is a group acting on a set S.

(i) The action of G on S is said to be *free* if for all $g\in G\setminus\{1\}$ and all $s\in S$ we have $g.s\neq s$ (i.e., non-trivial elements of G act without fixed points). In other words, the action is free if $\mathrm{Stab}_G(s)=\{1\}$, for every $s\in S$.

(ii) This action is called *transitive* if for all $s,t\in S$ there exists $g\in G$ such that $g.s=t$. This is equivalent to saying that there exists only one orbit of elements of S under the action of G.

(iii) This action is *faithful* if the corresponding homomorphism $G\to\mathrm{Sym}(S)$ is injective (in other words, if $g.s=s$ for all $s\in S$ then $g=1$ in G).

Note that a free group action on a set is necessarily faithful.

Example 3.4.2. (a) For any $n\in\mathbb{N}$ the symmetric group S_n acts transitively on the set $S=\{1,\dots,n\}$. This action is free if and only if $n\leq 2$ because stabilizers of points are quite large (one can check that $\mathrm{Stab}_{S_n}(i)\cong S_{n-1}$, for any $i\in\{1,\dots,n\}$). This action is obviously faithful.

(b) Consider the n-cycle $\alpha=(1\ 2\ \dots\ n)\in S_n$ and the natural action of the cyclic subgroup $C=\langle\alpha\rangle$ of S_n on the set $S=\{1,\dots,n\}$. This action is clearly transitive: if $i,j\in S$ then $\alpha^{j-i}.i=j$. It is also faithful, because $C\leqslant S_n$. Let us show that this action is free. Indeed, for any $s\in S$ we know that $|C.s|=|S|=n=|C|$, as C acts on S transitively, hence, by the Orbit-Stabilizer theorem (Exercise 3.3.9), $|\operatorname{Stab}_C(s)| = |C|/|C.s| = 1$. It follows that $\operatorname{Stab}_C(s) = \{1\}$, so C acts on S freely.

(c) The action of $\mathbb{Z}$ on $\mathbb{R}$ by translations is free (and, hence, faithful) but not transitive.

Definition 3.4.3. An action of a group G on a graph Γ is *free* if the corresponding action of G on $V\Gamma$ is free in the sense of Definition 3.4.1.

Exercise 3.4.4. Let G be a group with a subset X. Show that the natural action of G on the vertices of its Cayley graph $\Gamma(G, X)$ is both free and transitive.

Since the set of vertices of $\Gamma(G, X)$ is G, the next statement can be regarded as a converse to Exercise 3.4.4.

Exercise 3.4.5. Suppose that G acts freely and transitively on a non-empty set S. Show that for any given $s \in S$ the map $g \mapsto g.s$ defines a bijection between G and S.

Definition 3.4.6. A graph $\Gamma = (V, E, \alpha, \omega, ^{-})$ is *simple* if it has no loops or multiple edges. In other words, both of the following conditions must hold:

- for any $e \in E$ we must have $\alpha(e) \neq \omega(e)$;
- if $e, f \in E$ and $\alpha(e) = \alpha(f)$, $\omega(e) = \omega(f)$ then $e = f$.

Exercise 3.4.7. Let G be a group and let $X \subseteq G$. Find necessary and sufficient conditions on X ensuring that $\Gamma(G, X)$ is a simple graph.

Remark 3.4.8. Clearly, to define a simple graph it is sufficient to specify the set of vertices V and explain which vertices are adjacent (this corresponds to choosing a symmetric subset of $V \times V$ which will be the set of edges of the graph).

The next theorem provides a converse to the statement of Exercise 3.4.4 in the case of simple and connected graphs.

THEOREM 3.4.9. *Suppose that a group G acts on a non-empty simple connected graph $\Gamma = (V, E, \alpha, \omega, ^{-})$ freely and transitively on the vertices and without edge inversions. Then Γ is isomorphic to the Cayley graph $\Gamma(G, X)$, for some generating set X of G.*

More precisely, choose an orientation $E = E^+ \sqcup E^-$ on the edges of Γ so that the action of G preserves it (this is possible by Exercise 3.3.16). Fix any vertex $v \in V$ and set

$$X = \{x \in G \mid \exists\, e \in E^+ \text{ such that } \alpha(e) = v \text{ and } \omega(e) = x.v\} \subseteq G. \tag{3.1}$$

Then X generates G and $\Gamma(G, X)$ is isomorphic to Γ.

PROOF. Choose the subset $X \subseteq G$ as in the statement of the theorem (note that X depends on the choice of an orientation on Γ that is preserved by G).

Step 1. Let us first show that X generates G.

Consider any $g \in G$. Since Γ is connected, there is a path p in Γ such that $\alpha(p) = v$ and $\omega(p) = g.v$. We will now argue that $g \in \langle X \rangle$ by induction on the length of p. If $\|p\| = 0$ then $g.v = v$, so $g = 1 \in \langle X \rangle$ in G because the action of G on the vertices of Γ is free. Thus we can further assume that $p = e_1 \dots e_n$, where $n \in \mathbb{N}$ is positive.

Denote $u = \omega(e_{n-1}) = \alpha(e_n) \in V$. Since G acts on V transitively, there exists an element $h \in G$ such that $u = h.v$. Moreover, we know that $h \in \langle X \rangle$ by the induction hypothesis. Now, observe that $h^{-1}.e_n$ is an edge starting at $h^{-1}.\alpha(e_n) = h^{-1}.u = v$ and terminating at $h^{-1}.\omega(e_n) = h^{-1}.(g.v) = (h^{-1}g).v$.

If $e_n \in E^+$ then $h^{-1}.e_n \in E^+$, as the action of G on Γ is orientation-preserving, whence $h^{-1}g = x \in X$ by the definition of X in (3.1). It follows that $g = hx \in \langle X \rangle$.

If $e_n \in E^-$, then $\bar{e}_n \in E^+$ and so the edge $g^{-1}.\bar{e}_n \in E^+$ starts at $g^{-1}.\alpha(\bar{e}_n) = (g^{-1}g).v = v$ and terminates at $g^{-1}.\omega(\bar{e}_n) = (g^{-1}h).v$. Therefore, $g^{-1}h = y \in \langle X \rangle$ by the definition of X, so $g = hy^{-1} \in \langle X \rangle$.

Thus we have shown by induction that $g \in \langle X \rangle$, for all $g \in G$, which implies that $G = \langle X \rangle$.

Step 2. Now we will show that the graphs $\Gamma(G, X)$ and Γ are isomorphic.

Observe that because the action of G on V is free and transitive, according to Exercise 3.4.5 there is a bijection $\phi^0 : G \to V$ defined by

$$\phi^0(g) = g.v, \quad \text{for all } g \in G.$$

By the definition of X in (3.1), for every $x \in X$ there exists an edge $e_x \in E^+$ such that e_x starts at v and ends at $x.v$. Moreover, such an edge is unique because the graph Γ is simple. We can therefore define a map $\phi^1 : G \times X \to E$ by letting

$$\phi^1((g, x)) = g.e_x, \quad \text{for all } (g, x) \in G \times X.$$

Let us check that the map ϕ^1 is a bijection between the set of positively oriented edges $G \times X$, of $\Gamma(G, X)$, and E^+. Indeed, suppose that $g, h \in G$ and $x, y \in X$ satisfy $\phi^1((g, x)) = \phi^1((h, y))$, so that $g.e_x = h.e_y$. Then $g.v = h.v$, as $v = \alpha(e_x) = \alpha(e_y)$, whence $(g^{-1}h).v = v$. The latter implies that $g^{-1}h = 1$ in G because the action of G on V is free, thus $g = h$. We also deduce that $g.e_x = g.e_y$, so $e_x = e_y$ and the end vertices of this edge must be the same, i.e., $x.v = y.v$. As before, using freeness of the action, we conclude that $x = y$. Hence, $(g, x) = (h, y)$, so ϕ^1 is injective.

To show that ϕ^1 is surjective, choose any $e \in E^+$. By transitivity of the action, there must exist $g, h \in G$ such that $g.v = \alpha(e)$ and $h.v = \omega(e)$. Then the edge $g^{-1}.e \in E^+$ starts at v and ends at $(g^{-1}h).v$, hence $g^{-1}h = x \in X$, by the definition of X. Since the graph Γ is simple, it follows that $g^{-1}.e = e_x$, so $e = g.e_x = \phi^1((g, x))$. The latter shows that ϕ^1 is surjective.

Since both graphs $\Gamma(G, X)$ and Γ are oriented, we can extend the bijection ϕ^1 between the positively oriented edges to a full bijection between their edge sets by setting $\phi^1((g, x^{-1})) = \overline{\phi^1((g, x))}$, for all negatively oriented edges $(g, x^{-1}) \in G \times X^{-1}$ of $\Gamma(G, X)$. This also shows that the bijection ϕ^1 respects edge inversion (i.e., condition (ii) of Definition 3.3.1 is satisfied).

It remains to check that the maps ϕ^0 and ϕ^1 respect the incidence relation. If $(g, x) \in G \times X$ is a positively oriented edge of $\Gamma(G, X)$ then

$$\alpha\left(\phi^1((g, x))\right) = \alpha(g.e_x) = g.v = \phi^0(\alpha((g, x))), \text{ because } \alpha((g, x)) = g \text{ in } \Gamma(G, X).$$

Similarly, $\omega\left(\phi^1((g, x))\right) = \phi^0(\omega((g, x)))$. The incidence relation for negatively oriented edges follows from the definition of ϕ^1 on them. Hence, we have verified that $\phi = (\phi^0, \phi^1)$ is an isomorphism between $\Gamma(G, X)$ and Γ, and the theorem is proved. □

Remark 3.4.10. In fact, the isomorphism ϕ between the graphs $\Gamma(G, X)$ and Γ, produced in the proof of Theorem 3.4.9, is *equivariant* with respect to the actions of G on these graphs. This means that the G-actions on the vertices and edges commute with the maps ϕ^0 and ϕ^1: if $h \in G$, $g \in G$ is a vertex of $\Gamma(G, X)$ and $(g, x^\varepsilon) \in G \times X^{\pm 1}$ is an edge of $\Gamma(G, X)$ then

$$\phi^0(h.g) = h.\phi^0(g) \text{ and } \phi^1(h.(g, x^\varepsilon)) = h.\phi^1((g, x^\varepsilon)).$$

3.5. Cayley graphs of free groups

Recall that a connected graph is called a tree if it contains no cycles (see Definitions 3.1.7 and 3.1.11). The following theorem describes a connection between the algebraic properties of a group and the geometric properties of its Cayley graph.

THEOREM 3.5.1. *Let G be a group with a subset $X \subseteq G$. The Cayley graph $\Gamma(G, X)$ is a tree if and only if G is free and X is a free generating set of G.*

PROOF. Suppose that $\Gamma(G, X)$ is a tree. Then it is connected, so $G = \langle X \rangle$, by Lemma 3.2.8. Arguing by contradiction, assume that X is not a free generating set of G. Then, according to Corollary 1.2.20, there must exist a non-empty reduced word $w = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$ over $X^{\pm 1}$ such that w represents the identity element of G. Let p be the path starting at 1 and labelled by w in $\Gamma(G, X)$ (see Remark 3.2.4). Since $w = 1$ in G, p will be a non-trivial closed path (i.e., $\alpha(p) = 1 = \omega(p)$). Moreover, no edge in p is followed by its inverse because the word w is reduced, hence p is a cycle in

$\Gamma(G, X)$. This contradicts the assumption that $\Gamma(G, X)$ is a tree, so we have proved one direction of the theorem.

For the opposite direction, assume that G is free on X. Then X generates G, so $\Gamma(G, X)$ is connected by Lemma 3.2.8. If this Cayley graph is not a tree, it must contain a cycle $p = e_1 \dots e_n$, for some $n \in \mathbb{N}$. Let $g = \alpha(p) = \omega(p) \in G$ and let $\mathrm{Lab}(e_i) = x_i^{\varepsilon_i}$, where $x_i \in X$ and $\varepsilon_i \in \{\pm 1\}$, for $i = 1, \dots, n$, so that $w = \mathrm{Lab}(p) = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$. Since p is a cycle, we have $e_{i+1} \neq \overline{e}_i$ which means that $x^{\varepsilon_{i+1}} \neq x^{-\varepsilon_i}$, for all $i = 1, \dots, n-1$, i.e., the word w is reduced over $X^{\pm 1}$. Denote by h the element of G represented by w. Then $\omega(p) = gh$ by Remark 3.2.4, and, as p is a cycle, we see that $gh = \alpha(p) = g$, so $h = 1$ in G. Thus the non-empty reduced word w represents the trivial element of G, contradicting the assumption that X is a free generating set of G (see Corollary 1.2.20). Therefore $\Gamma(G, X)$ must be a tree, and the theorem is proved. ☐

Example 3.5.2. On Figure 3.12 we have sketched a part of the Cayley graph of the free group $F = F(\{x, y\})$ with respect to the free generating set $\{x, y\}$. As usual, we have only drawn the positively oriented edges. All horizontal (blue) edges are labelled by x and all vertical (red) edges are labelled by y. Thus we see that this Cayley graph is a 4-regular tree (which means that every vertex is the initial vertex of exactly 4 edges).

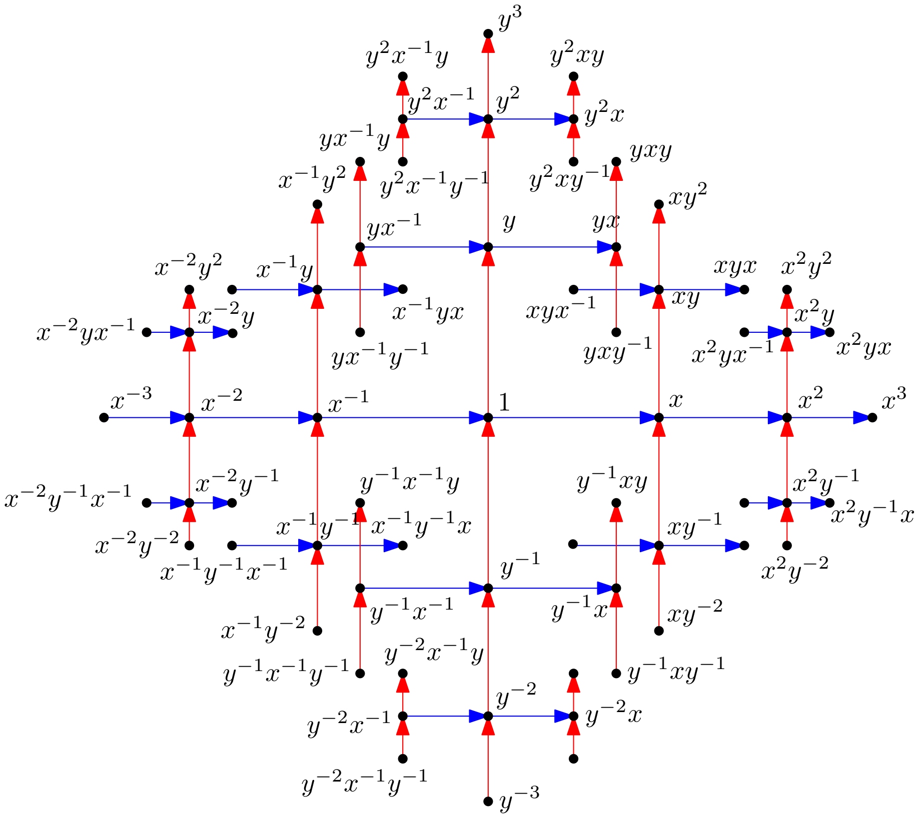


FIGURE 3.12. A part of the Cayley graph of the free group $F(\{x, y\})$ with respect to $\{x, y\}$.

By combining Theorem 3.5.1, Lemma 3.3.18 and Exercise 3.4.4 we get the following statement.

Corollary 3.5.3. *Any free group admits an action on a non-empty tree without edge inversions. Moreover, this action is free and transitive on the vertices of the tree.*

3.6. Spanning trees for group actions

Every connected graph Γ admits a *maximal tree*, which is a subgraph that is a tree and is not contained in an larger subtree of Γ. Moreover, this maximal tree contains every vertex of Γ. This can be easily proved for finite graphs, but for infinite graphs one needs to use Zorn's Lemma or an equivalent statement involving the Axiom of choice. In this section we will extend this fact to spanning trees of group actions on graphs.

Definition 3.6.1. Let G be a group acting on a connected graph Γ. A *spanning tree* for this action is a subgraph T of Γ such that the following two conditions hold:

- T is a tree (i.e., it has no cycles);
- T contains exactly one vertex from each G-orbit of vertices of Γ.

Example 3.6.2. (a) Given any $n \in \mathbb{N}$, $n \geq 3$, the dihedral group D_n acts on the n-gon graph P_n (see Example 3.3.13). Since this action is transitive on the vertices, there is only one orbit of vertices. Therefore, any spanning tree for this action consists of exactly one vertex (and any vertex of P_n gives rise to a spanning tree).

(b) Consider the reflections σ_0 and τ_0 in D_6 of the hexagon P_6, depicted on Figure 3.10. The action of the cyclic group $\langle\sigma_0\rangle \cong C_2$ on P_6 has 4 orbits of vertices:

$$\{v_1\},\ \{v_2, v_6\},\ \{v_3, v_5\} \text{ and } \{v_4\}.$$

Therefore the path subgraph Π, consisting of the path $e_1e_2e_3$ and its inverse, is a spanning tree for the action of $\langle\sigma_0\rangle$ on P_6.

Since the action of $\langle\tau_0\rangle$ on P_6 has only 3 orbits of vertices $\{v_1, v_6\}$, $\{v_2, v_5\}$ and $\{v_3, v_4\}$ the paths e_1e_2 or e_4e_5 both give rise to spanning trees for this action.

Exercise 3.6.3. Let $C = \langle(1, 1)\rangle$ be the "diagonal subgroup" of $\mathbb{Z}^2$, and consider the natural action (by translations) of C on the Cayley graph of $\mathbb{Z}^2$ with respect to the standard generating set $\{(1, 0), (0, 1)\}$ (see Example 3.2.7). Check that the "x-axis" subgraph T, whose vertex set is $VT = \{(n, 0) \mid n \in \mathbb{Z}\}$ and which includes all edges between these vertices, is a spanning tree for this action. Find a different spanning tree for the same action.

Proposition 3.6.4. *Every action of a group on a connected graph Γ admits a spanning tree.*

Proof. Suppose that a group G acts on connected graph Γ. Let $\mathcal{T}$ denote the set of all subgraphs of Γ such that every $T \in \mathcal{T}$ is a tree and distinct vertices of T are in different G-orbits. If Γ is empty then there is nothing to prove, so assume that there is at least one vertex $v \in V\Gamma$. Then $\{v\} \in \mathcal{T}$, so $\mathcal{T} \neq \emptyset$. We now order the elements of $\mathcal{T}$ by inclusion: given any $T_1, T_2 \in \mathcal{T}$ we write $T_1 \leq T_2$ if T_1 is a subgraph of T_2. It is easy to see that this defines a partial order on $\mathcal{T}$. Moreover, one can also verify that the conditions of Zorn's Lemma are satisfied, hence there exists a tree $T_0 \in \mathcal{T}$ which is maximal with respect to the above partial order (i.e., T_0 is not contained in a larger tree in $\mathcal{T}$). Let us show that T_0 is a spanning tree for the action of G on Γ.

Arguing by contradiction, suppose that T_0 does not contain a representative from some G-orbit of vertices of Γ. Then we can choose a vertex $v \in V\Gamma$ such that $v \notin G.VT_0$ and the distance $\mathrm{d}(v, T_0)$ is the smallest possible (here d denotes the path metric on Γ, see Definition 3.1.13). Choose a path $p = e_1 \dots e_n$ in Γ joining v with a vertex of T_0 such that $\|p\| = n = \mathrm{d}(v, T_0)$, and let $w = \omega(e_1)$ be the second vertex on this path. Since $\mathrm{d}(w, T_0) \leq n-1 < d(v, T_0)$, we can conclude that $w \in G.VT_0$, by the choice of v. Thus there exist $g \in G$ and a vertex $w' \in VT_0$ such that $w = g.w'$. Set $v' = g^{-1}.v$ and observe that

$$\mathrm{d}(v', T_0) \leq \mathrm{d}(v', w') = \mathrm{d}(g^{-1}.v, g^{-1}.w') = \mathrm{d}(v, w) = 1,$$

where we have used Remark 3.3.11 stating that the action of G on Γ is distance-preserving.

Note that $v' \in G.v$, whence $v' \notin G.VT_0$ because $v \notin G.VT_0$ by assumption. Thus, after replacing v with v' we can assume that $\mathrm{d}(v, T_0) = 1$, so there is an edge $e \in E\Gamma$ with $\alpha(e) = u \in VT_0$ and $\omega(e) = v$. Let T_1 be the subgraph of Γ obtained from T_0 by adding the edge e to it (thus $VT_1 = VT_0 \cup \{v\}$ and $ET_1 = ET_0 \cup \{e, \overline{e}\}$). Clearly T_1 is also a tree because $v \notin VT_0$, and $T_1 \in \mathcal{T}$ as $v' \notin G.VT_0$. The latter contradicts the maximality of T_0 in $\mathcal{T}$. Therefore, $V\Gamma = G.VT_0$, so T_0 is a spanning tree for the G-action on Γ. □

In fact, Proposition 3.6.4 generalizes the fact about the existence of maximal trees in connected graphs mentioned in the beginning of the section if one considers the action of the trivial group on a connected graph.

3.7. Groups acting freely on trees

We are finally in a position to prove the main result of this chapter.

THEOREM 3.7.1. *A group is free if and only if it admits a free action on a non-empty tree without edge inversions.*

PROOF. The"only if" direction has already been proved: see Corollary 3.5.3. Therefore, we only need to establish the "if" direction. Thus, let us assume that G is a group acting freely on a non-empty tree Γ. According to Proposition 3.6.4 there exists a spanning tree T for the action of G on Γ. Let us examine the induced action of G on the set of G-translates of the tree T in Γ.

Claim 3.7.2. *If $g, h \in G$ satisfy $g.T \cap h.T \neq \emptyset$ then $g = h$.*

PROOF OF CLAIM 3.7.2. Suppose that there is a vertex $u \in g.T \cap h.T$. Then $g^{-1}.u$ and $h^{-1}.u$ are two vertices of T belonging to the same G-orbit, hence $g^{-1}.u = h^{-1}.u$ because T is a spanning set for the G-action on Γ. It follows that $(hg^{-1}).u = u$, and since G acts freely, we can conclude that $hg^{-1} = 1$, i.e., $h = g$. □

Since T contains a representative from each G-orbit of vertices of Γ, Claim 3.7.2 implies that the set of vertices of Γ splits into the disjoint union of translats of the vertices of T:

$$V\Gamma = \bigsqcup_{g \in G} g.VT.$$

Let us define a new simple graph Δ whose vertices are G-translates of T and where two such translates are adjacent if and only if there is an edge of Γ joining one translate to the other. More precisely, we set

$$V\Delta = \{g.T \mid g \in G\} \quad \text{and} \quad E\Delta = \{(g.T, h.T) \in V\Delta \times V\Delta \mid \mathrm{d}_\Gamma(g.T, h.T) = 1\}. \tag{3.2}$$

Recall that d_Γ denotes the path metric on Γ, so the condition $\mathrm{d}_\Gamma(g.T, h.T) = 1$ means that $g.T \neq h.T$ and there is an edge $e \in E\Gamma$ such that $\alpha(e) \in g.VT$ and $\omega(e) \in h.VT$. Geometrically, the graph Δ is obtained from Γ by contracting all translates of T into vertices and by keeping all edges of Γ that did not belong to the orbit $G.ET$ (see Figure 3.13, where the translates of T are depicted in red and the edges connecting them are blue).

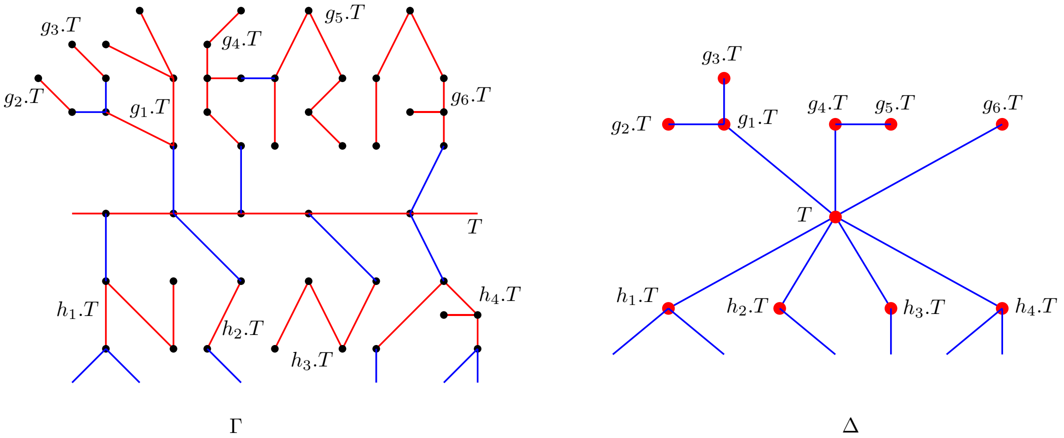


FIGURE 3.13. An example showing how translates of T (depicted in red) are contracted to get the graph Δ.

Claim 3.7.3. *The graph Δ is a tree.*

PROOF OF CLAIM 3.7.3. Exercise. [*Hint:* use the fact that T is connected to show that any cycle in Δ would give rise to a cycle in Γ.] □

Note that the action of G on Γ gives rise to a natural action $\circ$ of G on Δ. Indeed, we define

$$f \circ (g.T) = f.(g.T) = (fg).T \in V\Delta, \quad \text{for all } f \in G \text{ and } g.T \in V\Delta.$$

This action on the vertices of Δ preserves adjacency in Δ because G acts on Γ isometrically (see Remark 3.3.11): if $(g.T, h.T) \in E\Delta$ and $f \in G$ then

$$\mathrm{d}_\Gamma((fg).T, (fh).T) = \mathrm{d}_\Gamma(f.(g.T), f.(h.T)) = \mathrm{d}_\Gamma(g.T, h.T) = 1,$$

thus $f \circ (g.T, h.T) \in E\Delta$. Since Δ is a tree, it is a simple graph, therefore any group action on $V\Delta$ preserving adjacency can be uniquely extended to an action on the entire graph Δ. More precisely, for each edge $(g.T, h.T) \in E\Delta$, we set $f \circ (g.T, h.T) = (f \circ (g.T), f \circ (h.T)) = ((fg).T, (fh).T) \in E\Delta$, and the compatibility conditions from Definition 3.3.10 will be satisfied automatically.

Claim 3.7.4. *The above action of G on Δ is free and transitive on the vertices and is without edge inversions.*

Proof of Claim 3.7.4. If $f \in G$ fixes a vertex $g.T \in V\Delta$ then $f \circ (g.T) = (fg).T = g.T$, hence $fg = g$ by Claim 3.7.2, so $f = 1$. This shows that the action of G on T is free. It is also clearly transitive on the vertices, by the definition of $V\Delta$. Consider an edge $(g.T, h.T) \in E\Delta$ (see (3.2)). An element $f \in G$ inverting this edge must satisfy

$$f \circ (g.T) = f.(g.T) = h.T \text{ and } f \circ (h.T) = f.(h.T) = g.T.$$

Let e be the unique edge of Γ joining $g.T$ with $h.T$, i.e., $\alpha(e) \in g.VT$ with $\omega(e) \in h.VT$. This edge exists because $\mathrm{d}_\Gamma(g.T, h.T) = 1$ and it is unique because T is connected and Γ is a tree. Since f interchanges $g.T$ with $h.T$, it must send e to the unique edge of Γ joining $h.T$ with $g.T$, which clearly is the edge $\overline{e}$. Thus $f.e = \overline{e}$, contradicting the assumption that G acts on Γ without edge inversion. Therefore, the action of G on Δ must also be without edge inversions. ☐

We can now use Theorem 3.4.9 to conclude that Δ is isomorphic to the Cayley graph $\Gamma(G, X)$ of G, with respect to some generating subset $X \subseteq G$. Since Δ is a tree by Claim 3.7.3, the group G must be free and X must be a free generating set of G by Theorem 3.5.1. This completes the proof of Theorem 3.7.1. ☐

3.8. Nielsen-Schreier theorem revisited and Schreier index formula

In this section we will use the characterization of free groups as groups admitting free actions on trees to establish some properties of free groups. Let us start by giving a proof of the main result concerning subgroups of free groups, stated in Section 1.3.

Theorem 1.3.5 (Nielsen-Schreier Theorem). *Every subgroup of a free group is itself free.*

Proof. Let $F = F(X)$ be a free group on a set X. By Corollary 3.5.3, F admits a free action on a non-empty tree T without edge inversions. If $H \leqslant F$ is any subgroup then it has an induced action on the same tree T, which will still be free and without edge inversions. Therefore H must be a free group by Theorem 3.7.1. ☐

The next theorem gives another very useful result that can be used to compute ranks of finite index subgroups in free groups.

Theorem 3.8.1 (Schreier index formula). *Let F be a free group of finite rank $n \in \mathbb{N}_0$, and let $G \leqslant F$ be a subgroup of finite index $k \in \mathbb{N}$. Then G is free of rank $k(n-1)+1$. In particular, G is finitely generated.*

Remark 3.8.2. Note that if $\mathrm{rank}(F) \geq 2$, the formula from Theorem 3.8.1 can be re-written as

$$k = |F : G| = \frac{\mathrm{rank}(G) - 1}{\mathrm{rank}(F) - 1}.$$

In order to prove Theorem 3.8.1 we need to recall some basic notions from Graph Theory.

