Combinatorial Group Theory

Charles F. Miller III

7 March, 2004

Abstract

An early version of these notes was prepared for use by the participants in the Workshop on Algebra, Geometry and Topology held at the Australian National University, 22 January to 9 February, 1996. They have subsequently been updated and expanded many times for use by students in the subject 620-421 Combinatorial Group Theory at the University of Melbourne.

Contents

Chapter 1

Preliminaries

The reader is assumed to be familiar with the basics of group theory including the isomorphism theorems and the structure theory of finitely generated abelian groups. The groups we will be dealing with are usually infinite and non-abelian. In relation to several topics the reader will also be expected to be familiar with the theory of fundamental groups and covering spaces. In this chapter we establish some notational conventions and collect some frequently used facts from both elementary group theory and from the theory of covering spaces.

1.1 About groups

We normally write groups multiplicatively: if G is a group and $g, h \in G$ then their product is gh, the identity is $1 \in G$ and the inverse of g is g^{-1}. Two elements $g, h \in G$ are said to *commute* if $gh = hg$. Observe that

$$gh = hg \leftrightarrow h^{-1}gh = g \leftrightarrow g^{-1}h^{-1}gh = 1.$$

For any elements $u, v \in G$ we define their *commutator* $[u, v] = u^{-1}v^{-1}uv$ which measures whether or not u and v commute.

The element $v^{-1}uv$ is called the *conjugate* of u by v, while vuv^{-1} is the conjugate of u by v^{-1}. Two elements $u, w \in G$ are said to be *conjugate* in G if there exists and element $v \in G$ such that $w = v^{-1}uv$. Being conjugate is an equivalence relation. For fixed $v \in G$ the function $\iota_v : G \to G$ defined by $\iota_v(g) = v^{-1}gv$ is called *conjugation by* v. It is a homomorphism since

$$\iota_v(gh) = v^{-1}ghv = v^{-1}gvv^{-1}hv = \iota_v(g)\iota_v(h)$$

and in fact it is an automorphism of G (that is, an isomorphism of G with itself) called an *inner automorphism*.

If S is a subset of G, we denote by $\langle S\rangle$ the smallest subgroup of G containing S, called the *subgroup generated by* S. This subgroup $\langle S\rangle$ can be characterized as the set of all $g \in G$ which are equal to a (possibly empty) product of $s_i \in S$ and their inverses, that is

$$\{g \in G \mid g = s_1^{\epsilon_1} s_2^{\epsilon_2} \cdots s_k^{\epsilon_k} \text{ for some } s_i \in S \text{ and } \epsilon_i = \pm 1\}.$$

To see this, just observe that the collection of elements having such expressions contains S and is closed under products and also taking inverses since $(s_1^{\epsilon_1} s_2^{\epsilon_2} \cdots s_k^{\epsilon_k})^{-1} = s_k^{-\epsilon_k} \cdots s_2^{-\epsilon_2} s_1^{-\epsilon_1}$.

Similarly if R is a subset of G we denote by either $\mathrm{nm}_G(R)$ or sometimes $\langle R\rangle^G$ the smallest normal subgroup of G containing R, called the *normal closure of* R *in* G. This normal subgroup $\mathrm{nm}_G(R)$ can be characterized as the set of all $g \in G$ which are equal to a (possibly empty) product of conjugates of $r_i \in R$ and their inverses, that is the set of $g \in G$ such that

$$g = u_1^{-1} r_1^{\epsilon_1} u_1 \; u_2^{-1} r_2^{\epsilon_2} u_2 \cdots u_k^{-1} r_k^{\epsilon_k} u_k \text{ for some } r_i \in R, u_i \in G \text{ and } \epsilon_i = \pm 1.$$

Of course this is just the subgroup generated by all the conjugates of elements of R. To see that it is a normal subgroup we observe that, for g expressed as above,

$$\begin{aligned} v^{-1} g v &= v^{-1}(u_1^{-1} r_1^{\epsilon_1} u_1 \; u_2^{-1} r_2^{\epsilon_2} u_2 \cdots u_k^{-1} r_k^{\epsilon_k} u_k) v \\ &= v^{-1} u_1^{-1} r_1^{\epsilon_1} u_1 v \; v^{-1} u_2^{-1} r_2^{\epsilon_2} u_2 v \cdots v^{-1} u_k^{-1} r_k^{\epsilon_k} u_k v \\ &= (u_1 v)^{-1} r_1^{\epsilon_1} (u_1 v) \; (u_2 v)^{-1} r_2^{\epsilon_2} (u_2 v) \cdots (u_k v)^{-1} r_k^{\epsilon_k} (u_k v) \end{aligned}$$

One particularly important subgroup of any group G is its *commutator subgroup* $[G, G]$ which is defined to be the subgroup generated by all the commutators $[g, h]$ with $g, h \in G$. Since the conjugate of a commutator is again a commutator, $[G, G]$ is a normal subgroup and the quotient group $G/[G, G]$ is abelian because every commutator belongs to $[G, G]$. If we suppose $\phi : G \to A$ where A is abelian, then $\phi([g, h]) = [\phi(g), \phi(h)] =_A 1$ since A is abelian and so $[G, G] \subseteq \ker \phi$. Thus $G/[G, G]$ is the largest abelian quotient group of G. The subgroup $[G, G]$ has the additional property that it is *fully invariant*, meaning that it is mapped into itself by any homomorphism $\psi : G \to G$. (Note that a fully invariant subgroup is necessarily normal.)

Another subgroup of interest is the *center* $Z(G)$ consisting of all elements z such that $zg = gz$ for all $g \in G$. It is easy to see that $Z(G)$ is a normal subgroup of G and moreover that it is *characteristic*, meaning that it is invariant under any automorphism of G.

The commutator subgroup $[G, G]$ is sometimes called the *derived group* and denoted G'. Next one can define $G'' = [G', G'] = [[G, G], [G, G]]$ which is the commutator subgroup of G'. Inductively one defines a descending series of subgroups by $G^{(n+1)} = [G^{(n)}, G^{(n)}]$ called the *derived series* of G of which G' and G'' are the first two terms. The $G^{(k)}$ are fully invariant subgroups and the successive quotients $G^{(k)}/G^{(k+1)}$ are abelian. A group G is *solvable (or soluble) of derived length* $\leq n$ if $G^{(n)} = 1$. In this case G can be constructed by taking successive extensions by the abelian groups $G^{(k)}/G^{(k+1)}$. In the particular case that $G^{(2)} = G'' = 1$, the derived group G' of G is abelian and G is said to be *metabelian*. As an example, the multiplicative group of invertible $n \times n$ upper triangular matrices over any commutative ring is solvable of derived length $n - 1$.

Another commonly studied descending series of fully invariant subgroups of G is the *lower central series* defined by $\gamma_1(G) = G$ and $\gamma_{n+1} = [\gamma_n(G), G]$. Here, if H, K are subgroups of G, then $[H, K]$ is the sugroup generated by the commutators $[h, k]$ with $h \in H, k \in K$. Thus $\gamma_2(G) = [G, G]$ and $\gamma_3(G) = [[G, G], G]$. A group is said to be *nilpotent of class* c if $\gamma_{c+1}(G) = 1$. So a group G is nilpotent of class 2 if $[[G, G], G] = 1$ which is equivalent to saying that $[G, G] \subseteq Z(G)$ or "commutators are central in G". More generally $\gamma_{c+1}(G) = 1$ means the $\gamma_c(G)$ is contained in the center of G. Some examples: if p is a prime, a finite p-group is always nilpotent (of some class) and a finite nilpotent group is a direct product of its Sylow p-subgroups. The multiplicative group of $n \times n$ upper triangular matrices over $\mathbb{Z}$ (or any commutative ring) with 1's on the diagonal is nilpotent of class $n - 1$.

1.2 About fundamental groups and covering spaces

Useful references for fundamental groups and covering spaces are the textbooks by Massey [6] and Rotman [8]. We here summarize (with almost no proofs) some of the basic facts we will need.

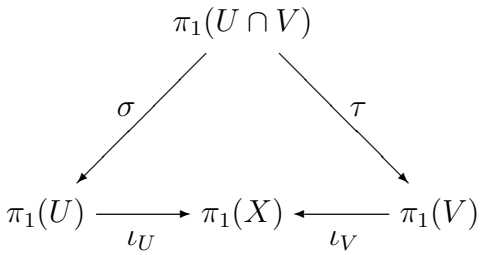
The spaces we will deal with will always be CW-complexes. These can

be thought of as cell complexes which are built inductively by attaching the boundaries of new standard n-cells to an existing complex. Thus a CW-complex X is a Hausdorff space which is the union of an increasing sequence of subspaces $X^0 \subset X^1 \subset X^2 \subset \cdots \subseteq X$. The initial *0-skeleton* X^0 is a discrete set of points. The *n-skeleton* X^n is obtained from X^{n-1} by attaching n-cells along their boundary (see [6] or [8] for details). X is endowed with the weak topology meaning that a subset A of X is closed if its intersection with each n-cell of X is closed. Since we are largely interested in fundamental groups, we will normally need only 2-dimensional CW-complexes, that is $X = X^2$.

We assume the reader is familiar with the fundamental group. If x_0 is a 0-cell of X, the fundamental group $\pi_1(X, x_0)$ is the collection of homotopy classes of loops beginning and ending at x_0 relative to this endpoint. The group multiplication is composition of paths meaning first go along one and then the other. Any path with endpoints in X^0 is homotopic relative to its endpoints to a path in X^1. An homotopy between two such paths in X_1 can be pushed into X^2 which is why we usually only need 2-dimensional complexes.

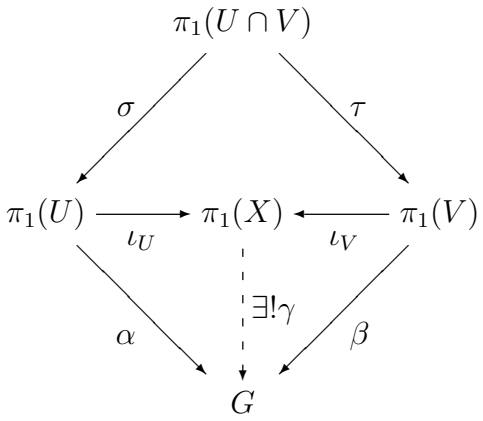
The main tool for computing the fundmental group is the following.

Theorem 1.1 (Seifert-vanKampen) *Let $X = U \cup V$ where U, V and $U \cap V$ are all non-empty, arc-wise connected open subsets of X. Choose a base point $x_0 \in U \cap V$ for all of their fundamental groups. Then the diagram of fundamental groups with maps induced by inclusions*



is a pushout. That is, if G is a group and $\alpha : \pi_1(U) \to G$ and $\beta : \pi_1(V) \to G$ are homomorphisms such that $\alpha\sigma = \beta\tau$ then there is a unique homomorphism

$\gamma : \pi_1(X) \to G$ *such that* $\alpha = \gamma\iota_U$ *and* $\beta = \gamma\iota_V$. *The diagram is:*



□

Usually there are ways around the requirement that U and V be open and it is enough to assume they are sub-complexes. Here is the special case which is of most interest to us.

Corollary 1.2 *Suppose that the CW-complex X is obtained from the connected CW-complex Y attaching a single 2-cell by identifying its boundary with the loop $\lambda \subseteq Y^1$ based at x_0. Then $\pi_1(X, x_0)$ is isomorphic to the quotient group of $\pi_1(Y, x_0)$ by the normal closure of the element $[\lambda] \in \pi_1(Y, x_0)$.*

Proof: Think of the 2-cell as the unit ball B^2 in $\mathbb{R}^2$. Let V be the image of the interior of B^2 in X and let U be Y union the image of B^2minus the origin. Now $U \cap V$ is V minus the image of the origin and so is topologically an open annulus with $\pi_1(U \cap V) \cong \mathbb{Z}$. Since V is topologically an open disk we have $\pi_1(v) = 1$. Hence, by the Seifert-vanKampen Theorem, $\pi_1(X)$ is the quotient of $\pi_1(U)$ by the normal closure of a loop corresponding to the generator of $\pi_1(U \cap V)$. But Y is a deformation retract of U and under the retraction the annular region $U \cap V$ maps simply onto λ. □

We now turn to covering spaces. For convenience we will assume all spaces are arc-wise connected (path connected) and locally arc-wise connected unless otherwise specified. Since we have in mind CW-complexes the local topology is always well behaved for our applications.

Let X be (path connected) topological space. A *covering space* is a path connected space $\tilde{X}$ together with a continuous mapping $p : \tilde{X} \to X$ such that for each $x \in X$ there is an open neighborhood $U = U_x$ of x that is *evenly*

covered by p, that is, $p^{-1}(U)$ is a disjoint union of open sets S_i called *sheets* such that $p|_{S_i} : S_i \to U$ is a homeomorphism for each i. The map p is often called the *covering projection*. If $x_0 \in X$ the discrete set of points $p^{-1}(x_0)$ is called the *fiber (or fibre)* over x_0.

By a *path* in a space X we mean a continuous map $\lambda : I \to X$ where $I = [0, 1]$, the unit interval. Some important properties of covering spaces are the following.

Theorem 1.3 (Unique Path Lifting) *Let $p : \tilde{X} \to X$ be a covering space and let $\tilde{x}_0 \in \tilde{X}$ and $x_0 \in X$ be such that $p(\tilde{x}_0) = x_0$. Then for any path $\lambda : I \to X$ with initial point x_0 there is a unique path $\tilde{\lambda} : I \to \tilde{X}$ with initial point $\tilde{x}_0$ such that $p\tilde{\lambda} = \lambda$.*

Theorem 1.4 (Covering Homotopy Lemma) *Let $p : \tilde{X} \to X$ be a covering space and let $\sigma, \tau : I \to \tilde{X}$ be two paths with the same initial point. If their images $p\sigma$ and $p\tau$ are homotopic (rel endpoints) in X, then σ and τ are homotopic (rel endpoints) in $\tilde{X}$ by a homotopy which projects to the one in X. In particular, σ and τ have the same end points.*

Corollary 1.5 *Let $p : \tilde{X} \to X$ be a covering space and let $\tilde{x}_0 \in \tilde{X}$ and $x_0 \in X$ be such that $p(\tilde{x}_0) = x_0$. Then the induced map*

$$p_* : \pi_1(\tilde{X}, \tilde{x}_0) \to \pi_1(X, x_0)$$

is a monomorphism. Thus a closed loop λ starting at x_0 in X lifts to a closed loop starting at $\tilde{x}_0$ if and only if $[\lambda]$ lies in the subgroup $p_*(\pi_1(\tilde{X}, \tilde{x}_0))$ of $\pi_1(X, x_0)$.

Let $Z = p^{-1}(x_0)$ be the fibre over a point $x_0 \in X$. If $g \in \pi_1(X, x_0)$, for each $z \in Z$ we define $zg \in Z$ as follows: g is represented by some closed loop $\lambda : I \to X$ at x_0. λ lifts uniquely to a path $\tilde{\lambda} : I \to \tilde{X}$ starting at z and ending at some point $\tilde{\lambda}(1)$ in Z. We define $zg = \tilde{\lambda}(1)$. One can check using the above results that this definition is independent of the choice of λ representing g. Moreover this defines a group action of $\pi_1(X, x_0)$ on the fiber over x_0, that is, $z1 = z$ and $(zg_1)g_2 = z(g_1g_2)$. In terms of this action we have the following.

Theorem 1.6 *Let $p : \tilde{X} \to X$ be a covering space and let $Z = p^{-1}(x_0)$ be the fibre over a point $x_0 \in X$. Then*

1. *$\pi_1(X, x_0)$ acts transitively on the fibre Z.*

2. *If $z \in Z$ then the stabilizer of z is $p_*(\pi_1(\tilde{X}, z))$.*

3. *The elements of Z are in one-one correspondence with the right cosets of $p_*(\pi_1(\tilde{X}, z))$ in $\pi_1(X, x_0)$.*

Consequently the fibres over any two points of X have the same number of elements. □

If $z_1, z_2 \in Z = p^{-1}(x_0)$ then the images $p_*(\pi_1(\tilde{X}, z_1))$ and $p_*(\pi_1(\tilde{X}, z_2))$ are conjugate subgroups of $\pi_1(X, x_0)$. To see this choose a path λ from z_1 to z_2 in $\tilde{X}$. If σ_2 is a closed loop at starting at z_2, then $\lambda\sigma_2\lambda^{-1}$ is a closed loop starting at z_1. Moreover every closed loop starting at z_1 is homotopic to one of this form (hint: insert $\lambda\lambda^{-1}$). Hence

$$[p\lambda]p_*(\pi_1(\tilde{X}, z_2))[p\lambda]^{-1} = p_*(\pi_1(\tilde{X}, z_1)).$$

One can also check that any conjugate subgroup of a given $p_*(\pi_1(\tilde{X}, z_1))$ arises in this way by lifting the conjugating element.

A covering space $p : \tilde{X} \to X$ is said to be *regular* if $p_*(\pi_1(\tilde{X}, \tilde{x}_0))$ is a normal subgroup of $\pi_1(X, x_0)$. Note that this is independent of the choice of base point in the fibre over x_0. If $\pi_1(\tilde{X}, \tilde{x}_0) = 1$, that is if $\tilde{X}$ is simply connected, then $\tilde{X}$ is called the *universal covering space* of X (it is unique). Of course the universal covering space is regular.

Unique path lifting has an important generalization which gives a criteria for lifting arbitrary maps to a covering space.

Theorem 1.7 (Lifting Criterion) *Let $p : \tilde{X} \to X$ be a covering space and let Y be a connected, locally path connected space. Suppose $y_0 \in Y$, $\tilde{x_0} \in \tilde{X}$ and $x_0 = p(\tilde{x}_0)$. Given a map of pointed spaces $\phi : (Y, y_0) \to (X, x_0)$, there exists a lifting $\tilde{\phi} : (Y, y_0) \to (\tilde{X}, \tilde{x}_0)$ with $\phi = p \circ \tilde{\phi}$ if and only if*

$$\phi_*(\pi_1(Y, y_0)) \subseteq p_*(\pi_1(\tilde{X}, \tilde{x}_0)).$$

If this lifting exists, it is unique.

Let $p : \tilde{X} \to X$ be a covering space. A homeomorphism $\phi : \tilde{X} \to \tilde{X}$ is said to be a *covering transformation* if $p \circ \phi = p$. The collection of covering transformations is clearly a group under composition. It is easy to check that

covering transformations are compatible with the action of $\pi_1(X, x_0)$ on the fibre $Z = p^{-1}(x_0)$ in the following sense: $\phi(zg) = \phi(z)g$ for $g \in \pi_1(X, x_0)$ and $z \in Z$.

Applying the above Lifting Criterion, one can show the following.

Corollary 1.8 *Let $p : \tilde{X} \to X$ and $z_1, z_2 \in Z = p^{-1}(x_0)$. Then there is a covering transformation ϕ such that $\phi(z_1) = \phi(z_2)$ if and only if*

$$p_*(\pi_1(\tilde{X}, z_1)) = p_*(\pi_1(\tilde{X}, z_2)).$$

Hence the group of covering transformations acts transitively on $Z = p^{-1}(x_0)$ if and only if $p : \tilde{X} \to X$ is a regular covering space.

Moreover, if $p : \tilde{X} \to X$ is a regular covering space then the group of covering transformation is isomorphic to the quotient group

$$\pi_1(X, x_0)/p_*(\pi_1(\tilde{X}, z))$$

for any $z \in p^{-1}(x_0)$. In particular, when $p : \tilde{X} \to X$ is the universal covering space, then the group of covering transformations is isomorphic to $\pi_1(X, x_0)$.

This bring us to the existence problem for covering spaces. On needs some local conditions on the topology of the space X. These conditions do hold for X a CW-complex, so we state this vaguely as the following.

Theorem 1.9 *Under suitable local conditions on X, given any conjugacy class of subgroups of $\pi_1(X, x_0)$ there exists a covering space $p : \tilde{X} \to X$ corresponding to the given conjugacy class. In particular, any subgroup of $\pi_1(X, x_0)$ is the image $p_*(\pi_1(\tilde{X}, z))$ for some choice of $z \in p^{-1}(x_0)$.*

This completes our sketch of covering space theory.

Chapter 2

Free groups and presentations

In introductory courses on abstract algebra one is likely to encounter the dihedral group D_3 consisting of the rigid motions of an equilateral triangle onto itself. The group has order 6 and is conveniently described by giving two generators which correspond to rotation through 120° and flipping about a central axis. These operations have orders 3 and 2 respectively and the group D_3 is described by the *presentation*

$$D_3 = \langle a, b \mid a^2 = 1, b^3 = 1, a^{-1}ba = b^{-1} \rangle$$

some equivalent version. Here the symbols a and b are called *generators* and the equations they are subjected to are called *defining relations*.

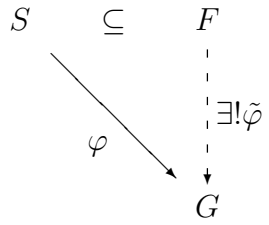
Combinatorial group theory is concerned with groups described by generators and defining relations and also with certain natural constructions for making new groups out of groups we already have in hand. Also one would like to (1) say something about their structure, their subgroups, and various properties they might enjoy and to (2) find algorithms for answering some natural questions about them and their elements. Combinatorial group theory has many connections with algebraic and geometric topology which have provided both motivation and methods for studying groups in this manner.

In order to begin our study of presentations we first need to discuss free groups. We will then introduce presentations in terms of generators and relations more formally and indicate their connection with algebraic topology via the fundamental group.

2.1 Free groups

In most algebraic contexts a *free* object is an object which has a *free basis*. The pattern for groups is typical of this sort of definition.

A subset S of a group F is said to be a *free basis* for F if, for every (set) function $\varphi : S \to G$ from the set S to a group G can be extended uniquely to a homomorphism $\tilde{\varphi} : F \to G$ so that $\tilde{\varphi}(s) = \varphi(s)$ for every $s \in S$.



A group F is said to be a *free group* if there is some subset which is a free basis for F.

Consider the infinite cyclic group C (written multiplicatively) which consists of all powers of a single element a,so

$$C = \{\ldots, a^{-2}, a^{-1}, 1 = a^0, a = a^1, a^2, a^3, \ldots\}$$

and multiplication is defined by $a^i \cdot a^j = a^{i+j}$ for $i, j \in \mathbb{Z}$. Then C is a free group with free basis the set with a single element $S = \{a\}$. For if $\varphi : S \to G$ is any function, say $\varphi(a) = g \in G$ then φ extends to a homomorphism $\tilde{\varphi} : C \to G$ by defining $\tilde{\varphi}(a^i) = g^i$. Moreover it is clear this is the only way to extend φ to a homomorphism. Notice that C also has another free basis, namely the singleton set $\{a^{-1}\}$ and that these two are the only free bases for C.

Similarly, the additive group of integers $\mathbb{Z}$ (which is of course isomorphic to C) is also a free group with either of the singleton sets $\{1\}$ or $\{-1\}$ as a free basis. As a slightly more exotic example, we note that the trivial group consisting of $\{1\}$ alone is a free group with the empty subset as free basis.

Beyond these familiar examples we have to do something to prove that free groups exist. In fact we can make a free group with any given set S as a basis in the way described below.

Theorem 2.1 *If S is any set, there is a free group F_S having S as a free basis.*

Suppose we are given a set S which we think of as just a set of symbols (S need not be countable or ordered). By a *word* on S we mean an expression of the form $a_1^{\epsilon_1}a_2^{\epsilon_2}\dots a_k^{\epsilon_k}$ where $\epsilon_i = \pm 1$ and $a_i \in S$ (not necessarily distinct symbols). That is, a word is a string of elements of S with exponents either $+1$ or -1. The intention is that a^{-1} and a^{+1} are to be mutually inverse group elements and one usually identifies a^{+1} with a.