Definition 3.8.3. Let $\Gamma = (V, E, \alpha, \omega, ^{-})$ be a graph. Given any vertex $v \in V$, the *degree* of v, $\deg_\Gamma(v)$, is the number of edges $e \in E$ starting at v, i.e., such that $\alpha(e) = v$. The graph Γ is said to be *m-regular*, for $m \in \mathbb{N}_0$, if $\deg_\Gamma(v) = m$, for each $v \in V$.

Example 3.8.4. (a) Let us revisit the three graphs on Figure 3.1. In the graph Γ_1 all vertices have degree 0. In Γ_2 we have

$$\deg_{\Gamma_2}(v_1) = \deg_{\Gamma_2}(v_4) = 1 \text{ and } \deg_{\Gamma_2}(v_2) = \deg_{\Gamma_2}(v_3) = 2$$

(because v_2 is the initial vertex of the edges $\overline{e}_1$ and e_2). Finally, in the graph Γ_3 we have

$$\deg_{\Gamma_2}(w_1) = 3, \ \deg_{\Gamma_2}(w_2) = 2 \ \text{ and } \ \deg_{\Gamma_2}(w_3) = 3.$$

(b) Let Γ be the Cayley graph of a group G with respect to some subset $X \subseteq G$. Then, according to Definition 3.2.1, every vertex in Γ has degree $2|X|$ (in particular, if $|X| = \infty$ then every vertex will have infinite degree). Thus, if $|X| < \infty$ then Γ is m-regular, for $m = 2|X|$.

Since every edge has exactly one initial vertex, according to our definition of a graph, the following statement can be easily proved by induction of the number of edges.

Exercise 3.8.5 (Degree sum formula)**.** If $\Gamma = (V, E, \alpha, \omega, ^{-})$ is a graph with finitely many edges then

$$\sum_{v \in V} \deg_\Gamma(v) = |E|.$$

Definition 3.8.6. A vertex of degree 1 in a graph is called a *leaf.*

Lemma 3.8.7. *Every finite tree with at least one edge has at least one leaf.*

Proof. Let T be a finite tree. Choose any vertex $u \in VT$ and let $v \in VT$ be any vertex such that $\mathrm{d}_T(u, v)$ is the largest possible. Suppose that v is not a leaf, then there must exist two distinct edges $e_1, e_2 \in ET$ such that $v = \alpha(e_1) = \alpha(e_2)$. Note that $e_2 \neq \overline{e}_1$ as T is a tree (so it does not contain loops). Let p_1 be the shortest path in T from u to $w_1 = \omega(e_1)$ and let p_2 be the shortest path in T from $w_2 = \omega(e_2)$ to u. Observe that the concatenation $q = p_1\overline{e}_1 e_2 p_2$ is a non-trivial closed path in T (see Figure 3.14).

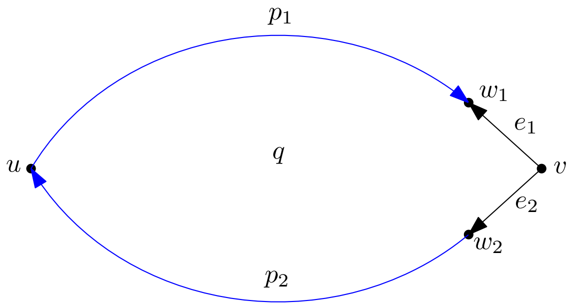


Figure 3.14. The cycle q from the proof of Lemma 3.8.7.

Since p_1 and p_2 are shortest paths between their endpoints, they cannot contain pairs of consecutive mutually inverse edges. Note that $\mathrm{d}_T(u, w_1) \leq \mathrm{d}_T(u, v)$, by the choice of v, hence p_1 cannot end with e_1 (because otherwise we would have $\mathrm{d}_T(u, v) \leq \|p_1\| - 1 = \mathrm{d}_T(u, w_1) - 1 < \mathrm{d}_T(u, v)$). Similarly, the path p_2 cannot start with $\overline{e}_2$. It follows that the closed path q does not contains consecutive pairs of mutually inverse edges, hence it must be a cycle. This contradicts the assumption that T is a tree, so the lemma is proved. □

In fact, the argument we used to prove Lemma 3.8.7 also applies to possibly infinite trees of finite diameter.

Exercise 3.8.8. (i) Improve the statement of Lemma 3.8.7 by showing that every tree with at least one edge and with finite diameter has at least two leaves.

(ii) Prove that if a finite tree has exactly two leaves then it is a simple path graph (consisting of $n \geq 2$ vertices $v_1, \dots, v_n$ and $2n-2$ edges $e_1, \dots, e_{n-1}, \bar{e}_1, \dots, \bar{e}_{n-1}$ such that $\alpha(e_i) = v_i$, $\omega(e_i) = v_{i+1}$, for $i = 1, \dots, n-1$).

Remark 3.8.9. If T is a tree and $v \in VT$ is a leaf then we can remove v (together with the edges e and $\bar{e}$ incident to it) to get another tree with fewer vertices. This observation allows one to prove statements about finite trees by induction of the number of vertices (or edges), which is often used in Graph Theory.

The approach suggested in Remark 3.8.9 can help in solving the following.

Exercise 3.8.10. Given a non-empty finite tree T, show that $|ET| = 2|VT| - 2$.

We are now in a position to prove the Schreier index formula.

Proof of Theorem 3.8.1. Let F be a free group with a free basis $X \subseteq F$ such that $|X| = n \in \mathbb{N}$. By Exercise 3.4.4, F acts freely on the Cayley graph $\Gamma = \Gamma(F, X)$ and this action is transitive on the vertices. Moreover, Γ is a $2n$-regular tree, by Example 3.8.4.(b) and Theorem 3.5.1.

Suppose that $G \leqslant F$ is a subgroup of index $k \in \mathbb{N}$. Let T be a spanning tree for the action of G on Γ induced by the natural action of F on Γ (see Proposition 3.6.4). Since $|F : G| = k$, we can write $F = \bigsqcup_{i=1}^{k} G f_i$, for some elements $f_1, \dots, f_k \in F$. By transitivity, there is a unique orbit of vertices under the action of F on Γ. And since this action is free, we can conclude that there are exactly k orbits of vertices for the induced action of G on Γ. In fact, given any vertex $v \in V\Gamma$, the vertices $f_1.v, \dots, f_k.v$ are the orbit representatives for the action of G on $V\Gamma$ (check this!).

By definition, the spanning tree T contains exactly one vertex from each G-orbit of vertices in Γ, hence $|VT| = k$. We will now analyse the proof of Theorem 3.7.1 to deduce the desired result. Recall that, by Claim 3.7.2, the subtrees $\{g.T\}_{g \in G}$ are pairwise disjoint and their union covers all vertices of Γ. Let r denote the number of edges e in Γ such that $\alpha(e) \in VT$ and $e \notin ET$. To calculate r, observe that the total number of edges starting at vertices of T is $2n|VT| = 2nk$, because Γ is $2n$-regular. On the other hand, the number of edges in T is $2k-2$ by Exercise 3.8.10. It follows that

$$r = 2nk - 2k + 2. \tag{3.3}$$

Recall that, in the proof of Theorem 3.7.1, after contracting each translate of T to a vertex we obtained a new graph Δ which turned out to be the Cayley graph $\Gamma(G, Y)$, for some free basis Y of G. The degree of the vertex of Δ corresponding to T is precisely the number r defined above. Since $\Delta \cong \Gamma(G, Y)$ is a $2|Y|$-regular graph, we conclude that $2|Y| = r$. Recalling (3.3), we obtain

$$\operatorname{rank}(G) = |Y| = r/2 = k(n-1) + 1,$$

which gives us the Schreier index formula. □

Example 3.8.11. Let F_5 denote the free group of rank 5 and let $H \leqslant F_5$ be a finite index subgroup such that $\operatorname{rank}(H) = 17$. Then we can find the index $i = |F : H|$ by the formula in Remark 3.8.2:

$$i = \frac{\operatorname{rank}(H) - 1}{\operatorname{rank}(F_5) - 1} = \frac{16}{4} = 4.$$

We can also deduce that F_5 contains no finite index subgroups isomorphic to the free group of rank 12, because $11 = 12 - 1$ is not divisible by $4 = 5 - 1$ (and the index of a subgroup is always an integer or infinity).

While the Schreier index formula is a statement for free groups, it has a corollary that applies to all finitely generated groups.

Corollary 3.8.12. *Finite index subgroups of finitely generated groups are themselves finitely generated.*

More precisely, let H be a group generated by a finite subset $X \subseteq H$, with $|X| = n \in \mathbb{N}_0$, and let $K \leqslant H$ be a subgroup of finite index $k \in \mathbb{N}$. Then K has a generating subset $Z \subseteq K$ such that $|Z| \leq k(n-1) + 1$.

Proof. Let $F = F(X)$ be the free group on X. By the Universal Property of free groups, there is an epimorphism $\psi : F \to H$, which extends the identity map $X \to X$. Denote $G = \psi^{-1}(K) \leqslant F$. The index of G in F is k, by Exercise 3.8.13 below, so, according to Theorem 3.8.1, G has a free generating set Y of cardinality $|Y| = k(n-1)+1$. Since ψ is surjective, we know that $\psi(G) = K$, whence K is generated by the set $Z = \psi(Y)$ and $|Z| \leq |Y| = k(n-1)+1$. □

Exercise 3.8.13. Suppose that $\psi : F \to H$ is a group epimorphism and $K \leqslant H$ is a subgroup of index $k \in \mathbb{N}$. Show that the full preimage $G = \psi^{-1}(K)$ has index k in F. [*Hint:* lift a coset decomposition $H = \bigsqcup_{i=1}^{k} Kh_i$, where $h_1, \dots, h_k \in H$, to a coset decomposition $F = \bigsqcup_{i=1}^{k} Gf_i$, for any $f_i \in \psi^{-1}(h_i)$, $i = 1, \dots, k$.]

Remark 3.8.14. Suppose that H is a cyclic group, so that it has a generating subset X with $|X| = n = 1$. Then Corollary 3.8.12 tells us that every finite index subgroup of H is cyclic, as it can be generated by at most $k(n-1)+1 = 1$ element. It is easy to see that each non-trivial subgroup of a cyclic group has finite index, thus Corollary 3.8.12 generalizes the fact that every subgroup of a cyclic group is cyclic (see Proposition 0.5.8).

Example 3.8.15. Let F be the free group on the set $\{x, y\}$ and consider the homomorphism $\phi : F \to S_3$ defined by $\phi(x) = (1\ 2)$ and $\phi(y) = (1\ 2\ 3)$. Since $(1\ 2)$ and $(1\ 2\ 3)$ generate the symmetric group S_3, we know that ϕ is surjective (by Lemma 0.5.11.(ii)). Therefore, after letting $N = \ker\phi \lhd F$, by the First Isomorphism Theorem we have $F/N \cong S_3$, in particular, $|F : N| = |S_3| = 6$. Schreier index formula (Theorem 3.8.1) tells us that N is free and

$$\operatorname{rank}(N) = |F : N|(\operatorname{rank}(F) - 1) + 1 = 6(2-1)+1 = 7.$$

Thus N is isomorphic to the free group of rank 7.

The subgroup $K = \langle (1\ 2) \rangle \leqslant S_3$ is cyclic of order 2, so it has index $|S_3 : K| = |S_3|/|K| = 3$ (by Lagrange's theorem). By Exercise 3.8.13, $G = \phi^{-1}(K)$ has index 3 in F, whence G is free of rank $3(2-1)+1 = 4$.

Example 3.8.16. The upper bound on the number of generators of the subgroup K in Corollary 3.8.12 is not always sharp. For example, consider the subgroup K of all the translations of $\mathbb{R}$ by integers in the infinite dihedral group D_∞. Then $K = \langle \rho_0 \rangle$ is cyclic, generated by the translation $\rho_0(x) = x - 1$ (see Example 2.5.5). Moreover, every element of D_∞ is either of the form ρ_0^m or of the form $\sigma_0\rho_0^m$, for some $m \in \mathbb{Z}$, where $\sigma_0 \in D_\infty$ is the reflection in 0. Therefore, $D_\infty = K \sqcup \sigma_0 K$, in particular $|D_\infty : K| = 2$. Since D_∞ is 2-generated, Corollary 3.8.12 tells us that we need at most $2(2-1)+1 = 3$ elements to generate K, but we know that K is cyclic, so it is actually 1-generated.

3.9. Finite groups and trees

In Corollary 3.5.3 we have seen that free groups admit free actions on non-empty trees. The purpose of this section is to show that for finite groups the situation is completely opposite: any action on a tree has a global fixed vertex. To establish this fact we will need some background.

The following exercise can be proved by induction on the number of vertices, as suggested in Remark 3.8.9.

Exercise 3.9.1. Let G be a group acting on a non-empty finite tree T without edge inversions. Then G fixes a vertex of T, that is there exists $v \in VT$ such that $g.v = v$, for all $g \in G$.

Definition 3.9.2. Given a connected graph Γ and two vertices $u, v \in V\Gamma$, a *geodesic* from u to v is any path p in Γ such that $\alpha(p) = u$, $\omega(p) = v$ and $\|p\| = \mathrm{d}_\Gamma(u, v)$. In other words, geodesics are shortest paths between their endpoints.

Exercise 3.9.3. If Γ is a tree then for any two vertices $u, v \in V\Gamma$ there exists a unique geodesic from u to v in Γ

Such graphs, where for any two vertices there is a unique geodesic from one to the other, are called *uniquely geodesic graphs.*

Exercise 3.9.4. Give an example of a connected uniquely geodesic graph that is not a tree.

Definition 3.9.5. If Γ is a connected graph we will use $[u,v]$ to denote any geodesic between vertices $u,v \in V\Gamma$. Given $n \in \mathbb{N}$, a *geodesic n-gon* P_n in Γ consists of the union of geodesics

$$[v_1,v_2] \cup [v_2,v_3] \cup \dots \cup [v_{n-1},v_n] \cup [v_n,v_1],$$

where $v_1,\dots,v_n \in V\Gamma$ are called *vertices* of P_n and the geodesics $[v_1,v_2],\dots,[v_n,v_1]$ are called *sides* of P_n.

A geodesic triangle with vertices u,v,w is a *tripod* if

$$[u,v] \subseteq [u,w] \cup [w,v], \;\; [v,w] \subseteq [v,u] \cup [u,w], \;\; \text{and} \;\; [w,u] \subseteq [w,v] \cup [v,u]$$

(see Figure 3.15).

In other words, a geodesic triangle is a tripod if each side is contained in the union of the two other sides. However, due to our definition of graphs and paths, we have to choose the directions of the side paths carefully for this to work (because, formally speaking, while the geodesic paths $[u,v]$ and $[v,u]$ share the same set of vertices, all of their edges are different).

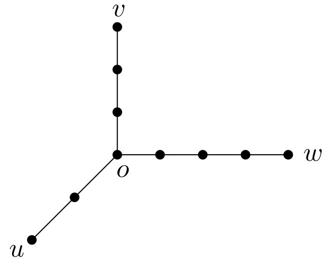


Figure 3.15. A tripod with vertices u,v,w in a graph.

The following characterization of trees is useful to know.

Proposition 3.9.6. *Let Γ be a connected graph. Then Γ is a tree if and only if every geodesic triangle in Γ is a tripod.*

Proof. Assume that Γ is a tree. Arguing by contradiction, suppose that there exist vertices $u,v,w \in V\Gamma$ such that the geodesic $[u,v]$ is not contained in the union of the geodesics $[u,w]$ and $[w,v]$. Without loss of generality we can assume that the perimeter of the geodesic triangle with vertices u,v,w (i.e., $\mathrm{d}_\Gamma(u,v)+\mathrm{d}_\Gamma(v,w)+\mathrm{d}_\Gamma(w,u)$) is the smallest possible. Then the concatenation $p = [u,v]\,[v,w]\,[w,u]$ is a non-trivial closed path in Γ, and this path cannot contain consecutive pairs of mutually inverse edges by the minimality assumption on the perimeter and because the sides of the triangle are geodesic. Therefore p is a cycle in Γ, contradicting the definition of a tree. Thus every geodesic triangle in Γ must be a tripod.

For the opposite implication, suppose that every geodesic triangle in a connected graph Γ is a tripod. If Γ is not a tree, then let $p = e_1 \dots e_n$ be a cycle of the shortest possible length $n \in \mathbb{N}$. Note that if $l \in \mathbb{N}$ and $l \leq n/2$ then the subpath $q = e_1 \dots e_l$ of p is a geodesic in Γ. Indeed, otherwise, choose the smallest l such that there exists a path r from $\omega(q)$ to $\alpha(q)$ of length less than $\|q\| = l$. Then the concatenation qr will be a cycle of length $\|q\|+\|r\| < 2l \leq n$, contradicting the minimality assumption on $n = \|p\|$. Similarly, and any $m \in \mathbb{N}$ such that $m \geq n/2$, the subpath $e_m \dots e_n$ of p must also be a geodesic in Γ.

If $n \in \mathbb{N}$ is even, then take $l = n/2$ and $m = l+1$, and if n is odd take $l = (n-1)/2$ and $m = l+1 = (n+1)/2$. Let $u = \alpha(p)$, $v = \alpha(e_m)$ and $w = \omega(e_m)$. Then, by the discussion in the previous paragraph, u,v,w are vertices of the geodesic triangle Δ, with sides $[u,v] = e_1 \dots e_l$, $[v,w] = e_m$ and $[w,u] = e_{m+1} \dots e_n$. By our assumption on Γ, Δ is a tripod, so the side $[v,w] = e_m$ must be contained in the union of $[v,u] = \overline{e}_l \dots \overline{e}_1$ with $[u,w] = \overline{e}_n \dots \overline{e}_{m+1}$. In the former case, e_m must equal $\overline{e}_l$ (because the path $[v,u]$ is geodesic) and in the latter case $e_m = \overline{e}_{m+1}$. This contradicts the assumption that p is a cycle, so the proof is complete. □

Given two vertices u,v in a graph Γ, the *discrete geodesic* $[u,v]^d$ between u and v is the set of vertices on a geodesic path $[x,y]$ in Γ (in other words, $[u,v]^d = [u,v] \cap V\Gamma$). The benefit of discrete geodesics is that they are not oriented (unlike the usual geodesics), so $[u,v]^d = [v,u]^d$, for any $u, v \in \Gamma$. The following fact is useful to know.

Exercise 3.9.7. Show that for any three vertices u, v, w in a tree Γ the intersection

$$[u,v]^d \cap [u,w]^d \cap [v,w]^d$$

consists of a single vertex $o \in V\Gamma$ (see Figure 3.15).

THEOREM 3.9.8. *If a finite group G acts on a non-empty tree Γ without edge inversions then G fixes some vertex $v \in VT$, that is, $g.v = v$, for all $g \in G$.*

PROOF. Choose any vertex $u \in VT$ and consider its orbit $G.u$. The *convex hull* T, of this orbit in Γ, is the subgraph of T consisting of all vertices and edges of the geodesic paths $[g.u, h.u]$, where $g, h \in G$ are arbitrary. Since $[g.u, h.u] \subseteq [g.u, u] \cup [u, h.u]$ by Proposition 3.9.6, we see that T is equal to the union of all geodesic paths $[g.u, u]$ and $[u, g.u]$, for $g \in G$. In particular, T is connected. Since Γ is a tree, T will also be a tree. Since G is finite, T is finite; it is non-empty because $u \in VT$.

Let us show that T is invariant under the action of G. Indeed, since G acts on Γ by isometries, we have $f.[a,b] = [f.a, f.b]$, for any vertices a, b in Γ and any element $f \in G$. Therefore, for every $f \in G$ we have

$$f.T = f.\left(\bigcup_{g,h\in G} [g.u, h.u]\right) = \bigcup_{g,h\in G} [(fg).u, (fh).u] = \bigcup_{g',h'\in G} [g'.u, h'.u] = T,$$

where we used the fact that when g and h run over arbitrary elements of G, the same is true for $g' = fg$ and $h' = fh$. Thus $f.T = T$, for all $f \in G$, so T is G-invariant. It follows that we have an action of G on the finite tree T, which is induced from the action of G on Γ. Hence, we can apply Exercise 3.9.1 to find a vertex $v \in VT$ such that $g.v = v$, for all $g \in G$. ☐

Corollary 3.9.9. *Every free group F is* torsion-free*. This means that if $f \in F$ is an element of finite order then $f = 1$ in F. Therefore, the only finite subgroup of F is the trivial subgroup.*

PROOF. By Corollary 3.5.3, the free group F admits a free action on a non-empty tree Γ without edge inversions. If $f \in F$ is an element of finite order, then the cyclic subgroup $G = \langle f \rangle \leqslant F$ is finite. Moreover the action of F on Γ induces an action of G on Γ (without edge inversions). By Theorem 3.9.8, the latter action must have a global fixed vertex, i.e., $f.v = v$, for some $v \in V\Gamma$. It follows that $f = 1$ because the action of F on Γ is free. ☐

3.10. Bass-Serre trees for free products

Free groups are not the only groups admitting interesting actions on trees. In this section we will construct trees for free products of two groups, and we will also discuss the theory of groups acting on trees with trivial edge stabilizers.

Let $G = A * B$ be the free product of two groups A and B (see Section 2.4). Recall that we can treat A and B as subgroups of G, and $A \cap B = \{1\}$ in G by Theorem 2.5.3.

Definition 3.10.1. The *Bass-Serre tree* for the free product G is the graph $\Gamma = (V, E, \alpha, \omega, ^{-})$ defined as follows. The vertex set V is the union of the sets of left cosets of G modulo A and B:

$$V = \{gA \mid g \in G\} \sqcup \{gB \mid g \in G\}.$$

The edges $E = E^+ \sqcup E^-$ come with a natural orientation, where the positively oriented edges are

$$E^+ = \{(gA, gB) \mid g \in G\}.$$

For an edge $e = (gA, gB) \in E^+$, we have $\alpha(e) = gA$ and $\omega(e) = gB$ (thus $\alpha(\overline{e}) = gB$ and $\omega(\overline{e}) = gA$).

Thus a vertex $u = gA \in V$, corresponding to a coset of A, is only adjacent to vertices which correspond to cosets of B. Note that for any $a \in A$ we have $gA = gaA$. It follows that there is a positively oriented edge from $u = gA$ to gaB, for every $a \in A$. Moreover, if $a, a' \in A$ and $a \neq a'$ then $gaB \neq ga'B$ (because otherwise we would have $aB = a'B$, which implies that $a^{-1}a' \in B$, so $a^{-1}a' \in A \cap B = \{1\}$, hence $a = a'$ in A). Thus the edges (gA, gaB) and $(gA, ga'B)$ are distinct.

Conversely, if $e \in E^+$ joins a coset gA with a coset hB, for some $g, h \in G$, then there must exist $f \in G$ such that $gA = fA$ and $hB = fB$. The former equation implies that $g^{-1}f = a \in A$, so $f = ga$, whence $hB = gaB$. We summarize this in the following remark.

Remark 3.10.2. Let $u = gA$ be a vertex of the Bass-Serre tree Γ, where $g \in G = A * B$. Then every positively oriented edge of Γ starting at u has the form (gA, gaB), for some $a \in A$. Moreover, if $a, a' \in A$ and $a \neq a'$ then the edges (gA, gaB) and $(gA, ga'B)$ have different terminal vertices (in particular, these edges are distinct). Thus, $\deg_\Gamma(gA) = |A|$.

Similarly, the positively oriented edges that terminate at a vertex gB of Γ are of the form (gbA, gB), for $b \in B$, and $gbA \neq gb'A$ if $b, b' \in B$ and $b \neq b'$. Therefore, $\deg_\Gamma(gB) = |B|$.

Recall that a graph is said to be *bipartite* if we can find a partition $V\Gamma = V_1 \sqcup V_2$ such that $V_i \neq \emptyset$ and no two vertices from V_i are adjacent, $i = 1, 2$ (that is, every edge starting at a vertex from V_1 must end at a vertex from V_2 and vice-versa). The following observation is an immediate consequence of Definition 3.10.1 (see Definition 3.4.6 for the definition of a simple graph).

Remark 3.10.3. The Bass-Serre tree for a free product $G = A * B$ is a simple bipartite graph.

Let us now construct Bass-Serre trees in some examples.

Example 3.10.4. (a) If A is any group and B is the trivial group then $G = A * B = A$, so there is only one left coset modulo A, but the set of left cosets modulo B, $\{aB \mid a \in A\}$, is in one-to-one correspondence with A. For example, if $A = \{a_1, \dots, a_n\}$ has $n \in \mathbb{N}$ elements then the Bass-Serre tree Γ is a *star* graph, with $|V\Gamma| = n + 1$ and $|E\Gamma| = 2n$ (see Figure 3.16).

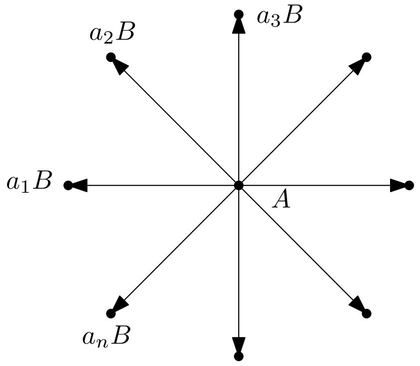


FIGURE 3.16. The Bass-Serre tree for $A * B$ when $A = \{a_1, \dots, a_n\}$ and $B = \{1\}$.

(b) Let us sketch the Bass-Serre tree Γ for $G = A * B$, where $A = \langle a \rangle \cong C_2$ and $B = \langle b \rangle \cong C_3$ (it is known that $G \cong C_2 * C_3$ is isomorphic to the modular group $\mathrm{PSL}(2, \mathbb{Z})$, see Proposition 5.2.2). Note that even though both groups A and B are finite, their free product G is infinite (see Exercise 2.5.6). It follows that $|G : A| = \infty$ and $|G : B| = \infty$, so the Bass-Serre tree Γ is an infinite graph.

To sketch Γ, we start by plotting the vertex A. By Remark 3.10.2, there are two positively oriented edges from this vertex to the vertices B and aB, as $A = \{1, a\}$. Since $B = \{1, b, b^2\}$, the vertex B is adjacent to the vertices A, bA and b^2A. On the other hand, the vertex aB is adjacent to the vertices $aA = A$, abA and ab^2A. On Figure 3.17 we have sketched the ball of radius 4 around the vertex A in the Bass-Serre tree Γ for G.

Lemma 3.10.5. *The free product $G = A * B$ has a natural action on the Bass-Serre tree Γ and this action preserves edge orientation. Vertex stabilizers in this action are conjugates of A and B*

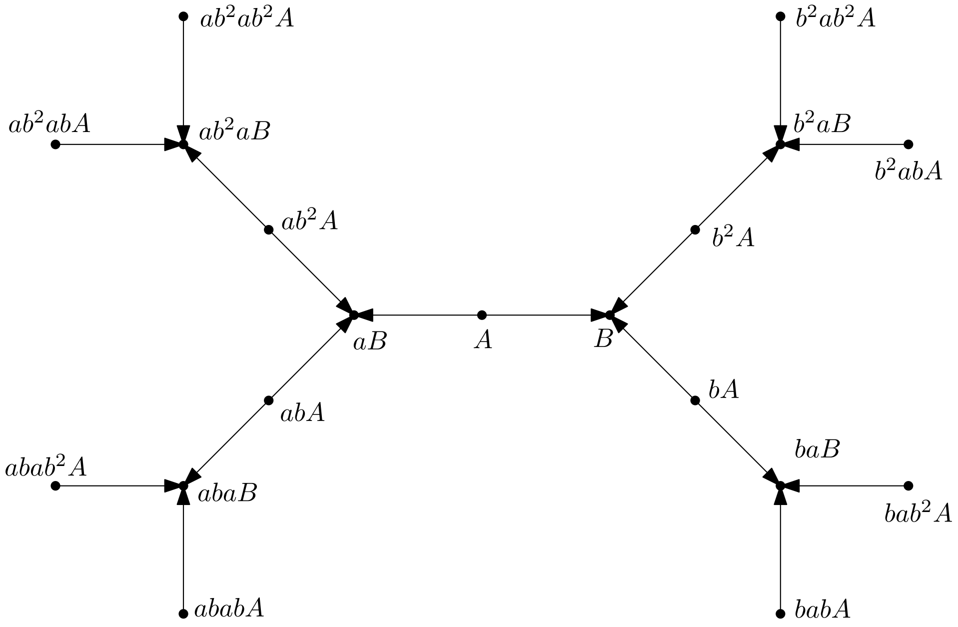


FIGURE 3.17. The Bass-Serre tree for $A * B$, where $A = \langle a \rangle \cong C_2$ and $B = \langle b \rangle \cong C_3$.

in G and edge stabilizers are trivial (i.e., $\mathrm{Stab}_G(e) = \{1\}$*, for every* $e \in \Gamma$*). In other words, the action of* G *on the set of edges of* Γ *is free.*

PROOF. The natural action of G on the vertices of Γ is induced by the left action of G on the left cosets modulo A and B:

$$h.gA = hgA \quad \text{and} \quad h.gB = hgB, \quad \text{for any } g, h \in G.$$

Similarly, for an edge $(gA, gB) \in E\Gamma$ and any $h \in G$ we set

$$h.(gA, gB) = (hgA, hgB).$$

It is easy to check that the above define actions of G on $V\Gamma$ and on $E\Gamma$ and these action are compatible in the sense of Definition 3.3.10. Therefore, we have defined a natural action of G on Γ. Evidently, this action preserves orientation of the edges of the graph.