Sometimes it is convenient to adopt a slightly different definition. Let $S^{-1} = \{a^{-1} \mid a \in S\}$ be thought of as the set of inverse symbols for the symbols of S. Then by a word on S we would mean just a string of symbols from $S \cup S^{-1}$. This rarely causes confusion and we use whichever version is convenient.

A word on S is said to be *(freely) reduced* if it does not contain a subword (consecutive substring) of the form aa^{-1} or of the form $a^{-1}a$; such substrings are called *inverse pairs* of generators. If a word w contains an inverse pair, say $w \equiv ua^{-1}av$ where u and v are subwords, then in any group containing w one must have $w = uv$ (here $\equiv$ means identical as words and $=$ means equal as group elements). In removing an inverse pair, the two symbols a and a^{-1} are said to *cancel* .

If we start with any word w by successively removing (canceling) inverse pairs we arrive in a finite number of steps at a freely reduced word w' which we call *a reduced form* of w. There can be several different ways to proceed with the cancellations, but one can show that end result w' does not depend on the order in which the inverse pairs are removed. This is a basic but non-trivial fact which certainly requires proof. It allows us to call w' *the* reduced form or free reduction of w and we write this as $w' = \rho(w)$. Here is the required result:

Lemma 2.2 *There is only one reduced form of a given word w on S.*

Proof: Let $x \in S \cup S^{-1}$ and write x^{-1} for the corresponding inverse symbol (so if $a \in S$ then a is the inverse symbol of $a^{-1} \in S^{-1}$). We use induction on the length of w as a string of symbols. If w is reduced there is nothing to show. So suppose $w \equiv uxx^{-1}v$ and focus on this particular occurrence of an inverse pair which we distinguish by underlining as in $ux\underline{x^{-1}}v$. If we can show that every reduced form w' of w can be obtained by canceling this occurrence first, then the lemma will follow by induction on the shorter word uv thus obtained.

Let w' be a reduced form of w. We know that w' is obtained by some sequence of cancellations. First suppose that the pair we are focusing on

xx^{-1} is cancelled at some step in this sequence. Then we can clearly rearrange the order so that this particular pair is canceled first. So this case is settled. Now xx^{-1} cannot remain in w' so at least one of these two symbols must be canceled at some step. The first cancellation must then look like $u_1x^{-1}\underline{xx^{-1}}v_1 \rightarrow u_1 \not{x}^{-1}\underline{\not{x}x^{-1}}v_1$ or $u_1\underline{xx^{-1}}xv_1 \rightarrow u_1\underline{x\not{x}^{-1}}\not{v}_1$. But in either case, the word obtained by cancellation is the same as that obtained by cancelling the original pair. So we may cancel the original pair at this stage instead. Hence we are back in the first case and the lemma is proved. □

We are now ready to define the free group F_S with free basis S. The elements of F_S are the freely reduced words on S (including the empty word which we denote by 1). Multiplication in F_S is defined by $u \cdot v = \rho(uv)$, that is the product is the free reduction of one word followed by the other (concatenation). One now must check the axioms for a group. The identity is the empty word 1 and the inverse of $a_1^{\epsilon_1}a_2^{\epsilon_2}\dots a_k^{\epsilon_k}$ is $a_k^{-\epsilon_k}a_{k-1}^{-\epsilon_{k-1}}\dots a_1^{-\epsilon_1}$. The associative law follows from the fact that the reduction of a word is independent of the order in which inverse pairs are removed.

To see that F_S is in fact free, consider any function $\varphi : S \to G$ where G is a group. We define

$$\tilde{\varphi}(a_1^{\epsilon_1}a_2^{\epsilon_2}\dots a_k^{\epsilon_k}) = \varphi(a_1)^{\epsilon_1}\varphi(a_2)^{\epsilon_2}\dots\varphi(a_k)^{\epsilon_k}.$$

This map $\tilde{\varphi}$ is easily checked to be a homomorphism. Moreover, it is clear the definition is forced upon us so that it is the unique homomorphism extending the function φ.

Here is one consequence of the above.

Corollary 2.3 *Every group is a quotient group of a free group. Thus if G is a group there is a free group F and a normal subgroup N such that $G \cong F/N$.*

To see this, given a group G we think of G as a set (forgetting its group operation for the moment) and form the free group F_G as above. Then the identity function $\varphi : G \to G$ from G as a set to G as a group (remember the group operation) extends uniquely to a homomorphism $\tilde{\varphi} : F_G \to G$. The homomorphism $\tilde{\varphi}$ is clearly surjective (since φ is a bijection), so $G \cong F_G/N$ where $N = \ker\tilde{\varphi}$.

In general, if S is a subset of a group G, the subgroup denoted by $\langle S\rangle$ (called the *subgroup generated by* S) is the image of the extension $\tilde{\varphi} : F_S \to G$ of the inclusion function. That is, the subgroup $\langle S\rangle$ generated by S consists

of those elements of G which are equal to some product of elements in S and their inverses. In particular, if $\langle S \rangle = G$, we say that S generates G.

The *rank* of a the free group F_S is the cardinality of the set S of generators. One can show this is an invariant of the free group F_S, that is if T is another free basis for F_S then S and T have the same cardinality (number of elements). If G is any group, then the *rank* of G is the cardinality of the smallest set of generators for G, that is the rank of the smallest free group F for which there is a surjection $\varphi : F \to G$. It is often difficult to determine the rank of a group (other than a free group).

Notation: Although we somewhat carefully distinguished between φ and $\tilde{\varphi}$ in the above, we often adopt the notational convention that the extending homomorphism is also denoted φ. This is common practice and usually causes no confusion. In some cases to be more precise we may resort to the original notation. We will also use notation such as $w \equiv uv$ to mean that the word w is identical to the word u followed by the word v. The equation $w = uv$ or $w =_G uv$ means that w is equal in some appropriate group G to the product of u and v. Also, if $x, y \in G$ their *commutator* is the element denoted $[x, y]$ which is by definition $[x, y] = x^{-1}y^{-1}xy$. As usual, two such elements *commute* if $xy =_G yx$ which is equivalent to $[x, y] =_G 1$.

Terminology We use the terms *surjective*,*epic* and *onto* for functions interchangeably. An *epimorphism* is a surjective homomorphism. Likewise, the terms *injective monic* and *one-one* are interchangeable, and a *monomorphism* is an injective homomorphism.

Theorem 2.4 (Characterization of freeness) *Let G be a group, and S a subset of G. Then G is free with basis S if and only if the following both hold:*

1. *S generates G; and*
2. *If w is a word on S and $w =_G 1$, then w is not freely reduced, that is w must contain an inverse pair.*

These conditions imply that every element of G is equal to a unique freely reduced word in G. For if $u =_G v$ and u and v are freely reduced, then uv^{-1} contains an inverse pair. Hence the last symbol of u is the same as the last symbol of v. So inductively u and v must be identical. Thus different freely reduced words represent different elements of G. Hence the obvious extension of the identity on S is an isomorphism from F_S onto G.

Recall that two elements say g and h in a group G are *conjugate* if there exists some $x \in G$ such that $g = x^{-1}hx$. A word w in the free group F_S with basis S is said to be *cyclically reduced* if every cyclic permutation of w is (freely) reduced. If w is (freely) reduced, cyclically reduced is equivalent to saying the first symbol of w is not the inverse of the last symbol of w.

Exercise 2.1 *Show that two cyclically reduced words u and v in a free group F_S are conjugate if and only if one is a cyclic permutation of the other.*

Exercise 2.2 *Show that in a free group two elements commute if and only if they are powers of a common element, that is $uv = vu$ implies that $u = w^m$ and $v = w^n$ for some w and $m, n \in \mathbb{Z}$.*

2.2 Presentations by generators and relations

As in the case of D_3 above, we want to describe groups by writing down some elements which generate the group and then imposing some equations on them. Such a piece of notation might look like

$$G = \langle a_1, a_2, \ldots \mid u_1 = v_1, u_2 = v_2, \ldots \rangle$$

where the a_i are symbols and the u_j and v_j are certain words in the a_i. (While we habitually use this sort of notation, it is not necessary for the set of generators to be countable or ordered.)

In any group, $u = v$ if and only if $uv^{-1} = 1$ so we can always write our presentations in the equivalent form

$$G = \langle a_1, a_2, \ldots \mid r_1 = 1, r_2 = 1, \ldots \rangle$$

where $r_i = u_i v_i^{-1}$. Although we will use presentations with equations of the form $u = v$, for our theoretical discussion it is convenient to assume our defining equations are of the form $r = 1$.

To be a bit more formal for a moment, a *presentation* $P = \langle S \mid D \rangle$ is a pair consisting of a set S called *generators* and a set D of words on S called *(defining) relators*. The *group presented by* P, denoted $gp(P)$ is the group F_S/N_D where F_S is the free group with free basis S and N_D is the normal closure of D in F_S, that is the smallest normal subgroup containing D. Thus if $r \in D$, then $r \in N_D$ and so $r =_{gp(P)} 1$. If $G = gp(P)$ we often

abuse notation and write $G = \langle S \mid D \rangle$ when it is not necessary to distinguish between a group and a description of that group.

A presentation $P = \langle S \mid D \rangle$ is said to be *finitely generated* if S is a finite set and to be *finitely related* if D is a finite set of words. If both S and D are finite, P is said to be a *finite presentation*. If $S = \{a_1, a_2, \ldots\}$ and $D = \{r_1, r_2, \ldots\}$ we use either the notation

$$P = \langle a_1, a_2, \ldots \mid r_1, r_2, \ldots \rangle$$

in which case the r_i are called *relators*, or the notation

$$P = \langle a_1, a_2, \ldots \mid r_1 = 1, r_2 = 1, \ldots \rangle$$

in which case the equations $r_i = 1$ are called relations. Usually we also extend the latter to allow

$$P = \langle a_1, a_2, \ldots \mid u_1 = v_1, u_2 = v_2, \ldots \rangle$$

where $r_i = u_i v_i^{-1}$.

Here are a few more examples of presentations.

1. The infinite cyclic group C written multiplicatively with generator a has presentation $C = \langle a \mid \rangle$ with an empty set of defining relators. More generally the free group F_S with free basis S has a presentation $F_S = \langle S \mid \emptyset \rangle$.

2. The finite cyclic group C_n of order n has a presentation $C_n = \langle a \mid a^n = 1 \rangle$.

3. The free abelian group of rank 2 has a presentation $\langle a, b \mid ab = ba \rangle$ or equivalently $\langle a, b \mid aba^{-1}b^{-1} = 1 \rangle$. In this group, which is isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$, every element is equal to a unique word the form $a^i b^j$ where $i, j \in \mathbb{Z}$ and multiplication is addition of the corresponding exponents.

4. The dihedral group D_n of order $2n$ consisting of the rigid motions of the regular n-gon has presentation

$$D_n = \langle a, b \mid a^2 = 1, b^n = 1, a^{-1}ba = b^{-1} \rangle.$$

5. The presentation $P = \langle a, b \mid ababa = 1 \rangle$ turns out to be a presentation of the infinite cyclic group. This is not quite obvious, but we will show this below.

6. Any finite group G has a finite presentation. For generators we can take all the group elements, say $\{a_1, \ldots, a_n\}$, and for relations we take all the equations from the multiplication table (these have the form $a_i a_j = a_k$ and are n^2 in number.)

Of course when we write down some defining relations for a group there can be consequences we don't expect. For instance consider the group mentioned above with presentation $G = \langle a, b \mid ababa = 1 \rangle$. Now $baaba =_G 1$ since this is just a conjugate of the given relator. Multiplying this by the inverse of the relator we obtain

$$1 =_G (baaba)(ababa)^{-1} = baabaa^{-1}b^{-1}a^{-1}b^{-1}a^{-1} = bab^{-1}a^{-1}$$

and so $ab =_G ba$. It follows that G is an abelian group which may not have been apparent at first.

Exercise 2.3 *Let $G = \langle a, t \mid t^{-1}at = a^2 \rangle$. Show that every element of G is equal to a word of the form $t^n a^k t^{-m}$ where $n \geq 0$ and $m \geq 0$. Let N denote the normal closure in G of the element a. Show that N is generated by elements of the form $t^n a t^{-n}$, and that N is abelian.*

Exercise 2.4 *Let $G = \langle a, b \mid a^{-1}ba = b^2, b^{-1}ab = a^2 \rangle$. Show that $a =_G 1$ and $b =_G 1$ and conclude that this is a presentation of the trivial group.*

Suppose that $G = \langle S \mid D \rangle$ is a group given by a presentation. There is a sort of theoretical characterization of those words w such that $w =_G 1$, namely

Lemma 2.5 *Let $G = \langle S \mid D \rangle$ be a group given by a presentation. If w is any word in the generators of G, then $w =_G 1$ if and only if as an element of the free group F_S there is an equation*

$$w =_{F_S} u_1 r_1^{\epsilon_1} u_1^{-1} u_2 r_2^{\epsilon_2} u_2^{-1} \ldots u_k r_k^{\epsilon_k} u_k^{-1}$$

for some words $u_i \in F_S$, $r_i \in D$ and $\epsilon_i = \pm 1$.

To see this, one simply observes that the set of words equal to such expressions contains D and is closed under conjugation, multiplication and inversion. Hence it is N_D, the smallest normal subgroup containing D.

2.3 Dehn's fundamental problems

Suppose we are studying groups given by presentations. We would like to know about the existence and nature of algorithms which decide

- *local properties* – whether or not elements of a group have certain properties or relationships;
- *global properties* – whether or not groups as a whole possess certain properties or relationships.

Such questions are called *decision problems*. The groups in question are assumed to be given by finite presentations or in some other explicit manner.

Historically the following three fundamental decision problems formulated by Max Dehn in 1911 have played a central role:

word problem: Let G be a group given by a finite presentation. Does there exist an algorithm to determine of an arbitrary word w in the generators of G whether or not $w =_G 1$?

conjugacy problem: Let G be a group given by a finite presentation. Does there exist an algorithm to determine of an arbitrary pair of words u and v in the generators of G whether or not u and v define conjugate elements of G?

isomorphism problem: Does there exist an algorithm to determine of an arbitrary pair of finite presentations whether or not the groups they present are isomorphic?

The word and conjugacy problems are decision problems about local properties while the isomorphism problem is a decision problem about a global relationship.

Motivation for studying these questions can be found in algebraic topology. For one of the more interesting algebraic invariants of a topological space is its fundamental group. If a connected topological space T is reasonably nice, for instance if T is a finite complex, then its fundamental group $\pi_1(T)$ is finitely presented and a presentation can be found from any reasonable description of T. The word problem for $\pi_1(T)$ then corresponds to the problem of determining whether or not a closed loop in T is contractible. The conjugacy problem for $\pi_1(T)$ corresponds to the problem of determining whether or not two closed loops are freely homotopic (intuitively whether one

can be deformed into the other). Since homeomorphic spaces have isomorphic fundamental groups, a solution to the isomorphism problem would give a method for discriminating between spaces (the homeomorphism problem).

2.4 Homomorphisms

One useful aspect of having a presentation for a group is that it gives us a method of checking whether a proposed map between groups is a homomorphism. Suppose that we have a group given by a presentation, say $G = \langle S \mid D \rangle$, and that $\psi : S \to H$ is a function. We want to know whether ψ can be extended to a homomorphism from G to H. Now we do know that ψ extends uniquely to a homomorphism $\tilde{\psi} : F_S \to H$. The map $\tilde{\psi}$ is of course just a formal extension of ψ to all (freely reduced) words.

Recall that $G = F_S/N_D$ where N_D is the normal closure of D. The original ψ extends to a homomorphism if and only if $N_D \subseteq \ker \tilde{\psi}$. Hence ψ extends to a homomorphism if and only if $\tilde{\psi}(r) =_H 1$ for all $r \in D$. We record this observation as

Theorem 2.6 *Let $G = \langle S \mid D \rangle$ and suppose that $\psi : S \to H$ is a function. Then ψ extends to a homomorphism $\psi : G \to H$ if and only if $\tilde{\psi}(r) =_H 1$ for all $r \in D$ where $\tilde{\psi} : F_S \to H$ is the formal extension of ψ to all words.*

As an illustration, consider the group $G = \langle a, b \mid ababa = 1 \rangle$ and the infinite cyclic group $C = \langle t \mid \rangle$ with generator t. Consider the function ψ defined by $\psi(a) = t^{-2}$ and $\psi(b) = t^3$. We ask whether ψ extends to a homomorphism. We compute that

$$\tilde{\psi}(ababa) = t^{-2}t^3t^{-2}t^3t^{-2} = t^{6-6} =_C 1$$

and so we can conclude that, yes, ψ extends to a homomorphism.

Continuing this example, the function $\varphi(t) = ab$ extends uniquely to a homomorphism from $\varphi : C \to G$ since C is free with basis t (or technically $\{t\}$). Now

$$(\psi \circ \varphi)(t) = \psi(\varphi(t)) = \psi(ab) = t^{-2}t^3 = t$$

and

$$(\varphi \circ \psi)(a) = \varphi(\psi(a)) = \varphi(t^{-2}) = (ab)^{-2} = a$$
$$(\varphi \circ \psi)(b) = \varphi(\psi(a)) = \varphi(t^3) = (ab)^3 = b$$

where last equations are follow easily from the relation $ababa = 1$, for instance $(ab)^3 = ababab = b$ and $a = (ab)^{-2}ababa$. It follows that the homomorphisms are mutually inverse and hence are both isomorphism. So G is isomorphic to the infinite cyclic group as we claimed earlier.

Here is another illustration using equation notation. Consider the group with presentation $G = \langle a, t \mid t^{-1}at = a^2 \rangle$. We ask whether the function ψ defined by $\psi(a) = a^2$ and $\psi(t) = t$ extends to a homomorphism from G to itself. To check this we simply compute $\tilde{\psi}$ of both sides of the defining relation and show they are equal.

$$\tilde{\psi}(t^{-1}at) = \psi(t)^{-1}\psi(a)\psi(t) = t^{-1}a^2t = (t^{-1}at)^2 = (a^2)^2 = \tilde{\psi}(a^2).$$

Hence ψ defines a homomorphism from G to itself which we again denote by ψ. Observe that this homomorphism is surjective. For its image contains t and a^2 and hence also a since $a = ta^2t^{-1}$ is a consequence of the defining relation.

Next consider the function φ defined by $\varphi(a) = tat^{-1}$ and $\varphi(t) = t$. One can check that φ extends to a homomorphism and that ψ and φ are mutually inverse. Hence they are both automorphisms of the group G.

This group is actually one of a family of interesting groups having presentations

$$G_{m,n} = \langle a, t \mid t^{-1}a^mt = a^n \rangle$$

which are known as *Baumslag-Solitar groups*. In this notation the above group is $G_{1,2}$.

Exercise 2.5 *Consider the group $G_{1,2} = \langle a, t \mid t^{-1}at = a^2 \rangle$. Define a function from given the generators of $G_{1,2}$ to the group $GL(2, \mathbb{Q})$ of 2×2 matrices with rational coefficients by*

$$a \mapsto \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad t \mapsto \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}.$$

Show that this function defines a homomorphism $\varphi : G_{1,2} \to GL(2, \mathbb{Q})$. Also show that φ is injective and hence $G_{1,2}$ is isomorphic to a group of matrices. (Hint: the results of an earlier exercise may be helpful.)

Again consider a presentation of a group $G = \langle S \mid D \rangle$ and let E be any set of words in F_S. Then the group presented by $\langle S \mid D \cup E \rangle$ is $F_S/N_{D \cup E}$ which is a quotient group of G. The quotient homomorphism from G to

$\langle S \mid D \cup E \rangle$ is defined by just sending the symbols in S to themselves. Moreover if $\theta : G \to H$ is surjective, then $H \cong F_S/N_{D\cup E}$ for some suitable set of words E; simply take E to be the set of words in the kernel of θ. We record this as follows:

Theorem 2.7 (von Dyck) *Let $G = \langle S \mid D \rangle$ be a presentation of a group and let E be any set of words in F_S. Then the group presented by $\langle S \mid D \cup E \rangle$ is a quotient group of G with quotient homomorphism defined by the identity map on S. Moreover, every quotient group of G is isomorphic to a quotient of this form.*

Suppose that we have a presentation

$$G = \langle a_1, a_2, \ldots \mid r_1 = 1, r_2 = 1, \ldots \rangle.$$

Then we can present the *abelianization* $G/[G,G]$ of G (which is the largest abelian quotient), by simply adding all the relations $a_i a_j = a_j a_i$, thus

$$G/[G,G] \cong \langle a_1, a_2, \ldots \mid a_i a_j = a_j a_i (\forall i, j), r_1 = 1, r_2 = 1, \ldots \rangle.$$

In case the given presentation is finite, one can then use the resulting presentation to compute the decomposition of $G/[G,G]$ as a direct sum of cyclic groups.

Exercise 2.6 *Suppose that $G = \langle a_1, \ldots a_n \mid r_1 = 1, \ldots r_m = 1 \rangle$ and that $H = \langle a_1, \ldots a_n \mid r_1 = 1, \ldots r_m = 1, r_{m+1} = 1 \rangle$. If G is finite, does the order of H divide the order of G ? Justify your answer.*

Exercise 2.7 *Consider the group with presentation $G = \langle a, b, c \mid c^{-1}ac = b, c^{-1}bc = a, c^4 = 1 \rangle$. Determine whether or not G is infinite.*

2.5 Presentations and fundamental groups

Consider a group G given by a presentation, say

$$G = \langle a_1, a_2, \ldots, a_n \mid r_1 = 1, r_2 = 1, \ldots, r_m = 1 \rangle.$$

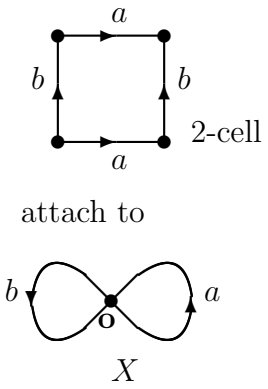
Such a presentation corresponds in a standard way to a 2-dimensional complex Y whose fundamental group is isomorphic to G, that is $G \cong \pi_1(Y, \mathbf{o})$.

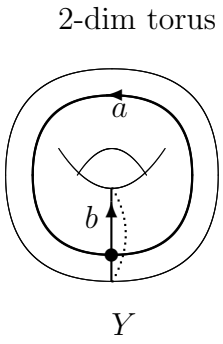
We review the features of this correspondence. Most of what we say can also be done for presentations with arbitrarily many generators and relations by resorting to CW-complexes. There are no substantial difficulties, just distracting topological niceties.

We start with a single 0-cell which we label as $\mathbf{o}$. For each a_i we take an oriented 1-cell, identify its ends with $\mathbf{o}$ and label its positive direction by the generator a_i. This gives us a space X consisting of a bouquet or wedge of n loops at $\mathbf{o}$. The fundamental group of this space is a free group with basis the (homotopy classes of the) closed loops labeled by the a_i. So we identify $\pi_1(X, \mathbf{o})$ with the free group $F = \langle a_1, a_2, \ldots, a_n \mid \rangle$.