Given a vertex $gA \in V\Gamma$, and element $h \in G$ stabilizes it if $hgA = gA$, which is equivalent to $h \in gAg^{-1}$. Thus $\mathrm{Stab}_G(gA) = gAg^{-1}$. Similarly, $\mathrm{Stab}_G(gB) = gBg^{-1}$. If $e = (gA, gB)$ is a positively oriented edge in Γ and $h \in G$ then $h.e = e$ if and only if $hgA = gA$ and $hgB = gB$. This can be re-written as

$$h \in gAg^{-1} \cap gBg^{-1} = g(A \cap B)g^{-1} = \{1\},$$

so $h = 1$ in G. Thus $\mathrm{Stab}_G(e) = \{1\}$; it follows that $\mathrm{Stab}_G(\overline{e}) = \{1\}$, whence the natural action of G on the set of edges of Γ is free. □

As the terminology of Definition 3.10.1 suggests, the Bass-Serre tree is indeed a tree.

THEOREM 3.10.6. *Let* $G = A * B$*, where* A *and* B *are groups. Then the graph* Γ *given by Definition 3.10.1 is a tree.*

PROOF. Arguing by contradiction, suppose that the graph $\Gamma = (V, E, \alpha, \omega, ^{-})$ contains a cycle p. Since Γ is bipartite, p must have a positive even length, so $p = e_1 e_2 \dots e_{2n}$, for some $n \in \mathbb{N}$. In particular, p will contain a vertex gA, for some $g \in G$, and, after renumbering, we can assume that $\alpha(p) = \omega(p) = gA$. By Remark 3.10.2, there are elements $a_1, \dots, a_n \in A$ and $b_1, \dots, b_n \in B$ such that

$$e_1 = (gA, ga_1B) \in E^+,\ e_2 = (ga_1B, ga_1b_1A) \in E^-, \dots, e_n = (ga_1b_1 \dots a_nB, ga_1b_1 \dots a_nb_nA) \in E^-.$$

Moreover, since p is a cycle, we must have $e_{j+1} \neq \overline{e}_j$, for $j = 1, \dots, 2n-1$, which implies that $a_i \neq 1$ in A and $b_i \neq 1$ in B, for all $i = 1, \dots, n$. Also, $\alpha(p) = \alpha(e_1)$ must be the same vertex as $\omega(p) = \omega(e_n)$, whence $gA = ga_1b_1 \dots a_nb_nA$, i.e., $a_1b_1 \dots a_nb_n = a_0 \in A$. Since $n > 0$, we see that $(a_0^{-1}a_1)b_1 \dots a_nb_n$ is a non-empty reduced word over the free product $G = A * B$, in the sense of Definition 2.5.2. We have shown that this word represents the identity element of G, which contradicts Theorem 2.5.3. Therefore the graph Γ cannot contain any cycles.

By Remark 2.4.6, every element of G can be written as a product of elements from A and B. This implies that Γ is connected (see Exercise 3.10.7 below). Therefore, Γ is a tree and the proof is complete. ☐

Exercise 3.10.7. Show that the graph Γ from Definition 3.10.1 is connected.

Remark 3.10.8. A Bass-Serre tree Γ can be defined for an arbitrary free product $G = *_{i \in I} A_i$. Vertices in this tree will again be left cosets $\{gA_i \mid g \in G,\ i \in I\}$, but the definition of the edges is less canonical, as it depends on the choice of a *quotient tree* for the action of G on Γ. The free product G will still act on Γ without edge inversions, with vertex stabilizers that are conjugates of A_i, $i \in I$, and with trivial edge stabilizers (cf. Lemma 3.10.5).

We can use the action on the Bass-Serre tree to establish the fact that every finite index subgroup of a free product must be contained in a conjugate of some free factor.

Corollary 3.10.9. *Let $G = *_{i \in I} A_i$ be the free product of a family of groups $\{A_i\}_{i \in I}$. If $M \leqslant G$ is a finite subgroup then there exist $g \in G$ and $i \in I$ such that $M \subseteq gA_ig^{-1}$.*

Proof. Let Γ be the Bass-Serre tree for G (as in Definition 3.10.1 in case $|I| = 2$ or as in Remark 3.10.8 if $|I| > 2$). Then G acts on Γ without edge inversions, inducing an action of the finite subgroup M on Γ without edge inversions. By Theorem 3.9.8, there is a vertex $v \in V\Gamma$ fixed by all elements of M. Now, according to Remark 3.10.8, $\mathrm{Stab}_G(v) = gA_ig^{-1}$, for some $g \in G$ and some $i \in I$, whence $M \subseteq \mathrm{Stab}_G(v) = gA_ig^{-1}$. ☐

In particular, Corollary 3.10.9 tells us that any element of $G = A * B$ that is not contained in a conjugate of A or B must have infinite order in G.

3.11. A theorem of Bass-Serre and applications

Lemma 3.10.5 and Theorem 3.10.6 show that free products of groups admit actions on trees that are free on the edges and are without edge inversions, which can be viewed as a generalization of the fact that free groups possess free actions on trees (see Corollary 3.5.3). And conversely, by Theorem 3.7.1, a group acting freely and without edge inversions on a tree must be free. This latter statement can also be generalized as follows.

Theorem 3.11.1 (Bass-Serre theorem). *Suppose that G is a group acting on a tree Γ without edge inversions and with trivial edge stabilizers. Then G splits as a free product*

$$G = (*_{i \in I} A_i) * F,$$

where I is some index set, $F \leqslant G$ is a free subgroup, and for each $i \in I$ there exists a vertex $v_i \in V\Gamma$ such that $A_i = \mathrm{Stab}_G(v_i) \leqslant G$.

Proof. Omitted. (See, for example, the book [J.-P. Serre, *Trees.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003.]) ☐

The following result, due to Aleksandr Kurosh, describes the structure of subgroups of free products, and is a generalization of the Nielsen-Schreier Theorem (Theorem 1.3.5) for free groups.

Theorem 3.11.2 (Kurosh subgroup theorem). *Let $G = *_{i \in I} A_i$ be the free product of a family of groups $\{A_i\}_{i \in I}$. If $H \leqslant G$ is any subgroup then $H \cong (*_{j \in J} H_j) * F$, where J is some index set, F is a free group, and for every $j \in J$ there exists $g_j \in G$ and $i \in I$ such that $H_j = H \cap g_jA_ig_j^{-1}$.*

Proof. As mentioned in Remark 3.10.8, the free product G admits an action on a tree Γ without edge inversions such that the vertex set of Γ is the set of all left cosets $\{gA_i \mid g \in G,\ i \in I\}$ and edge stabilizers are trivial. Since $H \leqslant G$, this action restricts to an action of H on Γ without edge inversions and with trivial edge stabilizers. For any vertex $v = gA_i$ of Γ, where $g \in G$ and $i \in I$, we have $\mathrm{Stab}_G(v) = gA_ig^{-1}$ (just like in the proof of Lemma 3.10.5), therefore

$$\mathrm{Stab}_H(v) = H \cap \mathrm{Stab}_G(v) = H \cap gA_ig^{-1}.$$

We can now apply Theorem 3.11.1 to obtain a desired free product decomposition for H. □

Example 3.11.3. Let A and B be groups, and consider their free product $G = A * B$ and their direct product $P = A \times B$. By the Universal Property of free products (Theorem 2.4.7) there exists a group homomorphism $\phi : G \to P$ such that $\phi(a) = (a, 1)$, for all $a \in A$, and $\phi(b) = (1, b)$, for all $b \in B$. Let $N = \ker\phi$ be the kernel of this homomorphism and let Γ be the Bass-Serre tree for the free product G (see Theorem 3.10.6). The natural action of G on Γ induces an action of N on Γ and this action is without edge inversions by Lemma 3.10.5. The latter statement also tells us that edge stabilizers under the action of G on Γ are of the form gAg^{-1} or gBg^{-1}, for some $g \in G$.

Given any $g \in G$, observe that $N \cap gAg^{-1} = \{1\}$ in G. Indeed, clearly $N \cap A = \{1\}$ because $N = \ker\phi$ and ϕ is injective on A, by definition. And since $N \lhd G$, we have $N = gNg^{-1}$, hence $N \cap gAg^{-1} = g(N \cap A)g^{-1} = g\{1\}g^{-1} = \{1\}$. Similarly, $N \cap gBg^{-1} = \{1\}$ in G. Therefore for any vertex $v \in V\Gamma$ we have

$$\mathrm{Stab}_N(v) = N \cap \mathrm{Stab}_G(v) = \{1\},$$

which means that the action of N on the tree Γ is free on the vertices. As a result we can apply Theorem 3.7.1 to conclude that N must be a free group.

Now, suppose that the groups A and B are finite. Then $P = A \times B$ is also finite, and $|P| = |A| \cdot |B|$. Since $P = A \times B$ is generated by $(A, 1) = \phi(A)$ and $(1, B) = \phi(B)$, we know that ϕ is surjective by Lemma 0.5.11.(ii) (in fact, $\phi(ab) = (a, b)$, for all $a \in A$ and $b \in B$). By the First Isomorphism Theorem, $G/N \cong P$, whence $|G : N| = |P| < \infty$, so N has finite index in P. Thus we have shown that the free group N has finite index in the free product $G = A * B$.

A group G is said to be *virtually free* if it has a free subgroup of finite index. We can use Kurosh subgroup theorem to generalize Example 3.11.3 as follows.

Proposition 3.11.4. *Let $A_1, \dots, A_n$ be a finite collection of non-trivial finite groups, with $n \in \mathbb{N}$. Then the free product $G = A_1 * \cdots * A_n$ is virtually free. In fact, the kernel N of the natural epimorphism $\phi : G \to A_1 \times \cdots \times A_n$ is free and finitely generated. Moreover, if $n \geq 3$ or if $n = 2$ and $|A_1| \geq 3$, $|A_2| \geq 2$ then $\mathrm{rank}(N) \geq 2$.*

Proof. Exercise. □

Example 3.11.5. We can actually compute the rank of the finite index free subgroup N explicitly. For instance, let $G = A*B$, where $A = \langle a \rangle \cong C_2$ and $B = \langle b \rangle \cong C_3$, as in Example 3.10.4.(b). Then Example 3.11.3 tells us that $N = \ker\phi$, where $\phi : A*B \to A \times B$, acts freely on the Bass-Serre tree Γ of G, depicted on Figure 3.17. We can therefore calculate $\mathrm{rank}(N)$ using a method similar to the argument in the proof of the Schreier Index formula (Theorem 3.8.1). More precisely, we will find a spanning tree for the action of N on Γ and then the number of edges that start in T but do not belong to T will be exactly $2\,\mathrm{rank}(N)$ (see the proof of Theorem 3.8.1).

In order to find a spanning tree for the action of N on Γ, we first need to determine the orbits for this action. Observe that two vertices $gA, hA \in V\Gamma$ are in the same N-orbit if and only if $N.gA = N.hA$, which is equivalent to $gAN = hAN$, as $N \lhd G$. Thus the number of different orbits of the form $N.gA$ is exactly the number of left cosets of the subgroup AN in G. Since $AN = \phi^{-1}((A, 1))$, Exercise 3.8.13 tells us that $|G : AN| = |(A \times B) : (A, 1)| = |B|$ and coset representatives for $(A, 1)$ in $A \times B$ can be "pulled back" to coset representatives of AN in G.

Since $B = \{1, b, b^2\}$, the elements $(1, 1)$, $(1, b)$ and $(1, b^2)$ represent the three distinct left cosets of $(A, 1)$ in $A \times B$, hence we can choose $1_G \in \psi^{-1}((1, 1))$, $b \in \phi^{-1}((1, b))$ and $b^2 \in \phi^{-1}((1, b^2))$ as representatives for the left cosets of AN in G. Thus there are $3 = |B|$ different N-orbits of vertices of the form $g.A$ in Γ:

$$N.A, \ N.bA \ \text{ and } \ N.b^2A.$$

Similarly, there are $2 = |A|$ different N-orbits of vertices of the form $g.B$ in Γ:

$$N.B \ \text{ and } \ N.aB.$$

It follows that as a spanning tree T for the action of N on Γ we can take the subtree of Γ containing 5 vertices A, B, bA, b^2A, aB and the edges joining them, see Figure 3.18. On this diagram we see that there are exactly 4 edges starting at a vertex of T but ending outside T, whence $\operatorname{rank}(N) = 4/2 = 2$, i.e., N is the free group of rank 2.

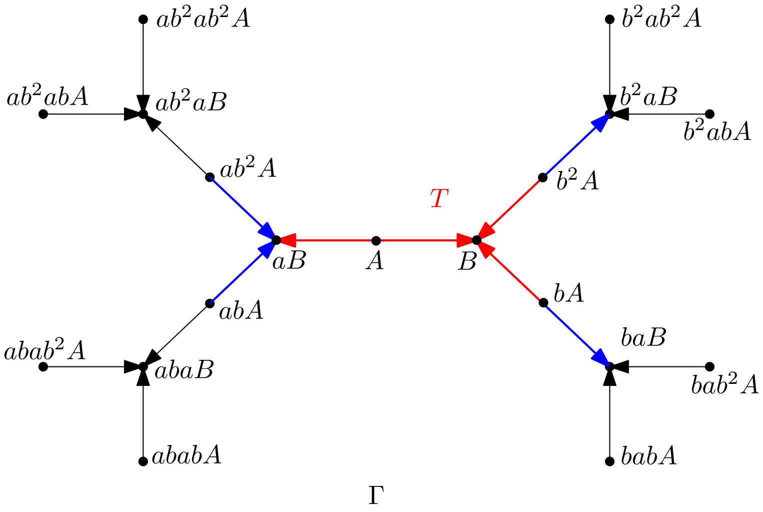


FIGURE 3.18. A spanning tree T for the action of N on the Bass-Serre tree Γ (highlighted in red) and the 4 edges emerging from it (highlighted in blue).

Example 3.11.5 can be generalised as follows.

Exercise 3.11.6. Show that if A and B are finite groups then the kernel $N = \ker \phi \lhd A * B$, of the natural homomorphism $\phi : A * B \to A \times B$, is free of rank $(|A| - 1)(|B| - 1)$.

CHAPTER 4

Hyperbolic spaces, groups and quasi-isometries

This chapter discusses a *hyperbolicity* condition for metric spaces introduced by Mikhail Gromov. Hyperbolic metric spaces generalize both trees, seen in Chapter 3, and the classical hyperbolic n-space from Hyperbolic Geometry. We will then study quasi-isometries, which induce a natural equivalence relation between metric spaces that preserves hyperbolicity. Finally, we will investigate groups that admit "nice" actions on hyperbolic spaces.

4.1. Rectifiable paths and path length

In this section we define rectifiable paths in a metric space and study some basic properties of such paths. Throughout this section (S, d) will denote a metric space.

Definition 4.1.1. A *path* p in S is a continuous function $p : [a, b] \to S$, where $[a, b]$ is a non-empty interval in $\mathbb{R}$. In this case we will say that p *starts* at $p(a) \in S$ and *ends* at $p(b) \in S$, writing $\alpha(p) = p(a)$ and $\omega(p) = p(b)$.

A *subpath* r of the path p is just a restriction of p to a non-empty subinterval $[c, d]$ of $[a, b]$ (i.e., $r : [c, d] \to \mathbb{R}$ is defined by $r(t) = p(t)$ for all $t \in [c, d]$). We will denote this subpath $p_{[c,d]}$.

A *curve* C in S is the image of a path $p : [a, b] \to S$, i.e., C is the subset of S defined by $C = \{p(t) \mid t \in [a, b]\}$. In this case the path p is called a *parametrization* of C.

Example 4.1.2. Consider the path $p : [0, 1] \to \mathbb{R}^2$, given by the formula $p(t) = (\cos(2\pi t), \sin(2\pi t))$. Clearly the curve defined by p is the unit circle $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. Of course there are many other parametrizations of the unit circle (e.g., $q : [\pi, 2\pi] \to \mathbb{R}^2$, $q(t) = (\sin(-2t), \cos(-2t))$ is another one).

Definition 4.1.3. A path $p : [a, b] \to S$ is said to be *rectifiable* if the quantity

$$\|p\| = \sup_{a=t_1 \le t_2 \le \dots \le t_n = b} \sum_{i=1}^{n-1} \mathrm{d}(p(t_i), p(t_{i+1}))$$

is finite. In this case $\|p\|$ is called the *length* of the path p.

Example 4.1.4. Suppose that $n \in \mathbb{N}$ and $p : [a, b] \to \mathbb{R}^n$ is a differentiable function with continuous derivative. If we equip $\mathbb{R}^n$ with the Euclidean metric, then a standard result from Analysis tells us that the path p is rectifiable and

$$\|p\| = \int_a^b \|p'(t)\| \, dt,$$

where $\|p\|$ denotes the length of p and $\|p'(t)\|$ is the Euclidean norm of the derivative $p'(t)$.

Lemma 4.1.5. *Suppose that $p : [a, b] \to S$ is a path and $c \in [a, b]$ is any point. Let $p_1 : [a, c] \to S$ and $p_2 : [c, b] \to S$ be the subpaths of p that are its restrictions to the subintervals $[a, c]$ and $[c, b]$ respectively. Then p is rectifiable if and only if both p_1 and p_2 are rectifiable. In this case we also have $\|p\| = \|p_1\| + \|p_2\|$.*

PROOF. Any partitions $a = u_1 \le \dots \le u_k = c$ and $c = v_1 \le \dots \le v_l = b$, of the intervals $[a, c]$ and $[c, b]$ respectively, can be joined together, giving rise to the partition $a = u_1 \le \dots \le u_k = v_1 \le \dots \le v_l = b$ of $[a, b]$. Consequently, we have

$$\sum_{i=1}^{k-1} \mathrm{d}(p(u_i), p(u_{i+1})) + \sum_{j=1}^{l-1} \mathrm{d}(p(v_j), p(v_{j+1})) \le \sup_{a=t_1 \le t_2 \le \dots \le t_n = b} \sum_{i=1}^{n-1} \mathrm{d}(p(t_i), p(t_{i+1})) = \|p\|.$$

Therefore

$$\sup_{a=u_1\le\cdots\le u_k=c}\sum_{i=1}^{k-1}\mathrm{d}(p(u_i),p(u_{i+1}))+\sup_{c=v_1\le\cdots\le v_l=b}\sum_{j=1}^{l-1}\mathrm{d}(p(v_j),p(v_{j+1}))\le\|p\|,$$

which implies that the paths p_1 and p_2 are rectifiable and $\|p_1\|+\|p_2\|\le\|p\|$.

To prove the opposite inequality, consider any partition $a=t_1\le t_2\le\cdots\le t_n=b$ of the interval $[a,b]$. Then there exists $k\in\{2,\dots,n\}$ such that $c\in[t_{k-1},t_k]$. Denote $u_1=t_1=a,\dots,u_{k-1}=t_{k-1}$, $u_k=c$ and $v_1=c$, $v_2=t_k,\dots,v_l=t_n=b$, where $l=n-k+1$, so that $a=u_1\le\cdots\le u_k=c$ and $c=v_1\le\cdots\le v_l=b$ are partitions of the intervals $[a,c]$ and $[c,b]$ respectively. The triangle inequality implies that $\mathrm{d}(p(t_{k-1}),p(t_k))\le\mathrm{d}(p(t_{k-1}),p(u_k))+\mathrm{d}(p(v_1),p(t_{k+1}))$, hence

$$\sum_{i=1}^{n-1}\mathrm{d}(p(t_i),p(t_{i+1}))\le\sum_{i=1}^{k-1}\mathrm{d}(p(u_i),p(u_{i+1}))+\sum_{j=1}^{l-1}\mathrm{d}(p(v_j),p(v_{j+1}))\le\|p_1\|+\|p_2\|.$$

It follows that p is rectifiable and $\|p\|\le\|p_1\|+\|p_2\|$, so $\|p\|=\|p_1\|+\|p_2\|$ and the lemma is proved. □

Exercise 4.1.6. (i) Prove that for a rectifiable path p in S we always have $\|p\|\ge\mathrm{d}(\alpha(p),\omega(p))$.

(ii) Show that any subpath of a rectifiable path is rectifiable.

The following exercise can be solved by noticing that any monotonic (i.e., increasing or decreasing) surjective function between two non-empty intervals $[a,b]$ and $[c,d]$ of $\mathbb{R}$ induces a surjection between the partitions of these intervals.

Exercise 4.1.7. Let $[a,b]$ and $[c,d]$ be non-empty intervals in $\mathbb{R}$ and let $f:[c,d]\to[a,b]$ be a continuous monotonic surjective function. Then for any rectifiable path $p:[a,b]\to S$ the path $p'=p\circ f:[c,d]\to S$ is rectifiable and satisfies $\|p'\|=\|p\|$.

Example 4.1.8. We can use Exercise 4.1.7 to show that for any $a,b,c,d\in\mathbb{R}$ such that $c<d$, every path $p:[a,b]\to S$ can be *re-parametrized* as a path $p':[c,d]\to S$ so that p and p' have the same start and end points, the same length and give rise to the same curve in S. Indeed, let $f:[c,d]\to[a,b]$ be the function defined by $f(t)=a+\dfrac{a-b}{d-c}(t-c)$, for all $t\in[c,d]$. Then f is continuous, increasing and surjective, so we can set $p'=p\circ f$.

Definition 4.1.9. Let $p_1:[a_1,b_1]\to S$ and $p_2:[a_2,b_2]\to S$ be two paths such that $p_1(b_1)=p_2(a_2)$, i.e., p_2 starts at the end of p_1. Then we can define the *concatenation* p_1p_2 of p_1 and p_2 as the path $q:[a_1,b_1+b_2-a_2]\to S$ given by the formula

$$q(t)=\begin{cases}p_1(t), & \text{for all } t\in[a_1,b_1]\\ p_2(t+a_2-b_1), & \text{for all } t\in[b_1,b_1+b_2-a_2]\end{cases}. \tag{4.1}$$

Note that, by Exercise 4.1.10.(i) below, q is a path in S starting at the start of p_1 and ending at the end of p_2. The curve for q is the union of the curves for p_1 and p_2.

More generally, if $n\ge2$ and $p_i:[a_i,b_i]\to S$, $i=1,\dots,n$, are paths in S satisfying $p_i(b_i)=p_{i+1}(a_{i+1})$, for $i=1,\dots,n-1$, then we can define the *concatenation*, $p_1\dots p_n$, of these paths inductively (by first concatenating p_1 with p_2, then concatenating the result with p_3, and so on).

Exercise 4.1.10. Let $p_i:[a_i,b_i]\to S$, $i=1,2$, be two paths in S such that $p_2(a_2)=p_1(b_1)$.

(i) Show that concatenation of p_1 and p_2, is another path in S (i.e., the function q defined by (4.1) is continuous).

(ii) Prove that if p_1 and p_2 are rectifiable then so is their concatenation q, and $\|q\|=\|p_1\|+\|p_2\|$. [*Hint:* observe that, by (4.1), $q_{[a_1,b_1]}=p_1$ and $q_{[b_1,b_1+b_2-a_2]}=p_2'$, where $p_2'=p_2\circ f$ is a re-parametrization of p_2, given by $f:[b_1,b_1+b_2-a_2]\to[a_2,b_2]$, $f:t\mapsto t+a_2-b_1$. Then apply Exercise 4.1.7 and Lemma 4.1.5.]

Definition 4.1.11. For a path $p : [a, b] \to S$ we define the *inverse* p^{-1} of p as the path $p^{-1} = p \circ f$, where $f : [a, b] \to [a, b]$ is the function given by $f(t) = b + a - t$, for all $t \in [a, b]$. Note that p^{-1} starts at the end of p and ends at the start of p. Thus p^{-1} gives rise to the same curve as p but traverses this curve in the opposite direction.

Remark 4.1.12. If p is a path in S then $(p^{-1})^{-1} = p$. Moreover, if p is rectifiable then so is p^{-1} and $\|p^{-1}\| = \|p\|$ (by Exercise 4.1.7).

4.2. Geodesic metric spaces

As before, throughout this section (S, d) denotes a metric space.

Definition 4.2.1. A *geodesic path* in the metric space S is an isometric embedding of an interval of $\mathbb{R}$ into S. In other words, $p : [a, b] \to S$ is geodesic if $\mathrm{d}(p(s), p(t)) = |s - t|$, for all $s, t \in [a, b]$. A *geodesic curve* is a curve in S that can be parametrized by a geodesic path.

Exercise 4.2.2. Suppose that $p : [a, b] \to S$ is a geodesic path.

(i) Show that p is rectifiable and $\|p\| = \mathrm{d}(\alpha(p), \omega(p)) = b - a$.
(ii) Prove that every subpath of p is geodesic.
(iii) Let $d = \mathrm{d}(\alpha(p), \omega(p)) = b - a$. Check that p can be re-parametrized as the geodesic path $q : [0, d] \to S$, defined by $q(t) = p(t + a)$.
(iii) Show that the inverse of p is also a geodesic path.

Lemma 4.2.3. *A curve C in S is geodesic if and only if there exists a parametrization $p : [a, b] \to S$ of this curve such that p is rectifiable and $\|p\| = \mathrm{d}(\alpha(p), \omega(p))$.*

Proof. If the curve C is geodesic then it is the image of an isometric embedding $p : [a, b] \to S$. In this case $\|p\| = \mathrm{d}(\alpha(p), \omega(p))$ by Exercise 4.2.2.

For the converse implication suppose that C is a curve in S parametrized by a path $p : [a, b] \to S$ satisfying $\|p\| = \mathrm{d}(\alpha(p), \omega(p))$.

Let us first show that

$$\|p_{[s,t]}\| = \mathrm{d}(p(s), p(t)), \quad \text{for any subinterval } [s, t] \text{ of } [a, b]. \tag{4.2}$$

Indeed, consider the subpaths $r = p_{[s,t]}$, $p_1 = p_{[a,s]}$ and $p_2 = p_{[t,b]}$ of p (thus p is the concatenation of p_1, r and p_2). By Exercise 4.1.6.(i), $\|r\| \geq \mathrm{d}(\alpha(r), \omega(r))$. If this inequality is strict, i.e. $\|r\| > \mathrm{d}(\alpha(r), \omega(r))$, then, by combining Lemma 4.1.5 with Exercise 4.1.6.(i), we obtain

$$\|p\| = \|p_1\| + \|r\| + \|p_2\| > \mathrm{d}(\alpha(p_1), \omega(p_1)) + \mathrm{d}(\alpha(r), \omega(r)) + \mathrm{d}(\alpha(p_2), \omega(p_2)) \geq \mathrm{d}(\alpha(p), \omega(p)),$$

where the last inequality used the fact that $\alpha(p) = \alpha(p_1)$, $\omega(p) = \omega(p_2)$ and the triangle inequality. But this contradicts our assumption that $\|p\| = \mathrm{d}(\alpha(p), \omega(p))$. Therefore, we must have $\|r\| = \mathrm{d}(\alpha(r), \omega(r))$, for every subpath r of p, so (4.2) is true.

In view of Lemma 4.1.5 and (4.2), for all $s, t \in [a, b]$, with $s \leq t$ we have

$$\mathrm{d}(p(s), p(t)) = \|p_{[s,t]}\| = \|p_{[a,t]}\| - \|p_{[a,s]}\| = \mathrm{d}(p(a), p(t)) - \mathrm{d}(p(a), p(s)). \tag{4.3}$$

In particular, by setting $t = b$ this inequality implies that

$$\mathrm{d}(p(a), p(s)) = \mathrm{d}(p(a), p(b)) - \mathrm{d}(p(s), p(b)) \leq \mathrm{d}(p(a), p(b)) = \|p\|, \quad \text{for every } s \in [a, b]. \tag{4.4}$$

The latter allows us to define a function $f : [a, b] \to [0, \|p\|]$ by

$$f(s) = \mathrm{d}(p(a), p(s)), \quad \text{for all } s \in [a, b].$$

This function is continuous because p is continuous and the triangle inequality holds in S. Since $f(a) = 0$ and $f(b) = \mathrm{d}(p(a), p(b)) = \|p\|$, the Intermediate Value Theorem tells us that this function is surjective. Thus, for every $u \in [0, \|p\|]$ there exists $s \in [a, b]$ such that $x = p(s) \in C$ satisfies $\mathrm{d}(p(a), x) = u$. Let's show that such a point $x \in C$ is unique. Indeed, if $y \in C$ is another point satisfying $\mathrm{d}(p(a), y) = u$ then $y = p(t)$, for some $t \in [a, b]$ (without loss of generality, assume that $s \leq t$), and, according to (4.3), we have

$$\mathrm{d}(x, y) = \mathrm{d}(p(a), p(t)) - \mathrm{d}(p(a), p(s)) = u - u = 0,$$

whence $x = p(s) = p(t) = y$.

We can therefore define a function $q : [0, \|p\|] \to S$ by letting $q(u)$ be the unique point of C satisfying $\mathrm{d}(p(a), q(u)) = u$. For any $x \in C$ we have $x = p(s)$, for some $s \in [a, b]$, so $u = \mathrm{d}(p(a), x) = \mathrm{d}(p(a), p(s)) \in [0, \|p\|]$, by (4.4), and $q(u) = x$. Whence the image of q is the entire curve C.

Let's check that q is an isometric embedding of the interval $[0, \|p\|]$ into S. If $u, v \in [a, b]$ then $q(u) = p(s)$ and $q(v) = p(t)$, for some $s, t \in [a, b]$. Without loss of generality we can suppose that $s \leq t$. Then by (4.3) we obtain

$$\mathrm{d}(q(u), q(v)) = \mathrm{d}(p(s), p(t)) = \mathrm{d}(p(a), p(t)) - \mathrm{d}(p(a), p(s)) = \mathrm{d}(p(a), q(v)) - \mathrm{d}(p(a), q(u)) = v - u,$$

by definition of q. It follows that q is an isometric embedding, so its image C is a geodesic curve. Thus the proof is complete. □

Lemma 4.2.3 tells us that geodesic curves C essentially arise from paths p that have the *shortest possible length* (i.e., such that $\|p\| = \mathrm{d}(\alpha(p), \omega(p))$). It shows that our definition of geodesic curves in general metric spaces is in line with Definition 3.9.2 of geodesics in graphs that we gave earlier.