We now add 2-cells corresponding to the defining relations. For each r_i take a 2-cell, thought of as a disk, and subdivide and label its boundary according to r_i. So if $r_i = a_{j_1}^{\epsilon_1} \ldots a_{j_k}^{\epsilon_k}$ we subdivide the boundary of the 2-cell into k 1-cells with orientation and labeling chosen so that reading in the counter-clockwise direction the label on the boundary is just r_i. We then attach or glue each of these 2-cells to X by identifying the oriented edges in the boundary labeled by a_i with the corresponding loop in X and the 0-cells in the boundary with $\mathbf{o}$. Call the resulting space Y. So Y has a single 0-cell $\mathbf{o}$, a 1-cell for each generator a_i and a 2-cell for each relation r_i. The Seifert-vanKampen Theorem tells us that the fundamental group of Y is just G, that is $G \cong \pi_1(Y, \mathbf{o})$.

As an illustration consider the group $G = \mathbb{Z} \oplus \mathbb{Z}$ which has presentation $G = \langle a, b \mid aba^{-1}b^{-1} = 1 \rangle$. In this case X consists of 2 loops at $\mathbf{o}$ labeled by a and b. There is a single 1-cell with 4 boundary edges labeled as shown. The space Y is then the 2-dimensional torus which is homeomorphic to $S^1 \times S^1$. As in general, $G \cong \pi_1(Y, \mathbf{o})$.





The universal covering space $\tilde{Y}$ of Y in this example can be identified with the Euclidean plane $\mathbb{R}^2$. With suitable identifications, the 0-skeleton of $\tilde{Y}$ consists of the lattice of points in the plane with integer coordinates and the 1-skeleton $\tilde{Y}^1$ of $\tilde{Y}$ is the grid of horizontal and vertical lines with one integer coordinate.

In the general situation, fix a 0-cell $\tilde{\mathbf{o}}$ in the universal cover $\tilde{Y}$ of Y as a base point. Any word w in the generators of G can be thought of as a closed path λ_w starting in X at $\mathbf{o}$. Now X is just the 1-skeleton of Y, so as a path in Y we can lift λ_w to a unique path $\tilde{\lambda}_w$ in the 1-skeleton $\tilde{Y}^1$ of $\tilde{Y}$ starting at $\tilde{\mathbf{o}}$. Now $\tilde{\lambda}_w$ is a closed path if and only if w belongs to the subgroup of G corresponding to $\tilde{Y}$, that is, if and only if $w =_G 1$. This means that the 1-skeleton $\tilde{Y}^1$ of $\tilde{Y}$ is just the covering space of X corresponding to the normal subgroup N_D of $F_S = \pi_1(X, \mathbf{o})$.

By a graph we mean a 1-dimensional CW-complex. The graph $\tilde{Y}^1$ we have just been considering is called the *Cayley graph* of G (with respect to the given generators) and is usually denoted $\Gamma = \Gamma_{G,S}$ where the decorative subscripts are used when necessary to show the group and generating set. (Notice that Γ does not depend on the particular defining relations used, but rather on the whole normal subgroup N_D.) The Cayley graph is of fundamental importance in the area known as geometric group theory.

2.6 Tietze transformations

There are some alterations one can make to a presentation which result in presentations of a group isomorphic to the original. These are called *Tietze*

transformations and we describe them as follows:

$T1$ (add consequences) If in F_S we have a collection of words $E \subseteq N_D$ (and hence $N_D = N_{D\cup E}$), replace $\langle S \mid D\rangle$ by $\langle S \mid D \cup E\rangle$.

$T1^{-1}$ (remove redundancies) If in F_S we have a collection of words E such that $N_{D\cup E} = N_D$ (and hence the relators in E are redundant), replace $\langle S \mid D \cup E\rangle$ by $\langle S \mid D\rangle$.

$T2$ (introduce abbreviations) If T is a collection of symbols disjoint from S and $\{u_t \mid t \in T\}$ is a set of words on S, replace $\langle S \mid D\rangle$ by $\langle S \cup T \mid D \cup \{t^{-1}u_t \mid t \in T\}\rangle$. (The effect here is to introduce abbreviations of the form $t = u_t$ where u_t is a word on the S symbols.)

$T2^{-1}$ (remove abbreviations) If T is a collection of symbols disjoint from S and $\{u_t \mid t \in T\}$ is a set of words on S and the words in D do not contain T symbols, replace $\langle S \cup T \mid D \cup \{t^{-1}u_t \mid t \in T\}\rangle$ by $\langle S \mid D\rangle$. (The effect is to remove abbreviations.)

If only one consequence is introduced (removed) or one abbreviation is introduced (removed) with a transformation, we term the move a *single step* transformation. In general we need to allow more the more general operations, but in finite situations a sequence of single step transformations suffice.

Theorem 2.8 *Suppose that the groups presented by the two presentations $\langle S \mid D\rangle$ and $\langle T \mid E\rangle$ are isomorphic. Then there is a sequence of Tietze transformations leading from one of these to the other. If these presentations are both finite the sequence can be taken to be a finite number of single step transformations.*

Proof: Roughly the proof proceeds as follows. Use the isomorphisms between the two groups to expand each of the given presentations to a common presentation containing both of the given presentations. Then since the inverse of a transformation is also a transformation, the result follows. Here are the details:

We write the two presentations as

$$\langle a_1, a_2, \ldots \mid r_1(\vec{a}) = 1, r_2(\vec{a}) = 1, \ldots\rangle$$

and

$$\langle b_1, b_2, \ldots \mid q_1(\vec{b}) = 1, q_2(\vec{b}) = 1, \ldots\rangle$$

where the notation $r_1(\vec{a})$ means that r_1 is a word on the symbols $a_1, a_2, \ldots$ We may suppose the isomorphisms are induced from maps of the free groups F_S and F_T defined by

$$\varphi(a_i) = u_i(\vec{b}) \text{ for } i = 1, 2, \ldots \text{ and } \psi(b_j) = v_j(\vec{a}) \text{ for } j = 1, 2, \ldots.$$

Starting with the first presentation, we apply transformations of type T2 to obtaint he presentation

$$\langle a_1, a_2, \ldots, b_1, b_2, \ldots \mid r_1(\vec{a}) = 1, r_2(\vec{a}) = 1, \ldots, b_1 = v_1(\vec{a}), b_2 = v_2(\vec{a}), \ldots \rangle.$$

Since ψ is a homomorphism each $q_k(\vec{b})$ is equal to 1 in this group and so using transformations T1 they can be added as relators to obtain the following presentation for G:

$$\langle a_1, a_2, \ldots, b_1, b_2, \ldots \mid r_1(\vec{a}) = 1, r_2(\vec{a}) = 1, \ldots, b_1 = v_1(\vec{a}), b_2 = v_2(\vec{a}), \ldots,$$

$$q_1(\vec{b}) = 1, q_2(\vec{b}) = 1, \ldots \rangle.$$

Now $\varphi \circ \psi(a_i) = a_i$ so the relations $a_i = u_i(b_1, b_2, \ldots)$ are consequences of our current set of relations so we can apply transformations T1 to obtain the presentation for G

$$\langle a_1, a_2, \ldots, b_1, b_2, \ldots \mid r_1(\vec{a}) = 1, r_2(\vec{a}) = 1, \ldots, b_1 = v_1(\vec{a}), b_2 = v_2(\vec{a}), \ldots,$$

$$q_1(\vec{b}) = 1, q_2(\vec{b}) = 1, \ldots, a_1 = u_1(\vec{b}), a_2 = u_2(\vec{b}), \ldots \rangle.$$

Now this presentation obtained starting from the first presentation is symmetric in the sense that it could also be obtained from the second presentation. Since the inverse of a Tietze transformation is again a Tietze transformation, this proves the theorem. □

An analogous result is actually true for a large class of algebraic systems and a similar proof can be used.

Exercise 2.8 *Use Tietze transformations, carefully showing every step, to transform the presentation*

$$\langle a, b, c \mid c^{-1}ac = b, c^{-1}bc = a, c^2 = 1 \rangle$$

into the presentation $\langle a, c \mid c^2 = 1 \rangle$.

2.7 Extraction principles

A group G is said to be *finitely generated* if there is some some finite set of elements $\{g_1, \ldots, g_n\}$ which generates G, that is, every element of G is equal to a product of the g_i and their inverses (see the Preliminaries). In this case we can take a set of n symbols, say $S = \{a_1, \ldots, a_n\}$ and define a homomorphism from F_S onto G by $a_i \to g_i$. If we let D be any set of words on S whose normal closure is the kernel of this map, it follows that $G \cong gp(\langle S \mid D\rangle)$. Thus $\langle S \mid D\rangle$ is presentation for G on a finite set of generating symbols. That is, G is finitely generated if and only if $G \cong gp(\langle S \mid D\rangle)$ for some finite set S.

Similarly G is said to be *finitely presented* if there is some presentation for G with a finite set of generators and a finite set of relations, that is $G \cong gp(\langle S \mid D\rangle)$ where both S and D are finite.

Suppose we are given an arbitrary presentation of a group G which satisfies one of these conditions. Is it possible to somehow extract a suitably finite "subpresentation" from the one we have. The answer is provided by two of the results below.

Theorem 2.9 *Suppose that G is a finitely generated group and that G is isomorphic to a group with presentation $\langle T \mid E\rangle$. Then there is a finite subset $T_0 \subseteq T$ and a collection of words D_0 on T_0 such that the inclusion of T_0 into T induces an isomorphism $\langle T_0 \mid D_0\rangle \cong \langle T \mid E\rangle$.*

Proof: We can identify G with $gp(\langle T \mid E\rangle)$. Since G is finitely generated there is some finite set of words $\{u_1(\vec{t}), \ldots, u_n(\vec{t})\}$ on T which generate G. Let T_0 be the set of all symbols from T which appear in these words $u_i(\vec{t})$. Then T_0 is finite and $\langle T_0\rangle$ contains all the $u_i(\vec{t})$ so that $\langle T_0\rangle = G$. Hence the inclusion of T_0 into T induces a homomorphism from F_{T_0} onto G. taking D_0 to be a set of normal generators for the kernel, the result follows. □

Even if G is finitely presented in the above result it may not be possible to choose D_0 a finite subset of the given relators E. An example is contained in an exercise below. However, in case the generating set was already finite, this can be done.

Theorem 2.10 *Suppose that G is a finitely presented group and that G is isomorphic to a group with presentation $\langle T \mid E\rangle$ where T is finite. Then there is a finite subset $E_0 \subseteq E$ such that in F_T the normal subgroups $N_{E_0} = N_E$ and hence $\langle T \mid E_0\rangle \cong \langle T \mid E\rangle$.*

Proof: By hypothesis G has some finite presentation

$$G \cong \langle a_1, \ldots, a_n \mid r_1(\vec{a}) = 1, \ldots, r_k(\vec{a}) = 1 \rangle$$

and we can suppose

$$\langle T \mid E \rangle = \langle b_1, \ldots, b_m \mid q_1(\vec{b}) = 1, q_2(\vec{b}) = 1, \ldots \rangle.$$

These are both presentations of G so we may suppose the isomorphisms are between the groups presented are defined by

$$\varphi(a_i) = u_i(\vec{b}) \text{ for } i = 1, 2, \ldots \text{ and } \psi(b_j) = v_j(\vec{a}) \text{ for } j = 1, 2, \ldots.$$

We now apply transformations of type T2 to obtaint he presentation

$$\langle a_1, \ldots, a_n, b_1, \ldots, b_m \mid r_1(\vec{a}) = 1, \ldots, r_k(\vec{a}) = 1,$$

$$b_1 = v_1(\vec{a}), \ldots, b_m = v_m(\vec{a}) \rangle.$$

From the relations in this presentation we can deduce the relations

$$a_1 = u_1(\vec{b}), \ldots, a_n = u_n(\vec{b}).$$

Using these we can replace the relations $r_i(\vec{a}) = 1$ and $b_j = v_j(\vec{a})$ by the equivalent relations $r_i(\vec{u}(\vec{b})) = 1$ and $b_j = v_j(\vec{u}(\vec{b}))$ to obtain the following presentation for G:

$$\langle a_1, \ldots, a_n, b_1, \ldots, b_m \mid r_1(\vec{u}(\vec{b})) = 1, \ldots, r_k(\vec{u}(\vec{b})) = 1, b_1 = v_1(\vec{u}(\vec{b})), \ldots$$

$$\ldots, b_m = v_m(\vec{u}(\vec{b})), a_1 = u_1(\vec{b}), \ldots, a_n = u_n(\vec{b}) \rangle.$$

Now we can eliminate the a_i's to obtain the finite presentation

$$G \cong \langle b_1, \ldots, b_m \mid r_1(\vec{u}(\vec{b})) = 1, \ldots, r_k(\vec{u}(\vec{b})) = 1,$$

$$b_1 = v_1(\vec{u}(\vec{b})), \ldots, b_m = v_m(\vec{u}(\vec{b})) \rangle.$$

We record the observation that this presents G on the generators $\{b_1, \ldots, b_m\}$ using $k + m$ defining relations.

Finally we observe that the $k + m$ relations in this finite presentation for G are all consequences of some finite subset $E_0 \subseteq E = \{q_j(\vec{b}) = 1\}$ and then $\langle T \mid E_0 \rangle$ is a presentation of G as required. □

Corollary 2.11 (B.H. Neumann) *Suppose that a group G has a presentation with n generators and k defining relations. If $b_1, \ldots, b_m$ is another set of generators of G, then G can be defined by at most $m + k$ relations on the $b_1, \ldots, b_m$.*

The following exercise gives an example in which simultaneously passing to finite subsets of generators and relations is not possible.

Exercise 2.9 *Consider the group G presented by*

$$\langle a, b, c_0, c_1, c_2, \ldots \mid a^4 = 1, b^3 = 1, c_0^{-1} b c_0 = a^2, c_1^{-1} c_0 c_1 = b,$$

$$c_2^{-1} c_1 c_2 = c_0, c_3^{-1} c_2 c_3 = c_1, \ldots \rangle.$$

Show that G is cyclic of order 2. Show that any finite sub-presentation of this, say $\langle T_0 \mid E_0 \rangle$, defines a group of order 3, or a group of order 4 or an infinite group.

Chapter 3

Construction of new groups

We are going to discuss several methods of constructing new groups from groups we already have in hand. In each case we write down a presentation of the resulting group and give some information on expressing elements in a normal (standard) form.

3.1 Direct products

A hopefully familiar construction is the direct product. Suppose we are given two groups H and K. We can construct their *direct product* $H \times K$ as the set of ordered pairs (h, k) with $h \in H, k \in H$ with multiplication defined by $(h_1, k_1) \cdot (h_2, k_2) = (h_1 h_2, k_1 k_2)$. The maps defined by $h \mapsto (h, 1)$ an $k \mapsto (1, k)$ embed H and K respectively into $H \times K$. Though of as subgroups in this way H and K are called the *direct factors* of $H \times K$. Observe that $(h, 1)(1, k) = (h, k) = (1, k)(h, 1)$ so the (images of) the direct factors commute in $H \times K$.

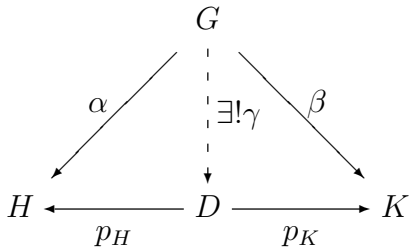
Suppose that H and K are given by presentations, say $H = \langle S \mid D \rangle$ and $K = \langle T \mid E \rangle$. By changing one of the alphabets if necessary, we can assume S and T are disjoint, that is $S \cap T = \emptyset$. Then a presentation for $H \times K$ can be obtained by joining these together and adding the relations which imply that elements of H commute with elements of K, that is

$$H \times K = \langle S, T \mid D, E,\ st = ts\ \forall s \in S, t \in T \rangle.$$

Here the notation means that generators are those symbols in S and those in T, and similarly for relations. We prefer this to the more set theoretic $S \cup T$.

The inclusion maps on generators induce embeddings of H and K into $H \times K$ presented in this way. The projection of $H \times K$ onto H, for instance, is defined by adding the relations $t = 1\ \forall t \in T$, thus "killing" the factor K. Observe that every element $w \in H \times K$ can be written by using just the commuting relations $st = ts$ in the form $w = uv$ where u is a word on S and v is a word on T. Moreover, if $w =_{H\times K} 1$ then $u =_H 1$ and $v =_K 1$. This last fact is equivalent to the following: if $w = u_1v_1 = u_2v_2$ where the u_i are words on S and the v_i are words on T, then $u_1 =_H u_2$ and $v_1 =_K v_2$.

There is a more abstract way (arrow theoretic or categorical) to define the direct product of H and K. It goes as follows. A group D is said to be the *direct product* of groups H and K if there are homomorphisms $p_H : D \to H, p_K : D \to K$ satisfying the following condition: for every group G, for every pair of homomorphisms $\alpha : G \to H, \beta : G \to K$ there is a unique homomorphism $\gamma : G \to D$ such that $\alpha = p_H \circ \gamma$ and $\beta = p_K \circ \gamma$.



One easily sees the group $H \times K$ has the properties required of D where γ is defined for any given α, β by $\gamma(g) = \alpha(g)\beta(g)$. A diagram chase using uniqueness now shows that $D \cong H \times K$.

But let's recover the description just using this arrow theoretic definition. Assume that D satisfies the above definition. If we take α to be the identity may on H and β to be the trivial map, we get $p_H \circ \gamma$ is the identity map on H and so p_H is surjective and γ is injective. Hence p_H maps the subgroup $\gamma(H)$ of D isomorphically onto H. For this same choice we have $p_K \circ \gamma$ is the trivial map and so $\gamma(H) \subseteq \ker p_K$.

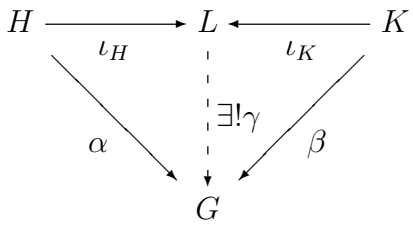
An analogous choice of α and β for K, shows that p_K maps a subgroup $\delta(K)$ of D isomorphically onto K and that $\delta(K) \subseteq \ker p_H$. Hence we can identify H with $\gamma(H)$ and K with $\delta(K)$ and so think of H and K as subgroups of D and p_H and p_K as retractions onto those subgroups. Now if $h \in H, k \in K$ observe that their commutator $[h, k] \in \ker p_H \cap \ker p_K$ What we expect is that $[h, k] =_D 1$ which follows if we can show this intersection is trivial.

Suppose $1 \neq x \in \ker p_H \cap \ker p_K$. Let C be the infinite cyclic group generated by a and take α and β to be the trivial maps from C to H and

K respectively. Now there are two maps which when composed with the projections give α and β: one is the trivial map, the other sends a to x. So this contradicts the uniqueness requirement. Thus $\ker p_H \cap \ker p_K = \{1\}$. Hence also $[H, K] = \{1\}$ and $D = HK$ as desired.

3.2 Free products

Suppose that H and K are two groups. A L is said to be the *free product* of H and K if there are homomorphisms $\iota_H : H \to L$ and $\iota_K : K \to L$ satisfying the following condition: for any pair of homomorphisms $\alpha : H \to G$ and $\beta : K \to G$ where G is any group, there is a unique homomorphism $\gamma : L \to G$ such that $\alpha = \gamma \circ \iota_H$ and $\beta = \gamma \circ \iota_K$.



An easy diagram chase show that the free product of H and K is unique (up to to isomorphism) and we will denote it by $H \star K$.

It is easy to see that free products exist because we can just write down a presentation for $H \star K$. Suppose that H and K are given by presentations, say $H = \langle S \mid D \rangle$ and $K = \langle T \mid E \rangle$. By changing one of the alphabets if necessary, we can assume S and T are disjoint, that is $S \cap T = \emptyset$. Then a presentation for $H \star K$ can be obtained by joining these together, thus

$$H \star K = \langle S \cup T \mid D \cup E \rangle.$$

The required maps ι_H and ι_K are just the homomorphisms induced by the inclusions on generators. Both of these are monomorphisms. For instance, if we define $\varphi : H \star K \to H$ by $s \mapsto s, t \mapsto 1 \ \forall s \in S, t \in T$, then φ defines a homomorphism and $\varphi \circ \iota_H$ is the identity on H. So ι_H is a monomorphism. It also follows from this argument that $H \cap K = \{1\}$.

Finally, given homomorphisms α and β as in the definition, the required γ is given by $\gamma(s) = \alpha(s)$ for $s \in S$ and $\gamma(t) = \beta(t)$ for $t \in T$. Then γ defines a homomorphism; since the definition was clearly forced on us, this is the unique such map.

In a sense the free product $H \star K$ is the "freest" group containing H and K. The subgroups H and K are called the *(free) factors* of $H \star K$. The construction can be generalized to any number of factors, say $H_j (j \in J)$, in which case we denote the free product by $\underset{j\in J}{\star} H_j$.

By an *alternating word or expression* in $H \star K$ we mean a product of the form $h_1 k_1 \cdots h_m k_m$ where each $h_i \in H$ and each $k_i \in K$; by convention, we allow the possibility that one or both of h_1 or k_m is not present so that all possible beginnings or ends are covered. That is, such an expression could have one of the four forms $h_1 k_1 \cdots h_m k_m$, or $k_1 \cdots h_m k_m$ where h_1 is not present, or $h_1 k_1 \cdots h_m$ where k_m is not present, or $k_1 \cdots h_m$ where neither h_1 nor k_m is present. The number of terms present is called the *length* of the word. We allow the empty expression which has length 0.

Such an alternating expression is said to be *reduced* if each $h_i \neq_H 1$ and each $k_i \neq_K 1$ when present. If such an alternating word is not reduced, it is equal in $H \star K$ to a shorter alternating expression obtained by removing one of the terms and regrouping. Thus if $h_i = 1$ the alternating expression $h_1 k_1 \cdots k_{i-1} h_i k_i \cdots h_m k_m$ we can replace this alterating expression by the alternating expression $h_1 k_1 \cdots h_{i-1}(k_{i-1} k_i) h_{i+1} \cdots h_m k_m$ having fewer alternations and stiil representing the same group element. Continuing in this way we eventually arrive at a reduced alternating expression representing the same group element as the original. Note that the empty expression is reduced.

Theorem 3.1 (Normal Form Theorem) *Every element of $H \star K$ is equal to a unique alternating expression of the form $h_1 k_1 \cdots h_m k_m$ with $h_i \neq_H 1$ and $k_i \neq_K 1$ when present. Here uniqueness means that if two such expressions are equal in $H \star K$, say*

$$h_1 k_1 \cdots h_m k_m =_{H\star K} h'_1 k'_1 \cdots h'_n k'_n$$

then $n = m$ and each $h_i =_H h'_i$ and each $k_i =_K k'_i$.

Proof: That any element is equal to an alternating expression is clear from the presentation. The uniqueness assertion is a non-trivial result and requires proof. To this end we let Ω denote the set of all reduced alternating expressions. With each element $h \in H$ we associate a permutation $\phi(h)$ in the group $\mathrm{Sym}(\Omega)$ of all permutations of Ω by the rule

$$\phi(h)(h_1 k_1 \cdots h_m k_m) = \begin{cases} k_1 h_2 \cdots h_m k_m & \text{if } h = h_1^{-1} \\ (h h_1) k_1 \cdots h_m k_m & \text{if } h \neq h_1^{-1} \end{cases}$$

where we understand the first case includes the possibility h_1 is not present and $h = 1$. Also in the second case h_1 is not necessarily present.

It is easy to check that $\phi(h^{-1}) = (\phi(h))^{-1}$ and that for $h, h' \in H$ we have $\phi(h)\phi(h') = \phi(hh')$. Thus the map $h \mapsto \phi(h)$ defines a homomorphism from H to $\mathrm{Sym}(\Omega)$. In an entirely analogous way we define an action of K on Ω and hence a homomorphism $\psi : K \to \mathrm{Sym}(\Omega)$. We thus obtain a homomorphism $\phi \star \psi : H \star K \to \mathrm{Sym}(\Omega)$.