Definition 4.2.4. We will say that a geodesic curve C in S *joins* a point $x \in S$ with a point $y \in S$ if there is a geodesic parametrization $p : [a, b] \to S$ of C such that $p(a) = x$ and $p(b) = y$. If $x, y \in S$, a *geodesic segment* $[x, y]$ in S is any geodesic curve C joining x to y.

Convention 4.2.5. Given two points $x, y \in S$, a geodesic segment between them may not be unique, so we use $[x, y]$ to denote *one* such geodesic segment. We will also assume that $[y, x] = [x, y]$ (as subsets of S). Generally, during a course of a proof we will fix a geodesic segment between any given pair of points and will use the same segment throughout that proof, unless specified otherwise.

Lemma 4.2.6. *Let $[x, y]$ be a geodesic segment in S, given as the image of a geodesic path $p : [a, b] \to S$. Then for any point $z \in [x, y]$, the subpaths of p starting at x and ending at z and starting at z and ending at y are geodesic, so their images are geodesic segments $[x, z]$ and $[z, y]$, respectively. Moreover,* $\mathrm{d}(x, y) = \mathrm{d}(x, z) + \mathrm{d}(z, y)$.

Proof. Recall that p is an isometric embedding of the interval $[a, b]$ in S. Since isometric embeddings are injective, there is a unique point $c \in [a, b]$ such that $p(c) = z$. Then the subpaths $p_{[a,c]}$ and $p_{[c,b]}$ are isometric embeddings of the intervals $[a, c]$ and $[c, b]$ in S, so their images are geodesic segments $[x, z]$ and $[z, y]$. We also have

$$\mathrm{d}(x, y) = \mathrm{d}(p(a), p(b)) = b - a = (c - a) + (b - c) = \mathrm{d}(p(a), p(c)) + \mathrm{d}(p(c), p(b)) = \mathrm{d}(x, z) + \mathrm{d}(z, y).$$

□

Exercise 4.2.7. Suppose that $[x, y]$ is a geodesic segment in S.

(i) Show that for every $d \in [0, \mathrm{d}(x, y)]$ there is a unique point $z \in [x, y]$ such that $\mathrm{d}(x, z) = d$.
(ii) Prove that if $u, v \in [x, y]$ are arbitrary points then $\mathrm{d}(u, v) \leq \mathrm{d}(x, y)$.

Definition 4.2.8. The metric space (S, d) is called *geodesic* if any two points in C can be joined by a geodesic segment. This metric space is *uniquely geodesic* is for all $x, y \in S$ there is a unique geodesic segment joining x to y.

Example 4.2.9. (a) The plane $\mathbb{R}^2$, equipped with the Euclidean metric, is a uniquely geodesic space. Geodesic segments between two points are just straight line segments connecting them.

(b) The punctured plane $\mathbb{R}^2 \setminus \{(0, 0)\}$, equipped with the Euclidean metric, is a connected metric space, but it is not geodesic. Indeed, the length of any path joining $(-1, 0)$ to $(1, 0)$ in this space is strictly greater than $\|(1, 0) - (-1, 0)\| = 2$.

(c) Equip the unit circle $\mathbb{S}^1 = \{e^{i\theta} \mid \theta \in [0, 2\pi)\} \subset \mathbb{C}$ with the arc-length metric $\mathrm{d}_{\mathbb{S}^1}$, which sets the distance between two points $z, w \in \mathbb{S}^1$ to be the length of a shortest arc in $\mathbb{S}^1$ joining these points. In other words,

$$\mathrm{d}(e^{i\theta_1}, e^{i\theta_2}) = \begin{cases} |\theta_2 - \theta_1|, & \text{if } |\theta_2 - \theta_1| \leq \pi \\ 2\pi - |\theta_2 - \theta_1|, & \text{if } |\theta_2 - \theta_1| > \pi \end{cases}.$$

Clearly this makes $\mathbb{S}^1$ into a geodesic space. However this space is not uniquely geodesic, because there are two geodesic segments joining any two antipodal points on the circle.

(d) Another example of a non-uniquely geodesic metric space that is useful to keep in mind is $\mathbb{R}^2$, equipped with the L^1-metric (a.k.a., taxicab metric), $\mathrm{d}_1(\cdot,\cdot)$, defined by

$$\mathrm{d}_1((x_1,y_1),(x_2,y_2)) = |x_1-y_1|+|x_2-y_2|, \quad \text{for all } (x_1,y_1),(x_2,y_2)\in\mathbb{R}^2.$$

(e) Let $\Gamma = (V,E,\alpha,\omega,^{-})$ be a non-empty connected graph without loops. Choose any orientation $E = E^+ \sqcup E^-$ on the edges of Γ. We can construct a geodesic metric space S_Γ by proclaiming that each positively oriented edge of Γ is isometric to the interval $[0,1]$ in $\mathbb{R}$. This can be done rigorously by defining S_Γ as the quotient of the space $E^+\times[0,1]$ by a suitable equivalence relation: see Section I.1.9 of the book [M.R. Bridson, A. Haefliger, *Metric spaces of non-positive curvature.* Fundamental Principles of Mathematical Sciences, 319. Springer-Verlag, Berlin, 1999].

The vertex set V of the graph Γ naturally embeds into S_Γ as a discrete subset and we will identify V with its image in S_Γ under this embedding. The metric d on S_Γ essentially is induced by the path metric d_Γ: if $u,v\in V\subseteq S_\Gamma$ then $\mathrm{d}(u,v)=\mathrm{d}_\Gamma(u,v)$ and this is naturally extended on the entire space S_Γ by treating each edge as a geodesic segment of length 1. This way, S_Γ becomes a *length space*, i.e., for any $x,y\in S_\Gamma$ we have

$$\mathrm{d}_{S_\Gamma}(x,y)=\min\{\|p\| \mid p \text{ is path connecting } x \text{ with } y \text{ in } S\}.$$

The space S_Γ, equipped with this metric, is a geodesic metric space and the natural embedding $V\to S_\Gamma$ is an isometry, when V is endowed with the path metric.

Remark 4.2.10. (a) Let C be the circle in $\mathbb{R}^2$ of circumference 1 (i.e., of radius $1/(2\pi)$), equipped with the arc-length metric, as Example 4.2.9.(c). The construction of the space S_Γ from Example 4.2.9.(e) can also be extended to the situation when Γ is a connected graph with loops. In this case, for every loop at a vertex $v\in V$ we glue an isometric copy of the circle C at the point corresponding to v in S_Γ.

(b) The geodesic space S_Γ is independent of the choice of an orientation on Γ (up to isometry).

Definition 4.2.11. Let $\Gamma = (V,E,\alpha,\omega,^{-})$ be a non-empty connected graph. Then the geodesic metric space S_Γ constructed in Example 4.2.9.(e) is called the *geometric realization* of Γ. The vertex set V admits a natural isometric embedding in S_Γ and we will abuse the notation by identifying V with its image under this embedding.

Example 4.2.12. (a) If Γ is a path graph of length $n\in\mathbb{N}$ (with n vertices and $2n-2$ edges joining consecutive vertices), then S_Γ is isometric to the interval $[0,n]$ in $\mathbb{R}$.

(b) If P_n is the n-gon graph with $n\geq 2$ vertices, then its geometric realization S_{P_n} is isometric to a circle in $\mathbb{R}^2$ with circumference n (of radius $n/(2\pi)$), equipped with the arc-length metric.

(c) If Γ is the Cayley graph $\Gamma(\mathbb{Z},\{1\})$, then S_Γ is isometric to the real line $\mathbb{R}$ (with the standard Euclidean metric).

Fact 4.2.13. (i) Any isomorphism ϕ between graphs Γ_1 and Γ_2 (see Definition 3.3.1) naturally gives rise to a bijective isometry $f: S_{\Gamma_1}\to S_{\Gamma_2}$ such that $f(v)=\phi^0(v)$, for all $v\in V\Gamma_1$

(ii) Suppose that G is a group acting on a non-empty graph Γ by automorphisms. Then this action induces an action of G on the geometric realization S_Γ by bijective isometries (which is simply a homomorphism from G to the group $\mathrm{Isom}(S_\Gamma)$, of all bijective isometries of S_Γ). Although this is true in general, it may be easier to see in the case when the action of G on Γ is without edge inversions, because in this case we can choose an orientation on Γ preserved by the action of G (see Exercise 3.3.16).

Example 4.2.14. (a) For example, the natural action of $\mathbb{Z}$ on its Cayley graph $\Gamma(\mathbb{Z},\{1\})$ gives rise to the standard action of $\mathbb{Z}$ on $\mathbb{R}$ by translations.

(b) For any $n\geq 2$, the Cayley graph $\Gamma(C_n,\{x\})$, of the finite cyclic group $C_n=\langle x\rangle$, is isomorphic to the n-gon graph P_n. Therefore, the natural action of C_n on P_n induces an action of C_n on a circle of circumference n by rotations. More precisely, x will rotate the circle by the angle $2\pi/n$.

4.3. Geodesic triangles

In this section we discuss some properties of geodesic triangles that will be later used to define hyperbolic spaces. As before, we fix a geodesic metric space (S, d).

Definition 4.3.1. Let $x, y, z \in S$ be any points. A *geodesic triangle* $\Delta = xyz$ in S, with vertices x, y, z, is the union of three geodesic segments $[x, y] \cup [y, z] \cup [z, x]$.

The *special points* of Δ are the three points $c_z \in [x, y]$, $c_x \in [y, z]$ and $c_y \in [z, x]$ such that

$$\mathrm{d}(x, c_z) = \mathrm{d}(x, c_y), \ \mathrm{d}(y, c_z) = \mathrm{d}(y, c_x) \ \text{ and } \mathrm{d}(z, c_x) = \mathrm{d}(z, c_y). \tag{4.5}$$

Example 4.3.2. In the case when S is the Euclidean plane, the special points of a geodesic triangle are precisely the points where the inscribed circle touches the sides of this triangle, see Figure 4.1.

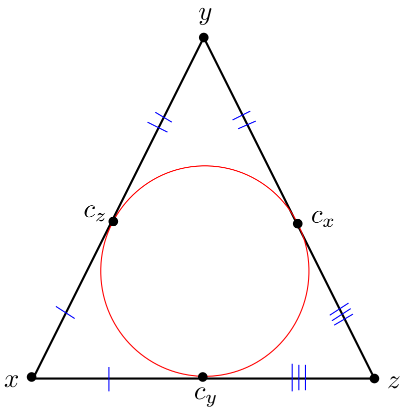


FIGURE 4.1. Special points of a geodesic triangle in $\mathbb{R}^2$.

Exercise 4.3.3. Prove that special points exist and are unique in any geodesic triangle with vertices $x, y, z \in S$. [*Hint:* use Lemma 4.2.6 to re-write (4.5) as a system of 3 linear equations with 3 unknowns λ, μ, ν, where $\lambda = \mathrm{d}(x, c_z)$, $\mu = \mathrm{d}(y, c_z)$ and $\nu = \mathrm{d}(z, c_x)$.]

Definition 4.3.4. Given 3 points $x, y, z \in S$ the *Gromov product* of x and y with respect to z is the real number $(x, y)_z$ defined by

$$(x, y)_z = \frac{1}{2}(\mathrm{d}(x, z) + \mathrm{d}(y, z) - \mathrm{d}(x, y)).$$

Exercise 4.3.5. Show that for any points $x, y, z \in S$ we have

$$0 \leq (x, y)_z = (y, x)_z \leq \min\{\mathrm{d}(x, z), \mathrm{d}(y, z)\}.$$

Remark 4.3.6. In the process of solving Exercise 4.3.3 one actually finds the following connection between the special points of a triangle with vertices $x, y, z \in S$ and the Gromov products of its vertices:

$$\mathrm{d}(x, c_z) = \mathrm{d}(x, c_y) = (y, z)_x, \ \mathrm{d}(y, c_z) = \mathrm{d}(y, c_x) = (x, z)_y \text{ and } \mathrm{d}(z, c_x) = \mathrm{d}(z, c_y) = (x, y)_z. \tag{4.6}$$

Definition 4.3.7. Let Δ be a geodesic triangle in S and let $u \in \Delta$ be a point on a side of this triangle. Let $x, y, z \in S$ denote the vertices of Δ so that $u \in [x, y]$ and $\mathrm{d}(x, u) \leq \mathrm{d}(x, c_z) = (y, z)_x$. We define the *twin points* of u in Δ as follows. If $u = c_z$ they u has two twin points $c_x \in [y, z]$ and $c_y \in [x, z]$, otherwise the twin of u is the unique point $v \in [x, z]$ such that $\mathrm{d}(x, u) = \mathrm{d}(x, v)$ (see Figure 4.2). Note that such point v exists and is unique by Exercise 4.2.7.

Thus a point u of a geodesic triangle Δ has two twins if and only if it is a special point of Δ; otherwise, u has a unique twin point in Δ.

Let us now recall a few basic notions from metric geometry.

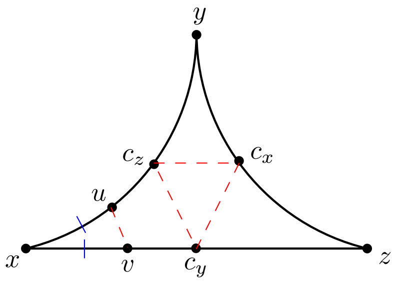


FIGURE 4.2. Twin point $v \in [x, z]$ of a point $u \in [x, z]$ in a geodesic triangle $\Delta = xyz$.

Definition 4.3.8. Let $x \in S$ be a point, let $A \subseteq S$ be a subset and let $\delta \in [0, \infty)$ be a real number. We define the (closed) *ball* of radius δ centered at x as

$$\mathcal{B}_\delta(x) = \{y \in S \mid \mathrm{d}(x, y) \leq \delta\}.$$

The δ-*neighborhood* of A in S is the subset

$$\mathcal{O}_\delta(A) = \bigcup_{x \in A} \mathcal{B}_\delta(x) \subseteq S.$$

The *diameter* of a A, $\mathrm{diam}(A)$, is 0 if $A = \emptyset$, otherwise it is given by

$$\mathrm{diam}(A) = \sup\{d(x, y) \mid x, y \in A\} \in [0, \infty].$$

We shall now give define three properties of geodesic triangles in S that measure their "slimness".

Definition 4.3.9. A geodesic triangle Δ in S is said to δ-*thin*, for some real number $\delta \geq 0$, if for any point $u \in \Delta$ the distance from u to any of its twin points is at most δ.

Example 4.3.10. If Δ is a geodesic triangle in the Euclidean plane $\mathbb{R}^2$ then it is $2r$-thin, where r is the radius of the circle inscribed in Δ.

Definition 4.3.11. A geodesic triangle Δ in S is said to δ-*slim*, for some real number $\delta \geq 0$, if every side of Δ is contained in the δ-neighborhood of the union of the two other sides. In other words, if $\Delta = xyz$ then it is δ-slim provided

$$[x, y] \subseteq \mathcal{O}_\delta([y, z] \cup [z, x]),\ \ [y, z] \subseteq \mathcal{O}_\delta([x, y] \cup [z, x]) \text{ and } [z, x] \subseteq \mathcal{O}_\delta([x, y] \cup [y, z])$$

(see Figure 4.3).

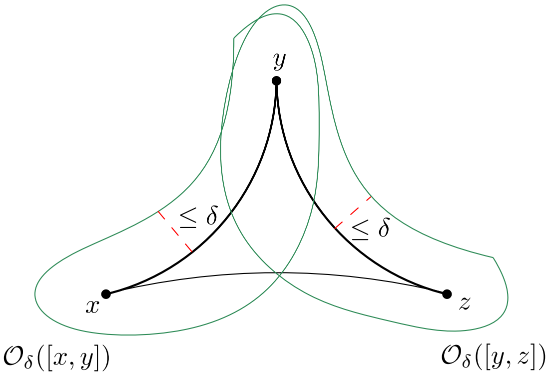


FIGURE 4.3. In a δ-slim triangle xyz the side $[z, x]$ is contained in $\mathcal{O}_\delta([x, y] \cup [y, z]) = \mathcal{O}_\delta([x, y]) \cup \mathcal{O}_\delta([y, z])$.

Definition 4.3.12. The *insize* of a geodesic triangle $\Delta = xyz$ in S is defined as

$$\mathrm{insize}(\Delta) = \mathrm{diam}(\{c_x, c_y, c_z\}) = \max\{\mathrm{d}(c_x, c_y), \mathrm{d}(c_y, c_z), \mathrm{d}(c_z, c_x)\} \in [0, \infty),$$

where c_x, c_y, c_z are the special points of Δ.

4.4. Hyperbolic metric spaces

In this section we will define Gromov hyperbolicity for metric spaces using uniform bounds on the "slimness" properties of geodesic triangles, introduced in Definitions 4.3.9, 4.3.11 and 4.3.12.

Definition 4.4.1. Let (S, d) be a geodesic metric space and let $\delta \geq 0$ be a real number. We will say that (S, d) satisfies the condition

- $\mathrm{Thin}(\delta)$ if all geodesic triangles in S are δ-thin;
- $\mathrm{Slim}(\delta)$ if all geodesic triangles in S are δ-slim;
- $\mathrm{Insize}(\delta)$ if for every geodesic triangle Δ in S we have $\mathrm{insize}(\Delta) \leq \delta$.

THEOREM 4.4.2. *Let δ be a non-negative number. For a geodesic metric space (S, d) the following implications hold.*

(i) $\mathrm{Thin}(\delta)$ *implies* $\mathrm{Slim}(\delta)$*;*
(ii) $\mathrm{Slim}(\delta)$ *implies* $\mathrm{Insize}(4\delta)$*;*
(iii) $\mathrm{Slim}(\delta)$ *implies* $\mathrm{Thin}(6\delta)$*.*

PROOF. Implication (i) is obvious, because a δ-thin triangle is δ-slim, by Definitions 4.3.9 and 4.3.11.

To prove (ii), suppose that every geodesic triangle in (S, d) is δ-slim. Consider any geodesic triangle Δ with vertices $x, y, z \in S$ and special points $c_x \in [y, z]$, $c_y \in [z, x]$ and $c_z \in [x, y]$ (see Definition 4.3.1). By δ-slimness, there exists a point $t \in [x, y] \cup [y, z]$ such that $\mathrm{d}(c_y, t) \leq \delta$. Without loss of generality, assume that $t \in [x, y]$. Then, using the triangle inequality, we get

$$|\mathrm{d}(x, c_y) - \mathrm{d}(x, t)| \leq \mathrm{d}(c_y, t) \leq \delta.$$

But, $\mathrm{d}(x, c_y) = \mathrm{d}(x, c_z)$ and $|\mathrm{d}(x, c_z) - \mathrm{d}(x, t)| = \mathrm{d}(t, c_z)$ by Lemma 4.2.6. Therefore, we can conclude that $\mathrm{d}(t, c_z) \leq \delta$, whence

$$\mathrm{d}(c_y, c_z) \leq \mathrm{d}(c_y, t) + \mathrm{d}(t, c_z) \leq 2\delta.$$

Arguing similarly, we get $\mathrm{d}(c_x, \{c_y, c_z\}) \leq 2\delta$, whence, by the triangle inequality,

$$\mathrm{insize}(\Delta) = \mathrm{diam}(\{c_x, c_y, c_z\}) \leq 4\delta.$$

(iii) Again, assume that the geodesic metric space (S, d) satisfies the condition $\mathrm{Slim}(\delta)$. Consider any geodesic triangle $\Delta = xyz$ in S with special points $c_x \in [y, z]$, $c_y \in [z, x]$ and $c_z \in [x, y]$. Let u be a point on Δ. Without loss of generality, we can assume that u belongs to the sub-segment $[x, c_z]$ of the side $[x, y]$ of Δ. Let $v \in [x, c_y]$ be the twin of u on the side $[x, z]$ of Δ. By condition $\mathrm{Slim}(\delta)$, the geodesic triangle with vertices x, c_z, c_y is δ-slim, so $u \in \mathcal{O}_\delta([x, c_y]) \cup \mathcal{O}_\delta([c_y, c_z])$ and $v \in \mathcal{O}_\delta([x, c_z]) \cup \mathcal{O}_\delta([c_y, c_z])$. If $u \in \mathcal{O}_\delta([x, c_y])$ then we could use the same argument as in the proof of (ii) to deduce that $\mathrm{d}(u, v) \leq 2\delta$, thus we can assume that $u \in \mathcal{O}_\delta([c_y, c_z])$, and, similarly, $v \in \mathcal{O}_\delta([c_y, c_z])$. Therefore, there exist points $a, b \in [c_y, c_z]$ such that $\mathrm{d}(u, a) \leq \delta$ and $\mathrm{d}(v, b) \leq \delta$ (see Figure 4.4).

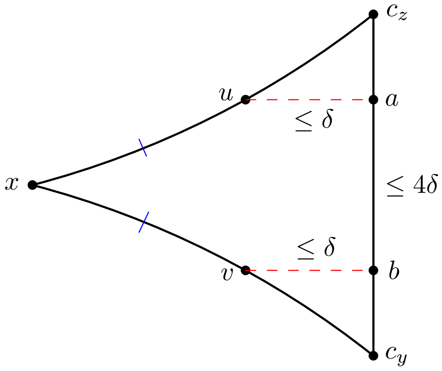


FIGURE 4.4. Illustration of the proof of implication (iii) in Theorem 4.4.2.

By Exercise 4.2.7, $\mathrm{d}(a,b) \le \mathrm{d}(c_y,c_z)$ and $\mathrm{d}(c_y,c_z) \le \mathrm{insize}(\Delta) \le 4\delta$, by implication (ii). Combining this with the triangle inequality we obtain

$$\mathrm{d}(u,v) \le \mathrm{d}(u,a) + \mathrm{d}(a,b) + \mathrm{d}(b,v) \le \delta + 4\delta + \delta = 6\delta.$$

This shows that the triangle Δ is δ-thin, which establishes (iii). □

Exercise 4.4.3. Prove that for a geodesic metric space (S,d) and any $\delta \ge 0$, the condition $\mathrm{Insize}(\delta)$ implies $\mathrm{Thin}(\delta)$ and $\mathrm{Slim}(\delta)$.

[*Hint:* let $\Delta = xyz$ be a geodesic triangle in S and $u \in [x,c_z]$ be a point. Let $p : [a,b] \to S$ be a parametrization of the geodesic segment $[x,y]$. Show that the Gromov product $(p(t),z)_x$ varies continuously with t (i.e., the function $f : [a,b] \to \mathbb{R}$ defined by $f(t) = (p(t),z)_x$ is continuous). Then use the Intermediate Value Theorem to conclude that there is $t \in [a,b]$ such that $\mathrm{d}(x,u) = (p(t),z)_x$, hence u is a special point of any geodesic triangle in S with vertices $x, p(t), z$, by Remark 4.3.6. Thus, $\mathrm{Insize}(\delta)$ implies that u is δ-close to its twin point on the side $[x,z]$.]

Exercise 4.4.4. Suppose that a geodesic metric space (S,d) satisfies $\mathrm{Slim}(\delta)$, for some $\delta \ge 0$. Prove that every geodesic quadrilateral in S is 2δ-slim, i.e., every side of the quadrilateral is contained in the 2δ-neighborhood of the union of the three other sides.

Definition 4.4.5. A geodesic metric space (S,d) is said to be *(Gromov) hyperbolic* if there exists $\delta \ge 0$ such that this space satisfies one of the conditions $\mathrm{Thin}(\delta)$ or $\mathrm{Slim}(\delta)$.

Remark 4.4.6. By Theorem 4.4.2, if a geodesic metric space (S,d) is hyperbolic then this space satisfies both $\mathrm{Slim}(\delta)$ and $\mathrm{Thin}(\delta)$, for some (sufficiently large) $\delta \in \mathbb{N}$.

Example 4.4.7. (a) Suppose that for a geodesic metric space (S,d) there exists $D \ge 0$ such that $\mathrm{diam}(S) \le D$ (i.e., S has bounded diameter). Then this space clearly satisfies $\mathrm{Slim}(D)$, so it is hyperbolic with $\delta = D$.

(b) If Γ is a tree then its geometric realization S_Γ (see Definition 4.2.11) is a geodesic metric space satisfying $\mathrm{Slim}(0)$ and $\mathrm{Thin}(0)$, in particular, it is hyperbolic. Indeed, this is true because any geodesic triangle in Γ (and, hence, in S_Γ) is a tripod, by Proposition 3.9.6.

(c) The hyperbolic plane $\mathbb{H}^2$, equipped with the standard hyperbolic metric $\mathrm{d}_{\mathbb{H}^2}$, is a Gromov hyperbolic metric space. One way to see this is to use the fact that every geodesic triangle Δ in $\mathbb{H}^2$ has an inscribed circle C and the points where this circle touches the sides of the triangle are precisely the special points of Δ, just as in the case of Euclidean triangles (cf. Example 4.3.2), see Figure 4.5.

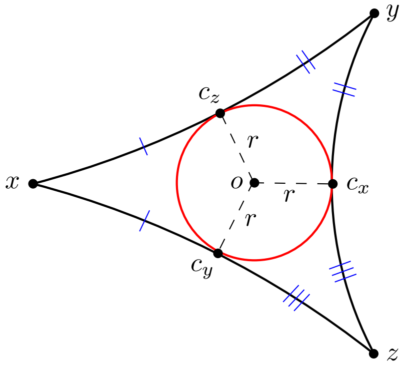


FIGURE 4.5. Inscribed circle in a geodesic triangle in $\mathbb{H}^2$.

The Gauss-Bonnet formula for geodesic triangles in $\mathbb{H}^2$ implies that $\mathrm{area}_{\mathbb{H}^2}(\Delta) \le \pi$, where $\mathrm{area}_{\mathbb{H}^2}$ denotes the hyperbolic area in $\mathbb{H}^2$. Thus, if D is the disk in $\mathbb{H}^2$, bounded by the inscribed circle C, then $\mathrm{area}_{\mathbb{H}^2}(D) \le \mathrm{area}_{\mathbb{H}^2}(\Delta) \le \pi$. On the other hand, from Hyperbolic Geometry we know that

$$\mathrm{area}_{\mathbb{H}^2}(D) = 4\pi \sinh^2(r/2),$$

where $r \in [0, \infty)$ is the (hyperbolic) radius of the disk D. Thus we have $4\pi \sinh^2(r/2) \leq \pi$, which implies that $\sinh(r/2) \leq 1/2$, so $e^{r/2} - e^{-r/2} \leq 1$. It is easy to see that the latter yields $r \leq 2 \ln \varphi$, where $\varphi = \dfrac{1+\sqrt{5}}{2}$ is the golden ratio.

Therefore, the hyperbolic plane $(\mathbb{H}^2, \mathrm{d}_{\mathbb{H}^2})$ satisfies the condition $\mathrm{Insize}(4 \ln \varphi)$, and, hence, $\mathrm{Slim}(4 \ln \varphi)$, by Exercise 4.4.3.

Example 4.4.8. (a) The most basic example of a non-hyperbolic geodesic metric space is the Euclidean plane $\mathbb{R}^2$. Indeed, for every $t > 0$ consider the geodesic triangle Δ_t in $\mathbb{R}^2$ with vertices $(-t, 0)$, $(t, 0)$ and $(0, t)$ (see Figure 4.6). Then the distance from $(0, 0)$, the midpoint of the side $[(-t, 0), (t, 0)]$, to the two other sides is exactly $t/\sqrt{2}$. So, for every $\delta \geq 0$, we can take $t > \sqrt{2}\delta$ to ensure that the triangle Δ_t is not δ-slim. Thus $\mathbb{R}^2$ does not satisfy $\mathrm{Slim}(\delta)$, for any $\delta \geq 0$, so it is not Gromov hyperbolic.

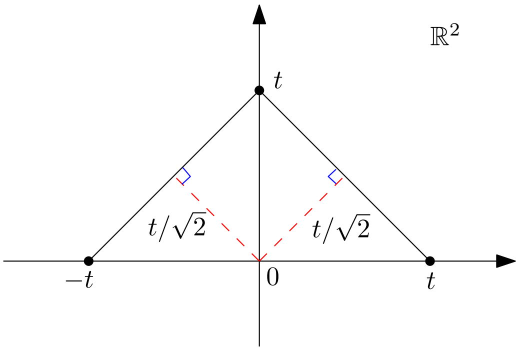


FIGURE 4.6. Illustration of Example 4.4.8.(a).

(b) Let Γ be the Cayley graph of $\mathbb{Z}^2$ with respect to the standard generating set $\{(1, 0), (0, 1)\}$ and let (S_Γ, d) be its geometric realization. Given two points $(k, l), (m, n) \in \mathbb{Z}^2$, we know that

$$\mathrm{d}((k, l), (m, n)) = \mathrm{d}_\Gamma((k, l), (m.n)) = |k - l| + |m - n|.$$

In other words, this distance is induced by the L^1-metric d_1 on $\mathbb{R}^2$ (see Example 4.2.9.(d)).

For every $n \in \mathbb{N}$ we can consider the geodesic triangle Δ_n in S_Γ with vertices $(0, 0)$, $(n, 0)$ and (n, n), where the sides $[(0, 0), (n, 0)]$ and $[(n, 0), (n, n)]$ are just straight line segments connecting the corresponding vertices, and the side $[(0, 0), (n, n)]$ is the union of the straight line segments from $(0, 0)$ to $(0, n)$ and from $(0, n)$ to (n, n) (this side is indeed a geodesic segment in S_Γ by Lemma 4.2.3, because its length, $2n$, equals the distance $\mathrm{d}((0, 0), (n, n)))$, see Figure 4.6.