Now if $h_1k_1 \cdots h_mk_m$ is a reduced alternating expression, then the permutation $\phi \star \psi(h_1k_1 \cdots h_mk_m)$ sends the empty expression to $h_1k_1 \cdots h_mk_m$. Hence $h_1k_1 \cdots h_mk_m \neq 1$ in $H \star K$ unless it is the empty expression.

By induction this implies uniqueness. For suppose

$$h_1k_1 \cdots h_mk_m =_{H\star K} h'_1k'_1 \cdots h'_nk'_n$$

where both sides are reduced alternating expressions. Then

$$1 = h'_1k'_1 \cdots h'_n(k'_nk_m^{-1})h_m^{-1} \cdots h_1^{-1}$$

and hence the right hand side is not reduced by what we have proved. Since the original expressions were reduced, we must have $k_m = k'_n$ and so by induction the two expressions are identical. □

The following is an alternate version of the normal form theorem which gives an equivalent characterization of free products. It follows immediately from the above proof.

Theorem 3.2 (Characterization of free products) *G is the free product of its subgroups H and K if and only if the following two conditions hold:*

1. *H and K generate G, that is every element of G is equal to an some alternating expression $h_1k_1 \cdots h_mk_m$; and*

2. *if $w \equiv h_1k_1 \cdots h_mk_m$ is an alternating expression and if $w =_G 1$ then for some i either $h_i =_H 1$ or $k_i =_K 1$.* □

As an example, consider the group $G = H \star K$ where $H = \langle a \mid a^2 = 1 \rangle$ and $K = \langle b \mid b^3 = 1 \rangle$. Then G has the presentation $G = \langle a, b \mid a^2 = 1, b^3 = 1 \rangle$. Several questions arise. Is G infinite ? What are the possible orders of the elements of G ? Clearly G has elements of finite order 2 and 3 (and 1 if you consider the identity element).

Consider the element ab. Could its order be finite ? A power of ab has the form $(ab)^n = abab\cdots ab$ which is an alternating word. By the normal form results, if $(ab)^n =_G 1$ then either $a =_H 1$ or $b =_K 1$ and neither is the case. Hence ab has infinite order in G. In fact this argument easily generalizes to show that if H and K are non-trivial groups then $H \star K$ contains an element of infinite order.

An alternating word $h_1k_1 \cdots h_mk_m$ is said to be *cyclically reduced* if it is reduced and either has length 1 or even length. It follows that a cyclically reduced w alternating word has first and last term from different factors and every cyclic permutation (as an alternating word) is reduced. Also if a cyclically reduced word w has length at least 2, the w has infinite order in $H \star K$, for the same reasons that ab had infinite order in our example.

But by cyclically permuting and reducing as often as possible one arrives at a cyclically reduced word. Hence any alternating word is conjugate to a cyclically reduced word in $H \star K$ Hence an element of finite order must be conjugate to an element of length 1, that is and element of H or K.

We record this observation as follows.

Lemma 3.3 *In the free product $H \star K$, every element of finite order is conjugate to an element of H or of K. If both H and K are non-trivial, then $H \star K$ has elements of infinite order; in fact every reduced alternating word of even length a has infinite order.*

It is reassuring that no new finite orders have been introduced in our construction since we didn't add any equations to force such new orders.

Recall that the *rank* of a group is the minimum number of elements in any generating set. One quite useful fact about free products is the following result whose proof we omit:

Theorem 3.4 (Grushko-Neumann) *The rank of the free product $H \star K$ is the sum of the ranks of H and K.*

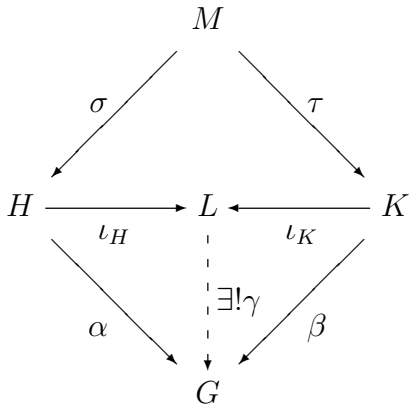
Recall that the centre of a group G is $Z(G) = \{x \in G \mid gx = xg \ \forall g \in G\}$.

Exercise 3.1 *Show that if both H and K are non-trivial, then $H \star K$ has a trivial centre.*

3.3 Free products with amalgamation

We want to generalize the free product construction to the following: suppose H and K have a isomorphic subgroups, so there are a pair of embeddings (monomorphisms) $\sigma : M \to H$ and $\tau : M \to K$. We want to form the "freest" group containing H and K in which their subgroups $\sigma(M)$ and $\tau(M)$ are identified, so hopefully $H \cap K = \sigma(M) = \tau(M)$.

The group L will be called the *free product of H and K with amalgamated subgroup M* if there are maps $\iota_H : H \to L$ and $\iota_K : K \to L$ such that $\iota_H \circ \sigma = \iota_K \circ \tau$ satisfying the following condition: for any pair of homomorphisms $\alpha : H \to G$ and $\beta : K \to G$ such that $\alpha \circ \sigma = \beta \circ \tau$ where G is any group, there is a unique homomorphism $\gamma : L \to G$ such that $\alpha = \gamma \circ \iota_H$ and $\beta = \gamma \circ \iota_K$.



An easy diagram chase shows that the free product L of H and K with amalgamated subgroup M is unique (up to isomorphism) and we will denote it by $L = H \underset{M}{\star} K$.

It is also easy to see that amalgamated free products exist because we can just write down a presentation for $H \underset{M}{\star} K$. Suppose that H and K are given by presentations, say $H = \langle S \mid D \rangle$ and $K = \langle T \mid E \rangle$. Also suppose that $M = \langle Q \mid V \rangle$ (only the generators of M are relevant here). By changing one of the alphabets if necessary, we can assume S and T are disjoint, that is $S \cap T = \emptyset$. Then a presentation for $H \underset{M}{\star} K$ can be obtained by joining these together and identifying the images of M, thus

$$H \underset{M}{\star} K = \langle S \cup T \mid D \cup E, \sigma(q) = \tau(q) \ \forall q \in Q \rangle.$$

The required maps ι_H and ι_K are just the homomorphisms induced by the inclusions on generators. Both of these are monomorphisms, but this is not obvious. Also one can show $H \cap K = \sigma(M) = \tau(M)$, but again this is not obvious.

Finally, given homomorphisms α and β as in the definition, the required γ is given by $\gamma(s) = \alpha(s)$ for $s \in S$ and $\gamma(t) = \beta(t)$ for $t \in T$. Then γ defines a homomorphism; since the definition was clearly forced on us, this is the unique such map.

In case the group M is the trivial subgroup, $H \underset{M}{\star} K$ reduces to the free product $H \star K$ discussed previously. We sometimes refer to $H \star K$ as the *ordinary* free product.

There is another convenient notation for the above that is frequently used. Namely, let $A = \sigma(M) \subseteq H$ and $B = \tau(M) \subseteq K$. Then these subgroups A and B are isomorphic via the homomorphism $\varphi = \tau \circ \sigma^{-1} : A \to B$. The amalgamated free product is then often denoted $H \underset{A=B}{\star} K$ and is presented in the following equivalent form:

$$H \underset{A=B}{\star} K = \langle S \cup T \mid D \cup E, a = \varphi(a)\ \forall a \in \sigma(Q) \rangle.$$

We will use both sorts of notation.

Here is a concrete example which the reader is urged to keep in mind in connection with the following discussion. Consider the two infinite cyclic groups $H = \langle c \mid \rangle$ and $K = \langle d \mid \rangle$ with their respective subgroups $A = \langle c^2 \rangle$ and $B = \langle d^3 \rangle$ which are isomorphic via the map $c^2 \mapsto d^3$. Then their amalgamated free product is

$$G = H \underset{A=B}{\star} K = \langle c, d \mid c^2 = d^3 \rangle.$$

To make effective use of an amalgamated free product we need to give a canonical expression or "normal form" for each of its elements and give methods for computing them and for proving expressions are equal. Let $G = H \underset{A=B}{\star} K$ be an amalgamated free product. The notion of an *alternating word or expression* is exactly as for ordinary free products. Clearly every element $g \in G$ is equal to some alternating expression. But some caution is required since we do not yet know that H and K are embedded in G, so we should think of expressions as sequences which have a natural image in G.

For an amalgamated free product we say that an alternating expression $h_1 k_1 \cdots h_m k_m$ is *reduced* if no $h_i \in A = \sigma(M)$ and no $k_i \in B = \tau(M)$

when present. Suppose that the alternating expression $h_1k_1 \cdots h_mk_m$ is not reduced, say with some if $h_i = a_i \in A$. In this expression

$$h_1k_1 \cdots k_{i-1}h_ik_i \cdots h_mk_m$$

we can replace $h_i = a_i \in A$ by the corresponding $b_i = \phi(a_i)$ to obtain (after consolidating) the alternating expression

$$h_1k_1 \cdots h_{i-1}(k_{i-1}b_ik_i)h_{i+1} \cdots h_mk_m$$

having fewer alternations and still representing the same group element. Similarly if $k_i \in B$ we can replace it by a corresponding a_i and consolidate to give an expression with fewer alternations equal to the original. So continuing in this way we eventually arrive at an alternating expression, representing the same group element as the original, which is either reduced or is an element of $A = B$.

In terms of our suggested example

$$G = H \underset{A=B}{\star} K = \langle c, d \mid c^2 = d^3 \rangle$$

we note that the two words c^3d^{-5} and cd^{-2} are both reduced. Now in G we have $c^3d^{-5} = cc^2d^{-5} = cd^3d^{-5} = cd^{-2}$ so these two words are equal in G, that is they represent the same group element. So "reduced" is not strong enough to give a unique canonical form.

To overcome this difficulty proceed as follows. We first choose a *transversal* Y for the right cosets of A in H, that is Y contains exactly one element (called the *coset representative*) from each right coset Ah where $h \in H$ subject to the condition that the representative chosen for A itself is 1. Similarly we choose a transversal Z for the right cosets of B in K. We define a *normal form* to be either an element of $A = B$ or an expression of the form $ah_1k_1 \cdots h_mk_m$ with $1 \neq h_i \in Y$ and $1 \neq k_i \in Z$ when present and $a \in A$. Here $h_1k_1 \cdots h_mk_m$ is an alternating expression meaning that as usual we allow it to any of the four forms $h_1k_1 \cdots h_mk_m$, or $k_1 \cdots h_mk_m$, or $h_1k_1 \cdots h_m$, or $h_1k_1 \cdots h_m$ (to avoid too many case distinctions). So a normal form is an alternating product of elements of the transversals Y and Z (all different from 1 so that the expression is reduced) preceded by an element of $A = B$. In the case h_1 is not present we usually use $b = \varphi(a)$ instead and write $bk_1 \cdots h_mk_m$. This is the same group element since $a = b$ in G.

We need to show that every element w of G is represented by such a normal form. First applying our reduction process, we may suppose we have a reduced alternating expression for w. We work from right to left across the word converting it into a normal form. So we can suppose $h_1k_1 \cdots k_{i-1}h_ik_i \cdots h_mk_m$ is reduced and that each term to the right of say h_i is an element in Y or Z different from 1. Now we write $h_i = ah_i'$ where $a \in A$ and $h_i' \in Y$. Note that $h_i' \neq 1$ since $h_i \notin A$ because the expression is reduced. Let $b = \phi(a) \in K$ and note that $a = b$ in the group. We now replace the current expression by the revised alternating expression $h_1k_1 \cdots h_{i-1}(k_{i-1}b)h_i'k_i \cdots h_mk_m$ which is still equal to the original group element. It is still reduced since $k_{i-1} \notin B$ implies $k_{i-1}b \notin B$. Continuing in this way we obtain a normal form for the original group element w. Note that normal form of w has the same length as the first reduced expression we obtained for w.

We summarize these procedures as follows:

Lemma 3.5 *Let $G = H \underset{A=B}{\star} K$ be an amalgamated free product. Choose tranversals Y for A in H and Z for B in K as above. Then, if w is any element of G,*

1. *w is equal in G to an alternating expression.*

2. *any alternating expression for w can be converted either to a reduced alternating expression or to an element of $A = B$ which equal to w in G.*

3. *any reduced alternating expression for w can be converted into a normal form having the same length which is equal to w in G.* □

Here is an exercise which may help clarify these processes.

Exercise 3.2 *Consider $G = \langle c, d \mid c^2 = d^3 \rangle$ with the above notation for subgroups. As a transversal for A in H take $Y = \{1, c\}$ and for B in K take $\mathbb{Z} = \{1, d, d^{-1}\}$. Compute the normal forms of the elements $c^3d^{-2}c^{-4}d^4$ and $c^{-3}dc^{-4}d^2cd^2$. Repeat the exercise with the transversals $Y = \{1, c^{-1}\}$ and $\mathbb{Z} = \{1, d^{-2}, d^5\}$.*

Two difficulties remain. We would like to know that when we obtain a normal form for $w \in G$ it is unique. Moreover it is not yet clear whether the images of H and K inside of G are actually isomorphic copies. This would

follow from uniqueness of normal forms: for $h_1 \neq_H h_2$ have normal forms $h_1 = a_1 h_1'$ and $h_2 = a_2 h_2'$ with $h_i \in Y$ which depend only on H and are different. Hence they represent different elements of G. Similarly for K and uniqueness implies they are both naturally embedded.

Here is one version of the result we are after.

Theorem 3.6 (Embedding and Reduction) *Let $G = H \underset{A=B}{\star} K$ be a free product with amalgamation. Then*

1. *The maps defined by inclusion of generators induce monomorphisms of H and K into G; and*

2. *if $w \equiv h_1 k_1 \cdots h_m k_m$ is an alternating expression and if $w = 1$ in G, then for some i either $h_i \in A$ viewed as an element of H or $k_i \in B$ viewed as an element of K.*

Proof: The above Lemma established the existence of a normal form representing each element of G. We proceed as in the case of ordinary free products by letting Ω be the set of all normal forms. With each element $h \in H$ we associate a permutation in the group $\mathrm{Sym}(\Omega)$ of all permutations of Ω by the rule

$$\theta(h)(ah_1 k_1 \cdots h_m k_m) = \begin{cases} b' k_1 h_2 \cdots h_m k_m & \text{if } hah_1 \in A = B \\ a' h_1' k_1 \cdots h_m k_m & \text{if } hah_1 \notin A = B \end{cases}$$

where $hah_1 = a' h_1'$ with $h_1' \in Y$ and $b' = \theta(a')$. Here we understand the first case includes the possibility h_1 is not present. Also in the second case h_1 is not necessarily present. Again is easy to check that $\theta(h^{-1}) = (\theta(h))^{-1}$ and that for $h, h' \in H$ we have $\theta(h)\theta(h') = \theta(hh')$. Thus the map $h \mapsto \theta(h)$ defines a homomorphism from H to $\mathrm{Sym}(\Omega)$.

In an analogous way we define an action of K on Ω by the rule of Ω by the rule for h_1 not present:

$$\psi(k)(ak_1 \cdots h_m k_m) = \begin{cases} b' h_2 \cdots h_m k_m & \text{if } kbk_1 \in A = B \\ b' k_1' h_2 \cdots h_m k_m & \text{if } kbk_1 \notin A = B \end{cases}$$

where $b = \varphi(a)$ and $kbk_1 = b' k_1'$ with $k_1' \in Z$ and $b' = \varphi(a')$; and for the case h_1 present by the rule:

$$\psi(k)(ah_1 k_1 \cdots h_m k_m) = \begin{cases} b' k_0' h_1 k_1 \cdots h_m k_m & \text{if } kb \in A = B \\ b' h_1 k_1 \cdots h_m k_m & \text{if } kb \in A = B \end{cases}$$

where $b = \varphi(a)$ and $kb = b'k'_0$ with $k'_0 \in Z$. Again one checks this defines a homomorphism $\psi : K \to \mathrm{Sym}(\Omega)$.

But we also observe that if $b = \varphi(a)$, the actions of a and b on Ω, that is $\theta(a) = \psi(b)$. By the definition of amalgamated free product, these two maps extend to a homomorphism $\theta \underset{A}{\star} \psi : G \to \mathrm{Sym}(\Omega)$.

Now if $ah_1k_1 \cdots h_mk_m$ is a normal form, then the permutation

$$\phi \underset{A}{\star} \psi(ah_1k_1 \cdots h_mk_m)$$

sends the empty expression to $ah_1k_1 \cdots h_mk_m$. Hence $ah_1k_1 \cdots h_mk_m \neq 1$ in G unless it is the empty expression.

If $1 \neq h \in H$ with say $h = ah_1$ where $h_1 \in Y$, then the image of h in G sends the empty string to ah_1. So the image of $h \neq_G 1$ and hence H is naturally embedded in G by the inclusion map. Similarly K is embedded in G. Since our construction of a normal form from a reduced expression preserves length, it follows that an alternating expression which represents the identity 1 cannot be reduced. This proves the second assertion of the theorem. □

Using this Theorem we can now show $H \cap K = A = B$. Clearly $H \cap K \supseteq A = B$. Suppose $g \in H \cap K$. Then $g = h = k$ for suitable words h and k in the generators of H and K respectively. Hence, applying the second assertion above to $hk^{-1} = 1$ we know either $h \in A$ or $k \in B$, and in either case it follows that $g \in A = B$.

Another consequence is the normal form theorem for amalgamated free products.

Theorem 3.7 (Normal Form Theorem) *Every element of* $G = H \underset{A=B}{\star} K$ *is equal to a unique* normal form $ah_1k_1 \cdots h_mk_m$ *with* $1 \neq h_i \in Y$ *where* $1 \neq k_i \in Z$ *when present and* $a \in A$. *Here* Y *and* Z *are the transversals chosen above. The uniqueness assertion means that if two such expressions are equal in* $H \star K$, *say*

$$ah_1k_1 \cdots h_mk_m =_G a'h'_1k'_1 \cdots h'_nk'_n$$

then $n = m$ *and each* $h_i = h'_i$ *and each* $k_i = k'_i$ *and* $a = a'$.

Proof: The existence of a normal form for any element follows from the above Lemma. Uniqueness of the normal forms follows easily from the previous Theorem. For if

$$ah_1k_1 \cdots h_mk_m =_G a'h'_1k'_1 \cdots h'_nk'_n$$

are two normal forms for the same group element, then

$$1 =_G (a'h'_1)k'_1 \cdots h'_n(k'_n k_m^{-1})h_m^{-1} \cdots k^{-1}(h_1^{-1}a^{-1})$$

is not reduced. Hence $k'_n k_m^{-1} \in B$ and so $k'_n = k_m$ since they belong to the transversal Z. Uniqueness follows by induction. □

In fact, as the proofs makes clear, these two theorems are easily shown to be equivalent without resort to permutations. The following is an alternate version which is often useful.

Theorem 3.8 (Characterization of amalgamated free products) *G is the free product of its subgroups H and K with amalgamated subgroup $M = H \cap K$ if and only if the following two conditions hold:*

1. *H and K generate G, that is every element of G is equal to an some alternating expression $h_1k_1 \cdots h_mk_m$; and*

2. *if $w \equiv h_1k_1 \cdots h_mk_m$ is an alternating expression and if $w =_G 1$ then for some i either $h_i \in M$ or $k_i \in M$.*

Continuing with our earlier example $G = \langle c, d \mid c^2 = d^3 \rangle$ observe that the element c^2 lies in the centre of G since $c^2d = d^3d = dd^3 = dc^2$ so that c^2 commutes with the generators of G and hence every element.

Also observe that G is not abelian. For consider the alternating word $w \equiv c^{-1}d^{-1}cd$. If it were the case that $w =_G 1$ then either $c =_H d^{2j}$ or $d =_K d^{3j}$ for some $j \in \mathbb{Z}$ by the normal form results. But neither is the case, so we conclude $w \neq_G 1$.

Exercise 3.3 *Determine the centre of the group $G = \langle c, d \mid c^2 = d^3 \rangle$.*

Exercise 3.4 *Define a suitable notion of cyclically reduced alternating words for amalgamated free products. Show that every element is conjugate to a cyclically reduced word.*

Exercise 3.5 *Show that in an amalgamated free product, any element of finite order is conjugate to an element of one of the factors.*

3.4 HNN extensions

Suppose that we are given a group G, say by a presentation $G = \langle S \mid D \rangle$ and a pair of isomorphic subgroups A and B with an isomorphism $\varphi : A \to B$. We want to find a larger group containing G in which the subgroups A and B are conjugate by an element realizing the isomorphism between them. Naturally we want to do this in the "freest" possible way.

It is easy to write down a candidate for the desired group, but not so easy to show it has the required properties and to give a normal form result. By the *HNN extension* of G with *associated subgroups* A and B via the isomorphism $\varphi : A \to B$ we mean the group $G \underset{\varphi}{\star}$ with presentation

$$G \underset{\varphi}{\star} = \langle S, p \mid D, p^{-1}ap = \varphi(a)\ \forall a \in A \rangle.$$

The additional generator p added here is called the *stable letter*. We note that it suffices to add only the relations $p^{-1}ap = \varphi(a)$ for a ranging over a set of generators for A. (Remark: HNN stands for Higman-Neumann-Neumann, the names of the authors who introduced this construction).

As for amalgamated free products, to make effective use HNN extensions we need to know that G is still embedded $G \underset{\varphi}{\star}$, to give a reduction process and to give a method of finding a "normal form" for each element. It is clear the p generates an infinite cyclic subgroup of $G \underset{\varphi}{\star}$ since the infinite cycle on p is a quotient group by the obvious map.

Here is a concrete example which the reader is urged to keep in mind in connection with the following discussion. Consider the infinite cyclic group $G = \langle c \mid \ \rangle$ with respective subgroups $A = \langle c^2 \rangle$ and $B = \langle c^3 \rangle$ which are isomorphic via the map $c^2 \mapsto c^3$. Then the corresponding HNN-extension is $G \underset{\varphi}{\star} = \langle c, p \mid p^{-1}c^2p = c^3 \rangle$

In general $G \underset{\varphi}{\star}$ is the quotient group of the ordinary free product $G \star \langle p \rangle$ of G with the infinite cycle on p obtained by adding the indicated relations. A *p-expression* is a sequence of the form

$$g_0 p^{\epsilon_1} g_1 p^{\epsilon_2} g_2 \cdots p^{\epsilon_m} g_m$$

where the g_i are elements of G and $\epsilon_i = \pm 1$. We allow the case where there are no p symbols appearing, and say that a p-expression *involves* p if at least one p or p^{-1} appears. Clearly any word in the generators of

$G \underset{\varphi}{\star}$ can be viewed as such a p-expression. If for some i the sub-expression $p^{-1}g_ip$ appears and if $g_i = a \in A$ then we can apply the defining relation $p^{-1}g_ip = p^{-1}ap = \varphi(a) = b \in B$ and hence replace the given expression by

$$g_0p^{\epsilon_1} \cdots p^{\epsilon_{i-1}} \ (g_{i-1}bg_{i+1}) \ p^{\epsilon_{i+2}} \cdots p^{\epsilon_m}g_m$$

which has two fewer p symbols and represents the same element of $G \underset{\varphi}{\star}$. Similarly if the sub-expression pg_ip^{-1} appears and if $g_i = b \in B$ we can apply the relation $pg_ip^{-1} = pbp^{-1} = \varphi^{-1}(b) = a \in A$ to obtain an expression with two fewer p symbols representing the same element of $G \underset{\varphi}{\star}$. Either of these two kind of replacements of sub-expressions of the indicated form are called *p-pinches* or sometimes *p-reductions*. If no p-pinches are possible, the p-expression is said to be *p-reduced*. Notice that we have described a process for *p-reducing* an a p-expression.