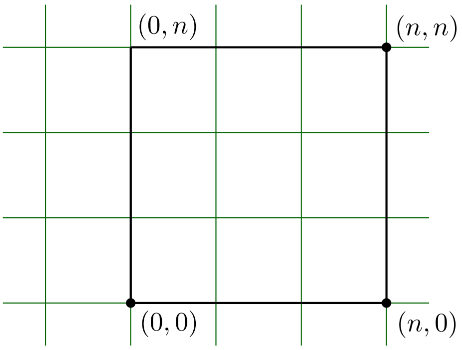


FIGURE 4.7. Geodesic triangle Δ_n in the geometric realization of the Cayley graph $\Gamma(\mathbb{Z}^2, \{(1, 0), (0, 1)\})$.

Evidently the distance from the midpoint $(0, n)$ of the third side of the triangle Δ_n to the union of the two other sides is exactly n, so it tends to infinity as $n \to \infty$. Therefore the geodesic metric space S_Γ cannot satisfy $\mathrm{Slim}(\delta)$, for any $\delta \geq 0$. Hence, it is not hyperbolic. The same example shows that $\mathbb{R}^2$, endowed with the L^1-metric, is not Gromov hyperbolic.

Exercise 4.4.9. It is possible to equip $\mathbb{R}^2$ with a metric so that it becomes a hyperbolic geodesic metric space. For example, show that this is the case for the SNCF metric (a.k.a., *French railway metric*) $\mathrm{d}(\cdot,\cdot)$ defined by

$$\mathrm{d}(\mathbf{x},\mathbf{y}) = \begin{cases} \|\mathbf{x}-\mathbf{y}\|, & \text{if the points } (0,0),\ \mathbf{x} \text{ and } \mathbf{y} \text{ are colinear} \\ \|\mathbf{x}\|+\|\mathbf{y}\|, & \text{otherwise} \end{cases},$$

for all $\mathbf{x},\mathbf{y} \in \mathbb{R}^2$, where $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^2$.

4.5. Hyperbolic groups

In this section we introduce hyperbolic groups whose geometry generalizes the geometry of free groups.

Definition 4.5.1. A group G is said to be *(Gromov) hyperbolic* if there is a finite generating set $X \subseteq G$ such that the geometric realization of the Cayley graph $\Gamma(G,X)$ is a hyperbolic metric space in the sense of Definition 4.4.5.

Example 4.5.2. (a) Any finite group G is hyperbolic, because its Cayley graph (with respect to any generating subset) has bounded diameter, see Example 4.4.7.(a).

(b) A free group F on a finite set X is hyperbolic because the Cayley graph of F with respect to X is a tree: see Theorem 3.5.1 and Example 4.4.7.(b). In particular, the free group F_k, of rank $k \in \mathbb{N}_0$, is hyperbolic (including $F_0 = \{1\}$ and $F_1 \cong \mathbb{Z}$).

(c) Let us show that the infinite dihedral group D_∞, defined in Example 2.5.5, is hyperbolic. To this end, we choose the generating subset $\{\sigma,\rho\}$ for this group, where $\sigma : x \mapsto -x$ and $\rho : x \mapsto x-1$, for all $x \in \mathbb{R}$. The resulting Cayley graph $\Gamma = \Gamma(D_\infty,\{\sigma,\rho\})$ is depicted on Figure 4.8, where the horizontal (blue) edges are labelled by ρ and the vertical (curved red) edges are labelled by σ).

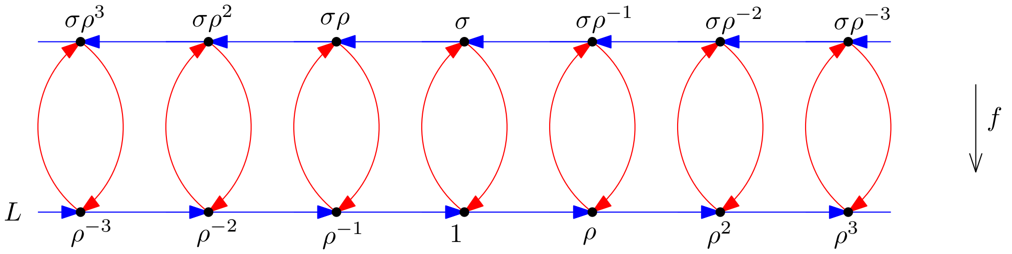


FIGURE 4.8. Cayley graph $\Gamma(D_\infty,\{\sigma,\rho\})$.

Let L be the subgraph of Γ which is the simplicial line on the vertices $\{\rho^n \mid n \in \mathbb{Z}\}$ (in fact, L is isomorphic to the Cayley graph of the infinite cyclic subgroup $\langle\rho\rangle \leqslant D_\infty$ with respect to the generating set $\{\rho\}$). Let S_Γ and S_L denote the geometric realizations of Γ and L, respectively.

Evidently there is a projection $f : S_\Gamma \to S_L$ which collapses every edge labelled by σ to its endpoint on L and sends each point x, of a ρ-edge, to itself if $x \in S_L$ or to the point $y \in S_L$ "directly below it". The latter means that if $x \in [\sigma\rho^{n-1},\sigma\rho^n]$, for some $n \in \mathbb{Z}$, is at distance $\alpha \in [0,1]$ from $\sigma\rho^n$ then $f(x) \in [\rho^{-n},\rho^{-n+1}]$ is the point at distance α from ρ^{-n} (in particular, $f(\sigma\rho^n) = \rho^{-n}$, for all $n \in \mathbb{Z}$).

Let us make two observations about the projection f:

(i) for any points $x,y \in S_\Gamma$, $f([x,y]) = [f(x),f(y)]$ in S_L;
(ii) for any $u \in S_\Gamma$, we have $\mathrm{d}_{S_\Gamma}(u,f(u)) \le 2$.

By (i), if Δ is a geodesic triangle in S_Γ with vertices x,y,z, then $f(\Delta)$ is a geodesic triangle in S_L with vertices $f(x),f(y),f(z)$. Consider any point $u \in \Delta$. Without loss of generality, we can assume that $u \in [x,y]$. Then $f(u) \in [f(x),f(y)]$ in S_L by observation (i). Since L is a tree, S_L satisfies Slim(0) (see Example 4.4.7.(b)), whence $f(u) \in [f(y),f(z)] \cup [f(z),f(x)]$ in S_L. In view of (i), the latter implies that there is a point $v \in [y,z] \cup [z,x]$ in S_Γ such that $f(u) = f(v)$. Therefore, we can combine the triangle inequality with observation (ii) to conclude that

$$\mathrm{d}_{S_\Gamma}(u,v) \le \mathrm{d}_{S_\Gamma}(u,f(u)) + \mathrm{d}_{S_\Gamma}(f(u),f(v)) + \mathrm{d}_{S_\Gamma}(f(v),v) \le 4.$$

It follows that $u \in \mathcal{O}_4([y,z] \cup [z,x])$ in S_Γ, and we have shown that S_Γ satisfies the condition Slim(4). Thus S_Γ is a hyperbolic metric space and D_∞ is a Gromov hyperbolic group.

4.6. Quasi-isometries

To find more examples of hyperbolic groups we will start investigating quasi-isometries between metric spaces, which are maps that do not distort distances "too much". An example of such a map is the projection f seen in Example 4.5.2.(c).

Definition 4.6.1. Suppose that (S, d_S) and (T, d_T) are two metric spaces. A map $f : S \to T$ is said to be a (λ, c)*-quasi-isometric embedding* (or simply a *quasi-isometric embedding*), for some constants $\lambda \geq 1$ and $c \geq 0$, if

$$\frac{1}{\lambda}\mathrm{d}_S(x,y) - c \leq \mathrm{d}_T(f(x), f(y)) \leq \lambda \mathrm{d}_S(x,y) + c, \quad \text{for all } x, y \in S. \tag{4.7}$$

The metric spaces (S, d_S) and (T, d_T) are *quasi-isometric*, which will be denoted $S \sim_{qi} T$, if there is a quasi-isometric embedding $f : S \to T$ such that the image of S is *qiasi-dense* in T. The latter means that there exists a constant $D \in [0, \infty)$ such that for every $t \in T$ there is $s \in S$ satisfying $\mathrm{d}_T(t, f(s)) \leq D$. In this case the map f will be called a (λ, c, D)*-quasi-isometry* (or simply a *quasi-isometry*) between S and T.

Example 4.6.2. (a) One can easily check that the projection $f : S_\Gamma \to S_L$, defined in Example 4.5.2.(c), is a quasi-isometry, with $\lambda = 1$, $c = 4$ and $D = 0$.

(b) Consider the integers $\mathbb{Z}$, endowed with the Euclidean metric from $\mathbb{R}$. Then the inclusion $i : \mathbb{Z} \to \mathbb{R}$ is a quasi-isometry because it preserves distances (so, $\lambda = 1$ and $c = 0$) and its image is quasi-dense with $D = 1/2$.

(c) As observed in Example 4.4.8.(b), the standard metric on geometric realization S_Γ of the Cayley graph $\Gamma = \Gamma(\mathbb{Z}^2, \{(1,0), (0,1)\})$ is induced by the L^1-metric d_1 on $\mathbb{R}^2$. Therefore the natural embedding $f : S_\Gamma \to \mathbb{R}^2$ preserves distances. The image $f(S_\Gamma)$ is quasi-dense in $\mathbb{R}^2$ with $D = 1/2$, whence $(S_\Gamma, \mathrm{d}_{S_\Gamma}) \sim_{qi} (\mathbb{R}^2, \mathrm{d}_1)$.

(d) Let $\mathrm{d}_1(\cdot,\cdot)$ be the L^1-metric on $\mathbb{R}^2$ (see Example 4.2.9.(d)) and let $\mathrm{d}(\cdot,\cdot)$ be the Euclidean metric in $\mathbb{R}^2$. Let us show that the identity map $\mathrm{Id}_{\mathbb{R}^2} : \mathbb{R}^2 \to \mathbb{R}^2$ induces a quasi-isometry between $(\mathbb{R}^2, \mathrm{d}_1)$ and $(\mathbb{R}^2, \mathrm{d})$. Since $\mathrm{Id}_{\mathbb{R}^2}$ is surjective, its image is quasi-dense with $D = 0$. Recall that

$$\mathrm{d}_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1 \quad \text{and} \quad \mathrm{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|, \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^2,$$

where $\|(x_1, x_2)\|_1 = |x_1| + |x_2|$ is the L^1-norm and $\|(x_1, x_2)\| = \sqrt{x_1^2 + x_2^2}$ is the Euclidean norm on $\mathbb{R}^2$. Note that for all $a, b \in \mathbb{R}$ the following inequalities hold:

$$\frac{1}{\sqrt{2}}(|a| + |b|) \leq \sqrt{a^2 + b^2} \leq |a| + |b|.$$

It follows that

$$\frac{1}{\sqrt{2}}\mathrm{d}_1(\mathbf{x}, \mathbf{y}) \leq \mathrm{d}(\mathbf{x}, \mathbf{y}) \leq \mathrm{d}_1(\mathbf{x}, \mathbf{y}), \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^2. \tag{4.8}$$

Therefore, $\mathrm{Id}_{\mathbb{R}^2}$ is a quasi-isometry between $(\mathbb{R}^2, \mathrm{d}_1)$ and $(\mathbb{R}^2, \mathrm{d})$ with $\lambda = \sqrt{2}$ and $c = D = 0$.

Example 4.6.3. The function $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^3$, is not a quasi-isometry, when both copies of $\mathbb{R}$ are equipped with the Euclidean metric $\mathrm{d}(\cdot,\cdot)$, because there exist no $\lambda \geq 1$ and $c \geq 0$ such that

$$\mathrm{d}(g(0), g(x)) = |x^3| \leq \lambda|x| + c = \lambda \mathrm{d}(0, x) + c, \quad \text{for all } x \in \mathbb{R}.$$

Exercise 4.6.4. Let $\mathbb{R}$ be equipped with the Euclidean metric. Which of the following functions are quasi-isometries?

(i) $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = ax + b$, where $a, b \in \mathbb{R}$ and $a \neq 0$.
(ii) $g : \mathbb{R} \to [0, \infty)$, defined by $g(x) = |x|$.
(iii) $h : \mathbb{R} \to \mathbb{R}$, defined by $h(x) = 3x + \sin x$.
(iv) $j : \mathbb{R} \to \mathbb{R}$, defined by $j(x) = x \cos x$.
(v) $k : \mathbb{R} \to \mathbb{R}$, defined by $k(x) = \lfloor x \rfloor$, where $\lfloor x \rfloor \in \mathbb{Z}$ is the largest integer such that $\lfloor x \rfloor \leq x$.

Exercise 4.6.5. Suppose that $f : S \to T$ is a quasi-isometry between metric spaces (S, d_S) and (T, d_T). Show that f has a *quasi-inverse* $g : T \to S$. That is, there exists a quasi-isometry $g : T \to S$ and a constant $E \geq 0$ such that

$$\mathrm{d}_S\big((g \circ f)(s), s\big) \leq E, \quad \text{for all } s \in S, \quad \text{and} \quad \mathrm{d}_T\big((f \circ g)(t), t\big) \leq E, \quad \text{for all } t \in T$$

(i.e., the compositions $g \circ f$ and $f \circ g$ are, in a certain sense, at bounded distances from the identity maps Id_S and Id_T, respectively).

[*Hint:* since the image of f is quasi-dense in T, for every $t \in T$ we can choose some $s = s_t \in S$ with $\mathrm{d}_T(t, s) \leq D$ and then define $g : T \to S$ by $g(t) = s_t$. In general, this uses the axiom of choice.]

Lemma 4.6.6. *Quasi-isometries define an equivalence relation between metric spaces. In other words, for all metric spaces (S, d_S), (T, d_T) and (U, d_U) the following hold:*

(i) $S \sim_{qi} S$ *(*reflexive*);*
(ii) $S \sim_{qi} T$ *implies* $T \sim_{qi} S$ *(*symmetric*);*
(iii) *if* $S \sim_{qi} T$ *and* $T \sim_{qi} U$ *then* $S \sim_{qi} U$ *(*transitive*).*

Proof. The identity map $\mathrm{Id}_S : S \to S$ is obviously a quasi-isometry of S with itself, so (i) holds. Claim (ii) is an immediate consequence of Exercise 4.6.5, thus it remains to prove (iii). Suppose that we have a (λ_1, c_1, D_1)-quasi-isometry $f_1 : S \to T$ and a (λ_2, c_2, D_2)-quasi-isometry $f_2 : T \to U$. Then $f = f_2 \circ f_1 : S \to U$ is a function satisfying

$$\frac{1}{\lambda_2}\mathrm{d}_T(f_1(x), f_1(y)) - c_2 \leq \mathrm{d}_U(f(x), f(y)) \leq \lambda_2 \mathrm{d}_T(f_1(x), f_1(y)) + c_2, \quad \text{for all } x, y \in S.$$

We also know that

$$\frac{1}{\lambda_1}\mathrm{d}_S(x, y) - c_1 \leq \mathrm{d}_T(f_1(x), f_1(y)) \leq \lambda_1 \mathrm{d}_S(x, y) + c_1, \quad \text{for all } x, y \in S.$$

Therefore the function f satisfies (4.7) with $\lambda = \lambda_1\lambda_2$ and $c = \lambda_2 c_1 + c_2$. To see that the image of f is quasi-dense in U, take any $u \in U$ and let $t \in T$ be a point such that $\mathrm{d}_U(u, f_2(t)) \leq D_2$. There exists $s \in S$ satisfying $\mathrm{d}_T(t, f_1(s)) \leq D_1$. Combining the triangle inequality in U with the fact that f_2 is a (λ_2, c_2, D_2)-quasi-isometry, we achieve

$$\mathrm{d}_U(u, f(s)) \leq \mathrm{d}_U(u, f_2(t)) + \mathrm{d}_U(f_2(t), f(s)) \leq D_2 + \lambda_2 \mathrm{d}_T(t, f_1(s)) + c_2 \leq D_2 + \lambda_2 D_1 + c_2.$$

Thus the image of f is quasi-dense in U with the constant $D = D_2 + \lambda_2 D_1 + c_2 < \infty$. Therefore, $f : S \to U$ is a quasi-isometry, so $S \sim_{qi} U$, as required. □

Exercise 4.6.7. Suppose that (S, d_S), (T, d_T) and (U, d_U) are metric spaces. Given quasi-isometric embeddings $f : S \to T$ and $g : T \to U$, prove that the composition $g \circ f : S \to U$ is also a quasi-isometric embedding.

4.7. Word metric on groups

In this section we introduce word metrics on groups. This will allow us to treat groups as metric spaces, which is a fundamental idea of Geometric Group Theory.

Definition 4.7.1. Let G be a group and let X be a generating subset of G. For any $g \in G$ we define the *word length* of g with respect to X, denoted $|g|_X$, as the length of a shortest word w over $X^{\pm 1}$ representing g in G. The *word metric* on G, corresponding to the generating set X, is defined as

$$\mathrm{d}_X(g, h) = |g^{-1}h|_X, \quad \text{for all } g, h \in G.$$

The next exercise shows that, in some sense, the word length function on a group behaves similarly to a norm on a vector space.

Exercise 4.7.2. Suppose that X is a generating set of a group G. Prove that the word length with respect to X satisfies the following properties:

(i) $|1_G|_X = 0$;
(ii) $|g^{-1}|_X = |g|_X$, for all $g \in G$;

(ii) $|gh|_X \leq |g|_X + |h|_X$.

Lemma 4.7.3. *If G is a group with a generating subset $X \subseteq G$ then the word metric $\mathrm{d}_X(\cdot,\cdot)$ is a metric on G. In fact, it is induced from the path metric on the Cayley graph $\Gamma(G,X)$ under the natural identification of G with $V\Gamma$.*

Proof. The first statement that d_X is a metric on G directly follows from the second statement that it is induced by the path metric on $\Gamma = \Gamma(G,X)$, that is $\mathrm{d}_X(g,h) = \mathrm{d}_\Gamma(g,h)$, for all $g,h \in G$. The latter can be justified as follows. If $g,h \in G$ then $d_\Gamma(g,h)$ is defined as the length of a shortest path p in Γ joining g with h. The label $\mathrm{Lab}(p)$ is a word w over $X^{\pm 1}$ representing the element $g^{-1}h$ in G, by Remark 3.2.4, and $\|p\| = \|w\|$. Therefore,

$$\mathrm{d}_\Gamma(g,h) = \min\{\|w\| \mid w \text{ is a word over } X^{\pm 1} \text{ representing } g^{-1}h \text{ in } G\} = |g^{-1}h|_X = \mathrm{d}_X(g,h),$$

for all $g,h \in G$. □

Example 4.7.4. (a) If G is any group and $X = G$ then the Cayley graph $\Gamma(G,X)$ is just a complete graph on its vertices (i.e., any two vertices are adjacent), so

$$\mathrm{d}_X(g,h) = \begin{cases} 0, & \text{if } g = h \\ 1, & \text{if } g \neq h \end{cases}$$

is the discrete metric on G.

(b) If $G = \mathbb{Z}^n$, for some $n \in \mathbb{N}$ and X is the standard basis

$$X = \{(1,0,\dots,0),(0,1,\dots,0),\dots,(0,\dots,0,1)\} \subseteq \mathbb{Z}^n,$$

then for any element $(a_1,\dots,a_n) \in \mathbb{Z}^n$ we have

$$|(a_1,\dots,a_n)|_X = |a_1| + \dots + |a_n| = \|(a_1,\dots,a_n)\|_1,$$

i.e., the word length with respect to X on $\mathbb{Z}^n$ is simply the L^1-norm on $\mathbb{Z}^n$. Therefore the word metric on $\mathbb{Z}^n$ corresponding to X will be the L^1-metric:

$$\mathrm{d}_X\big((a_1,\dots,a_n),(b_1,\dots,b_n)\big) = \sum_{i=1}^n |b_i - a_i|, \quad \text{for all } (a_1,\dots,a_n),(b_1,\dots,b_n) \in \mathbb{Z}^n.$$

(c) If X is a non-empty set and $F = F(X)$ is the free group on X then $|f|_X$ simply measures the length of the (unique) reduced word w over $X^{\pm 1}$ representing f in F. Thus $\mathrm{d}_X(g,h)$ is the length of the reduced word representing $g^{-1}h$ in G.

Remark 4.7.5. The standard action of a group G on itself by left multiplication preserves the word metric corresponding to any generating subset $X \subseteq G$. Indeed, if $f,g,h \in G$ then

$$\mathrm{d}_X(fg,fh) = |(fg)^{-1}(fh)|_X = |g^{-1}h|_X = \mathrm{d}_X(g,h).$$

The following important fact is an immediate consequence of Lemma 4.7.3 and the definitions.

Proposition 4.7.6. *Let G be a group with a generating subset $X \subseteq G$, and let S_Γ be the geometric realization of the Cayley graph $\Gamma = \Gamma(G,X)$. The natural inclusion $\iota : G \to S_\Gamma$, sending each $g \in G$ to the corresponding vertex of $\Gamma(G,X)$, is a $(1,0,1/2)$-quasi-isometry between (G,d_X) and $(S_\Gamma,\mathrm{d}_{S_\Gamma})$. Moreover, this quasi-isometry is G-equivariant:*

$$\iota(fg) = f.\iota(g), \quad \textit{for all } f,g \in G.$$

Exercise 4.7.7. Let X,Y be two finite generating subsets of a group G. Show that there exists $\lambda \geq 1$ such that

$$\frac{1}{\lambda}|g|_X \leq |g|_Y \leq \lambda|g|_X, \quad \text{for all } g \in G.$$

[*Hint:* take $\lambda = \max\{\alpha,\beta,1\}$, where $\alpha = \max\{|y|_X \mid y \in Y\}$ and $\beta = \max\{|x|_Y \mid x \in X\}$.]

Proposition 4.7.8. *Suppose that X,Y are two finite generating subsets of a group G. Then the identity map $\mathrm{Id}_G : G \to G$ defines a quasi-isometry between (G,d_X) and (G,d_Y).*

Proof. This is an immediate consequence of Definition 4.7.1 and Exercise 4.7.7. □

Thus replacing one finite generating set of a group by another one does not change the geometry of the group (and of its Cayley graph) too much.

Corollary 4.7.9. *Let S_1 and S_2 be geometric realizations of Cayley graphs $\Gamma(G,X)$ and $\Gamma(G,Y)$ of a group G with respect to two finite generating subsets X and Y, respectively. Then $S_1 \sim_{qi} S_2$.*

PROOF. By Proposition 4.7.6, we have $(G,\mathrm{d}_X) \sim_{qi} (S_1,\mathrm{d}_{S_1})$ and $(G,\mathrm{d}_Y) \sim_{qi} (S_2,\mathrm{d}_{S_2})$. Since $(G,\mathrm{d}_X) \sim_{qi} (G,\mathrm{d}_Y)$, by Proposition 4.7.8, we can apply Lemma 4.6.6 to conclude that $(S_1,\mathrm{d}_{S_1}) \sim_{qi} (S_2,\mathrm{d}_{S_2})$. ☐

4.8. The Schwarz-Milnor lemma

Recall that an action of a group G on a metric space (S,d) *by isometries* is simply a homomorphism from G to the group $\mathrm{Isom}(S)$, of all bijective self-isometries of S. In the current section we will discuss a method to obtain quasi-isometries between a group (equipped with a word metric) and a space on which the group acts by isometries. This result was originally proved by Albert Schwarz and was later independently re-discovered by John Milnor. Before stating it we need to give auxiliary definitions.

Definition 4.8.1. Suppose that a group G acts on a metric space (S,d) by isometries. This action is called *proper* if there exists a basepoint $s \in S$ such that for all $r \geq 0$ we have

$$|\{g \in G \mid \mathrm{d}(s,g.s) \leq r)| < \infty.$$

The action of G on S is said to be *co-bounded* if there exist $s \in S$ and a constant $R \geq 0$ such that every point of S is at distance at most R from the orbit $G.s$, i.e., $S = \mathcal{O}_R(G.s)$ (in other words, this says that the orbit $G.s$ is quasi-dense in S).

Recall that a *(closed) ball* in a metric space (S,d) is a subset B of the form

$$B = \mathcal{B}_r(x) = \{y \in S \mid \mathrm{d}(x,y) \leq r\},$$

where x is the *center* of the ball and r is its *radius* (see Definition 4.3.8).

Exercise 4.8.2. Let G be a group acting on a metric space (S,d) by isometries.

(a) Prove that this action is proper if and only if for every ball B in S we have

$$|\{g \in G \mid g.B \cap B \neq \emptyset\}| < \infty.$$

(b) Show that the action is co-bounded if and only if there is a ball B in S such that S is covered by the G-translates of B, i.e., $S = \bigcup_{g\in G} g.B$.

In particular, properness and co-boundedness of the action are independent of the choice of a basepoint $s \in S$.

THEOREM 4.8.3 (Schwarz-Milnor lemma). *Let G be a group acting by isometries on a non-empty geodesic metric space (S,d). If this action is proper and co-bounded then G is generated by a finite subset $X \subseteq G$ and $(G,\mathrm{d}_X) \sim_{qi} (S,\mathrm{d})$. More precisely, for any basepoint $s \in S$ the orbit map $f : G \to S$, $g \overset{f}{\mapsto} g.s$, defines a quasi-isometry between (G,d_X) and (S,d).*

PROOF. Choose any basepoint $s \in S$. Then the action of G on S is proper and co-bounded with respect to s, by Exercise 4.8.2. Therefore, there exist $R \geq 0$ such that $G.s$ is R-quasi-dense in S. Define the subset X of G by

$$X = \{g \in G \mid \mathrm{d}(s,g.s) \leq 2R+1\}.$$

Note that $|X| < \infty$ because the action of G on S is proper.

Claim 4.8.4. *The subset X generates G. Moreover, for every $g \in G$ we have*

$$|g|_X \leq \mathrm{d}(s,g.s) + 1. \tag{4.9}$$

Proof of Claim 4.8.4. Consider any element $g \in G$ and choose a geodesic segment $[s, g.s]$ in S. Let $p : [0, d] \to S$ be a geodesic path whose image is this segment, where $d = \mathrm{d}(s, g.s) = \|p\|$ (cf. Exercise 4.2.2), and set $n = \lfloor d \rfloor \in \mathbb{N}_0$. By definition, p is an isometric embedding of the interval $[0, d]$ into S, so if we denote $s_i = p(i) \in S$, for $i = 0, \dots, n$, and $s_{n+1} = p(d) = g.s \in S$, then we will have

$$\mathrm{d}(s_{i-1}, s_i) = i - (i-1) = 1, \quad \text{for } i = 1, \dots, n, \quad \text{and } \mathrm{d}(s_n, s_{n+1}) = d - n = d - \lfloor d \rfloor \le 1.$$

Using the co-boundedness of the action of G on S, we can find elements $g_0 = 1$, $g_1, \dots, g_n$, $g_{n+1} = g$ in G such that $\mathrm{d}(s_i, g_i.s) \le R$, for $i = 0, \dots, n+1$, see Figure 4.9.

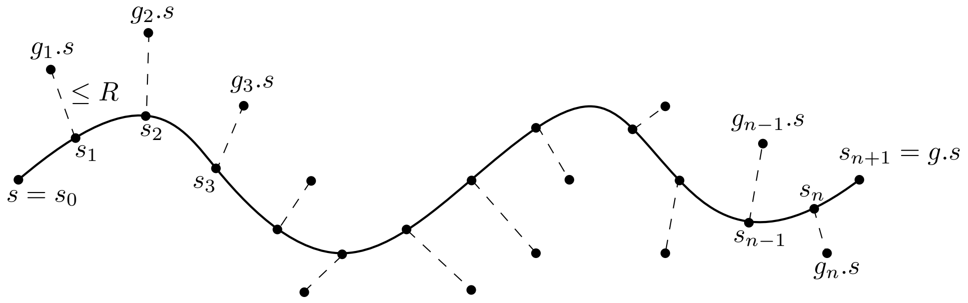


Figure 4.9. Illustration of the proof of Claim 4.8.4.

Observe that $\mathrm{d}(s, g_1.s) \le R$ and, since G acts on S by isometries, we have

$$\mathrm{d}(s, g_i^{-1} g_{i+1}.s) = \mathrm{d}(g_i.s, g_{i+1}.s) \le \mathrm{d}(g_i.s, s_i) + \mathrm{d}(s_i, s_{i+1}) + \mathrm{d}(s_{i+1}, g_{i+1}.s) \le 2R + 1,$$

for $i = 1, \dots, n$. From the definition of X, it follows that $g_1, g_1^{-1}g_2, \dots, g_n^{-1}g_{n+1} \in X$. Whence,

$$g = g_{n+1} = g_1(g_1^{-1}g_2)(g_2^{-1}g_3)\cdots(g_n^{-1}g_{n+1}) \in \langle X \rangle, \tag{4.10}$$

thus X is a generating set of G. Inequality (4.9) is an immediate consequence of (4.10) and the definition of $n = \lfloor \mathrm{d}(s, g, s) \rfloor$. □

Claim 4.8.5. *For every $g \in G$ we have*

$$\mathrm{d}(s, g.s) \le (2R + 1)|g|_X. \tag{4.11}$$

Proof of Claim 4.8.5. Exercise. [*Hint:* use induction on $|g|_X$ together with the observation that $\mathrm{d}(s, x^\varepsilon.s) = \mathrm{d}(x^{-\varepsilon}.s, s) \le 2R + 1$, for every $x^\varepsilon \in X^{\pm 1}$.] □

Let us now show that the map $f : G \to S$, given by $f(g) = g.s$, for all $g \in G$, is a quasi-isometry between (G, d_X) and (S, d). Consider arbitrary $h_1, h_2 \in G$ and set $g = h_1^{-1}h_2 \in G$, so that $\mathrm{d}_X(h_1, h_2) = |g|_X$. Inequality (4.9) implies that

$$\mathrm{d}(f(h_1), f(h_2)) = \mathrm{d}(h_1.s, h_2.s) = \mathrm{d}(s, (h_1^{-1}h_2).s) = \mathrm{d}(s, g.s) \ge |g|_X - 1 = \mathrm{d}_X(h_1, h_2) - 1.$$

On the other hand, (4.11) gives

$$\mathrm{d}(f(h_1), f(h_2)) = \mathrm{d}(h_1.s, h_2.s) = \mathrm{d}(s, g.s) \le (2R + 1)|g|_X = (2R + 1)\mathrm{d}_X(h_1, h_2).$$

These two inequalities show that f is a $(2R + 1, 1, R)$-quasi-isometry between (G, d_X) and (S, d), completing the proof. □

Corollary 4.8.6. *Suppose that a group G acts properly and co-boundedly by isometries on a non-empty metric space (S, d). Then G is finitely generated and for every finite generating set Y of G we have $(G, \mathrm{d}_Y) \sim_{qi} (S, \mathrm{d})$ and $(S_\Gamma, \mathrm{d}_{S_\Gamma}) \sim_{qi} (S, \mathrm{d})$, where $(S_\Gamma, \mathrm{d}_{S_\Gamma})$ is the geometric realization of the Cayley graph $\Gamma = \Gamma(G, Y)$.*

Proof. By Theorem 4.8.3, there is a finite generating set X of G such that $(G, \mathrm{d}_X) \sim_{qi} (S, \mathrm{d})$. If Y is any other finite generating subset of G then $(G, \mathrm{d}_Y) \sim_{qi} (G, \mathrm{d}_X)$, by Proposition 4.7.8, and $(G, \mathrm{d}_Y) \sim_{qi} (S_\Gamma, \mathrm{d}_{S_\Gamma})$, where $(S_\Gamma, \mathrm{d}_{S_\Gamma})$ is the geometric realization of $\Gamma = \Gamma(G, Y)$, by Proposition 4.7.6. The statement of the corollary now follows from Lemma 4.6.6. □

Example 4.8.7. Given any $n \in \mathbb{N}$, we can consider the natural action of $\mathbb{Z}^n$ on the Euclidean n-space $(\mathbb{R}^n, \mathrm{d})$ by translations:

$$\mathbf{z}.\mathbf{x} = \mathbf{x} + \mathbf{z}, \quad \text{for all } \mathbf{z} = (z_1, \dots, z_n) \in \mathbb{Z}^n \text{ and } \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Evidently, this action is by isometries. Note that for every $\mathbf{x} \in \mathbb{R}^n$, we have $\mathrm{d}(\mathbf{x}, \lfloor \mathbf{x} \rfloor) \leq \sqrt{n}$, where $\lfloor (x_1, \dots, x_n) \rfloor = (\lfloor x_1 \rfloor, \dots, \lfloor x_n \rfloor) \in \mathbb{Z}^n$. Hence the orbit $\mathbb{Z}^n.\mathbf{0}$ is $\sqrt{n}$-quasi-dense in $\mathbb{R}^n$, where $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^n$. Thus the action of $\mathbb{Z}^n$ on $\mathbb{R}^n$ is co-bounded.

Clearly, for each $r \geq 0$ there are only finitely many $\mathbf{z} \in \mathbb{Z}^n$ satisfying

$$\mathrm{d}(\mathbf{0}, \mathbf{z}.\mathbf{0}) = \|\mathbf{z}\| \leq r,$$

where $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^n$. Thus the action is proper.

So, if Y is any finite generating set of $\mathbb{Z}^n$ then Corollary 4.8.6 tells us that $(\mathbb{Z}^n, \mathrm{d}_Y)$ is quasi-isometric to $(\mathbb{R}^n, \mathrm{d})$.

Definition 4.8.8. Let G and H be two finitely generated groups. We will say that these two groups are *quasi-isometric*, writing $G \sim_{qi} H$, if $(G, \mathrm{d}_X) \sim_{qi} (H, \mathrm{d}_Y)$, for some finite generating subsets X of G and Y of H.

By Proposition 4.7.8, quasi-isometry between two finitely generated groups is independent of the choices of finite generating sets for them. If H is a finite index subgroup of a finitely generated group, then we know that H is finitely generated, as a consequence of Schreier index formula (see Corollary 3.8.12). We will now use Schwarz-Milnor lemma to give a different proof of this fact and establish a quasi-isometry between G and H.

Proposition 4.8.9. *Let G be a finitely generated group and let $H \leqslant G$ be a finite index subgroup. Then H is finitely generated and $H \sim_{qi} G$.*

Proof. Suppose that X is a finite generating subset of G. Let $\Gamma = \Gamma(G, X)$ be the Cayley graph of G with respect to X and let $(S_\Gamma, \mathrm{d}_{S_\Gamma})$ be its geometric realization. As we know, S_Γ is a geodesic metric space and the left action of G on itself by left multiplication extends to an action of G on $(S_\Gamma, \mathrm{d}_{S_\Gamma})$ by isometries (se Example 4.2.9.(d) and Fact 4.2.13). This induces an action of H on S_Γ by isometries, which is defined by $h.g = hg$ for all $h \in H$ and all $g \in G = V\Gamma \subseteq S_\Gamma$. Let us check that this action is proper and co-bounded.

Since $|G : H| < \infty$, we have $G = \bigsqcup_{i=1}^{k} Hg_i$. Let $R = \max\{|g_i|_X \mid i = 1, \dots, k\} < \infty$ and observe that for any point $x \in S_\Gamma$ there is a vertex $g \in V\Gamma$ such that $\mathrm{d}_{S_\Gamma}(x, g) \leq 1/2$ (by definition of a geometric realization), and $g = hg_i$, for some $h \in H$ and $i \in \{1, \dots, n\}$. It follows that

$$\mathrm{d}_{S_\Gamma}(x, h) \leq \mathrm{d}_{S_\Gamma}(x, g) + \mathrm{d}_{S_\Gamma}(g, h) \leq \frac{1}{2} + \mathrm{d}_{S_\Gamma}(hg_i, h) = \frac{1}{2} + \mathrm{d}_X(h, hg_i) = \frac{1}{2} + |g_i|_X \leq \frac{1}{2} + R.$$

Thus the H-orbit of 1_G is $(R + 1/2)$-quasi-dense in S_Γ, and so the action of H on S_Γ is co-bounded.

Since the set X is finite, the Cayley graph $\Gamma(G, X)$ is a $2|X|$-regular graph, so, for every $r \geq 0$, it contains only finitely many vertices $g \in G$ satisfying $\mathrm{d}_\Gamma(1_G, g) \leq r$. Therefore, there are finitely many elements $h \in H$ such that $\mathrm{d}_{S_\Gamma}(1_G, h.1_G) \leq r$, i.e., the action of H on S_Γ is proper.

Thus, all assumptions of Theorem 4.8.3 are satisfied, so H is generated by a finite subset Y and $(H, \mathrm{d}_Y) \sim_{qi} (S_\Gamma, \mathrm{d}_{S_\Gamma})$. Applying Proposition 4.7.6 and Lemma 4.6.6, we can conclude that $G \sim_{qi} H$. □

Recall that a group is said to be *virtually cyclic* if it has a cyclic subgroup of finite index. Evidently, any finite group is virtually cyclic. The infinite dihedral group D_∞ contains the cyclic subgroup of translations of $\mathbb{R}$ by integers as an index 2 subgroup, so D_∞ is virtually cyclic. Proposition 4.8.9 together with Exercise 0.5.10 immediately give the following fact.

Corollary 4.8.10. *Any infinite virtually cyclic group is quasi-isometric to $\mathbb{Z}$. In particular, all infinite virtually cyclic groups are quasi-isometric to each other.*

Exercise 4.8.11. Prove that for every $n \geq 2$ the free group of rank n is quasi-isometric to the free group of rank 2. Conclude that any two non-abelian finitely generated free groups are quasi-isometric to each other.

4.9. Invariance of hyperbolicity under quasi-isometries

In this section we will show that hyperbolicity of a geodesic matric space is preserved by quasi-isometries. The standard way of doing this is via quasi-geodesics that we will now define.

Definition 4.9.1. Suppose that $\lambda \geq 1$ and $c \geq 0$ are real numbers and (S, d) is a metric space. A (λ, c)*-quasi-geodesic* (or simply a *quasi-geodesic*) in S is a (λ, c)-quasi-isometric embedding $f : [a, b] \to S$, where $[a, b]$ is a non-empty interval of $\mathbb{R}$, equipped with the standard Euclidean metric.

A (λ, c)*-quasi-geodesic path* (or simply a *quasi-geodesic path*) in S is a (continuous) path that is a (λ, c)-quasi-geodesic.

Note that quasi-isometric embeddings are not necessarily continuous, but the next exercise tells us that quasi-geodesics are uniformly close to (continuous) quasi-geodesic paths.

Exercise 4.9.2. For all real numbers $\lambda \geq 1$ and $c \geq 0$ there exist real numbers $\lambda' \geq 1$, $c' \geq 0$ and $\varepsilon \geq 0$ such that the following holds. Suppose that (S, d) is a geodesic metric space and $f : [a, b] \to S$ is a (λ, c)-quasi-geodesic. Then there exists a (λ', c')-quasi-geodesic path $f^t : [a, b] \to S$ such that the images of f and f^t are contained within ε-neighborhoods of each other in S.

[*Hint:* Consider the subdivision $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$ of the interval $[a, b]$, where $\{t_1, \ldots, t_{n-1}\} = (a, b) \cap \mathbb{Z}$. Define f^t as the concatenation of geodesic paths p_i, $i = 1, \ldots, n$, where p_i is starts at $f(t_{i-1})$ and ends at $f(t_i)$. In other words, f^t is a parametrization of a broken line in S whose vertices are on the image of f.]

The following result is very important in the theory of Gromov hyperbolic spaces: in hyperbolic metric spaces images of quasi-geodesics are uniformly close to images of geodesics with the same endpoints.

Theorem 4.9.3 (Stability of quasi-geodesics, a.k.a. Morse lemma). *Given constants $\delta \geq 0$, $\lambda \geq 1$ and $c \geq 0$ there exists a constant $\nu = \nu(\delta, \lambda, c) \geq 0$ such that the following holds. Let (S, d) be a hyperbolic metric space satisfying* $\mathrm{Slim}(\delta)$ *and let $f : [a, b] \to S$ be a (λ, c)-quasi-geodesic in S with image $C \subseteq S$. Then for any geodesic segment $E = [f(a), f(b)] \subseteq S$ we have*

$$C \subseteq \mathcal{O}_\nu(E) \quad \text{and} \quad E \subseteq \mathcal{O}_\nu(C) \quad \text{in } S.$$

Proof. We do not include a proof in these notes, due to its length and technicality. However, you can find it, for example, in Theorem 1.7 of Chapter III.H in the book [M.R. Bridson, A. Haefliger, *Metric spaces of non-positive curvature.* Fundamental Principles of Mathematical Sciences, 319. Springer-Verlag, Berlin, 1999]). ☐

Theorem 4.9.4 (Invariance of hyperbolicity under quasi-isometries). *Let (S, d_S) and (T, d_T) be geodesic metric spaces and let $f : S \to T$ be a quasi-isometric embedding of S in T. If T is Gromov hyperbolic then so is S. In particular, hyperbolicity is a quasi-isometry invariant for geodesic matric spaces.*

Proof. Assume that the metric space T satisfies the condition $\mathrm{Slim}(\delta)$, for some $\delta \geq 0$. Suppose that $f : S \to T$ is a (λ, c)-quasi-isometric embedding, in the sense of Definition 4.6.1, for some $\lambda \geq 1$ and $c \geq 0$. Let $\nu = \nu(\delta, \lambda, c) \geq 0$ be the constant from Theorem 4.9.3.

Consider a geodesic triangle $\Delta = xyz$ in S and any point u on the side $[x, y]$ of this triangle. Evidently, $C_1 = f([x, y])$, $C_2 = f([y, z])$ and $C_3 = f([z, x])$ are images of (λ, c)-quasi-geodesics in T, hence, by Theorem 4.9.3, there is $u' \in [f(x), f(y)]$ in T such that $\mathrm{d}_T(f(u), u') \leq \nu$. The geodesic triangle in T with vertices $f(x)$, $f(y)$ and $f(z)$ is δ-slim, so

$$u' \in \mathcal{O}_\delta([f(y), f(z)]) \cup \mathcal{O}_\delta([f(z), f(x)]) \subseteq \mathcal{O}_{\delta+\nu}(C_2) \cup \mathcal{O}_{\delta+\nu}(C_1),$$

where the second inclusion follows from Theorem 4.9.3. Therefore, there exists a point $v' \in C_1 \cup C_2$ such that $\mathrm{d}_T(u', v') \leq \nu + \delta$ (see Figure 4.10), hence

$$\mathrm{d}_T(f(u), v') \leq \mathrm{d}_T(f(u), u') + \mathrm{d}_T(u', v') \leq 2\nu + \delta. \tag{4.12}$$

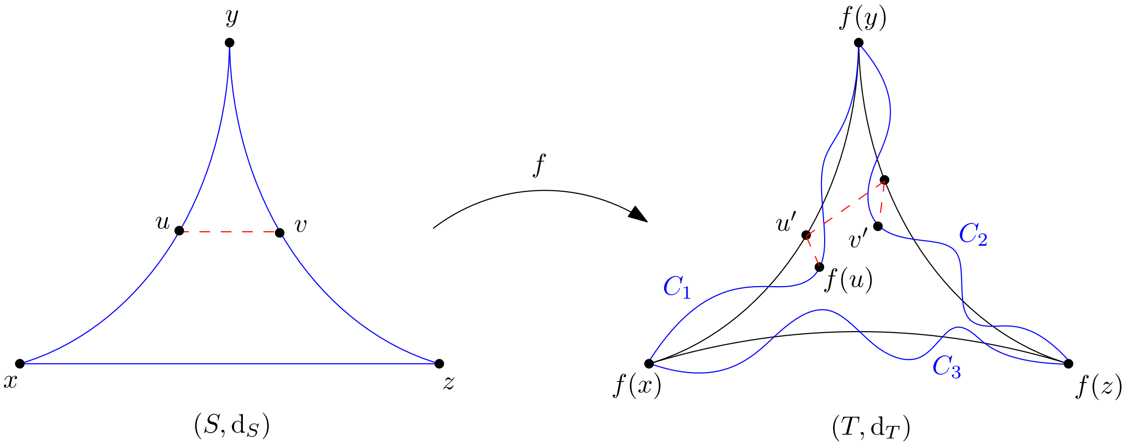


FIGURE 4.10. Illustration of the proof of Theorem 4.9.4.

Since $v' \in C_1 \cup C_2$, there is $v \in [y, z] \cup [z, x]$ in S such that $f(v) = v'$, and since f satisfies (4.7), we can use (4.12) to obtain

$$\mathrm{d}_S(u, v) \leq \lambda\big(\mathrm{d}_T(f(u), f(v)) + c\big) \leq \lambda(2\nu + \delta + c).$$

Thus, we have shown that $[x, y] \subseteq \mathcal{O}_\gamma([y, z] \cup [x, z])$ in S, where $\gamma = \lambda(2\nu + \delta + c) \geq 0$. Since this is true for any side of every geodesic triangle in S, we can conclude that S satisfies $\mathrm{Slim}(\gamma)$, so it is Gromov hyperbolic. ☐

Corollary 4.9.5. *Let G be a group, let $X, Y \subseteq G$ be two finite generating subsets of G, and let $\Gamma_1 = \Gamma(G, X)$ and $\Gamma_2 = \Gamma(G, Y)$ be the Cayley graphs of G corresponding to X and Y, respectively. If the geometric realization S_{Γ_1} is Gromov hyperbolic then so is S_{Γ_2}. In particular, hyperbolicity of G is independent of the choice of a finite generating set for this group.*

PROOF. This is an immediate consequence of Theorem 4.9.4 and Corollary 4.7.9. ☐

In fact, we can prove a more general statement.

Corollary 4.9.6. *Suppose that G is a hyperbolic group. If H is another finitely generated group that is quasi-isometric to G (in the sense of Definition 4.8.8), then H is also hyperbolic. In particular, every finite index subgroup/supergroup of G is hyperbolic.*

PROOF. The first statement follows from Theorem 4.9.4, Proposition 4.7.6 and Lemma 4.6.6. The second statement is true by Proposition 4.8.9. ☐

Corollary 4.9.7. *The free abelian group $\mathbb{Z}^2$ is not hyperbolic.*

PROOF. In Example 4.4.8.(b) we saw that the geometric realization of the Cayley graph of $\mathbb{Z}^2$ with respect to the standard generating set $\{(1, 0), (0, 1)\}$ is not hyperbolic. Therefore, $\mathbb{Z}^2$ is not a hyperbolic group by Corollary 4.9.5. ☐

4.10. More examples of hyperbolic groups

In this section we use Schwarz-Milnor lemma (Theorem 4.8.3) to find more examples of hyperbolic groups.

Example 4.10.1. Suppose that G is a finitely generated virtually free group (i.e., there is a finite index free subgroup $F \leqslant G$). Then G is hyperbolic by Corollary 4.9.6, because F is hyperbolic (see Example 4.5.2.(b)). In particular the group $C_2 * C_3$, studied in Example 3.11.5 and, more generally, any free product of finitely many finite groups is hyperbolic (it is virtually free by Proposition 3.11.4).

In fact, we can establish the following stronger fact by analysing geodesic triangles in Cayley graphs of free products.

Exercise 4.10.2. Let A and B be hyperbolic groups. Prove that the free product $G = A * B$ is a hyperbolic group.

[*Hint:* Choose some finite generating subsets $X \subseteq A$ and $Y \subseteq B$, and let S_{Γ_1}, S_{Γ_2} and S_Γ be the geometric realizations of the Cayley graphs $\Gamma_1 = \Gamma(A, X)$, $\Gamma_2 = \Gamma(B, Y)$ and $\Gamma = \Gamma(G, X \cup Y)$, respectively. Note that we have natural (isometric) embeddings $S_{\Gamma_i} \hookrightarrow S_\Gamma$, for $i = 1, 2$. Using the normal form theorem for free products (Theorem 2.5.3), show that any geodesic triangle in S_Γ looks as the triangle in Figure 4.11, where the geodesic triangle in the middle and each geodesic bigon is completely contained in S_{Γ_i}, for some $i = 1, 2$. Therefore, if S_{Γ_i} satisfies $\mathrm{Slim}(\delta_i)$, for some $\delta_i \geq 0$, $i = 1, 2$, then S_Γ will satisfy $\mathrm{Slim}(\delta)$, for $\delta = \max\{\delta_1, \delta_2\}$.]

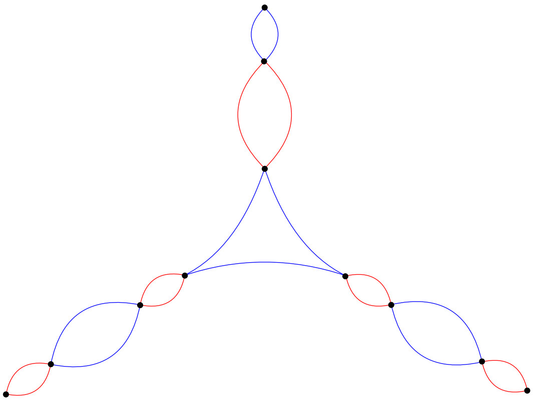


FIGURE 4.11. The structure of a geodesic triangle in S_Γ: the blue triangle and bigons are geodesic in S_{Γ_i}, and the red bigons are geodesic in S_{Γ_j}, where $\{i, j\} = \{1, 2\}$.

Example 4.10.3 (Triangle groups)**.** Given three natural numbers $k, l, m \geq 2$, the triangle group $T(k, l, m)$ is defined by the presentation

$$T(k, l, m) = \langle a, b, c \mid a^2 = b^2 = c^2 = 1,\ (ab)^k = (bc)^l = (ca)^m = 1 \rangle. \tag{4.13}$$

Let $\Sigma = 1/k + 1/l + 1/m \in \mathbb{R}$. We will consider a geodesic triangle $\Delta(k, l, m)$, whose angles are π/k, π/l and π/m (adding up to $\pi\Sigma$), in the geodesic metric space (S, d), which is either the 2-sphere $\mathbb{S}^2$, the Euclidean plane $\mathbb{R}^2$ or the hyperbolic plane $\mathbb{H}^2$. The triangle group $T(k, l, m)$ will then act on S by isometries. It will be generated by the reflections in the three sides of the triangle $\Delta(k, l, m)$, and will preserve the tessellation of S by triangles congruent to $\Delta(k, l, m)$.

The following classification of $T(k, l, m)$, depending on the value of $\Sigma = 1/k + 1/l + 1/m$, can be found, for example, in Section 7.2 of the book [John G. Ratcliffe, *Foundations of hyperbolic manifolds.* Graduate Texts in Mathematics, 149. Springer-Verlag, New York, 1994.]

Fact 4.10.4.
- (i) If $\Sigma > 1$ then the geodesic triangle $\Delta(k, l, m)$, with the angles π/k, π/l and π/m, exists on the 2-sphere $\mathbb{S}^2$ (here $\mathbb{S}^2 = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| = 1\}$ is the unit sphere in $\mathbb{R}^3$ centered at the origin, equipped with the arc length metric $\mathrm{d}_{\mathbb{S}^2}$). In this case presentation (4.13) defines a *finite* group which acts faithfully by isometries on $(\mathbb{S}^2, \mathrm{d}_{\mathbb{S}^2})$. The full list of *spherical triangle groups* (up to isomorphism) consists of one infinite family $T(2, 2, m)$, where $m \in \{2, 3, \dots\}$, and the groups $T(2, 3, 3)$, $T(2, 3, 4)$ and $T(2, 3, 5)$.
- (ii) If $\Sigma = 1$ (so that $\pi/k + \pi/l + \pi/m = \pi$) then the geodesic triangle $\Delta(k, l, m)$ exists in the Euclidean plane $\mathbb{R}^2$. The triangle group $T(k, l, m)$ acts on $\mathbb{R}^2$ faithfully and by isometries. This action is proper and co-bounded (in the sense of Definition 4.8.1), so, by Schwarz-Milnor lemma (Theorem 4.8.3), $T(k, l, m)$, equipped with the word metric

(corresponding to the generating set $\{a, b, c\}$) is quasi-isometric to $\mathbb{R}^2$, equipped with the Euclidean metric (in fact, $T(k, l, m)$ has finite index subgroup isomorphic to $\mathbb{Z}^2$). There are exactly 3 *Euclidean triangle groups*, up to isomorphisms: $T(2, 3, 6)$, $T(2, 4, 4)$ and $T(3, 3, 3)$.

(iii) If $\Sigma < 1$ then the geodesic triangle $\Delta(k, l, m)$ can be constructed in the hyperbolic plane $\mathbb{H}^2$, equipped with the hyperbolic metric $\mathrm{d}_{\mathbb{H}^2}$. In this case $T(k, l, m)$ will act properly and co-boundedly by isometries on the hyperbolic plane $\mathbb{H}^2$, so this group (and the geometric realization of its Cayley graph $\Gamma(T(k, l, m), \{a, b, c\})$) will be quasi-isometric to $(\mathbb{H}^2, \mathrm{d}_{\mathbb{H}^2})$, by Schwarz-Milnor lemma. Recall, from Example 4.4.7.(c), that the hyperbolic plane $(\mathbb{H}^2, \mathrm{d}_{\mathbb{H}^2})$ is Gromov hyperbolic. Therefore, in view of Theorem 4.9.4, every *hyperbolic triangle group* $T(k, l, m)$ is Gromov hyperbolic, in the sense of Definition 4.5.1. One can show that for two different ordered triples of natural numbers (k, l, m) and (k', l', m') the triangle groups $T(k, l, m)$ and $T(k', l', m')$ are not isomorphic, therefore, there are infinitely many hyperbolic triangle groups. All of these groups are infinite and are not virtually cyclic.

Example 4.10.5. Hyperbolic triangle groups discussed in Example 4.10.3 are generalized by Fuchsian groups, which are discrete subgroups of orientation-preserving isometries of $\mathbb{H}^2$. Another generalization are groups acting properly and co-boundedly on the hyperbolic n-space $\mathbb{H}^n$, for $n \geq 2$, such as fundamental groups of closed hyperbolic manifolds. All of these groups are Gromov hyperbolic.

Example 4.10.6. Confirming a conjecture of Gromov, Alexander Ol'shanskii proved that in a certain statistical sense most finitely presented groups are hyperbolic. More precisely, he showed the following. Fix a pair of natural numbers $k \geq 2$ and $l \in \mathbb{N}$, and let $\mathbf{n} = (n_1, \dots, n_l) \in \mathbb{N}^l$ be an l-tuple of natural numbers. Consider the set $\mathcal{P}(k, l, \mathbf{n})$ consisting of group presentations

$$P = \langle x_1, \dots, x_k \mid r_1 = 1, \ \dots, \ r_l = 1 \rangle,$$

where $r_1, \dots, r_l \in F(x_1, \dots, x_k)$, $\|r_i\| = n_i$, for $i = 1, \dots, l$. Then the probability that a randomly selected presentation $P \in \mathcal{P}(k, l, \mathbf{n})$ defines a hyperbolic group tends to 1 exponentially fast, as $n = \min\{n_1, \dots, n_l\} \to \infty$.

Thus, by studying hyperbolic groups we, effectively, study "generic" finitely presented groups.

4.11. Local geodesics in hyperbolic graphs

This section investigates properties of k-local geodesics in hyperbolic graphs, which will give us a tool for showing that every hyperbolic group is finitely presented and has solvable word problem. In this section we will restrict ourselves to working on a connected graph Γ, with the path metric d_Γ defined on its vertex set $V\Gamma$. As usual, S_Γ will denote the geometric realization of Γ, equipped with the natural extension d_{S_Γ} of the path metric.

Definition 4.11.1. Let $k \in \mathbb{N}$ be a natural number. A path p in Γ is said to be *k-local geodesic* if every subpath p', of p, with $\|p'\| \leq k$, is a geodesic (i.e., $\|p'\| = \mathrm{d}(\alpha(p'), \omega(p'))$).

Note that here we are working with *combinatorial* paths and subpaths in Γ, as in Definition 3.1.7. Images of combinatorial geodesics from Γ in S_Γ are geodesic segments, by Lemma 4.2.3. The distance from a vertex $x \in V\Gamma$ to a combinatorial path q in Γ is defined naturally as

$$\mathrm{d}_\Gamma(x, q) = \min\{\mathrm{d}_\Gamma(x, v) \mid v \text{ is a vertex of } q\} \in \mathbb{N}_0.$$

Lemma 4.11.2. *Suppose that the geometric realization* $(S_\Gamma, \mathrm{d}_{S_\Gamma})$ *satisfies the condition* $\mathrm{Slim}(\delta)$, *for some* $\delta \geq 0$, *and* $k > 4\delta$ *is a natural number. If* p *is a* $2k$*-local geodesic in* Γ *and* q *is a geodesic such that* $\alpha(q) = \alpha(p)$ *and* $\omega(q) = \omega(p)$ *then every vertex of* p *is contained within the* 2δ*-neighborhood of a vertex of* q *in* $(\Gamma, \mathrm{d}_\Gamma)$.

Proof. Choose a vertex x on p maximizing the distance to the (set of vertices of) path q, let $a = \alpha(q) = \alpha(p)$ and $b = \omega(q) = \omega(p)$. Suppose, first, that the two subpaths of p from a to x

and from x to b contain at least k edges each. Then there are vertices y, z on p such that x is the midpoint of the subpath p', of p, starting at y and ending at z, and the length of p' is $2k$.

Choose vertices y' and z' that are vertices of q closest to y and z, respectively. The path p' is geodesic, because p is $2k$-local geodesic, therefore it is a side of a geodesic quadrilateral in Γ with vertices y, z, z', y'. Since S_Γ satisfies $\mathrm{Slim}(\delta)$, geodesic quadrilaterals are 2δ-slim, by Exercise 4.4.4. Therefore, there is a vertex $u \in [y, y'] \cup [y', z'] \cup [z', z]$ such that $\mathrm{d}_\Gamma(x, u) \le 2\delta$, where we choose $[y', z']$ as a subpath of the geodesic q.

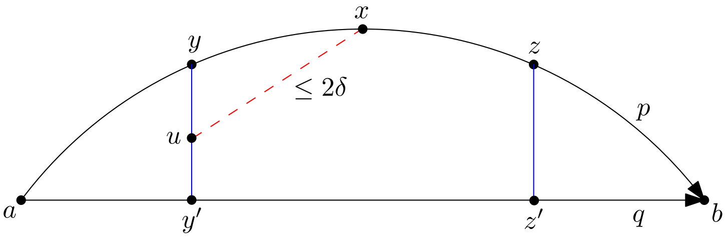


FIGURE 4.12. Illustration of the proof of Lemma 4.11.2.

Assume, first, that $u \in [y, y']$ (see Figure 4.12). Since x is the midpoint of the geodesic path p' and $\|p'\| = 2k$, we have $\mathrm{d}_\Gamma(x, y) = k > 4\delta$, hence

$$\mathrm{d}_\Gamma(y, u) \ge \mathrm{d}_\Gamma(x, y) - \mathrm{d}_\Gamma(x, u) > 4\delta - 2\delta \ge \mathrm{d}_\Gamma(x, u). \tag{4.14}$$

Note that $\mathrm{d}_\Gamma(y, y') = \mathrm{d}_\Gamma(y, u) + \mathrm{d}_\Gamma(u, y')$, because $u \in [y, y']$ and $[y, y']$ is geodesic. Therefore, (4.14) can be combined with the triangle inequality to yield

$$\mathrm{d}_\Gamma(y, y') = \mathrm{d}_\Gamma(y, u) + \mathrm{d}_\Gamma(u, y') > \mathrm{d}_\Gamma(x, u) + \mathrm{d}_\Gamma(u, y') \ge \mathrm{d}_\Gamma(x, y').$$

We have thus shown that

$$\mathrm{d}_\Gamma(x, q) \le \mathrm{d}_\Gamma(x, y') < \mathrm{d}_\Gamma(y, y') = \mathrm{d}_\Gamma(y, q),$$

which contradicts to the definition of x as the vertex of p that is farthest away from q.