If $g'_0p^{\delta_1}g'_1p^{\delta_2}g'_2 \cdots p^{\delta_n}g'_n$ another p-expression such as the above, then the two are said to be *p-parallel* if $n = m$ and $\epsilon_1 = \delta_1, \ldots, \epsilon_m = \delta_m$. It turns out that any two p-reduced expressions representing the same element of $G \underset{\varphi}{\star}$ must be p-parallel, but this requires proof. (It follows from the normal form results below.)

In terms of our suggested example $G \underset{\varphi}{\star} = \langle c, p \mid p^{-1}c^2p = c^3 \rangle$ we note that the two words $c^3pc^{-5}p^{-1}$ and $cpc^{-2}p^{-1}$ are both p-reduced. Now in $G \underset{\varphi}{\star}$ we have

$$c^3pc^{-5}p^{-1} = c^3pc^{-3}c^{-2}p^{-1} = c^3c^{-2}pc^{-2}p^{-1} = cpc^{-2}p^{-1}$$

so these two words are equal in G, that is they represent the same group element. So "p-reduced" is not strong enough to give a unique canonical form.

To overcome this difficulty proceed as for amalgamated free products. We first choose a *transversal* Y for the right cosets of A in G subject to the condition that the representative chosen for A itself is 1. Similarly we choose a transversal Z for the right cosets of B in G. We define a *normal form* to be a p-expression

$$g_0p^{\epsilon_1}g_1p^{\epsilon_2}g_2 \ldots p^{\epsilon_m}g_m$$

where if $\epsilon_i = -1$, then $g_i \in Y$; if $\epsilon_i = +1$, then $g_i \in Z$; and if $g_i = 1$ then $\epsilon_i \neq -\epsilon_{i+1}$. So g_0 is arbitary, but each of $g_1, \ldots, g_m$ must be in the appropriate transversal Y or Z. The last condition means there are no inverse pairs of

p symbols separated by the trivial coset representative 1. Observe that a normal form is necessarily p–reduced.

We need to show that every element w of G is represented by such a normal form. First applying our reduction process, we may suppose we have a p-reduced expression for w. We work from right to left across the word converting it into a normal form. So we can suppose

$$g_0 p^{\epsilon_1} g_1 p^{\epsilon_2} g_2 \dots p^{\epsilon_m} g_m$$

is p-reduced and that the portion of the expression to the right of g_i satisfies the normal form conditions. In case $\epsilon_i = -1$ in G we can uniquely write $g_i = a y_i$ where $a \in A$ and $y_i \in Y$. In the above expression replace $p^{-1} g_i = p^{-1} a y_i = b p^{-1} y_i$ where $b = \phi(a) \in B$ to obtain

$$g_0 p^{\epsilon_1} \cdots p^{\epsilon_{i-1}} \ (g_{i-1} b) p^{-1} y_i \ p^{\epsilon_{i+1}} \cdots p^{\epsilon_m} g_m.$$

Note that if $y_i = 1$ then $g_i \in A$ so that $\epsilon_{i+1} = -1$ because the expression was p-reduced. Also note that if $\epsilon_{i-1} = 1$, then $g_{i-1} \notin B$ because our expression was p-reduced. Hence in this case we also have $g_{i-1} b \notin B$ so our new expression is p-reduced and a larger right hand portion satisfies the normal form conditions. The case $\epsilon_i = -1$ is similar: in G write $g_i = b z_i$ where $b \in B$ and $z_i \in Z$. Replace $p g_i = p b z_i = a p z_i$ where $a \ \varphi^{-1}(b)$ to obtain

$$g_0 p^{\epsilon_1} \cdots p^{\epsilon_{i-1}} \ (g_{i-1} a) p z_i \ p^{\epsilon_{i+1}} \cdots p^{\epsilon_m} g_m.$$

By similar considerations this expression is again p-reduced and a larger right hand portion satisfies the normal form conditions. After m steps we arrive at a p-reduced expression in normal form which is p-parallel to the original.

We summarize these procedures as follows:

Lemma 3.9 *Let $G \overset{\varphi}{\star}$ be an HNN extension. Choose transversals Y for A in G and Z for B in G as above. Then, if w is any element of G,*

1. *w is equal in G to a p-expression.*
2. *any p-expression for w can be converted to a p-reduced expression which is equal to w in G.*
3. *any p-reduced expression for w can be converted into a normal form which is p-parallel to the given expression and is equal to w in G.* □

Here is an exercise which may help clarify these processes.

Exercise 3.6 *Consider* $G\underset{\varphi}{\star} = \langle c, p \mid p^{-1}c^2p = c^3\rangle$ *with the above notation for subgroups. As a transversal for* A *in* G *take* $Y = \{1, c\}$ *and for* B *in* G *take* $\mathbb{Z} = \{1, c, c^{-1}\}$. *Compute the normal forms of the elements* $c^3pc^{-2}p^{-1}c^{-4}pc^4p^{-1}$ *and* $c^{-3}pcp^{-1}c^{-4}pc^2p^{-1}cpd^2p^{-1}$. *Repeat the exercise with the transversals* $Y = \{1, c^{-1}\}$ *and* $\mathbb{Z} = \{1, c^{-2}, c^5\}$.

Here is the main result about this construction.

Theorem 3.10 (Embedding and Reduction) *Let* $G\underset{\varphi}{\star}$ *be the HNN extension of* G *with associated subgroups* A *and* B *via the isomorphism* $\varphi : A \to B$. *Then*

1. *(Higman,Neumann,Neumann) The identity map on generators induces and embedding of* G *into* $G\underset{\varphi}{\star}$, *and* p *generates an infinite cyclic subgroup of* $G\underset{\varphi}{\star}$.

2. *(Britton's Lemma) Let* w *be any word of* $G\underset{\varphi}{\star}$ *which involves* p, *that is either* p *or* p^{-1} *appears as a subword. If* $w =_{G\underset{\varphi}{\star}} 1$, *then* w *contains a subword of the form (i)* $p^{-1}cp$ *or (ii)* pcp^{-1}, *where* c *is a word on* S, *and such that, in case (i)* c *is equal in* G *to an element of* A, *and in case (ii),* c *is equal in* G *to an element of* B.

Remark: The second assertion, known as Britton's Lemma, says that if w is a p-expression which involves p and if $w = 1$ in $G\underset{\varphi}{\star}$, then w is not p-reduced and hence some p-pinch can be applied. So by repeatedly applying p-reductions we will eventually arrive at a word of G (no p symbols) which is the equal to 1 because of the relations of G.

Proof: The above Lemma established the existence of a normal form representing each element of $G\underset{\varphi}{\star}$. We proceed as in the case of ordinary and amalgamated free products by letting Ω be the set of all normal forms. With each element $g \in G$ we associate a permutation in the group $\mathrm{Sym}(\Omega)$ of all permutations of Ω by the rule

$$\theta(g)(g_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m) = (gg_0)p^{\epsilon_1}g_1p^{\epsilon_1}\cdots p^{\epsilon_m}g_m.$$

It is easy to check that θ defines a homomorphism from G to $\mathrm{Sym}(\Omega)$.

With the symbol p we associate a permutation

$$\psi(p)(g_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m) = \begin{cases} apz_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m & \text{if } \epsilon_1 = 1 \\ apz_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m & \text{if } z_0 \neq 1 \text{ and } \epsilon_1 = -1 \\ (ag_1)p^{\epsilon_2}\cdots p^{\epsilon_m}g_m & \text{if } z_0 = 1 \text{ and } \epsilon_1 = -1 \end{cases}$$

where $g_0 = bz_0$ with $b \in B, z_0 \in Z$ and $a = \varphi^{-1}(b)$. Similarly with p^{-1} we associate

$$\psi(p^{-1})(g_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m) = \begin{cases} bpy_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m & \text{if } \epsilon_1 = 1 \\ bpy_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m & \text{if } y_0 \neq 1 \text{ and } \epsilon_1 = 1 \\ (bg_1)p^{\epsilon_2}\cdots p^{\epsilon_m}g_m & \text{if } y_0 = 1 \text{ and } \epsilon_1 = 1 \end{cases}$$

where $g_0 = ay_0$ with $a \in A, y_0 \in Y$ and $b = \varphi(a)$. A routine check shows that $\psi(p) \circ \psi(p^{-1})$ and $\psi(p^{-1}) \circ \psi(p)$ are both the identity. Hence they both define permutations an determine a homomorphism ψ from the infinite cycle on p to $\mathrm{Sym}(\Omega)$.

Now these homomorphisms extend to a homomorphism

$$\theta \star \psi : G \star \langle p \rangle \to \mathrm{Sym}(\Omega)$$

and one can check that $\theta \star \psi(p\varphi(a))$ and $\theta \star \psi(ap)$ are the same permutation of Ω, that is the relations of $G \underset{\varphi}{\star}$ are sent to the identity permutation. Now if $g_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m$ is a non-trivial normal form, it is clear that

$$\theta \star \psi(g_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m)(1) = g_0p^{\epsilon_1}g_1p^{\epsilon_2}\cdots p^{\epsilon_m}g_m.$$

Hence such a normal form is not equal to 1 in $G \underset{\varphi}{\star}$. In particular it follows that G is embedded in $G \underset{\varphi}{\star}$ and that a p-reduced word which involves p is not equal to 1. This proves the theorem. □

As another illustration, consider the infinite cyclic group $G = \langle a \mid \rangle$ and its two isomorphic subgroups $A = \langle a \rangle$ and $B = \langle a^2 \rangle$. The the corresponding HNN extension has presentation $G \underset{\varphi}{\star} = \langle a, p \mid p^{-1}ap = a^2 \rangle$. Let $w \equiv a^{-1}pap^{-1}$. We claim that $w \neq_{G \underset{\varphi}{\star}} 1$ and hence $G \underset{\varphi}{\star}$ is not abelian. For if $w =_{G \underset{\varphi}{\star}} 1$, then by Britton's Lemma it must contain a p-pinch. But there is only possibility, at the subword pap^{-1} with $a \in B$. But clearly $a \notin B$ since B consists of the even powers of a. So there is no such pinch and $w \neq_{G \underset{\varphi}{\star}} 1$.

The unique normal form theorem is now easily proved.

Theorem 3.11 (Unique normal form) *Every element in* $G \underset{\varphi}{\star}$ *is equal to a unique expression of the form*

$$g_0 p^{\epsilon_1} g_1 p^{\epsilon_2} g_2 \dots p^{\epsilon_n} g_n$$

where if $\epsilon_i = -1$*, then* $g_i \in Y$*; if* $\epsilon_i = +1$*, then* $g_i \in Z$*; and if* $g_i = 1$ *then* $\epsilon_i \neq -\epsilon_{i+1}$*. Here* Y *and* Z *are the transversal chosen above.*

Proof: We already know the existence of a normal form for each element. Suppose $g_0 p^{\epsilon_1} g_1 \cdots p^{\epsilon_m} g_m$ and $g'_0 p^{\delta_1} g'_1 \cdots p^{\delta_n} g'_n$ are two normal forms representing the same element. Then

$$1 = g'_0 p^{\delta_1} g'_1 \cdots p^{\delta_n} (g'_n \; g_1^{-1}) p^{-\epsilon_m} \cdots g_1^{-1} p^{-\epsilon_1} g_0^{-1}$$

and so there must be a p-pinch. Since the normal forms are reduced this implies $\delta_n = \epsilon_m$. Moreover for $\delta_n = -1$ we must have $g'_n \; g_m^{-1} \in A$ and hence $g'_n = g_m$ while for $\delta_n = 1$ have $g'_n \; g_m^{-1} \in B$ and hence $g'_n = g_m$ since the g_i and g'_i belong to transversals. By induction this proves uniqueness. □

Exercise 3.7 *Define a suitable notion of cyclically p-reduced words for HNN extensions. Show that every element is conjugate to a cyclically p-reduced word.*

Exercise 3.8 *Show that in an HNN extension* $G \underset{\varphi}{\star}$*, any element of finite order is conjugate to an element of* G*.*

Finally, we remark that the whole discussion above for HNN extensions applies more generally to adding any number of stable letters $p_1, p_2, \dots$ conjugating pairs of isomorphic subgroups $(A_1, B_1), (A_2, B_2), \dots$ onto each other.

3.5 HNN related to amalgams

The reader will have noticed the similarity in our treatment of HNN extensions and amalgamated free products. There are many connections and parallels between between these two constructions and we now spell out a few of them. Often facts about one can easily be deduced from facts about the other.

Let $G\underset{\varphi}{\star} = \langle S, p \mid D, p^{-1}ap = \varphi(a)\ \forall a \in A\rangle$ be an HNN extension and consider the subgroup L generated by the two subgroups G and pGp^{-1} of $G\underset{\varphi}{\star}$. Observe that these two subgroups intersect in $A = pBp^{-1}$. Every element of L can be written as an alternating expression (from alternate subgroups) of the form

$$g_0pg_1p^{-1}g_2pg_3p^{-1}\cdots pg_{2n-1}p^{-1}g_{2n}.$$

If such an expression is equal to 1 in $G\underset{\varphi}{\star}$, then by Britton's Lemma for some i either $g_{2i} \in A$ or $g_{2i-1} \in B$. Hence this is not a reduced alternating expression. It follows that L is isomorphic to the free product of G and pGp^{-1} with amalgamated subgroup $A = pBp^{-1}$. We record this as follows.

Lemma 3.12 (Higman) *Let $G\underset{\varphi}{\star} = \langle S, p \mid D, p^{-1}ap = \varphi(a)\ \forall a \in A\rangle$ be an HNN extension. Then the subgroups G and pGp^{-1} generate their free product with amalgamated subgroup $A = pBp^{-1}$.* □

Again let $G\underset{\varphi}{\star} = \langle S, p \mid D, p^{-1}ap = \varphi(a)\ \forall a \in A\rangle$ be an HNN extension. We form the ordinary free product of this group with the infinite cyclic group on x and perform a sequence of Tietze transformations as follows:

$$\begin{aligned}
(G\underset{\varphi}{\star}) \star \langle x\rangle &= \langle S, p, x \mid D, p^{-1}ap = \varphi(a)\ \forall a \in A\rangle \\
&\cong \langle S, p, x, y \mid D,\ y = p^{-1}x,\ p^{-1}ap = \varphi(a)\ \forall a \in A\rangle \\
&\cong \langle S, p, x, y \mid D,\ p = xy^{-1},\ p^{-1}ap = \varphi(a)\ \forall a \in A\rangle \\
&\cong \langle S, p, x, y \mid D,\ p = xy^{-1},\ yx^{-1}axy^{-1} = \varphi(a)\ \forall a \in A\rangle \\
&\cong \langle S, p, x, y \mid D,\ p = xy^{-1},\ x^{-1}ax = y^{-1}\varphi(a)y\ \forall a \in A\rangle \\
&\cong \langle S, x, y \mid D,\ x^{-1}ax = y^{-1}\varphi(a)y\ \forall a \in A\rangle \\
&\cong (G \star \langle x\rangle) \underset{G\star x^{-1}Ax = G\star y^{-1}By}{\star} (G \star \langle y\rangle).
\end{aligned}$$

That is, we observe that in $G \star \langle x\rangle$ the subgroups G and $x^{-1}Ax$ generate their ordinary free product. This subgroup is isomorphic to the subgroup of $G \star \langle y\rangle$ generated by G and $y^{-1}By$. So $(G\underset{\varphi}{\star}) \star \langle x\rangle$ is isomorphic to the free product of $G \star \langle x\rangle$ and $G \star \langle y\rangle$ amalgamating the indicated subgroups.

Notice that from the theory of amalgamated free products it follows that G is embedded in $G\underset{\varphi}{\star}$. It is also possible to easily deduce Britton's Lemma

from the reduction results for amalgamated free products. So using the above observations we could reduce the study many aspects of HNN extensions to the study of amalgamated free products.

But we can also go the other way. Suppose H and K are groups with isomorphic subgroups A and $B = \varphi(a)$. First form the ordinary free product $H \star K$ and then the HNN extension conjugating A to B, which can be written as

$$L = (H \star K)\underset{\varphi}{\star} = \langle H, K, p \mid p^{-1}ap = \varphi(a)\ \forall a \in A\rangle.$$

We define a map $\theta : H \to L$ by $\theta(h) = p^{-1}hp$ and a map $\psi : K \to L$ by $\psi(k) = k$. Then for $a \in A$ we have $\theta(a) = p^{-1}ap = \varphi(a) \in B$. But in $H \underset{A=B}{\star} K$ we have $a = \varphi(a) \in K$ so that θ and ψ agree on the amalgamated subgroup. Hence they define a homomorphism from $H \underset{A=B}{\star} K$ to L. Now using the theory of HNN extensions it follows that both H and K are embedded in $H \underset{A=B}{\star} K$.

The reduction theorem for amalgamated free products follow from Britton's Lemma. To see this, suppose $w = h_1k_1 \cdots h_mk_m$ is an alternating expression in $H \underset{A=B}{\star} K$. The image of this expression in L is

$$p^{-1}h_1pk_1p^{-1}h_2p \cdots p^{-1}h_mpk_m.$$

If $w = 1$ then its image is equal to 1 and so by Britton's Lemma must contain a p-pinch. Hence some $h_i \in A$ or some $k_i \in B$. So w was not reduced in $H \underset{A=B}{\star} K$. This proves the reduction theorem. So the study of many aspects of amalgamated free products reduces to the the study of HNN extensions.

3.6 Semi-direct products and wreath products

A subgroup H of a group G is a *retract* it there is a homomorphism $\rho : G \to H$ such that $\rho(h) = h$ for all $h \in H$. Such a ρ is called a *retraction* from G onto H. If we put $K = \ker(\rho)$, then K is a normal subgroup of G such that $G = KH$ and $K \cap H = 1$. Conversely given a pair of subgroups H and K satisfying these conditions, then H is a retract of G. We also say that G is the *split extension* of K by H since the exact sequence

$$1 \to K \to G \underset{\rho}{\rightleftarrows} H \to 1$$

splits because $\rho \circ$ inclusion is the identity on H.

Observe that H acts on K by conjugation $k \mapsto h^{-1}kh$. This defines a homomorphism $\theta : H \to \mathrm{Aut}(K)$, the group of automorphisms of K. Conversely, given two groups H and K and a homomorphism $\theta : H \to \mathrm{Aut}(K)$ we can construct a group $K \rtimes H$ as the set of pairs hk with $k \in K$ and $h \in H$ with multiplication defined by the formula:

$$(h_1k_1) \cdot (h_2k_2) = (h_1h_2)(\theta(h_1)(k_1)k_2).$$

It is routine to check that this is a group containing H and K as subgroups with $H \cap K = 1$. One can verify that $h^{-1}kh = \theta(h)(k)$ as expected. The group $K \rtimes H$ constructed in this way from external data is called the *semi-direct product* of K and H. Of course H is a retract of $K \rtimes H$ which is a split extension as above.

One important example of these kinds of groups occurs whenever we have a homomorphism $\psi : G \to F'$ where F' is a free group with basis $x'_1, x'_2, \ldots$. For each x'_i pick a lift $x_i \in G$ such that $\psi(x_i) = x'_i$. Then the subgroup F of G generated by $x_1, x_2, \ldots$ is free and is a retract of G under $\rho(g) =$ lift of $\psi(g)$. Thus any group G mapping onto a free group is a split extension the kernel by (a copy of) that free group. In particular, this is true for maps onto the infinite cyclic group.

We now describe another construction construction called the *wreath product* of two groups. Suppose that A and H are two groups. Begin by making an isomorphic copy A_h of A for each $h \in H$. Next form the (restricted) direct product of all of these groups $B = \oplus_{h \in H} A_h$. Now we can let H act on the right as an automorphism of B by defining $(a_h) \cdot g = a_{hg} \in A_{hg}$ where $g \in H$ and $a_h \in A_h$. Finally we define the *wreath product of A and H* by $A \wr H = A\mathrm{wr}H = B \rtimes H$. Observe that $A \wr H$ is generated by A_1 together with H. In particular if A and H are finitely generated, then $A \wr H$ is finitely generated.

To gain a little more insight into this construction, we consider the case in which both A and H are infinite cyclic groups, say $A = \langle a \rangle$ and $H = \langle t \rangle$. We write a_i instead of the more elaborate a_{t^i}. Then B can be presented as $B = \langle a_i (i \in \mathbb{Z}) \mid a_i a_j = a_j a_i (i, j \in \mathbb{Z}) \rangle$. Now t acts on B by sending $a_i \mapsto a_{i+1}$, so we can write down a presentation for the wreath product as

$$A \wr H = \langle a \rangle \wr \langle t \rangle = \langle a_i (i \in \mathbb{Z}), t \mid t^{-1} a_i t = a_{i+1}, a_i a_j = a_j a_i (i, j) \in \mathbb{Z}) \rangle.$$

But this group is clearly generated by a_0 and t and in terms of these gener-

ators we have the following presentation:

$$A \wr H = \langle a \rangle \wr \langle t \rangle = \langle a_0, t \mid a_0 t^{-i} a_0 t^i = t^{-i} a_0 t^i a_0 (i \in \mathbb{Z}) \rangle.$$

Exercise 3.9 *Show that* $\langle a \rangle \wr \langle t \rangle$ *as above is not finitely presented.*

Chapter 4

Properties, embeddings and examples

In this chapter we will utilize some of the constructions above to build new groups with interesting properties. In a sense we are exercising the mathematical muscles the results the preceding chapters have developed.

At the outset one might ask a number of naive questions such as the following. Is every subgroup of a finitely generated group, finitely generated? How many finitely generated groups are there (up to isomorphism of course)? Is every finitely generated subgroup of a finitely presented group again finitely presented? These questions among others will be answered in the next two sections.

4.1 Countable groups embed in 2-generator groups

Let $F = \langle a, b \mid \rangle$ be a free group on two generators a and b. Consider the set of elements $a^{-i}ba^i$ $(i \geq 0)$. One can show that any freely reduced word in these elements is not equal to 1 in F, because in forming such an expression and then reducing in F the central b of each term survives. Hence, by our characterization of freeness, these elements are a free basis for the subgroup they generate. (Notice that this answers one of our naive questions because it shows a subgroup of a finitely generated group need not be finitely generated.)

We use this observation to prove the following remarkable fact.

Theorem 4.1 (Higman, Neumann and Neumann) *Any countable group can be embedded in a group with two generators.*

For let C be a countable group with presentation $C = \langle c_1, c_2, \ldots \mid D \rangle$ on a countable set of generators. First form the group $L = C \star F$ where $F = \langle a, b \mid \rangle$ is the free group as above. Now the two subgroups

$$A = \langle b, c_1 a^{-1} b a, c_2 a^{-2} b a^2, c_3 a^{-3} b a^3, \ldots \rangle$$

$$B = \langle a, b^{-1} a b, b^{-2} a b^2, b^{-3} a b^3, \ldots \rangle$$

are both free with free bases the listed generators by our previous discussion. So we can form the HNN extension

$$G = \langle a, b, c_1, c_2, \ldots, t \mid D, t^{-1} b t = a, t^{-1} c_i a^{-i} b a^i t = b^{-i} a b^i \ (i \geq 1) \rangle$$

in which the stable letter t conjugates the basis for A to the basis for B.