Similarly, we will obtain a contradiction if $u \in [z', z]$, hence we must have $u \in [y', z']$, so u belongs to q and $\mathrm{d}_\Gamma(x, q) \le 2\delta$, as required.

In the case when a subpath of p from a to x contains fewer thank k edges but a subpath from x to b has at least k edges, the argument is simpler, because the subpath of starting at a and ending at z would have length at most $2k$, so it would be geodesic, and the triangle with vertices a, z, z' would be δ-slim in S_Γ. The two remaining cases can be treated similarly, so the proof is complete. ☐

4.12. Dehn presentations of hyperbolic groups

In this section we will show that every hyperbolic group G is finitely presented. Moreover, G possesses especially nice "Dehn" presentations, which will allow us to solve the Word Problem in the next section.

Definition 4.12.1. Let r be a word over some alphabet. A *prefix* of r is any initial subword u, i.e., $r = uv$, for another subword v of r. We will say that u is a dominant prefix of r if $\|u\| > \|r\|/2$.

Note that a dominant prefix is always non-empty, by definition.

For example, if $X = \{x, y\}^{\pm 1}$ and $r = x^{-3}y^5xy^{-8}$, then the words x^{-1}, $x^{-3}y^2$ and $x^{-3}y^5x$ are all prefixes of r, but only the last one is a dominant prefix.

Definition 4.12.2. A presentation $\langle X \mid R \rangle$ of a group G is said to be a *Dehn presentation* of G if for every non-empty reduced word w over $X^{\pm 1}$, representing the identity element in G, w contains a subword that is dominant prefix of some defining relator $r \in R$.

THEOREM 4.12.3. *Every Gromov hyperbolic group has a finite Dehn presentation.*

Proof. Let G be a Gromov hyperbolic group and let $X \subseteq G$ be any finite generating set of G. By Corollary 4.9.5, there exists $\delta \geq 0$ such that the geometric realization $(S_\Gamma, \mathrm{d}_{S_\Gamma})$, of the Cayley graph $\Gamma = \Gamma(G, X)$, satisfies the condition $\mathrm{Slim}(\delta)$. Let $F(X)$ be the free group on X and let $\phi : F(X) \to G$ be the epimorphism extending the identity map on X (it exists by the Universal Property of the free group $F(X)$, see Theorem 1.2.17, and it is surjective by Lemma 0.5.11.(ii)). Define the subset R of $F(X)$ as follows:

$$R = \{r \in \ker\phi \mid \|r\| \leq 16\lfloor\delta + 1\rfloor\}. \tag{4.15}$$

Claim 4.12.4. *Suppose that $w \in \ker\phi$ is a non-empty reduced word in $F(X)$ then w contains a subword u which is a dominant prefix of a word $r \in R$.*

Proof of Claim 4.12.4. Set $k = 4\lfloor\delta + 1\rfloor \in \mathbb{N}$, so that $k > 4\delta$. Consider the path p in $\Gamma = \Gamma(G, X)$ starting at 1 and labelled by w. Since $\phi(w) = 1$, we know that p is a closed path, thus the geodesic between its endpoints is the trivial path consisting of the single vertex 1.

Suppose, first, that p is $2k$-local geodesic. Then, by Lemma 4.11.2, every vertex of p is at distance at most 2δ from 1 in $\Gamma(G, X)$. If $\|p\| > 2k$ then p contains an initial subpath q of length $2k$, whence q must be geodesic and

$$\mathrm{d}_\Gamma(1, \omega(q)) = \mathrm{d}_\Gamma(\alpha(q), \omega(q)) = 2k > 2\delta,$$

giving a contradiction. Therefore, we must have $\|p\| \leq 2k$, which means that p is a geodesic path (as it is $2k$-local geodesic). Since $\alpha(p) = \omega(p)$, the latter implies that p is the trivial path, consisting of the single vertex 1, so that $\mathrm{Lab}(p) = w$ is the empty word, contradicting our assumptions.

Thus, we can conclude that p cannot be $2k$-local geodesic, so it contains a non-geodesic subpath q such that $\|q\| \leq 2k$. Without loss of generality, we can assume that q is a shortest such non-geodesic subpath. Then q is non-trivial and satisfies $\mathrm{d}_\Gamma(\alpha(q), \omega(q)) < \|q\|$. Let s be a geodesic path in Γ starting at $\omega(q)$ and ending at $\alpha(q)$ (see Figure 4.13), so that qs is a closed path in Γ of length

$$\|qs\| = \|q\| + \|s\| = \|q\| + \mathrm{d}_\Gamma(\alpha(q), \omega(q)) < 2\|q\| \leq 4k = 16\lfloor\delta + 1\rfloor.$$

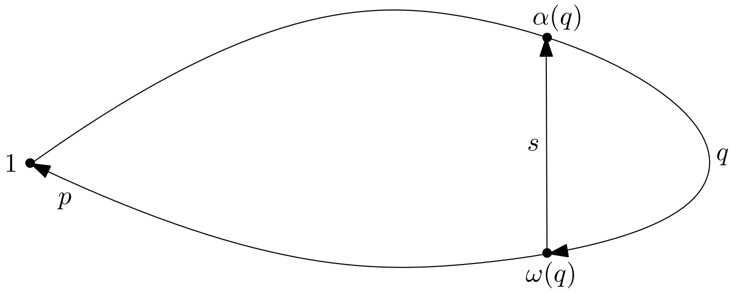


Figure 4.13. The paths p, q and s in $\Gamma(G, X)$.

Note that $r = \mathrm{Lab}(qs)$ represents the trivial element of G, since qs is a closed path in $\Gamma(G, X)$. Moreover, the word r is reduced over $X^{\pm 1}$, because s is geodesic, q is a subword of the reduced word $w \in F(X)$ and any cancellation between $\mathrm{Lab}(q)$ and $\mathrm{Lab}(s)$ would lead to a non-geodesic subpath of p shorter than q. Since $\|r\| = \|qs\| \leq 16\lfloor\delta + 1\rfloor$, we can conclude that $r \in R$, by (4.15). Observe that $u = \mathrm{Lab}(q)$ is a dominant prefix of r, by construction, as $\|q\| > \|s\|$. Since q is a subpath of p, we see that u is a desired subword of w, so the claim is proved. □

Claim 4.12.5. *$\langle X \mid R\rangle$ is a finite presentation of G.*

Proof. Note that since X is a finite set, there are only finitely many words of length at most $16\lfloor\delta + 1\rfloor$ over $X^{\pm 1}$, so $|R| < \infty$.

We need to show that $\ker\phi$ is the normal closure of R in $F(X)$ (see Definition 2.1.2). Evidently, $R \subseteq \ker\phi$, by the definition of R in (4.15), so $\langle\!\langle R\rangle\!\rangle^{F(X)} \subseteq \ker\phi$. For the opposite inclusion, pick a reduced word $w \in \ker\phi$. We will show that $w \in \langle\!\langle R\rangle\!\rangle^{F(X)}$ by induction on the length $\|w\|$.

If $\|w\| = 0$ then $w = 1 \in \langle\langle R\rangle\rangle^{F(X)}$, so the base of induction holds. Suppose that $\|w\| = n > 0$ and we have already shown that $w' \in \langle\langle R\rangle\rangle^{F(X)}$, for all reduced words $w' \in \ker\phi$ of length less than n. By Claim 4.12.4, we can write $w = sut$, where s, u, t are subwords of w, and u is a dominant prefix of a word $r \in R$, so that $r = uv$, where $\|u\| > \|v\|$. Observe that $u = r \cdot v^{-1}$ in $F(X)$, whence

$$w = s \cdot r \cdot v^{-1} \cdot t = s \cdot r \cdot s^{-1} \cdot s \cdot v^{-1} \cdot t = s \cdot r \cdot s^{-1} \cdot w' \text{ in } F(X),$$

where w' is the reduction of the word $sv^{-1}t$. Since w and $s \cdot r \cdot s^{-1}$ are both in $\ker\phi$, we see that $w' \in \ker\phi$. Moreover, the inequality $\|v^{-1}\| = \|v\| < \|u\|$ implies that

$$\|w'\| \leq \|s\| + \|v^{-1}\| + \|t\| < \|s\| + \|u\| + \|t\| = \|w\|,$$

so $w' \in \langle\langle R\rangle\rangle^{F(X)}$, by induction. Recall that normal closures are subgroups and $s \cdot r \cdot s^{-1} \in \langle\langle R\rangle\rangle^{F(X)}$, hence $w = s \cdot r \cdot s^{-1} \cdot w' \in \langle\langle R\rangle\rangle^{F(X)}$. This establishes the step of induction, completing the proof. □

Clearly, Claims 4.12.4 and 4.12.5 imply that $\langle X \mid R\rangle$ is a finite Dehn presentation of G, so the proof of Theorem 4.12.3 is finished. □

4.13. Solution of the Word Problem in hyperbolic groups

In 1911 Max Dehn proposed three fundamental problems in Group Theory: the Word Problem, the Conjugacy Problem and the Isomorphism Problem.

Definition 4.13.1. Given a group G, generated by a subset X, the *Word Problem in G is solvable* if there exists an algorithm taking on input pairs of words u, v over $X^{\pm 1}$ and deciding whether these two words represent the same element of G.

The group G is said to have *solvable Conjugacy Problem* if there is an algorithm taking on input pairs of words u, v over $X^{\pm 1}$ and deciding whether the elements represented by them are conjugate in G.

Remark 4.13.2. Note that equation $u = v$ in a group G is equivalent to the equation $u^{-1}v = 1$. So to solve the Word Problem in G, generated by a subset X, it is sufficient to be able to determine whether a given word over $X^{\pm 1}$ represents the identity element 1 in G. Since being equal to 1 is equivalent to being conjugate to 1, a group with solvable Conjugacy Problem also has a solvable Word Problem.

Definition 4.13.3. Let $\mathfrak{G}$ be a family of finitely presented groups. The *Isomorphism Problem is solvable in $\mathfrak{G}$* if there exists an algorithm taking on input pairs of finite presentations $\langle X_1 \mid R_1\rangle$, $\langle X_2 \mid R_2\rangle$, of groups $G_1, G_2 \in \mathfrak{G}$, and deciding whether $G_1 \cong G_2$.

The notion of an "algorithm" appearing in the above definitions can be formalized using Turing machines. Intuitively speaking, we can think of them as computer programs. A priori, solvability of the Word Problem may depend on the choice of a generating set in a group. Our first lemma shows that this is not an issue for finite generating sets of finitely generated groups.

Lemma 4.13.4. *Let G be a group and let $X, Y \subseteq G$ be two finite generating subsets of G. If the Word Problem is solvable in G with respect to X then it is also solvable with respect to Y.*

Proof. Let $Y = \{y_1, \dots, y_k\}$, and, for each $i = 1, \dots, k$, let u_i be a word over $X^{\pm 1}$ representing y_i in G. By Remark 4.13.2, to solve the Word Problem in G with respect to Y, it is sufficient to construct an algorithm determining whether any given word w over $Y^{\pm 1}$ represents the identity element in G. So, take a word w over $Y^{\pm 1}$ and re-write it as a word $v = v(w)$ over $X^{\pm 1}$, by replacing each letter $y_i^{\varepsilon} \in Y^{\pm 1}$ in w by the word u_i^{ε}, for $i = 1, \dots, k$ and $\varepsilon \in \{\pm 1\}$. Since the Word Problem in G with respect to X is solvable, we can determine whether this new word v represents the identity element in G. Clearly, this will also tell us whether the original word w over $Y^{\pm 1}$ represents 1 in G, as required. □

Example 4.13.5. (a) Consider the free group $F = F(X)$ on a set X. Let us describe an easy algorithm to solve the Word Problem in F. Given any word w over $X^{\pm 1}$, let w' be the reduced word obtained from w by applying elementary reductions (see Definition 1.2.8). By definition of the group $F = F(X)$ (cf. Definition 1.2.13), $w = 1$ in F if and only if w' is the empty word.

(b) Let $G = \mathbb{Z}^n$ be the free abelian group of rank $n \in \mathbb{N}_0$, and let $X = \{\mathbf{e_1}, \mathbf{e_2}, \dots, \mathbf{e_n}\}$ be the standard basis of G. Given any word w over $X^{\pm 1}$, using the commutativity of G we can re-write it in the form $k_1\mathbf{e_1} + k_2\mathbf{e_2} + \cdots + k_n\mathbf{e_n}$, for some $k_1, k_2, \dots, k_n \in \mathbb{Z}$ (we use the additive notation on G). And we know that $w = \mathbf{0} = (0, \dots, 0)$ in $G = \mathbb{Z}^n$ if and only if $(k_1, \dots, k_n) = \mathbf{0}$. Thus the Word Problem is solvable in G.

Dehn presentations of hyperbolic groups obtained in Section 4.12 provide an easy and efficient solution for the Word Problem in such groups, via an application of *Dehn's algorithm.*

Theorem 4.13.6 (Dehn's algorithm). *Let G be a group with a finite Dehn presentation $\langle X \mid R\rangle$. Then G has solvable Word Problem.*

Proof. According to Remark 4.13.2 and Lemma 4.13.4, it is enough to construct an algorithm deciding whether a word over $X^{\pm 1}$ represents 1 in G. So, suppose that we are given a word w_0 over $X^{\pm 1}$. First, we apply all possible elementary reductions to w_0 to obtain a reduced word w_0' from it. Evidently, w_0' represents the same element as w_0 in G. If w_0' is the empty word, then $w_0 = 1$ in G and the algorithm stops. So, we can further assume that w_0' is non-empty. Now, list all dominant prefixes of words $r \in R$ (there are finitely many of them because R is finite) and check if any of them is a subword of w_0'. If none of them are subwords of w_0', then $w_0 = w_0' \neq 1$ in G, according to Definition 4.12.2 of a Dehn presentation, and we can stop the algorithm. Thus, we can suppose that w_0' is the concatenation sut, of words s, u, t over $X^{\pm 1}$, where u is a dominant prefix of some word $r \in R$. Then r is the concatenation uv, where v is a word over $X^{\pm 1}$, with $\|v\| < \|u\|$. Since $r = 1$ we have $u = v^{-1}$ in G, whence w_0' represents the same element of G as the word $w_1 = sv^{-1}t$. Note that $\|w_1\| < \|w_0'\| \leq \|w_0\|$ because $\|v^{-1}\| = \|v\| < \|u\|$, so we have reduced our problem to working with a shorter word w_1. After repeating the same steps with w_1 as with w_0, we will either decide whether $w_1 = 1$ in G or reduce our problem to a shorter word w_2, with $\|w_2\| < \|w_1\|$. Clearly, proceeding this way, the algorithm will eventually terminate, telling us whether or not $w_0 = 1$ in G, as required. ☐

Combined with Theorem 4.12.3, Theorem 4.13.6 immediately gives the following.

Corollary 4.13.7. *Every hyperbolic group has solvable Word Problem.*

The Conjugacy Problem can also be solved in hyperbolic groups fairly easily (see, for example, Theorem 2.8 in Section III.Γ.2 of [M.R. Bridson, A. Haefliger, *Metric spaces of non-positive curvature.* Fundamental Principles of Mathematical Sciences, 319. Springer-Verlag, Berlin, 1999]). It is a much deeper and harder result that the Isomorphism Problem is solvable in the class of all hyperbolic groups. The proof of this fact has only been completed recently by François Dahmani and Vincent Guirardel, based on prior work of Zlil Sela and of Dahmani with Daniel Groves.

On the other hand, we should note that there exist finitely presented groups with unsolvable Word Problem. Such examples were first constructed by Pyotr Novikov in 1955, and independently, by William Boone in 1958 (an explicit finite presentation of such a group can be found here). Once such examples were produced, they were used to construct finitely presented groups with solvable Word Problem but unsolvable Conjugacy Problem, and to show that the Isomorphism Problem is undecidable in the class of all finitely presented groups. In fact, a celebrated Adian-Rabin theorem implies that many important group properties cannot be detected from finite group presentations. For instance, there is no algorithm determining whether a finite group presentation determines the trivial group, i.e., the *Triviality Problem* is unsolvable in the class of all finitely presented groups.

4.14. Undistorted subgroups of hyperbolic groups

When one studies subgroups hyperbolic groups, one usually starts by looking at undistorted (a.k.a., *quasiconvex*) subgroups that are quasi-isometrically embedded in the group. These subgroups possess many nice properties. In particular, they are hyperbolic themselves.

Definition 4.14.1. A finitely generated subgroup H of a finitely generated group G is said to be *undistorted* if, for some choices of finite generating subsets $Y \subseteq H$ and $X \subseteq G$, there exists $\mu \geq 1$ such that

$$|h|_Y \leq \mu |h|_X, \quad \text{for all } h \in H,$$

where $|\cdot|_X$ and $|\cdot|_Y$ denote the word lengths with respect to X and Y (see Definition 4.7.1).

One source of undistorted subgroups comes from retracts.

Definition 4.14.2. A subgroup H of a group G is called a *retract* if there is a homomorphism $\rho : G \to H$ such that $\rho(h) = h$, for all $h \in H$ (i.e., the restriction of ρ to H is the identity map Id_H). In this case the map ρ is said to a be *retraction of* G *onto* H.

Lemma 4.14.3. *If G is a finitely generated group then any retract H of G is finitely generated and undistorted in G.*

Proof. Suppose that $G = \langle X \rangle$, for some finite subset $X \subseteq G$, and let $\rho : G \to H$ be a retraction of G onto H. Note that $Y = \rho(X) \subseteq G$ is a finite generating set of H, by Lemma 0.5.11.(v), in particular, H is finitely generated. Consider any $h \in H$. By the definition of the word length, $|h|_X = n$ implies that $h = x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}$, for some $x_1, \dots, x_n \in X$ and some $\varepsilon_1, \dots, \varepsilon_n \in \{\pm 1\}$. Therefore,

$$h = \rho(h) = \rho(x_1)^{\varepsilon_1} \dots \rho(x_n)^{\varepsilon_n}, \quad \text{where } \rho(x_1), \dots, \rho(x_n) \in Y.$$

Hence, $|h|_Y \leq n = |h|_X$, thus H is undistorted in G. □

Example 4.14.4. Basic examples of retracts are free and direct factors.

(a) Suppose that $G = A * B$ is a free product of groups A and B. By the universal property of free products (Theorem 2.4.7), there is a homomorphism $\rho : G \to A$ such that $\rho(a) = a$, for all $a \in A$, and $\rho(b) = 1$, for all $b \in B$. Clearly, ρ is a retraction of G onto A.

(b) Similarly, if $G = A \times B$ is a direct product, then the map $\rho : G \to A$, defined by $\rho((a, b)) = a$, for all $a \in A$, is a retraction.

Exercise 4.14.5. Suppose that H is a subgroup of a group G, and Y, X are some finite generating sets of H and G respectively.

- (i) Prove that if H is undistorted in G (with respect to Y and X), then it is undistorted with respect to any other choice of finite generating subsets in H and G. [*Hint:* use Exercise 4.7.7.]
- (ii) Show that H is undistorted in G if and only if the inclusion map $i : H \to G$ is a quasi-isometric embedding between the metric spaces (H, d_Y) and (G, d_X). [*Hint:* by Exercise 4.7.7 you can assume that $Y \subseteq X$, whence $\mathrm{d}_X(h_1, h_2) \leq \mathrm{d}_Y(h_1, h_2)$, for all $h_1, h_2 \in H$.]
- (iii) Assume that G is hyperbolic and H is undistorted in G. Prove that H is itself hyperbolic. [*Hint:* Combine claim (ii) with Exercise 4.6.7 and Theorem 4.9.4.]
- (iv) Suppose that F is a finitely generated undistorted subgroup of H and H is undistorted in G. Show that F is undistorted in G.

Claim (iii) of Exercise 4.14.5 gives us a rich source of hyperbolic subgroups in hyperbolic groups. We can also strengthen Corollary 4.9.7 as follows.

Corollary 4.14.6. *If $n \geq 2$ then $\mathbb{Z}^n$ cannot be embedded as an undistorted subgroup in any hyperbolic group.*

Proof. Arguing by contradiction, assume that $\mathbb{Z}^n$ is an undistorted subgroup of a hyperbolic group G, for some $n \geq 2$. Since $n \geq 2$, we know that $\mathbb{Z}^n \cong A \times B$, where $A = \mathbb{Z}^2$ and $B = \mathbb{Z}^{n-2}$, thus $A = \mathbb{Z}^2$ is a retract of $\mathbb{Z}^n$, by Example 4.14.4.(b). Lemma 4.14.3 and claim (iv) of Exercise 4.14.5 imply that $\mathbb{Z}^2$ is an undistorted subgroup of G, whence $\mathbb{Z}^2$ must be hyperbolic by claim (iii) of the same exercise. The latter contradicts Corollary 4.9.7, so the statement is proved. □

Definition 4.14.7. A subgroup H of a group G is said to be a *virtual retract* if there is a finite index subgroup $K \leqslant G$ such that $H \subseteq K$ and H is a retract of K.

Virtual retracts play an important role in Geometric Group Theory, for example they are always quasi-isometrically embedded.

Exercise 4.14.8. If H is a virtual retract of a finitely generated group G then H is finitely generated and undistorted in G.

[*Hint:* combine Proposition 4.8.9 with Exercise 4.14.5.]

Many subgroups of groups studied in Geometric Group Theory are virtual retracts. We mention just two results in this direction, without proof. The first one is a theorem of Marshall Hall Jr., and the second one is an easy consequence of the classification of finitely generated abelian groups.

THEOREM 4.14.9 (M. Hall's theorem). *If H is a finitely generated subgroup of a finite rank free group F, then H is a free factor of a finite index subgroup $K \leqslant F$. In particular, H is a virtual retract of F.*

Proposition 4.14.10. *Suppose that G is a finitely generated abelian group. Then for every subgroup $H \leqslant G$ there is a finite index subgroup $K \leqslant G$, containing H, such that H is a direct factor of K. In particular, H is a virtual retract of G.*

Definition 4.14.11. Let (S, d) be a hyperbolic geodesic metric space. A subset $Q \subseteq S$ is said to be *quasi-convex* if there exists $\eta \geq 0$ such that for arbitrary points $x, y \in Q$ every geodesic segment joining these points is contained in the η-neighborhood of Q in S.

If G is a hyperbolic group and S_Γ is the geometric realization of the Cayley graph of this group with respect to some generating subset $X \subseteq G$, then a subset $Q \subseteq G$ is *quasi-convex in G* if its image is quasi-convex in $(S_\Gamma, \mathrm{d}_{S_\Gamma})$, under the natural inclusion of G "as the vertex set" of S_Γ.

Example 4.14.12. (a) If $\mathbb{R}$ is equipped with the standard Euclidean metric, then it is hyperbolic and the subsets $\mathbb{Z}$ and $\mathbb{N}$ are quasi-convex in it.

(b) Any subtree of a tree is quasi-convex in the geometric realization of this tree.

The next theorem lists important properties of quasi-convex subgroups in hyperbolic groups.

THEOREM 4.14.13. *Let G be a hyperbolic group.*

(a) *A subgroup $H \leqslant G$ is quasi-convex in G if and only if H is finitely generated and undistorted in G.*
(b) *The intersection of finitely many quasi-convex subgroups of G is again quasi-convex.*
(c) *For every $g \in G$ the centralizer $\mathrm{C}_G(g) = \{h \in G \mid hg = gh\} \leqslant G$ is quasi-convex in G.*

PROOF. Omitted. See Section III.Γ.3 of the book [M.R. Bridson, A. Haefliger, *Metric spaces of non-positive curvature.* Fundamental Principles of Mathematical Sciences, 319. Springer-Verlag, Berlin, 1999]. □

Combined with Exercise 4.14.5.(i), Theorem 4.14.13 implies that quasi-convexity of a subgroup of a hyperbolic group G is independent of the choice of a finite generating set $X \subseteq G$. The tool of quasi-convex subgroups allows us to improve Corollary 4.14.6 as follows.

THEOREM 4.14.14. *The free abelian group $\mathbb{Z}^2$ cannot be a subgroup of a hyperbolic group G.*

PROOF. Suppose that $A = \langle a, b\rangle \cong \mathbb{Z}^2$ is a subgroup of a hyperbolic group G. Let $C = \mathrm{C}_G(A)$ be the centralizer of A in G. Then $C = \mathrm{C}_G(a) \cap \mathrm{C}_G(b)$, so, according to Theorem 4.14.13, C is a quasi-convex subgroup of G, and $C = \langle Y \rangle$, for some finite subset $Y \subseteq C$. Applying the same theorem one more time, we can conclude that

$$H = \mathrm{Z}(C) = C \cap \mathrm{C}_G(C) = C \cap \bigcap_{y \in Y} \mathrm{C}_G(y)$$

is a finitely generated undistorted subgroup of G. Note that H is abelian (as it is the center of C) and $A \subseteq H$, by construction, therefore A is undistorted in H, by Proposition 4.14.10 and Exercise 4.14.8. Exercise 4.14.5.(iv) implies that A must be undistorted in G, contradicting Corollary 4.14.6. This contradiction shows that G cannot contain free abelian subgroups of rank (at least) two. □

Remark 4.14.15. Finitely generated free groups are prime examples of hyperbolic groups, and every finitely generated subgroup of a free group is free (by Theorem 1.3.5) and undistorted (by Theorem 4.14.9 and Exercise 4.14.8). However, it is not true that every finitely generated (or even finitely presented) subgroup of a general hyperbolic group is hyperbolic. Indeed, hyperbolic groups can contain *distorted* finitely generated subgroups that are not hyperbolic. The first construction of such subgroups was given by Eliyahu Rips, using small cancellation theory.

CHAPTER 5

Additional topics

In this chapter we discuss additional material regarding quasi-isometries between regular trees, free products, presentation of $\mathrm{SL}(2,\mathbb{Z})$ and residual finiteness of groups. In particular, we will give a proof of Theorem 1.3.7.

5.1. Classification of regular trees up to quasi-isometry

In this section we present an application of Group Theory to Graph Theory, by using groups to give a classification of regular trees up to quasi-isometry.

Definition 5.1.1. Let Γ_1 and Γ_2 be connected graphs, endowed with the path metrics d_{Γ_1} and d_{Γ_2}. We will say that these graphs are quasi-isometric, writing $\Gamma_1 \sim_{qi} \Gamma_2$, if $(V\Gamma_1, \mathrm{d}_{\Gamma_1})$ is quasi-isometric to $(V\Gamma_2, \mathrm{d}_{\Gamma_2})$.

Similarly, if (S, d) is a metric space, we will say that Γ_1 is quasi-isometric to (S, d) if $(V\Gamma_1, \mathrm{d}_{\Gamma_1})$ is quasi-isometric to (S, d), in which case we will write $\Gamma_1 \sim_{qi} S$.

Evidently, for every connected graph Γ the natural embedding of $V\Gamma$, equipped with the path metric d_Γ, into the geometric realization S_Γ is a $(1, 0, 1/2)$-quasi-isometry. Combined with Lemma 4.6.6, this observation immediately implies the following.

Remark 5.1.2. Suppose that Γ_1 and Γ_2 are connected graphs. Then $\Gamma_1 \sim_{qi} \Gamma_2$ if and only if $S_{\Gamma_1} \sim_{qi} S_{\Gamma_2}$, where, as usual, S_{Γ_i} denotes the geometric realization of Γ_i, for $i = 1, 2$.

For every $n \in N_0$ we will denote by T_n the (non-empty) n-regular tree. In other words, this is a tree where each vertex has degree n (see Definition 3.8.3).

Example 5.1.3. The tree T_0 consists of a single vertex (without edges), the tree T_1 consists of two vertices and a single edge joining them, the tree T_2 is the simplicial line (isomorphic to the Cayley graph $\Gamma(\mathbb{Z}, \{1\})$, see Figure 3.5), the tree T_3 is the 3-regular tree, displayed on Figure 5.1, and the tree T_4 is the 4-regular tree isomorphic to the Cayley graph of the free group of rank 2 with respect to a free basis, see Figure 3.12.

Remark 5.1.4. The discussion in Section 3.10 and, in particular, Remark 3.10.2, easily imply that for every $n \in \mathbb{N}$ the tree T_n is isomorphic to the Bass-Serre tree of the free product $C_n * C_n$.

Theorem 5.1.5 (Classification of regular trees up to quasi-isometry). *There are exactly* 3 *quasi-isometry classes of non-empty n-regular trees, for $n \in \mathbb{N}_0$:*

(i) $\{T_0, T_1\}$;
(ii) $\{T_2\}$;
(iii) $\{T_n \mid n \geq 3\}$.

In particular,

$$T_0 \sim_{qi} T_1 \not\sim_{qi} T_2 \not\sim_{qi} T_3 \sim_{qi} T_n, \quad \text{for all } n \geq 3.$$

The proof of Theorem 5.1.5 will occupy the rest of this section. To classify a family of metric spaces up to quasi-isometry one usually uses *quasi-isometry invariants*, that are properties of metric spaces preserved by quasi-isometries. One such property is hyperbolicity, see Theorem 4.9.4, however this invariant does not help in distinguishing trees because all of them are hyperbolic. Another simple invariant is *boundedness*. Recall that a metric space (S, d) is said to be bounded if it has finite diameter, i.e., if $\sup\{\mathrm{d}(x, y) \mid x, y \in S\} < \infty$.

Exercise 5.1.6. Show that

- any non-empty bounded metric space is quasi-isometric to the metric space consisting of a single point; hence, any two non-empty bounded metric spaces are quasi-isometric to each other;
- any metric space quasi-isometric to a bounded space is bounded.