We can rewrite these added defining relations of G to put them in the equivalent form

$$c_i =_G t b^{-i} a b^i t^{-1} a^{-i} b^{-1} a^i$$

so the group G is generated by $\{t, a, b\}$. But since $b = tat^{-1}$, the group G is even generated by a and t alone. So if we substitute tat^{-1} for b in the above we get equations of the form $c_i =_G u_i(a, t)$ where the u_i are (suitably complicated) words on a and t. Now the words in D are words on the c_i alone so if we rewrite them in terms of the u_i we obtain a new set of words, say $\overline{D}$. Applying Tietze transformations to eliminate the other symbols, it follows that G can be presented as $G \cong \langle a, t \mid \overline{D} \rangle$.

Now our previous results on free products imply that C is embedded in L and our results on HNN extensions imply that L is embedded in G. Hence C is embedded in G which is a two generator group. This completes the proof.

Notice that G can be presented with the same number of relations as C. Also observe that by properties of free products and HNN extensions, any element of finite order in G is conjugate to an element of C. Hence the group G has an element of finite order k if and only if C has and element of order k. So have actually proved slightly more, namely:

Corollary 4.2 *If C is a countable group having a presentation with n generators and m defining relations, then C can be embedded in a group G having 2 generators and m defining relations. Moreover, G can be constructed so that any element of finite order in G is conjugate to an element of C.*

Using this more detailed version, we can also determine how two generator groups there are (up to isomorphism).

Corollary 4.3 *There are continuously many non-isomorphic two generator groups.*

Observe that there are at most continuously many such groups since there are at most continuously many presentations on two generating symbols. To construct lots of non-isomorphic groups, we start with an arbitrary infinite set P of primes, say $P = \{p_1, p_2, p_3, \ldots\}$. Consider the group with presentation

$$C_P = \langle c_1, c_2, \ldots \mid c_1^{p_1} = 1, c_2^{p_2} = 1, \ldots, c_i^{p_i} = 1, \ldots \rangle$$

and let G_P be the group constructed for C_P in the previous corollary. Observe that C_P is just the free product of the infinitely many cyclic groups of order $p_i \in P$. Hence C_P and thus G_P contain an element of finite order k if and only if $k \in P$. So if Q is a different set of primes then G_P and G_Q are not isomorphic. Since there are continuously many ways to choose such a set of primes, there are continuously many such non-isomorphic G_P. This completes the proof.

Suppose that $G = \langle S \mid D \rangle$ is a group and that u and v are two words which have the same order as elements. Then we can form the HNN-extension $G \underset{\varphi}{\star} = \langle S, t \mid D, t^{-1}ut = v \rangle$ in which they are conjugate. More generally, if the pairs we have a set of such pairs u_i and v_i which have the same order, we can form the HNN extension

$$G \underset{\varphi}{\star} = \langle S, t_1, t_2, \ldots \mid D, t_1^{-1}u_1t_1 = v_1, t_2^{-1}u_2t_2 = v_2, \ldots \rangle$$

in which these pairs become conjugate. This observation may be helpful for the following exercise.

Exercise 4.1 (Higman,Neumann and Neumann) *Show that any countable group can be embedded in a countable group in which any two elements of the same order are conjugate. Also show that there is a countable torsion free group in which any two non-trivial elements are conjugate.*

4.2 Non-finite presentability of subgroups

We are going to give an example of a finitely presented group G having a finitely generated subgroup L which is not finitely presented. But to do this we need a method of showing that a group is not finitely presented. The following gives appropriate criteria for HHN extensions and amalgamated free products.

Theorem 4.4 (G. Baumslag[13])

1. *If $G = H \underset{M}{\star} K$ is an amalgamated free product where H and K are finitely presented groups, and if M is not finitely generated, then G is not finitely presented.*

2. *Let $G = \langle S \mid D \rangle$ be a finitely presented group with isomorphic subgroups via the isomorphism $\psi : A \to B$. If A is not finitely generated, then the corresponding HNN extension*

$$G\underset{\psi}{\star} = \langle S, t \mid D, t^{-1}at = \psi(a)\ (a \in A)\rangle$$

is not finitely presented.

We prove the second assertion. The proof of the first is similar using facts about amalgamated free products. We can assume that $G = \langle S \mid D\rangle$ is a finite presentation so that the given presentation of $G\underset{\psi}{\star}$ is finite except for the $t^{-1}at = \psi(a)$ relations. Assume on the contrary that $G\underset{\psi}{\star}$ is finitely presented. Then, by the extraction theorem, some finite subset $\{D, t^{-1}a_1t = \psi(a_1), \ldots, t^{-1}a_nt = \psi(a_n)\}$ of the given relations suffice.

Let $A_0 = \langle a_1, \ldots, a_n\rangle$ be the subgroup of A generated by the a_i that appear in these relations. Form the HNN extension of G with associated subgroups A_0 and $\psi(A_0)$ which can be presented as,

$$H = \langle S, t \mid D, t^{-1}a_1t = \psi(a_1), \ldots, t^{-1}a_nt = \psi(a_n)\rangle.$$

Since A was not finitely generated we know there is some $x \in A \setminus A_0$. Then by Britton's Lemma applied to H we know $t^{-1}xt\psi(x)^{-1} \neq_H 1$. But above we saw $t^{-1}xt\psi(x)^{-1} = 1$ is a consequence of the given relations, which is a contradiction. Hence $G\underset{\psi}{\star}$ is not finitely presented.

We are now going to construct an example of a finitely presented group (namely $F \times F$ where F is free of rank 2) which has a finitely generated subgroup L which is not finitely presented. We are going to use the following fact.

Exercise 4.2 *Let $F = \langle a, b \mid \rangle$ be the free group on a and b and consider the subgroup $K = \langle a^i b^{-i} \ (i \in \mathbb{Z}) \rangle$. Show that the listed elements are a free basis for K. Also show that K is normal and is in fact the kernel of the quotient map from F onto the group $\langle a, b \mid a = b \rangle$.*

Let F and K be as in the previous exercise. Form the HNN extension corresponding to the identity isomorphism on K which has presentation

$$L = \langle a, b, t \mid t^{-1} a^i b^{-i} t = a^i b^{-i} \ (i \in \mathbb{Z}) \rangle.$$

Since the associated subgroup K is not finitely generated, by the previous theorem L is not finitely presented.

Observe the K is a normal subgroup of L. If we let $\varphi : L \to G$ where $G = \langle a, b, t \mid a = b \rangle$ be the map induced by the identity on generators, then $K = \ker \varphi$. Observe that G is just a free group on two generators and φ is surjective. It is convenient to change notation slightly and write $G = \langle s, t \mid \rangle$ where we rename the image of a and b by the letter s.

Let $\psi : L \to F$ be the map which is the identity on F and sends the stable letter t to $1 \in F$. Clearly $\ker \psi \cap F = \{1\}$ so that $\ker \varphi \cap \ker \psi = \{1\}$. Hence the map $\gamma : L \to F \times G$ defined by $\gamma(x) = (\psi(x), \varphi(x))$ is a monomorphism and so L embeds in $F \times G$. In terms of the given generators we have $\gamma(a) = (a, s)$, $\gamma(b) = (b, s)$ and $\gamma(t) = (1, t)$. We summarize this as follows.

Theorem 4.5 *Let $F = \langle a, b \mid \rangle$ and $G = \langle s, t \mid \rangle$ be two free groups. Their direct product $D = F \times G$ is finitely presented, for instance by*

$$D \cong \langle a, b, s, t \mid as = sa, bs = sb, at = ta, bt = tb \rangle$$

but the subgroup L of D generated by the three elements $\{(a, s), (b, s), (1, t)\}$ is not finitely presented.

Later we will discuss a remarkable theorem of Graham Higman which actually characterizes those finitely generated groups which are subgroups of finitely presented groups.

4.3 Hopfian and residually finite groups

A group G is said to be *hopfian* if $G/N \cong G$ implies $N = \{1\}$, that is, every epimorphism $\alpha : G \to G$ is an automorphism.

Being hopfian has aspects of a finiteness property. Clearly any finite group is hopfian since a function from a finite set onto itself is a bijection. A free group F_S with an infinite basis $S = \{a_1, a_2, \ldots\}$ is not hopfian since $\alpha(a_1) = 1, \alpha(a_{i+1}) = a_i$ $(i \geq 1)$ defines a homomorphism $\alpha : F_S \to F_S$ which is surjective but not injective.

We will eventually see that all finitely generated free groups are hopfian and all finitely generated abelian groups are hopfian. It is also known that all finitely generated groups of matrices are hopfian. So one might ask whether all finitely presented groups are hopfian since they satisfy one sort of finiteness condition. The answer is know as the following simple example shows.

Theorem 4.6 (Baumslag-Solitar) *The group with presentation*

$$G = \langle a, t \mid t^{-1}a^2t = a^3 \rangle$$

is non-hopfian

To prove this we define $\psi(t) = t$ and $\psi(a) = a^2$. To see that ψ defines a homomorphism we observe that $\psi(t^{-1}a^2t) = t^{-1}a^4t =_G a^6 = \psi(a^3)$. Also since $a =_G t^{-1}a^2ta^{-2}$ the homomorphism ψ is surjective and hence an epimorphism.

Now $[t^{-1}at, a] \equiv t^{-1}a^{-1}ta^{-1}t^{-1}ata \neq_G 1$ by Britton's Lemma since there are no t-pinches possible. But

$$\psi([t^{-1}at, a]) = [t^{-1}a^2t, a^2] =_G [a^3, a^2] = 1$$

so that $[t^{-1}at, a]$ is a non-trivial element in the kernel of ψ. Hence ψ is not an isomorphism as desired.

Exercise 4.3 (G. Higman) *Show that the group with presentation*

$$H = \langle a, p, q \mid p^{-1}ap = a^2, q^{-1}aq = a^2 \rangle$$

is non-hopfian.

There is a large class of finitely generated groups which turn out to be hopfian, namely, the *residually finite* groups. Before proving this result, we briefly discuss residual properties in general.

Let $\mathcal{P}$ be a property of groups which is abstract in the sense that it depends only on the isomorphism type of the group and not on the way it is presented or defined. For example "being finite" is such a property. A group G is said to be *residually-$\mathcal{P}$* if for every $1 \neq g \in G$ there is a surjective homomorphism $\psi : G \to H$ where $H \in \mathcal{P}$ with $\psi(g) \neq_H 1$. Equivalently, if for every $1 \neq g \in G$ there is a a normal subgroup N_g of G such that $g \notin N_g$ and $G/N_g \in \mathcal{P}$. Thus the fact that $g \neq_G 1$ is witnessed in some quotient group of G which enjoys the property.

For example the infinite cyclic group $C = \langle a \mid \rangle$ is residually finite. To see this observe that if a^n $(n \neq 0)$ is any nontrivial element, then $a^n \notin \langle a^{2n} \rangle$ and $C/\langle a^{2n} \rangle$ is finite. Also any finite group is residually finite.

Exercise 4.4 *Show that the direct product of two residually-$\mathcal{P}$ groups is residually-$\mathcal{P}$*

A property $\mathcal{P}$ is said to be *hereditary* if every subgroup H of a group $G \in \mathcal{P}$ also has property $\mathcal{P}$. Examples of hereditary properties are "being finite", "being abelian", "being nilpotent" "being solvable" and so on.

Exercise 4.5 *Show that residually abelian is the same as abelian. Show that residually-(residually-$\mathcal{P}$) is the same as residually-$\mathcal{P}$.*

Exercise 4.6 *Show that if $\mathcal{P}$ is a hereditary property, the G is residually-$\mathcal{P}$ if and only if G is isomorphic to a subgroup of an (unrestricted) direct product of groups with property $\mathcal{P}$.*

We want to investigate residual finiteness, but to do so we need a few facts about subgroups of finite index in groups. We pose these as exercises.

Exercise 4.7 *Let G be a group and let H and K be two subgroups of finite index in G. Show that $H \cap K$ has finite index in G.*

Exercise 4.8 *Let G be a group and H a subgroup of finite index in G. Show that H contains a subgroup N which is normal in G and $[G : N] < \infty$.*

Exercise 4.9 *Show that if G is a finitely generated group and $1 \leq k \in \mathbb{N}$, then there are at most finitely many subgroups of G of index k.*

Exercise 4.10 *Show that G is residually finite if and only if the intersection of all of the subgroups of finite index in G is the trivial subgroup $\{1\}$.*

The following useful result provides more examples of residually finite groups.

Theorem 4.7 *If H and K are residually finite, then their free product $H \star K$ is residually finite.*

We first observe that it suffices to consider the case in which both H and K are finite groups. For if $w \equiv h_1 k_1 \dots h_n k_n \neq_{H\star K} 1$ is a reduced word, since H and K are residually finite they have normal subgroups finite index N and M respectively such that the $h_i \notin N$ and the $k_i \notin M$. Hence the image of w in the quotient group $(H/N) \star (K/M)$ is reduced and has the same length as w and so is not equal to 1. So it suffices to show this latter, which is the free product of two finite groups, is residually finite. Hence we can assume from now on that H and K are finite groups.

Now assume $w \equiv h_1 k_1 \dots h_n k_n \neq_{H\star K} 1$ is a reduced word of length m. Let Ω_m be the collection of all elements of $H \star K$ of length at most m, so that Ω_m is a finite set. We let H act on Ω_m by the following rule: if $u \in \Omega_m$ the $h \cdot u = hu$ if the length of hu after reduction is $\leq m$; otherwise $h \cdot u = u$. Once can check this is an action and therefore it defines a homomorphism $\alpha : H \to Sym(\Omega_m)$, the group of all permutations of Ω_m. Similarly define an action of K on Ω_m which gives a homomorphism $\beta : K \to Sym(\Omega_m)$. Hence, by the definition of free product, there is a homomorphism $\gamma : H \star K \to Sym(\Omega_m)$ which extends these. Thus the two actions extend to an action of $H \star K$ on Ω_m.

Now the element w acting on $1 \in \Omega_m$ yields w, that is $w \cdot 1 = w$ and so w acts non-trivially. Thus $\gamma(w) \neq 1$ in $Sym(\Omega_m)$ and this later is a finite group. Thus for any non-trivial element w we have found a homomorphism to a finite group which sends w to a non-trivial element. Hence $H \star K$ is residually finite. This completes the proof.

Since a free group is a free product of infinite cyclic groups, we conclude

Corollary 4.8 *Free groups are residually finite.*

Linear groups provide a rich source of residually finite groups because of the following which we state without proof.

Theorem 4.9 (Malcev) *A finitely generated linear group (group of matrices) is residually finite.*

At last we are ready to prove an assertion we made in connection with the hopfian property.

Theorem 4.10 (Malcev) *Finitely generated residually finite groups are hopfian.*

For suppose that G is finitely generated and residually finite. Let $\psi : G \to G$ be an epimorphism. Let H be a subgroup of finite index, say $n = [G : H]$. Then $\psi^{-1}(H)$ is again a subgroup of finite index n in G (which contains $\ker \psi$). If K is another subgroup of G having this same index n, so $H \neq K$, then $\psi^{-1}(H) \neq \psi^{-1}(K)$. Hence ψ^{-1} defines an injection from the set Θ_n of subgroups of G having index n into itself. But by the exercises above Θ_n is a finite set and so ψ^{-1} must be a bijection. Thus

$$\ker \psi \subseteq \bigcap_{H \in \Theta_n} H.$$

Hence $\ker \psi$ lies in the intersection of all of the subgroups of finite index in G. Since G is residually finite, that intersection is the trivial subgroup, so ψ is an isomorphism. This completes the proof.

4.4 Local and poly properties

Let P be a property of groups. A group L is said to be *locally-P* if every finitely generated subgroup of L has the property P. Thus a group is *locally finite* if every finitely generated subgroup is finite. For example, if C_n denotes the cyclic group of order n, then the infinite direct sum $A = \bigoplus_{i=1}^{\infty} C_n$ is an infinite, locally finite group. Similarly $\mathbb{Q}/\mathbb{Z}$ and S_ω, the group of all permutations of $\mathbb{N}$ which move only finitely many symbols are both locally finite groups.

Exercise 4.11 *Suppose that the group G has a normal subgroup N such that both N and G/N are locally finite. Show that G is locally finite.*

Observe that a locally abelian group is always abelian. In contrast to this observation, locally nilpotent (respectively solvable) groups need not be nilpotent (respectively solvable). To see this, let H_n be a sequence of nilpotent groups of class exactly n. Specifically we can take H_n to be the

group of upper triangular matrices in $GL(n+1,\mathbb{Z})$ with 1's on the diagonal. Then the direct sum $G = \bigoplus_{i=1}^{\infty} H_n$ is a locally nilpotent group which is not nilpotent. Since the H_n also have increasing derived length, it also follows that G is locally solvable but not solvable.

A group L is said to be *coherent* if all of its finitely generated subgroups are finitely related, that is if it is locally finitely presented. Locally finite groups are coherent. Since finitely generated abelian groups are a direct sum of cyclic groups, it follows that abelian groups are coherent. Also free groups are coherent (since subgroups of free groups are free - see Theorem 5.1). In Theorem 4.5 we showed that the direct product of two free groups is not coherent by exhibiting a finitely generated subgroup which is not finitely related.

The property P is said to be a *poly property* if P is preserved by extensions, that is N and G/N both having P implies G has P. For example being finite is a poly property and, by the previous exercise, being locally finite is a poly property. The following is an easy but useful fact.

Theorem 4.11 (P. Hall) *Being finitely presented (respectively finitely generated) is a poly property.*

Proof: Suppose that G is a group with a normal subgroup N such that N and G/N have presentations

$$N = \langle a_1, \ldots, a_n \mid r_1 = 1, \ldots, r_m = 1 \rangle$$

$$G/N = \langle x_1, \ldots, x_p \mid s_1 = 1, \ldots, s_q = 1 \rangle.$$

Pick elements of G, again denoted x_i, which map onto the x_i in G/N. Since N is a normal subgroup of G, there are words v_{ij} on $a_1, \ldots, a_n$ such that $x_i^{-1} a_j x_i = v_{ij}$ in G. Similarly there are word u_k in $a_1, \ldots, a_n$ such that $s_k = u_k$ in G. In terms of this data we can present G as

$$G = \langle a_1, \ldots, a_n, x_1, \ldots, x_p \mid r_1 = 1, \ldots, r_m = 1, s_k = u_k \ (1 \leq k \leq q)$$

$$x_i^{-1} a_j x_i = v_{ij} \ (1 \leq i \leq p, 1 \leq j \leq n) \ \rangle.$$

It is clear that the x_i's and a_j's generate G, and we leave to the reader the verification that the listed relations suffice to define G. □

Exercise 4.12 *Complete the details of the above proof by showing that any word in the given generators which is equal to the identity is a consequence of the listed relations.*

If P is a property, a group G is said to be *poly-P* if there is a finite series of subgroups $G = G_0 \rhd G_1 \rhd \cdots G_{n-1} \rhd G_n = 1$ such that G_i/G_{i+1} has property P. Observe that if P is a poly property, then "being poly-P" is the same as "being P" .

Exercise 4.13 *Show that a group is poly-abelian if and only if it is solvable.*

4.5 Finitely presented coherent by cyclic groups

The purpose of this section is to give a very simple proof of the following theorem due largely to Bieri and Strebel [16] and to deduce a number of consequences which we will describe in due course.

Theorem 4.12 *Suppose that G is a group having a normal subgroup L with G/L infinite cyclic. If G is finitely presented then G is an HNN extension of a finitely generated subgroup H of L with finitely generated associated subgroups. If in addition L is coherent, then G is an HNN extension of a finitely presented subgroup H of L with finitely presented associated subgroups.*

Proof: We can suppose G is presented in the form

$$G = \langle t, b_1, \ldots, b_n \mid r_1 = 1, \ldots, r_m = 1 \rangle$$

where the b_i represent elements of L and the r_j have exponent sum 0 on t. It follows that all of the r_j are freely equal to words in the elements $\beta_{ik} = t^{-k} b_i t^k$. So, replacing each of the r_j by $t^{-\ell} r_j t^{\ell}$ we can assume that all of the β_{ik} that arise have $k \geq 0$ by simply choosing ℓ sufficiently large.

Since the r_j are finite in number there is a maximum value, say δ of k required so that all the r_j are freely equal to words in the β_{ik} for $k = 0, \ldots, \delta$ and $i = 1, \ldots, n$.

We now perform Tietze transformations on our presentation to obtain a presentation of the form

$$G = \langle t, \beta_{ik} \ (1 \leq i \leq n, \ 0 \leq k \leq \delta) \mid q_1 = 1, \ldots, q_m,$$

$$t^{-1} \beta_{ik} t = \beta_{i \ k+1} \ (1 \leq i \leq n, \ 0 \leq k \leq \delta - 1) \rangle$$

where the q_j are the r_j expressed as words in the β_{ik}.

We denote the subgroup generated by the β_{ik} by H. Notice that H is finitely generated, but not necessarily finitely related. We may enlarge the

set of relations q_j so that they give a presentation for H on the generators β_{ik}. Now in H the subgroup C_0 generated by the elements β_{ik} $(1 \leq i \leq n,\ 0 \leq k \leq \delta - 1)$ is isomorphic to the subgroup C_1 generated by β_{ik} $(1 \leq i \leq n,\ 1 \leq k \leq \delta)$ since they are conjugate in G by t. Thus we have exhibited G as an HNN extension of the finitely generated group H. Under the additional assumption L is coherent, it is immediate that H, C_0 and C_1 are also finitely presented.

Recall that the HNN extension $G = \langle H, t \mid t^{-1}C_0t = C_1\rangle$ is said to be *ascending* if either $C_0 = H$ or $C_1 = H$. Let L be the normal closure of H in the HNN extension G. The structure of L is very different in the ascending and non-ascending cases.

If G is ascending with say $C_0 = H$ then we have

$$\cdots \supseteq t^2Ht^{-2} \supseteq tHt^{-1} \supseteq H \supseteq t^{-1}Ht \supseteq \cdots$$

so that $L = \cup_{n\in\mathbb{Z}} t^{-n}Ht^n$. Thus L is the union of copies of H.

On the other hand if G is not ascending so that $C_0 \neq H \neq C_1$ then L has the structure of a two-way infinite amalgamated free product

$$\cdots tHt^{-1} \underset{tC_1t^{-1}=C_0}{\star} H \underset{C_1=t^{-1}C_0t}{\star} t^{-1}Ht \underset{t^{-1}C_1t=t^{-2}C_0t^2}{\star} t^{-2}Ht^2 \cdots$$

where the amalgamations are proper (see for example [18] or [19]). From this one can easily deduce that L has non-abelian free subgroups. Here is a very elementary proof of this folklore fact which doesn't rely on this structural description.

Since the HNN extension is not ascending there are elements $h_0 \in H \setminus C_0$ and $h_0 \in H \setminus C_0$. Then of course $h_0^{-1} \notin C_0$ and $h_1^{-1} \notin C_1$. Define two elements $u = h_0t^{-2}h_1t^2$ and $v = th_0t^{-2}h_1t = tut^{-1}$. We claim the u and v freely generate a free subgroup of G. Consider the following six products (shown freely reduced):

1. $uv = h_0t^{-2}h_1t^2\ th_0t^{-2}h_1t$
2. $uv^{-1} = h_0t^{-2}h_1\ t\ h_1^{-1}t^2h_0^{-1}t^{-1}$
3. $vu = th_0t^{-2}h_1t\ h_0t^{-2}h_1t^2$
4. $v^{-1}u = t^{-1}h_1^{-1}t^2h_0^{-1}t^{-1}\ h_0t^{-2}h_1t^2$
5. $u^2 = uu = h_0t^{-2}h_1t^2\ h_0t^{-2}h_1t^2$

6. $v^2 = vv = th_0t^{-2}h_1t\ th_0t^{-2}h_1t$

Notice that only in case (2) has there been any free reduction and in that case the final t of u cancels with the initial t^{-1} of v^{-1}. In the other case the concatenation is freely reduced as written. Also observe that none of u, v nor these six products contains a t-pinch, that is a subword of the form $t^{-1}c_0t$ or tc_1t^{-1} with $c_i \in C_i$. So by Britton's lemma [**?**], none is equal to a word with fewer t symbols. Moreover it is clear that any non-empty freely reduced word $w(u, v)$ in u and v has only small free reductions from case (2) giving a non-trivial word involving t which is t-reduced an so not equal to 1. Again by Britton's lemma $w \neq 1$. Hence our claim holds, completing the proof.

Our discussion can be summarized as follows.

Lemma 4.13 *Suppose $G = \langle H, t \mid t^{-1}C_0t = C_1\rangle$ is an HNN extension and let L be the normal closure of H in G. Then*

1. *if G is ascending, then L is is a union of subgroups isomorphic to H;*
2. *if G is not ascending, then L contains non-abelian free subgroups.* □

Next we observe

Corollary 4.14 *Suppose that G is a group having a locally finite, normal subgroup L with G/L infinite cyclic. If G is finitely presented, then L is finite.*

Proof: Theorem 1 applies here and we continue the notation in its proof. Notice that the resulting H is finite. Now suppose that $C_1 \neq H$. Then $C_0 \neq H$ since C_0 and C_1 have the same order. So G is not ascending and Lemma 4.13 implies L contains non-abelian free subgroups. But by hypothesis L is locally finite, a contradiction. Hence $C_0 = C_1 = H$ and H is normal, so it must be all of L.

The following special case of Theorem 1 applies in particular to finitely presented free-by-cyclic groups:

Corollary 4.15 *Suppose that G is a group having a locally free, normal subgroup L with G/L infinite cyclic. If G is finitely presented, then G is an HNN extension of a finitely generated free subgroup H of L with finitely generated associated subgroups.* □

Corollary 4.16 *Suppose that G is a group having an abelian normal subgroup L with G/L infinite cyclic. If G is finitely presented, then G is an ascending HNN extension of a finitely generated abelian subgroup H of L.*

Except for the ascending assertion this is a special case of Theorem 1. Since such a group G must be solvable, the ascending assertion is an immediate consequence of Lemma 4.13.

Here is a generalization of the Corollary 5 which has essentially the same proof.

Corollary 4.17 *Suppose that G is a group having an locally polycyclic normal subgroup L with G/L infinite cyclic. If G is finitely presented, then G is an ascending HNN extension of a (finitely generated) polycyclic subgroup H of L.* □

By way of a concrete example of the last several corollaries, we recall the groups $BS_{1,p} = \langle t, b \mid t^{-1}bt = b^p \rangle$. They are visibly ascending HNN extensions of the infinite cyclic group. The normal closure of the element b in these groups is isomorphic to the additive group of the p-adic rationals $\mathbb{Z}[\frac{1}{p}]$ which is locally infinite cyclic.

Conversely suppose that G has a locally infinite cyclic, normal subgroup L with G/L infinite cyclic. Now locally infinite cyclic is the same as the conjunction of locally free and abelian. So Corollaries 4.15 and 4.16 imply that G is an ascending HNN extension on an infinite cyclic group H. Hence G must be one of the groups $BS_{1,p}$.

Finally we give here an example to show that the conclusion Theorem 1 when L is not coherent cannot be strengthened to conclude that H is finitely presented. To this end, let $H = \langle a \rangle \wr \langle s \rangle$ be the wreath product of the infinite cyclic group on a by the infinite cyclic group on s. Then H is not finitely presented by Corollary 4.16 (or see [12]).

Put $C_0 = H$ and $C_1 = gp(as^{-1}as, s)$. Then $C_1 = \langle as^{-1}as \rangle \wr \langle s \rangle$ is again the wreath product of two infinite cyclic groups and so we can form the HNN extension

$$G = \langle H, t \mid t^{-1}at = as^{-1}as, t^{-1}st = s \rangle$$

with associated subgroups C_0 and C_1. But now (by the following exercise or [14]) G is finitely presented.

Exercise 4.14 *Suppose* $H = \langle a \rangle \wr \langle s \rangle$ *and as above define*

$$G = \langle H, t \mid t^{-1}at = as^{-1}as, t^{-1}st = s \rangle.$$

Show that G *is finitely presented.*

Chapter 5

Subgroup Theory

We are going to discuss the structure of subgroups of various kinds of groups introduced above.

5.1 Subgroups of Free Groups

We first consider subgroups of free groups. The general case is easily dealt with, but in the case of a finitely generated subgroup algorithms and more information are available.

5.1.1 The general case

The fundamental group of a graph Y is easy to compute. We first choose a maximal tree T in Y. Contracting T to the base point yields a bouquet of circles at the base point, one circle for each edge of Y not in T. Hence the fundamental group of Y is free with basis the loops formed by connecting the edges not in T to the base point by paths in T from their vertices.

Using this observation we prove the following *Nielsen-Schreier* theorem.

Theorem 5.1 *Subgroups of free groups are free.*

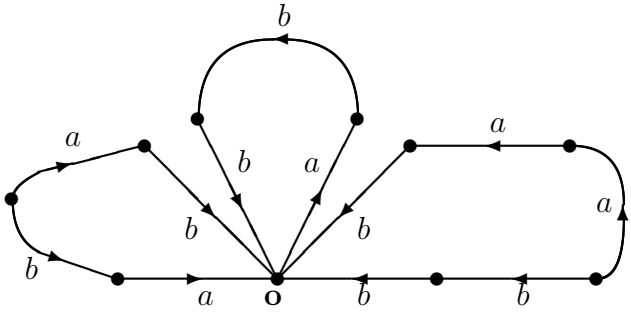
The fundamental group of any graph is free. A free F group is the fundamental group of a suitable graph X. A subgroup H of F is the fundamental group of a corresponding covering space Y of X. But Y is again a graph, so H is free.

5.1.2 Finitely generated subgroups of free groups

Let F be a free group and H the subgroup of F generated by a finite set of words $H = \langle w_1, \ldots, w_m \rangle$. We know that H is actually a free group, but the w_i may not be a free basis. There are graphical and computational methods available to construct a basis for H which give algorithms for deciding membership in H and computing coset representatives. It is convenient to discuss an extended example of these methods rather than give a theoretical description.

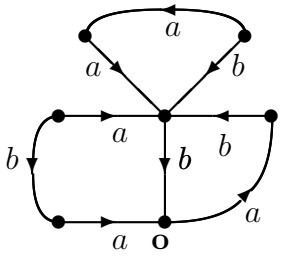
Suppose for example that we are interested in the free group F with two free generators a and b. Consider the subgroup H of F generated by the three words $w_1 = b^{-2}a^2b, w_2 = ab^2, w_3 = b^{-1}a^{-1}ba$.

We begin by forming three loops in the plane joined at a common vertex labeled **o**. We then subdivide each loop and orient and label the resulting edges according the the words w_i in the usual manner. This gives us the following initial (oriented and labeled) graph.

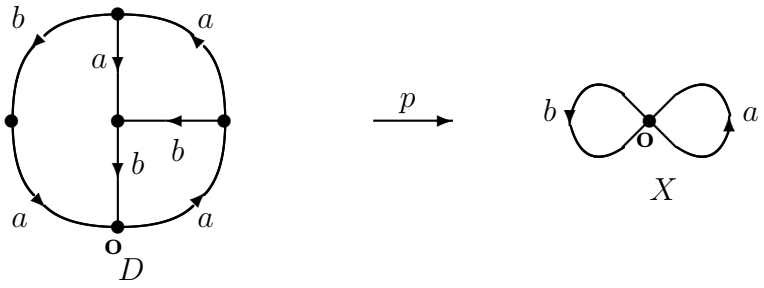


Initial graph

Now observe that there are four edges labeled b coming into the vertex **o**. We identify these edges and also identify the vertices they come from to reduce the graph to the following:



In this graph there are two edges labeled a and two edges labeled b entering the central vertex. If we first identify the b edges and then the a edges, we obtain (after two steps) the graph D shown at the left below.



This graph D now has the property that at any vertex at most one edge arriving at that vertex is labeled by a given generator and at most one edge leaving that vertex is labeled by a given generator. We say that such a graph is *reduced*.

Observe that there is a continuous map p from this graph D to the standard space X with fundamental group F consisting of two oriented loops at $\mathbf{o}$ labeled by a and b. The map p sends all the vertices to $\mathbf{o}$ and the oriented edges homeomorphically onto the corresponding oriented loops. Observe that the induced map on fundamental groups has $p_*(\pi_1(D,\mathbf{o})) = H \subseteq F = \pi_1(X,\mathbf{o})$ since the image is generated by the words w_i.

Let $\tilde{X}$ denote the universal covering space of X. So $\tilde{X}$ is a tree and at each vertex there is exactly one edge leaving and one edge entering labeled by each of the generators a and b. As it stands D is not a covering space of X. But we can enlarge D to a covering space Y of X by adding infinite branches from $\tilde{X}$ corresponding to missing edges. For example at $\mathbf{o}$ there is no outgoing edge labeled b. So in $\tilde{X}$ we remove an outgoing branch labeled b (that is, the component obtained by cutting that edge) and glue this branch on to D at $\mathbf{o}$. The map p extends to this added branch in the obvious way. Repeating this for each missing edge at each vertex of D we obtain a space Y and an extension of p. Now (Y,p) is a covering space since at each vertex there is exactly one edge leaving and one edge entering labeled by each of the generators a and b. Moreover, since Y is formed by attaching trees at various vertices of D, the graph D is a deformation retract of Y and $p_*(\pi_1(Y,\mathbf{o})) = p_*(\pi_1(D,\mathbf{o})) = H$.

In general, starting with a finite set of generators for a subgroup H of F, we obtain in this way a graph D_H and a covering space $Y_H \supseteq D_H$ cor-

responding to H. Now the index of H in F is the number of vertices of Y_H which is infinite unless $Y_H = D_H$. Thus the graph D_H constructed in this way is a covering space of X if and only if H has finite index in F.

Thus one way to think of D is as the *core* of the covering space Y corresponding to H obtained by removing all infinite tree branches. But there is also way to regard D as the graph of a finite state automaton for determining membership in H.

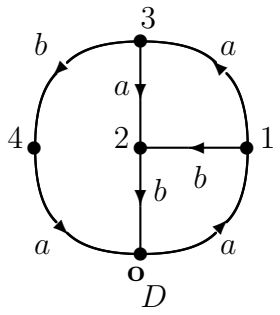
For suppose u is a freely reduced word in the generators of F. By covering space theory, $u \in H$ if and only if tracing out the word u starting from $\mathbf{o} \in Y$ along the corresponding labeled edges we return to $\mathbf{o}$ exactly at the end of u. Since u is freely reduced the corresponding path must must consist of a path in D possibly followed by a path off into one of the added branches at the end. Thus $u \in H$ if and only if u corresponds to a path entirely in D which returns to $\mathbf{o}$ at the end.

To determine membership of a freely reduced word u in H, we start tracing out u in D from $\mathbf{o}$. If there is ever no edge corresponding to the next symbol in u, then we know $u \notin H$ (in Y we would go off into one of the infinite branches here). If we finish the path corresponding to u and are not at $\mathbf{o}$, then again $u \notin H$. Finally if we finish at $\mathbf{o}$, then $u \in H$ so the test was successful.

Remark: D can also be used to compute a unique coset representative with respect to H for any word in F.

So the vertices of D can be regarded as the states of a finite state automaton and the edges as giving the transition instructions depending on the next symbol being read.

The information needed here is conveniently described in tabular form which we call the *automaton table*. First we number the vertices of D in some manner, for instance as shown below (with 0 the number of $\mathbf{o}$).



The table is to have one row for each vertex. The columns of the table correspond to the generators of F and their inverses. The number j in row i in the column headed by generator a means there is an edge with label a from vertex i to vertex j. The number j in row i in the column headed by inverse a^{-1} of a generator means there is an edge with label a from vertex j to vertex i. If there is no edge at vertex i with the given label a -1 is placed in the table.

The automaton table for the graph D of our example is shown at the left below:

Automaton Table

	Label			
Vertex	a	a^{-1}	b	b^{-1}
o	1	4	-1	2
1	3	**o**	2	-1
2	-1	3	**o**	1
3	2	1	4	-1
4	**o**	-1	-1	3

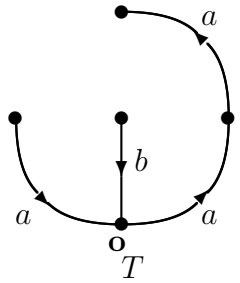
Edge Table

Edge	Label	Initial	Terminal
1	a	0	1
2	b	2	0
3	a	4	0
4	b	1	2
5	a	1	3
6	a	3	2
7	b	3	4

With any graph we can also associate what we term an *edge table*. This is a table with one row for each edge of the graph. Each edge is considered in the positive orientation of the labeling generator, and the row in the table for the edge contains the labeling generator and the numbers of the initial and terminal vertices. The edge table for our example D is shown at the right above.

It is computationally quite easy to pass between these two tables if the graph is reduced. To even write down the automaton table we must have a reduced graph. But the edge table makes sense in general, and the reduction process described before can be carried out on a computer using the edge table for the graph.

In order to write down a *Nielsen basis* for the subgroup H we first select a maximal tree T in D with root **o**. First we choose an edge from each vertex at distance 1 from **o** back to **o** - this gives a subtree T_1. Then choose an edge from each vertex at distance 2 from **o** back to T_1 - this gives a larger tree T_2; and so on. There is a simple procedure for doing this by passing through the above automaton table d times where d is the maximum distance of any vertex from **o**. Such a maximal tree for our example is shown below.



T

In such a maximal tree T there is a unique reduced path from any vertex to **o**. This information is conveniently recorded in the following table which completely describes T.

Vertex	Previous	Label	Distance
o	-	-	0
1	**o**	a^{-1}	1
2	**o**	b	1
3	1	a^{-1}	2
4	**o**	a	1

As usual, H is free on the omitted edges, and a Nielsen basis for H is

$$u_1 = a\underline{b}b, \quad u_2 = aa\underline{a}b, \quad u_3 = aa\underline{b}a$$

where the underlined symbol corresponds to the edge omitted from T. This underlined symbol is isolated in the sense it remains uncancelled when forming freely reduced words in the u_i.

Exercise 5.1 *Let F be the free group with free basis $\{a, b\}$ and consider the subgroup $H = \langle aba^2, ba, a^2b, b^2, a^3\rangle$ generated by the listed elements. Determine the automata graph D_H of H. Also find the corresponding automaton table and edge table. Finally find a Nielsen basis for H. What is the index of H in F? Is H normal?*

Exercise 5.2 *Repeat the previous exercise for the subgroup*

$$K = \langle a^2, ab^2, ab^{-2}, b^{-1}ab^{-1}, b^{-1}a^{-1}b^{-1}\rangle.$$

Exercise 5.3 *If L is a subgroup of index 2 in F, use the geometry of D_L to explain why L is normal. Also, by drawing an appropriate covering space, show that if $2 < n \in \mathbb{Z}$, there is a subgroup of index n which is not normal.*

5.1.3 More on subgroups of free groups

We use the methods of the previous section collect few more useful facts about finitely generated subgroups of free groups.

First, suppose that $N \neq 1$ is a finitely generated normal subgroup of F. Construct the corresponding finite graph D_N as in the previous section. Then $D_N \subseteq Y_N$ where Y_N is the covering space of X corresponding to N which is obtained by adding infinite tree branches to certain vertices of D_N. Now $D_N = Y_N$ if and only if N has finite index in F and no infinite branches need to be added. By covering space theory, F/N acts transitively on Y_N as a group of homeomorphisms, so Y_N must look the same viewed from each vertex.. But this is not possible if Y_N contains an infinite tree branch because there are closed loops at the base point of length at most $2k + 1$ where k is the number of vertices of D_N. This proves the following fact:

Theorem 5.2 *A non-trivial, finitely generated, normal subgroup of a free group has finite index.* □

If we are given two finite sets of elements of F which generate subgroups H and K, we can find the corresponding reduced automata graphs D_H and D_K as above. We also have two maps $p_H : D_H \to X$ and $p_K : D_K \to X$.

Now form the pull-back Z of these two maps. Then Z is a graph with vertex set the collection of pairs (u, v) where u is a vertex of D_H and v is a vertex of D_K. There is an edge labeled by a generator a from vertex (u_1, v_1) to vertex (u_2, v_2) in Z if and only if there is both an edge labeled by a in D_H from u_1 to u_2 and an edge labeled by a in D_K from v_1 to v_2.

The component of the vertex $(\mathbf{0}_H, \mathbf{0}_K)$ in the pull-back graph Z of these two graphs is then a reduced graph which accepts exactly those words in $H \cap K$.

Note that Z may have vertices of degree 1. For instance, if $H = \langle a^2 \rangle$ and $K = \langle ab \rangle$ so $H \cap K = 1$, one finds that Z is a graph with 2 vertices joined by a single edge with label a. But after successively removing all vertices of degree 1 and their incident edges, we obtain the the graph $D_{H \cap K}$. In particular, $H \cap K$ is finitely generated and this gives an algorithm for finding its automata graph and hence a set of generators. In particular we have proved the following.

Theorem 5.3 (Howson) *The intersection of two finitely generated subgroups of a free group is again finitely generated.* □

Exercise 5.4 *Let F, H and K be as in the previous exercises. Determine the automata graph $D_{H\cap K}$ of their intersection. Also find the corresponding automaton table and edge table. Finally find a Nielsen basis for $H \cap K$. What is the index of $H \cap K$ in F?*

Suppose that H is a finitely generated subgroup of the free group F. As in the previous section we associate with H a graph D_H and corresponding automaton and edge tables. Observe that an edge of D_H with label a generator, say a, gives rise to exactly one non-negative entry in each of the a and a^{-1} columns of the automaton table. Hence, for any -1 in the a column, there must be a (not necessarily unique) -1 in the a^{-1} column, and vice-versa.

Suppose the $u_1, \ldots, u_m$ is a basis for H, say computed using D_H. In light of the above observations we successively add new edges to D_H with appropriate labels until we obtain a covering space of D_L of X corresponding to a subgroup L of finite index. The vertices of D_L are the same as those of D_H and if we use the same tree to compute a basis for L we see that L has basis $u_1, \ldots, u_m, v_1, \ldots, v_k$ where each v_i corresponds to a unique added edge. If we define $k = \langle v_1, \ldots, v_k \rangle$, it follows that L decomposes as the ordinary free product $L = H * K$. So we have shown the following:

Theorem 5.4 (M. Hall [17] - see also [21]) *If H is a finitely generated subgroup of the free group F, then H is a free factor of a subgroup of finite index in F.* □

Suppose that a group G is generated by n elements. Then G is the homomorphic image of a free group on n generators, say $F = \langle a_1, \ldots, a_n \mid \rangle$ with $\varphi : F \to G$. If H is a subgroup of index k in G, then $\varphi^{-1}(H)$ is a subgroup of index k in F. Moreover the isomorphism theorems tell us that φ^{-1} gives a bijection between the subgroups of G of index k and the subgroups of F of index k which contain $\ker\varphi$. We have seen that the subgroups of index k in F correspond uniquely to certain reduced graphs D with k vertices. There are clearly only finitely many such graphs, so we conclude the following:

Proposition 5.5 *If G is a finitely generated group and $k > 0$, then the number of subgroups of G having index exactly k is finite.* □

5.1.4 Subgroups of a direct product with a free group

In an earlier section we gave an example of a finitely generated subgroup of a direct product of two free groups which was not finitely presented (Theorem 4.5). It turns out this is not at at all accidental. Baumslag and Roseblade [15] have shown that, with only the "obvious" exceptions, the finitely generated subgroups of a direct product of free groups are never finitely presented. Using the results of the previous section, we can now give a very simple proof of this result [20].

Theorem 5.6 *Let $A \times F$ be the direct product of a group A with a free group F. Suppose that $G \leq A \times F$ is a subgroup which intersects F non-trivially.*

1. *If G is finitely generated, then G has a subgroup G_0 of finite index which is an HNN-extension of the form*

$$G_0 = \langle D, t \mid t^{-1}bt = b,\ \forall b \in L \rangle$$

   where D is finitely generated and $L = G \cap A \leq D$.

2. *If G is finitely presented, then $L = G \cap A$ is finitely generated.*

Proof: Since subgroups of free groups are free, we may assume that both of the projection induced maps $p_A : G \to A$ and $p_F : G \to F$ are surjective. Since G intersects F non-trivially, there is some $1 \neq t \in G \cap F$. By Marshall Hall's Theorem 5.4, F contains a subgroup M of finite index which has the cyclic group $\langle t \rangle$ as a free factor. Thus M has a basis of the form $\{t, s_1, s_2, \ldots, s_n\}$. For each $i = 1, \ldots, n$ pick a lift $\hat{s}_i \in p_F^{-1}(s_i)$.

If we put $L = A \cap G$, then $G_0 = p_F^{-1}(M)$ is a subgroup of finite index in G which has the structure of an HNN-extension

$$G_0 = \langle L, t, \hat{s}_1, \ldots, \hat{s}_n \mid t^{-1}bt = b, \hat{s}_i^{-1}b\hat{s}_i = \phi_i(b), i = 1, \ldots, n, \forall b \in L \rangle$$

where ϕ_i is the automorphism of L induced by conjugation by $\hat{s}_i$.

Suppose now that G is finitely generated (so that $n < \infty$ in the above). Then G_0 is also finitely generated and hence is generated by a finite set of elements $a_1, \ldots, a_k$ in L together with $t, \hat{s}_1, \ldots, \hat{s}_n$. Notice that since t acts trivially by conjugation, L is generated by the a_i's together with their conjugates by words in the $\hat{s}_i$. Let D be the subgroup generated by the a_i

and $\hat{s}_i$, that is $D = \langle a_1, \ldots, a_k, \hat{s}_1, \ldots, \hat{s}_n \rangle$. Of course L is a subgroup of D. Now G_0 also has the structure of an HNN-extension of D, namely

$$G_0 = \langle D, t \mid t^{-1}bt = b, \ \forall b \in L \rangle.$$

Assume now that G is finitely presented. Then G_0 is also finitely presented since it has finite index in G. So Baumslag's extraction principle for HNN extensions (Theorem 4.4) implies that L must also be finitely generated. This completes the proof of the theorem. □

The theorem of Baumslag and Roseblade is an easy consequence of Theorem 5.6.

Corollary 5.7 (Baumslag and Roseblade [15]) *Let $F_1 \times F_2$ be the direct product of two free groups F_1 and F_2. Suppose that $G \leq F_1 \times F_2$ is a subgroup and define $L_i = G \cap F_i$.*

1. *If either $L_i = 1$ then G is free.*
2. *If both L_i are non-trivial and one of them is finitely generated, then $L_1 \times L_2$ has finite index in G.*
3. *Otherwise, G is not finitely presented.*

Proof: Since subgroups of free groups are free, we can assume both of the induced projections $p_i : G \to F_i$ are surjective. The subgroups L_i are normal in G and, since the p_i are surjective, each L_i is also normal in F_i. If one of the L_i is trivial, say $L_1 = 1$, then the projection p_2 is an isomorphism and hence G is free.

Suppose both L_i are non-trivial and one of them, say L_1, is finitely generated. Then L_1 is a non-trivial, finitely generated normal subgroup of F_1 and hence has finite index. Thus $p_1^{-1}(L_1) = L_1 \times L_2$ has finite index in G.

The final assertion now follows immediately from Theorem 5.6. □

5.2 Subgroups of presented groups

Suppose we are given a group G by a presentation. It is convenient to write this presentation as

$$G = \langle a_1, a_2, \ldots \mid r_1 = 1, r_2 = 1, \ldots \rangle$$

although the sets of generators and relations need not be countable. Our notation will be cumulative for this section and somewhat at variance from that used in other sections.