Lemma 5.1.7. *$T_0 \sim_{qi} T_1$ and $T_1 \not\sim_{qi} T_n$, for any $n \geq 2$.*

PROOF. Note that both T_0 and T_1 are bounded, while T_n is unbounded, for any $n \geq 2$. Thus the claim of the lemma follows from Exercise 5.1.6. ☐

Another quasi-isometry invariant is *growth*. It is particularly useful in discrete metric spaces such as (vertex sets of) graphs. Roughly speaking, this invariant counts the number of points in balls of radius k, for all $k \in \mathbb{N}$. We will use this invariant implicitly in the next lemma (note that the cardinality of balls in T_2 grows linearly with the radius, while the cardinality of balls in T_3 grows exponentially).

Lemma 5.1.8. *The simplicial line T_2 is not quasi-isometric to the 3-regular tree T_3.*

PROOF. Arguing by contradiction, assume that $T_2 \sim_{qi} T_3$. Then there is a quasi-isometric embedding $f : VT_3 \to VT_2$, where VT_i is equipped with the path metric d_i, for $i = 2, 3$. This means that there exist $\lambda \geq 1$ and $c \geq 0$ such that

$$\frac{1}{\lambda}\mathrm{d}_3(u,v) - c \leq \mathrm{d}_2(f(u), f(v)) \leq \lambda \mathrm{d}_3(u,v) + c, \tag{5.1}$$

for all vertices $u, v \in VT_3$.

Fix a vertex $o \in VT_3$ and observe that, for every $r \in \mathbb{N}$, the sphere $\mathcal{S}_r(o) = \{x \in VT_3 \mid \mathrm{d}_3(o,x) = r\}$ contains exactly $3 \cdot 2^{r-1}$ vertices. For every vertex $u \in VT_3$ we will say that a vertex $w \in VT_3$ is a *successor* of u if u belongs to the unique geodesic path from o to w in T_3. For each vertex $u \in \mathcal{S}_r(o)$ choose a single successor $w \in \mathcal{S}_{2r}(o)$ and denote by $W_r \subseteq VT_3$ the resulting subset of $\mathcal{S}_{2r}(o) \subseteq VT_3$, so that $|W_r| = |\mathcal{S}_r(o)| = 3 \cdot 2^{r-1}$ (see Figure 5.1 for an illustration of the case $r = 2$).

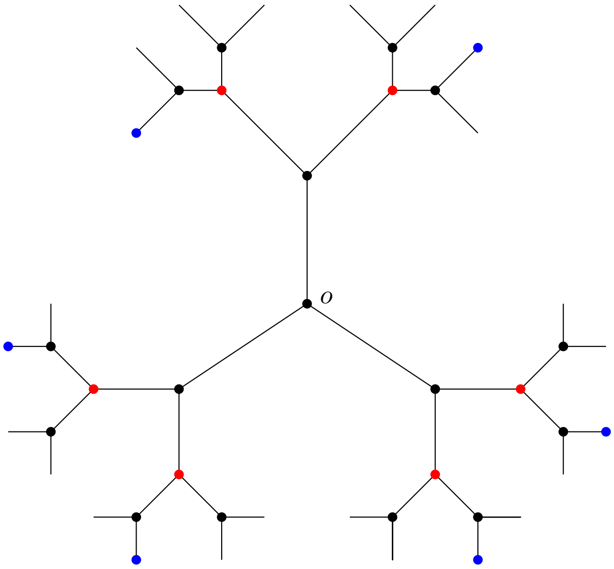


FIGURE 5.1. The sphere of radius 2 (in red) and the corresponding set W_2 (in blue) in T_3.

Observe that since T_3 is a tree, we have

$$\mathrm{d}_3(o, w_1) = 2r \text{ and } \mathrm{d}_3(w_1, w_2) \geq 2r + 2 > 2r, \text{ whenever } w_1, w_2 \in W_r \text{ are distinct.} \tag{5.2}$$

Combining this with (5.1), we can conclude that

$$\mathrm{d}_2(f(0), f(w)) \le 2\lambda r + c, \quad \text{for every } w \in W_r,$$

hence $f(W_r) \subseteq B_R$, where B_R is the ball of radius $R = 2\lambda r + c$ centered at o in T_2. Note that since T_2 is the simplicial line, we have $|B_R| = 2R + 1 = 4\lambda r + 2c + 1$. Thus the cardinality of B_R grows linearly with r, while the cardinality $|W_r| = 3 \cdot 2^{r-1}$ grows exponentially fast. Therefore, there exists $r \in \mathbb{N}$ such that $|W_r| > |B_R|$ and

$$r > \frac{1}{2}\lambda c, \text{ so that } \frac{2r}{\lambda} - c > 0. \tag{5.3}$$

Since $f(W_r) \subseteq B_R$ and $|W_r| > |B_R|$, there must be two distinct vertices $w_1, w_2 \in W_r$ such that $f(w_1) = f(w_2)$. In view of (5.3) and (5.2), the latter implies that

$$\mathrm{d}_2(f(w_1), f(w_2)) = 0 < \frac{2r}{\lambda} - c < \frac{1}{\lambda}\mathrm{d}_3(w_1, w_2) - c,$$

contradicting (5.1). Therefore, we have shown that T_3 is not quasi-isometric to T_2. □

Remark 5.1.9. The argument from the proof of Lemma 5.1.8 actually shows the stronger statement that T_3 cannot be quasi-isometrically embedded in T_2.

Proving that certain metric spaces are quasi-isometric requires different methods. One such method is the Schwarz-Milnor lemma, discussed in Section 4.8. It plays and important role in our proof of the following statement.

Lemma 5.1.10. *If $m, n \in \mathbb{N}$ and $m, n \ge 3$ then $T_m \sim_{qi} T_n$.*

The proof of this lemma will use the following two exercises.

Exercise 5.1.11. Given any integer $n \ge 2$, consider the free product $G_n = C_n * C_n$, equipped with the word metric d_X, corresponding to some finite generating set $X \subseteq G$. Prove that (G_n, d_X) is quasi-isometric to T_n. [*Hint: use Remark 5.1.4 and the Schwarz-Milnor lemma (Theorem 4.8.3).*]

Exercise 5.1.12. Show that for any integers $k, l \ge 2$ the free group of rank k is quasi-isometric to the free group of rank l. [*Hint: show that for each $k \ge 2$ the free group F_2 has a finite index subgroup isomorphic to F_k.*]

Proof of Lemma 5.1.10. By Exercise 5.1.11, for every $n \ge 2$ the tree T_n is quasi-isometric to the free product $G_n = C_n * C_n$ (equipped with the word metric corresponding to some finite generating set). On the other hand, from Example 3.11.3 we know that G_n has a free subgroup N_n of finite index. Moreover, according to Proposition 3.11.4, if $n \ge 3$ then the rank of N_n is finite and is at least 2 (in fact, $\mathrm{rank}(N_n) = (n-1)^2$ by Exercise 3.11.6). Proposition 4.8.9 tells us that $G_n \sim_{qi} N_n$.

Now, if $m \ge 2$ is another integer, then $N_m \sim_{qi} N_n$ by Exercise 5.1.12, whence

$$T_m \sim_{qi} G_m \sim_{qi} N_m \sim_{qi} N_n \sim_{qi} G_n \sim_{qi} T_n.$$

Thus, in view of Lemma 4.6.6, we can conclude that $T_m \sim_{qi} T_n$. □

Evidently, the complete classification of regular trees stated in Theorem 5.1.5 follows immediately from Lemmas 5.1.7,5.1.8 and 5.1.10.

Exercise 5.1.13. Use Theorem 5.1.5 to show that $\mathbb{Z} \not\sim_{qi} F_2$, where F_2 denotes the free group of rank 2 (this fact can be compared with Exercise 5.1.12).

5.2. Ping-pong for free products

The following statement gives a very useful “table tennis” criterion for showing that specific subgroups of a group G generate their free product.

Theorem 5.2.1 (Ping-pong lemma for free products). *Let G be a group acting on a set Ω. Suppose that H_1, H_2 are subgroups of G, with $|H_1| \geq 3$, and $\Omega_1, \Omega_2 \subseteq \Omega$ are subsets, with $\Omega_1 \not\subseteq \Omega_2$, such that*

$$h_1.\Omega_1 \subseteq \Omega_2 \quad \text{and} \quad h_2.\Omega_2 \subseteq \Omega_1, \quad \text{for all } h_1 \in H_1 \setminus \{1\},\ h_2 \in H_2 \setminus \{1\}. \tag{5.4}$$

Then the subgroup $\langle H_1, H_2 \rangle \leqslant G$, generated by H_1 and H_2 in G, is naturally isomorphic to the free product $H_1 * H_2$.

Proof. Denote by H the subgroup of G generated by H_1 and H_2. By the Universal property of free products (Theorem 2.4.7), there is a group homomorphism $\phi : H_1 * H_2 \to H$ such that the restriction of ϕ to H_i is the identity map, for each $i = 1, 2$. Note that ϕ is surjective, by Lemma 0.5.11.(ii), so we only need to prove that it is injective.

Consider a non-trivial element $g \in H_1 * H_2$; we will show that $g \notin \ker \phi$. As we know, g has a non-empty normal form w, which is a freely reduced word over the alphabet $H_1 \sqcup H_2$ (see Section 2.4), and we can consider several possibilities.

Case 1: w starts and ends with letters from H_1, i.e., $w = a_0b_1a_1b_2 \dots b_ka_k$, where $k \in \mathbb{N}_0$, $a_0, \dots, a_k \in H_1 \setminus \{1\}$ and $b_1, \dots, b_k \in H_2 \setminus \{1\}$. Then

$$\phi(g) = \phi(a_0b_1a_1b_2 \dots b_ka_k) = a_0b_1a_1b_2 \dots b_ka_k \in H.$$

Conditions (5.4) ensure that

$$\begin{aligned} \phi(g).\Omega_1 \quad &= (a_0b_1a_1b_2 \dots b_ka_k).\Omega_1 \subseteq (a_0b_1a_1b_2 \dots b_k).\Omega_2 \\ &\subseteq (a_0b_1a_1b_2 \dots a_{k-1}).\Omega_1 \subseteq \cdots \subseteq a_0.\Omega_1 \subseteq \Omega_2. \end{aligned}$$

Since Ω_1 is not contained in Ω_2, by assumptions, we can conclude that $\phi(g) \neq 1$ in H, i.e., $g \notin \ker \phi$.

Case 2: w starts with a letter from H_1 and ends with a letter from H_2, i.e., $w = a_1b_1a_1b_2 \dots a_kb_k$, where $k \in \mathbb{N}$, $a_1, \dots, a_k \in H_1 \setminus \{1\}$ and $b_1, \dots, b_k \in H_2 \setminus \{1\}$. Since $|H_1| \geq 3$ we can choose some $a \in H_1 \setminus \{1, a_1\}$. Then $(a^{-1}a_1)b_1a_1b_2 \dots a_kb_ka$ is a non-empty reduced word in the free product $H_1 * H_2$ (in the sense of Definition 2.5.2), representing the element $a^{-1}ga \in H_1 * H_2$. This new word fits the assumptions of Case 1, so $\phi(a^{-1}ga) \neq 1$ in H, hence $g \notin \ker \phi$.

The remaining two cases, when w starts with a letter from H_2 and ends with a letter from H_1, and when w starts and ends with letters from H_2, can be dealt with in a similar fashion (we can reduce each of them to Case 1 by conjugating with a suitable element of H_1). Therefore, we have shown that $\ker \phi = \{1\}$, i.e., ϕ is injective. Thus ϕ is an isomorphism between $H_1 * H_2$ and $H = \langle H_1, H_2 \rangle \leqslant G$, as required. □

To give an application of the Ping-Pong lemma, let us use it to show that $\mathrm{PSL}(2, \mathbb{Z}) \cong C_2 * C_3$. Recall that $\mathrm{PSL}(2, \mathbb{Z})$ is defined as the quotient of $\mathrm{SL}(2, \mathbb{Z})$ by its center $Z = \{I_2, -I_2\}$. Thus we have a natural epimorphism $\mathrm{SL}(2, \mathbb{Z}) \to \mathrm{PSL}(2, \mathbb{Z})$, and we let $b, c \in \mathrm{PSL}(2, \mathbb{Z})$ denote the images of the matrices $B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$, respectively.

Proposition 5.2.2. *The group $\mathrm{PSL}(2, \mathbb{Z})$ has the presentation $\langle b, c \mid b^2 = 1,\ c^3 = 1 \rangle$. In particular, $\mathrm{PSL}(2, \mathbb{Z}) \cong C_2 * C_3$.*

Proof. Note that $\mathrm{SL}(2, \mathbb{R})$ acts on the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ by *Möbius transformations*. This action is defined as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.z = \frac{az + b}{cz + d}, \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}) \text{ and all } z \in \mathbb{H}. \tag{5.5}$$

It is easy to check that the kernel of this action is $Z = \{I_2, -I_2\}$, which is central in $\mathrm{SL}(2, \mathbb{R})$. The restriction of this action to $\mathrm{SL}(2, \mathbb{Z})$ gives a homomorphism $\psi : \mathrm{SL}(2, \mathbb{Z}) \to \mathrm{Homeo}(\mathbb{H})$, where $\mathrm{Homeo}(\mathbb{H})$ denotes the group of all self-homeomorphisms of $\mathbb{H}$ ($\mathbb{H}$ is equipped with the topology induced from $\mathbb{R}^2$), such that $\ker \psi = Z$. By the First Isomorphism Theorem, $\psi\big(\mathrm{SL}(2, \mathbb{Z})\big) \cong \mathrm{SL}(2, \mathbb{Z})/Z = \mathrm{PSL}(2, \mathbb{Z})$. Thus we can think of $\mathrm{PSL}(2, \mathbb{Z})$ as the image of $\mathrm{SL}(2, \mathbb{Z})$ in $\mathrm{Homeo}(\mathbb{H})$, which consists of Möbius transformations given by matrices from $\mathrm{SL}(2, \mathbb{Z})$ (see (5.5)).

Recall, from Example 0.5.14, that $\mathrm{SL}(2,\mathbb{Z})$ is generated by the matrices

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

It follows that $\{B, C\}$ is also a generating set of $\mathrm{SL}(2,\mathbb{Z})$, where $C = AB = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$. Therefore, according to Lemma 0.5.11.(v), $\mathrm{PSL}(2,\mathbb{Z}) = \langle b, c\rangle$, where $b = \psi(B)$ and $c = \psi(C)$ are the Möbius transformations of $\mathbb{H}$ given by

$$b : z \mapsto -\frac{1}{z} \text{ and } c : z \mapsto \frac{z-1}{z} = 1 - \frac{1}{z}.$$

Note that $b^2 = c^3 = \mathrm{Id}_{\mathbb{H}}$, because $B^2 = C^3 = -I_2 \in Z$, and since $C^2 = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$, we have

$$c^2 : z \mapsto \frac{-1}{z-1} = \frac{1}{1-z}.$$

Now, consider the following subsets of $\mathbb{H}$:

$$\Omega_1 = \{z \in \mathbb{H} \mid \mathrm{Re}(z) < 0)\} \text{ and } \Omega_2 = \{z \in \mathbb{H} \mid \mathrm{Re}(z) > 0)\}.$$

Observe that $b.\Omega_2 \subseteq \Omega_1$, because $b(x+iy) = \dfrac{-x+iy}{x^2+y^2}$, for all $z = x+iy \in \mathbb{H}$. Similarly, easy calculations show that $c^k.\Omega_1 \subseteq \Omega_2$, for $k = 1, 2$ (see Figure 5.2).

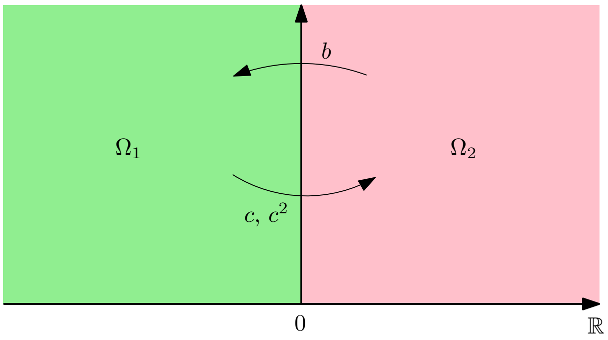


FIGURE 5.2. The actions of $\langle c\rangle$ and $\langle b\rangle$ on Ω_1 and Ω_2.

Therefore, the assumptions of the Ping-Pong lemma (Theorem 5.2.1) are satisfied by the subgroups $H_1 = \langle c\rangle \cong C_3$ and $H_2 = \langle b\rangle \cong C_2$, so these subgroups generate their free product in $\mathrm{Homeo}(\mathbb{H})$. Hence,

$$\mathrm{PSL}(2,\mathbb{Z}) = \langle b, c\rangle \cong C_2 * C_3,$$

and this isomorphism sends b to the generator of C_2 and c to a generator of C_3. It follows that $\mathrm{PSL}(2,\mathbb{Z})$ has the presentation $\langle b, c \mid b^2 = 1,\ c^3 = 1\rangle$. □

We can now use the previous result to produce a presentation for $\mathrm{SL}(2,\mathbb{Z})$.

Corollary 5.2.3. *The group* $\mathrm{SL}(2,\mathbb{Z})$ *has the presentation*

$$\langle x, y \mid x^4 = 1,\ x^2 = y^3\rangle. \tag{5.6}$$

PROOF. As shown in Proposition 5.2.2, $\mathrm{SL}(2,\mathbb{Z})$ is generated by the matrices $B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$, $B^2 = C^3 = -I_2$ and $B^4 = I_2$. Moreover, if $\psi : \mathrm{SL}(2,\mathbb{Z}) \to \mathrm{PSL}(2,\mathbb{Z})$ is the natural homomorphism then $b = \psi(B)$ and $c = \psi(C)$ satisfy $b^2 = c^3 = 1$ in $\mathrm{PSL}(2,\mathbb{Z})$, and $\mathrm{PSL}(2,\mathbb{Z}) \cong \langle b\rangle_2 * \langle c\rangle_3$.

Let G be the group defined by the presentation (5.6). By von Dyck's Theorem (Theorem 2.2.1), there is a homomorphism $\phi : G \to \mathrm{SL}(2,\mathbb{Z})$, such that $\phi(x) = B$ and $\phi(y) = C$. Moreover, ϕ is surjective by Lemma 0.5.11.(ii).

Consider any non-trivial element $g \in G$. We need to show that $\phi(g) \neq 1$. Note that since $x^2 = y^3 \in \langle x \rangle \cap \langle y \rangle$ in G, $x^2 = y^3$ commutes with both generators $x, y \in G$, so we can move all even powers of x and all powers of y divisible by 3 to the beginning of any word over $\{x,y\}^{\pm 1}$ representing g in G. Thus, we can write $g = zw$, where $z \in \langle x^2 \rangle \leqslant G$ and w is an alternating product of elements from $\{x\}$ and $\{y, y^2\}$ (i.e., $w = xy^{\varepsilon_1}xy^{\varepsilon_2}\ldots$ or $w = y^{\varepsilon_1}xy^{\varepsilon_2}x\ldots$, where $\varepsilon_i \in \{1,2\}$).

First, suppose that $g = z \in \langle x^2 \rangle$ in G. Since $x^4 = 1$ and $g \neq 1$, we can conclude that $z = x^2$, hence $\phi(g) = \phi(x^2) = B^2 = -I_2$ is non-trivial in $\mathrm{SL}(2,\mathbb{Z})$.

Therefore, we can assume that w is non-trivial. Then its image under the homomorphism $\psi \circ \phi : G \to \mathrm{PSL}(2,\mathbb{Z})$ is a non-empty reduced word in the free product $\langle b \rangle * \langle c \rangle$, hence it will be non-trivial in $\mathrm{PSL}(2,\mathbb{Z})$, by Theorem 2.5.3. But $\phi(z) \in \{I_2, -I_2\} = \ker \psi$, so $(\psi \circ \phi)(g) = (\psi \circ \phi)(w) \neq 1$ in $\mathrm{PSL}(2,\mathbb{Z})$, which implies that $\phi(g) \neq 1$ in $\mathrm{SL}(2,\mathbb{Z})$.

Thus, we have shown that $\ker \phi = \{1\}$, so $\phi : G \to \mathrm{SL}(2,\mathbb{Z})$ is an isomorphism and the proof is complete. $\square$

Exercise 5.2.4. Consider the matrices $D = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $E = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ in $\mathrm{SL}(2,\mathbb{Z})$. Prove that these matrices freely generate a free subgroup of $\mathrm{SL}(2,\mathbb{Z})$.

[*Hint:* apply the Ping-Pong lemma (Theorem 5.2.1) to the natural action of $\mathrm{SL}(2,\mathbb{Z})$ on $\mathbb{R}^2$, with $\Omega_1 = \{(x,y)^T \in \mathbb{R}^2 \mid |x| < |y|\}$ and $\Omega_2 = \{(x,y)^T \in \mathbb{R}^2 \mid |x| > |y|\}$.]

5.3. Residually finite groups

Sometimes an infinite group G can be "approximated" by finite groups and this may help in answering various questions about G. In this section we investigate one such approximation property, called residual finiteness.

Definition 5.3.1. A group G is said to be *residually finite* if for every non-trivial element $g \in G$ there exists a finite group M and a homomorphism $\phi : G \to M$ such that $\phi(g) \neq 1$ in M.

Evidently, every finite group is residually finite.

Exercise 5.3.2. Prove that a group G is residually finite if and only if the intersection of all finite index normal subgroups of G is trivial.

Exercise 5.3.3. Show that if G is a residually finite group then for every $g \in G \setminus \{1\}$ there exists $n \in \mathbb{N}$ and a homomorphism $\phi : G \to S_n$ such that $\phi(g) \neq 1$ in S_n.

Remark 5.3.4. Definition 5.3.1 immediately implies that every subgroup of a residually finite group is residually finite.

Exercise 5.3.5. Let G be a group.

- Prove that for every finite index subgroup K of G there is a finite index normal subgroup $N \lhd G$ such that $N \subseteq K$.

  [*Hint:* Consider the natural action of G on the finite set of left cosets $\{hK \mid h \in G\}$, defined by $g.hK = ghK$, and show that the kernel of this action has the desired properties.]
- Show that if G contains a residually finite subgroup of finite index then G is residually finite.

An important source of residually finite groups comes from finitely generated subgroups of $\mathrm{GL}(n,\mathbb{K})$, where $\mathbb{K}$ is a field. We demonstrate this in the following special case.

Lemma 5.3.6. *For every $n \in \mathbb{N}$, every subgroup of $\mathrm{GL}(n,\mathbb{Z})$ is residually finite.*

PROOF. By Remark 5.3.4, it is sufficient to show that $\mathrm{GL}(n,\mathbb{Z})$ is residually finite itself.

Choose any non-trivial matrix $A \in \mathrm{GL}(n,\mathbb{Z})$. Since $A \neq I_n$, we can find a sufficiently large prime $p \in \mathbb{N}$ such that $A \not\equiv I_n \mod p$ (for example, this will be true as long as $p > \max\{2, |a_{ij}| \mid 1 \leq i,j \leq n\}$, where $A = (a_{ij})$). This means that $\phi_p(A)$ is non-trivial in the group $\mathrm{GL}(n,\mathbb{Z}_p)$, where $\phi_p : \mathrm{GL}(n,\mathbb{Z}) \to \mathrm{GL}(n,\mathbb{Z}_p)$ is the homomorphism sending each matrix $A = (a_{ij}) \in \mathrm{GL}(n,\mathbb{Z})$ to the matrix $A = ([a_{ij}]_p) \in \mathrm{GL}(n,\mathbb{Z}_p)$, and $[a]_p$ denotes the residue of $a \in \mathbb{Z}$ modulo p. Since $|\mathrm{GL}(n,\mathbb{Z}_p)| < \infty$, we can conclude that $\mathrm{GL}(n,\mathbb{Z})$ is residually finite, as claimed. $\square$

Corollary 5.3.7. *Every free group of countable rank (finite or infinite) is residually finite.*

Proof. By Exercise 5.2.4, $\mathrm{GL}(2,\mathbb{Z})$ contains a free subgroup F_2, of rank 2, and, by Proposition 1.3.3, F_2 contains free subgroups of all countable ranks. So, we can use Lemma 5.3.6 to conclude that any countable rank free group is residually finite. ☐

Exercise 5.3.8. Show that a free group F of arbitrary rank can be "approximated" by finite rank free groups, and use this to conclude that F is residually finite.

Remark 5.3.9. Exercises 5.3.8 and 5.3.5 together imply that every virtually free group is residually finite. In particular, the free product of finitely many finite groups (such as D_∞ or $\mathrm{PSL}(2,\mathbb{Z})$) is residually finite, see Proposition 3.11.4.

Let us now discuss two applications of residual finiteness.

Definition 5.3.10. A group G is called *Hopfian* if every surjective homomorphism $\psi : G \to G$ is injective i.e., it is an automorphism of G.

The following fact was first observed by Russian mathematician Anatoliy Mal'cev.

Proposition 5.3.11. *A finitely generated residually finite group is Hopfian.*

Proof. Let G be a finitely generated residually finite group. Arguing by contradiction, suppose that there is a non-injective epimorphism $\psi : G \to G$. Take any $g \in \ker\psi \setminus \{1\}$, then, by residual finiteness, there is a finite group M and a homomorphism $\phi : G \to M$ such that $\phi(g) \neq 1$ in M.

For every $n \in \mathbb{N}$ we can consider the homomorphism $\xi_n = \phi \circ \psi^n : G \to M$. Let us show that $\xi_m \neq \xi_n$, provided $m, n \in \mathbb{N}$ and $m < n$. Indeed, $\psi^m : G \to G$ is surjective, as a composition of surjective maps, so there is $h \in G$ such that $\psi^m(h) = g$. Then $\psi^n(h) = \psi^{n-m}(g) = 1$ in G, as $g \in \ker\psi$ and $n > m$. Hence,

$$\xi_m(h) = \phi(\psi^m(h)) = \phi(g) \neq 1 \text{ and } \xi_n(h) = \phi(\psi^n(h)) = 1 \text{ in } M, \text{ so } \xi_m \neq \xi_n.$$

Therefore, we have found an infinite sequence $\{\xi_n\}_{n\in\mathbb{N}}$ of pairwise distinct group homomorphisms from G to M. However, since G is generated by k elements, for some $k \in \mathbb{N}_0$, the number of such homomorphisms cannot exceed $|M|^k$ (see Exercise 0.5.13). This yields a contradiction, showing that G must be Hopfian. ☐

By combining Proposition 5.3.11 with Corollary 5.3.7 we immediately achieve the following.

Corollary 5.3.12. *Free groups of finite rank are Hopfian.*

We can now prove Theorem 1.3.7, first stated in Chapter 1.

Theorem 1.3.7. *Let F be the free group of rank $n \in \mathbb{N}_0$. If $Y \subseteq F$ is a generating set such that $|Y| = n$ then Y is a free basis of F.*

Proof. By the assumptions, F has a free basis X, with $|X| = n = |Y|$. Fix any bijection $\phi : X \to Y$, and let $\hat{\phi} : F \to F$ be the unique homomorphism extending this bijection, which exists by the Universal Property of free groups (see Definition 1.1.1). Note that $\hat{\phi}$ is surjective by Lemma 0.5.11.(ii), because $F = \langle Y \rangle$. Hence, $\hat{\phi}$ must be an automorphism of F, by Corollary 5.3.12. Therefore, we can apply Exercise 1.5.2 to conclude that Y is a free generating set of F. ☐

Our second application of residual finiteness is also often attributed to Mal'cev. Recall that the Word Problem in a group was described in Definition 4.13.1.

Proposition 5.3.13. *A finitely presented residually finite group has solvable Word Problem.*

Proof. Let G be a residually finite group with a finite presentation $\langle X \mid R\rangle$. The algorithm to solve the Word Problem in G takes on input a word w over $X^{\pm 1}$ and determines whether w represents the identity element in G (see Remark 4.13.2). It does so by running simultaneously the following two procedures on w (without loss of generality, we can assume that w is a freely reduced word).

Procedure 1: start writing down (on an infinite tape) all reduced words $\{u_1 = 1, u_2, u_3, \dots\}$ in the normal closure $\langle\langle R \rangle\rangle^{F(X)}$, by considering all possible products of words of the form $fr^{\pm 1}f^{-1}$ in $F(X)$, where $f \in F(X)$ and $r \in R$ are arbitrary. After a new word u_n has been written down, compare it with w. If $w = u_n$ in $F(X)$ then stop the algorithm and output that w is trivial in G. Otherwise, proceed to forming and checking u_{n+1}.

Procedure 2: for each $n \in \mathbb{N}$ write down all possible set maps $X \to S_n$ (there are $(n!)^{|X|} < \infty$ of them). Using von Dyck's Theorem 2.2.1, for each of these maps we can check if it extends to a group homomorphism $G \to S_n$; if it does, then we can also check whether the image of the element $g \in G$, represented by w, is non-trivial in S_n under this homomorphism. If $\phi(g) \neq 1$ for some homomorphism $\phi : G \to S_n$ then the algorithm stops and outputs that w is non-trivial in G.

Clearly, Procedure 1 stops if and only if $w = 1$ in G. Furthermore, since G is residually finite, Exercise 5.3.3 tells us that Procedure 2 stops if and only if $w \neq 1$ in G. Thus the above algorithm will necessarily terminate after finite time on each input $w \in F(X)$, deciding whether or not w is trivial in G. □

The algorithm described in Proposition 5.3.13 is not practical, as it would take too long from the computational viewpoint. On the other hand, the solution of the Word Problem for hyperbolic groups discussed in Section 4.13 runs in linear time on the input, and is, therefore, much more efficient.

The following famous question, attributed to Gromov, is still open.

Open Question 5.3.14. Is every hyperbolic group residually finite?