Suppose that H is a subgroup of G. We want to find a presentation for H. To this end, we inductively choose a set of words T in the generators of G which are a transversal for the right cosets Hx of H in G. First we choose the empty word 1 as the representative of H. By the *length* of a coset Hx we mean the length of the shortest word in the generators of G representing an element of Hx. Suppose that representatives have been chosen for all cosets of length n. For each coset of length $n+1$, select an element of length $n+1$, say $a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n} a_{i_{n+1}}^{\epsilon_{n+1}}$. Then as representative of this coset, choose the element

$$\overline{a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}} a_{i_{n+1}}^{\epsilon_{n+1}}$$

where $\overline{a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}}$ is the representative already chosen for $a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}$.

If $x \in G$ we denote the coset chosen representative of x by $\overline{x}$ so $\overline{x} \in T$. We use the letters K and M as variables ranging over these the chosen representatives in T. Observe that by construction an initial segment of a representative is again a representative.

Topological interpretation: We can realize G in the standard way as the fundamental group of a 2-complex X with a single 0-cell, so $G \cong \pi_1(X, \mathbf{o})$. Let Y be the covering space corresponding to H so $H = \pi_1(Y, \tilde{\mathbf{o}})$. We then choose an "expanding maximal tree" T based at $\tilde{\mathbf{o}}$. Each coset of H corresponds to a 0-cell of Y, and the unique reduced path in T joining $\tilde{\mathbf{o}}$ to a 0-cell can be taken as a coset representative. Again an initial segment of a representative is a representative.

We want to find a presentation for H. To this end, for each $K \in T$ and each generator a_i of G we introduce a new symbol s_{K,a_i}. Each such symbol is to represent the element $Ka_i(\overline{Ka_i})^{-1}$ of H. We now define a *rewriting process* τ which assigns to each word in the a_i a corresponding word in the s_{K,a_i}. Given a word $w \equiv a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}$ in the generators of G, we define $\tau(w)$ as follows: $\tau(w)$ is the word obtained from w by replacing the j-th symbol $a_{i_j}^{\epsilon_j}$ by the following rule: if $\epsilon_j = 1$ replace $a_{i_j}^{\epsilon_j}$ by in $s_{K,a_{i_j}}$ where

$$K = \overline{a_{i_1}^{\epsilon_1} \dots a_{i_{j-1}}^{\epsilon_{j-1}}},$$

and if $\epsilon_j = -1$ replace $a_{i_j}^{\epsilon_j}$ by in $s_{K,a_{i_j}}^{-1}$ where

$$K = \overline{a_{i_1}^{\epsilon_1} \dots a_{i_j}^{\epsilon_j}}.$$

Remark: In the topological interpretation the path corresponding to $Ka_i(\overline{Ka_i})^{-1}$ goes out along K in the maximal tree T, then across the edge labeled a_i then back to $\tilde{\mathbf{o}}$ in the tree T. If the edge labeled a_i is in T then the path back to $\tilde{\mathbf{o}}$ is the same as that going out, so the whole path is homotopically trivial. Otherwise it is a genuine loop in the 1-skeleton of $\tilde{Y}$. The set of such loops generate $\pi_1(Y, \tilde{\mathbf{o}})$. The effect of τ is to express any path as a product of these followed by the unique path in T back to $\tilde{\mathbf{o}}$. If the path is closed, it expresses the path in terms of the generators for $\pi_1(Y, \tilde{\mathbf{o}})$.

Observe that it can happen that Ma_i and $\overline{Ma_i}$ are freely equal (topologically they are homotopic in T). In this case we must have $s_{M,a_i} =_H 1$.

With the above notation we are now ready to specify how to present subgroups of groups given by presentations.

Theorem 5.8 (Reidemeister-Schreier) *Suppose that G is a group given by a presentation*

$$G = \langle a_1, a_2, \ldots \mid r_1 = 1, r_2 = 1, \ldots \rangle$$

and that H is a subgroup of G. Choose a transversal T and introduce a rewriting process τ as above. Then H can be presented as follows:

1. generators*: s_{K,a_i} for each $K \in T$ and each a_i*
2. relators*: $s_{M,a_i} = 1$ whenever Ma_i and $\overline{Ma_i}$ are freely equal; and $\tau(Kr_jK^{-1}) = 1$ for each $K \in T$ and each relator r_j in the presentation for G.*

In case the index of H in G is finite (so the number of $K \in T$ is finite) and the number of a_i's is finite (so G is finitely generated) we can draw the following consequences.

Corollary 5.9 *1. If G is a finitely generated group and H a subgroup of finite index, then H is finitely generated.*

2. If G is a finitely presented group and H a subgroup of finite index, then H is finitely presented.

Here are some exercises in carrying out the Reidemesister-Schreier procedure for presenting subgroups.

Exercise 5.5 *Let $G = S_3 = D_3 = \langle a, b \mid a^2 = b^3 = 1, aba = b^{-1} \rangle$ which is the non-abelian group of order 6. Use the Reidemeister-Schreier method to obtain a presentation for $H = \langle b \rangle$ the normal subgroup of order 3. Then use Tietze transformations to simplify the presentation you get. Repeat the exercise for the subgroup $\langle a \rangle$ which has order 2 but is not normal.*

Exercise 5.6 *Let $G = \langle a, b \mid b^{-1}ab = a^2 \rangle$ and let H be the normal closure of a, that is the kernel of the map from G onto the infinite cyclic group $\langle b \rangle$. Use the Reidemeister-Schreier method to obtain a presentation for H. Then use Tietze transformations to simplify the presentation you get.*

Exercise 5.7 *Let F be the free group with basis $\{a, b\}$ and let $H = [F, F]$ be its commutator subgroup. Find a free basis for H by using the Reidemeister-Schreier method to obtain a presentation for H.*

5.3 Subgroups of free products

The subgroups of an (ordinary) free product have a particularly simple structure as described in the following.

Theorem 5.10 (Kurosh Subgroup Theorem) *Let $G = \underset{i \in I}{\star} H_i$ be the free product of a collection of groups H_i. If A is a subgroup of G, then A decomposes as a free product of the form*

$$A = F \star \left(\underset{i \in I}{\star} \left(\underset{j \in J(i)}{\star} A \cap u_j H_i u_j^{-1} \right) \right)$$

where F is a free group. That is, A is the free product of a free group and of various subgroups which are the intersections of A with conjugates of the H_i.

A fairly straight forward proof using covering spaces can be found in Massey's text [6]. It also follows from the more general results described in the next section.

Note that applying the Kurosh Subgroup Theorem to the free product of infinite cyclic groups implies that subgroups of free groups are free. Here are a few other special cases.

Corollary 5.11 *If a subgroup A of a free product $G = \underset{i \in I}{\star} H_i$ intersects each conjugate of an H_i trivially, then A is a free group. In particular if A is a normal subgroup which intersects each H_i trivially, then A is free.*

Corollary 5.12 *The kernel of the epimorphism $H \star K \to H \times K$ is a free group.*

5.4 Groups acting on trees

The subgroup structure of amalgamated free products and HNN extensions is more difficult to describe. In fact one considers a more general notion of a *graph of groups* which is a graph together with a collection of groups corresponding to vertices and a collection of (sub)groups corresponding to edges end embeddings of the edge groups into the vertex groups of their initial and terminal vertices. There is a notion of the *fundamental group* of such a graph of groups; ordinary free products, amalgamated free products and HNN extensions are all special cases.

By considering the universal cover, such a graph of groups acts on a suitable tree (all actions here are without inversions). The stabilizers of vertices are the vertex groups and the stabilizers of edges are the edge groups. Conversely any group acting on a tree has a description in these terms. In particular, a group is free if and only if it acts freely on a tree. Hence subgroups of free groups are free.

Using these methods rather detailed structural information about subgroups of all these objects can be obtained. The reader is referred to the book by Serre [11] or the notes by Scott and Wall [10] for two different (but related) accounts of this theory.

Chapter 6

Decision Problems

6.1 The word and conjugacy problems

A finite presentation π of a group is a piece of notation such as

$$\langle x_1, \ldots, x_n \mid r_1 = 1, \ldots, r_m = 1 \rangle$$

where the x_i are letters in some fixed alphabet and the r_j are words in the x_i and their inverses x_i^{-1}. The group presented by π, denoted $gp(\pi)$, is the quotient group of the free group on the x_i by the normal closure of the r_j. Usually it is not necessary to distinguish so carefully between a group and its presentation and we often write simply

$$G = \langle x_1, \ldots, x_n \mid r_1 = 1, \ldots, r_m = 1 \rangle$$

to mean the G is the group defined by the given presentation.

It is convenient to introduce some notation for several decision problems we will consider. Suppose that G is a finitely presented group defined by a presentation as above. Then the *word problem* for G is the decision problem

$$WP(G) = (?w \in G)(w =_G 1).$$

Here the "?" is intended as a sort of quantifier and should be read as "the problem of deciding for an arbitrary word w in G whether or not ...." A closely related problem is the *equality problem*:

$$EqP(G) = (?w_1, w_2 \in G)(w_1 =_G w_2).$$

Of course, $w_1 =_G w_2$ if and only if $w_1 w_2^{-1} =_G 1$ so that an algorithm for solving either of $WP(G)$ or $EqP(G)$ easily yields an algorithm for solving the other. On the other hand, from the viewpoint of computational complexity, these problems are subtly different.

Again using this "?" quantifier, the *conjugacy problem* for G is

$$CP(G) = (?u, v \in G)(\exists x \in G)(x^{-1}ux =_G v).$$

If H is a finitely generated subgroup of G and if H given by say a finite set of words which generate it, then the *generalized word problem* for H in G is the problem of deciding for an arbitrary word w in G whether or not w lies in the subgroup H, that is

$$GWP(H, G) = (?w \in G)(w \in H).$$

When the subgroup H is an arbitrary finitely generated subgroup rather than a fixed one we write simply $GWP(G)$.

On the face of it, each of these algorithmic problems appears to depend on the given presentation. We will show below that the solvability of each of these problems is independent of the finite presentation chosen. It can happen that for a particular finitely presented group each of the above problems is solvable. For instance, if G is a finite group given by a multiplication table presentation, it is easy to describe algorithms for solving $WP(G)$, $CP(G)$ and $GWP(G)$. Similarly, if $F = \langle x_1, \ldots, x_n \mid \rangle$ is a finitely generated free group $WP(F)$ is solved by freely reducing and $CP(F)$ is solved by cyclically permuting and freely reducing. The $GWP(H, F)$ for finitely generated subgroups H of F is more difficult and its solution is due to Nielsen (see [5]).

Finally, in terms of the "?" notation, the *isomorphism problem* for finitely presented groups is

$$IsoP = (?\pi_1, \pi_2 \text{ finite presentations})(gp(\pi_1) \cong gp(\pi_2)).$$

We assume the reader is familiar with the rudiments of the theory of algorithms and recursive functions. Thus a set of objects is *recursive* if there is an algorithm for deciding membership in the set. A set S of objects is *recursively enumerable* if there is an algorithm for listing all the objects in S. It is easy to see that every recursive set is recursively enumerable. Moreover, a set S is recursive if and only if both S and its complement are recursively enumerable. A diagonal argument can be used to prove the important result

that there exists a set which is recursively enumerable but not recursive. This fact is in a sense the source of all undecidability results in mathematics.

Each of the above decision problems is recursively enumerable in the sense that the collection of questions for which the answer is "Yes" is recursively enumerable. For instance, the set of words w of G such that $w =_G 1$ is recursively enumerable. For it is the set of words freely equal to a product of conjugates of the given finite set of defining relations and this set can (in principle) be systematically listed. Thus $WP(G)$ is recursively enumerable. Now $WP(G)$ is recursively solvable (decidable) exactly when the set of words $\{w \in G \mid w =_G 1\}$ is recursive. So $WP(G)$ is recursively solvable if and only if $\{w \in G \mid w \neq_G 1\}$ is recursively enumerable.

Similarly, one can systematically list all true equations between words of G and all true conjugacy equations so that $EqP(G)$ and $CP(G)$ are recursively enumerable. $GWP(H, G)$ is recursively enumerable since one can list the set of all true equations between words of G and words in the generators of H. Finally, if two presentations present isomorphic groups, then one can be obtained from the other by a finite sequence of Tietze transformations. Since the set of presentations obtainable from a given one by a finite sequence of Tietze transformations is recursively enumerable, it follows that $IsoP$ is recursively enumerable.

We recall the notion of Turing reducibility. If A and B are two sets of objects, we write $A \leq_T B$ if an (hypothetical) algorithm to answer questions about membership in B would yield an algorithm to answer questions about A. Thus the decision problem for A is reducible to that for B. One way to make this precise is through the theory of recursive functions. Recursive functions can be defined as the collection of functions obtained from certain base functions (like multiplication and addition) by closing under the usual operations of composition, minimalization and recursion. A function is said to be B-recursive if it is among the functions obtained from the base functions together with the characteristic function for B by closing under the usual operations. Then $A \leq_T B$ is defined to mean that that the characteristic function of A is B-recursive. Of course, if B is already recursive (that is, membership in B is decidable) and if $A \leq_T B$ then A is also recursive.

Now the relation $\leq_T$ is a partial order so we can form the corresponding equivalence relation. Two sets of objects A and B are *Turing equivalent* $A \equiv_T B$ if each is Turing reducible to the other, that is both $A \leq_T B$ and $B \leq_T A$. In terms of this notation there are some obvious relationships

among our decision problems:

$$EqP(G) \equiv_T WP(G) \leq_T CP(G)$$

$$WP(G) \equiv_T GWP(1, G) \leq_T GWP(G).$$

We have already observed the first equivalence. Since $w =_G 1$ if and only if w and 1 are conjugate in G it follows that $WP(G) \leq_T CP(G)$. The other assertions are clear.

A *recursive presentation* is a presentation of the form

$$\langle x_1, \ldots, x_n \mid r_1 = 1, r_2 = 1, \ldots \rangle$$

where $r_1, r_2, \ldots$ is a recursively enumerable set of words. A finitely generated group G is *recursively presented* if it has a recursive presentation. Of course finitely presented groups are recursively presented but the converse is false. The word problem and conjugacy problem are defined for recursively presented groups as before and they are still recursively enumerable problems.

Lemma 6.1 *Let G be a finitely generated group given by a recursive presentation*

$$G = \langle x_1, \ldots, x_n \mid r_1 = 1, r_2 = 1, \ldots \rangle.$$

Suppose that H is a finitely generated group with generators $y_1, \ldots, y_m$ and that $\phi : H \to G$ is an injective homomorphism. Then H has a recursive presentation of the form

$$H = \langle y_1, \ldots, y_n \mid q_1 = 1, q_2 = 1, \ldots \rangle$$

where $q_1, q_2, \ldots$ is a recursively enumerable set of words in $y_1, \ldots, y_m$. Moreover, $WP(H) \leq_T WP(G)$.

Proof: Let $F = \langle y_1, \ldots, y_m \mid \rangle$ be the free group with basis $y_1, \ldots, y_m$. Now we can write $\phi(y_i) = u_i$ $(i = 1, \ldots, m)$ where the u_i are certain words on $x_1, \ldots, x_n$. There is then a unique homomorphism $\psi : F \to G$ such that $\psi(y_i) = u_i$ $(i = 1, \ldots, m)$ and since ϕ is injective we have $H \cong F/ker\ \psi$. Now the set of all formal products of the words u_i and their inverses is a recursively enumerable set of words of G. The set of words of G equal to the identity is also recursively enumerable. Hence the intersection of these two sets is a recursively enumerable set of words, and it follows that $ker\ \psi$ is a

recursively enumerable set of words on $y_1, \ldots, y_m$. The first claim follows by taking $q_1, q_2, \ldots$ to be a recursive enumeration of $ker\ \psi$.

For the second claim, suppose that we have an algorithm A_G to solve the word problem for G. We describe an algorithm to solve the word problem for H as follows: let $w(y_1, \ldots, y_m)$ be an arbitrary word in the generators of H. Since ϕ is injective, $w =_H 1$ if and only if $\phi(w) =_G 1$. Now $\phi(w) = w(u_1, \ldots, u_m)$ so we can apply the algorithm A_G to decide whether or not $w(u_1, \ldots, u_m) =_G 1$. If so, then $w =_H 1$; if not, then $w \neq_H 1$. This algorithm solves the word problem for H. Thus $WP(H) \leq_T WP(G)$ completing the proof.

Lemma 6.2 *For finitely presented groups (respectively finitely generated, recursively presented groups), the word problem, conjugacy problem and generalized word problem are algebraic invariants. That is, for any two presentations π_1 and π_2 of the same group on a finite set of generators, $WP(\pi_1) \equiv_T WP(\pi_2)$, $CP(\pi_1) \equiv_T CP(\pi_2)$ and $GWP(\pi_1) \equiv_T GWP(\pi_2)$.*

Proof: The proof is in each case similar to the proof of the second part of the previous lemma except that ϕ is an isomorphism. We omit the details.

The main local unsolvability result is the following:

Theorem 6.3 (Novikov-Boone) *There exists a finitely presented group whose word problem is recursively unsolvable.*

The original proofs of this result proceed along the following lines: start with a Turing machine T whose halting problem is unsolvable. That is, the problem of deciding whether the machine started with an arbitrary tape in a certain state will eventually halt is unsolvable. Constructions of Markov and of Post, associate to such a Turing machine a certain semigroup $S(T)$ whose defining relations mimic the transition rules defining the Turing machine T. They show a code word incorporating a tape and state of T is equal in $S(T)$ to a particular fixed halting word, say q_0, if and only if T halts when started with that tape and state.

Groups $G(T)$ having unsolvable word problem are constructed by in turn mimicking the defining relations of $S(T)$ inside a group. The construction is not so direct as the Markov-Post construction and involves starting with free groups and performing a number of HNN-extensions and/or free products with amalgamation. Nevertheless, there is a direct coding of a tape and

state of T as a word w of $G(T)$ so that $w =_{G(T)} 1$ if and only if the machine T halts when started with that tape and state. Since T has an unsolvable halting problem, it follows that $G(T)$ has unsolvable word problem.

A readable account of the Novikov-Boone Theorem along these lines can be found in the textbook by Rotman [9].

In view of the previously noted relationships among our various decision problems, the Novikov-Boone Theorem has the following immediate corollary:

Corollary 6.4 *There exists a finitely presented group G such that $WP(G)$, $CP(G)$ and $GWP(G)$ are all recursively unsolvable.*

We turn now to briefly consider other local decision problems concerning elements in a group.

The structure of finitely generated abelian groups can be completely determined from a finite presentation of such a group, and in particular one can solve the word problem for such groups. Consequently, if G is an arbitrary finitely presented group one can effectively determine the structure of its abelianization $G/[G,G]$. So for instance, there is an algorithm to decide whether G is perfect, that is $G = [G,G]$. Moreover, since one can solve the word problem for $G/[G,G]$ it follows that one can decide of a arbitrary word w of G whether or not $w \in [G,G]$.

However, it would seem that any property of elements a finitely presented group which is not determined by the abelianization $G/[G,G]$ will be recursively unrecognizable. The following result show a few common properties of elements are not recognizable.

Theorem 6.5 (Baumslag, Boone and Neumann) *There is a finitely presented group G such that there is no algorithm to determine whether or not a word in the given generators represents*

1. *an element of the center of G;*
2. *an element which commutes with a given element of G;*
3. *an n-th power, where $n > 1$ is a fixed integer;*
4. *an element whose class of conjugates is finite;*
5. *a commutator;*

6. an element of finite order > 1.

Proof: Fix a finitely presented group U having unsolvable word problem. Define G to be the ordinary free product of U with a cyclic group of order 3 and an infinite cyclic group, that is,

$$G = U * \langle s \mid \rangle * \langle t \mid t^3 = 1 \rangle.$$

We use the commutator notation $[x, y] = x^{-1}y^{-1}xy$. In the following, w is a variable for an arbitrary word in the generators of U.

The center of G is trivial so w lies in the center of G if and only if $w =_U 1$. So there is no algorithm to determine whether an arbitrary word of G lies in the center. This gives the first assertion. Similarly, w is permutable with s if and only if $w =_U 1$ which establishes the second assertion. The element $s^n[t, w]$ is an n-th power if and only if $w =_U 1$ establishing the third assertion. The conjugacy class of w is finite if and only if $w =_U 1$ since if $w \neq_U 1$ the conjugates $s^{-i}ws^i$ would all be distinct. This gives the fourth assertion. For the fifth assertion, note that $[s, t]w$ is a commutator if and only if $w =_U 1$. Finally for the sixth assertion, observe that tw has infinite order if and only if $w \neq_U 1$, while if $w =_U 1$ then tw has order 3. This completes the proof.

6.2 Higman's embedding theorem

In contrast to the difficulties encountered for finitely presented groups, it is easy to give examples of finitely generated, recursively presented groups with unsolvable word problem. For example, let $S \subset \mathbb{N}$ be a recursively enumerable set of natural numbers which is not recursive. Define the recursively presented group

$$H_S = \langle a, b, c, d \mid a^{-i}ba^i = c^{-i}dc^i \ \forall i \in S \rangle.$$

Now H_S can be described as the free product with amalgamation of the free group $\langle a, b \mid \rangle$ and the free group $\langle c, d \mid \rangle$ amalgamating the subgroup (freely) generated by the left hand sides of the indicated equations with the subgroup (freely) generated by the right hand sides. It follows from the normal form theorem for amalgamated free products that $a^{-i}ba^ic^{-i}d^{-1}c^i =_{H_S} 1$ if and only if $i \in S$. Thus $S \leq_T WP(H_S)$ and so $WP(H_S)$ is recursively unsolvable.

Using this observation Graham Higman gave a very different proof of the unsolvability of the word problem. Indeed he proved the following remarkable result:

Theorem 6.6 (Higman Embedding Theorem) *A finitely generated group H can be embedded in a finitely presented group if and only if H is recursively presented.*

That finitely generated subgroups of finitely presented groups are recursively presented is contained in our first lemma above. The difficult part of this theorem is to show that a recursively presented group can be embedded in a finitely presented group.

The Novikov-Boone Theorem is an easy corollary. For let H_S be the finitely generated, recursively presented group with unsolvable word problem constructed above. By Higman's Embedding Theorem, H_S can be embedded in a finitely presented group, say G_S. Then by an earlier lemma, $WP(H_S) \leq_T WP(G_S)$ and so G_S has unsolvable word problem.

Higman's Embedding Theorem has a number of other remarkable aspects. It provides a complete characterization of the finitely generated subgroups of finitely presented groups - namely they are the recursively presented groups. It also provides a direct connection between a purely algebraic notion and a notion from recursive function theory. Another consequence is the existence of universal finitely presented groups.

Corollary 6.7 (Higman) *There exists a universal finitely presented group; that is, there exists a finitely presented group G which contains an isomorphic copy of every finitely presented group.*

To prove this one systematically enumerates all finite presentations on a fixed countable alphabet. The free product of all of these can be embedded in a two generator group which will be recursively presented. This group can then be embedded in a finitely presented group which is the desired universal group.

6.3 The isomorphism problem and recognizing properties

In this section the existence of a finitely presented group with unsolvable word problem is applied to obtain a number of global unsolvability results.

Consider the problem of recognizing whether a finitely presented group has a certain property of interest. For example, can one determine from a

presentation whether a group is finite? or abelian? It is natural to require that the property to be recognized is *abstract* in the sense that whether a group G enjoys the property is independent of the presentation of G.

An abstract property P of finitely presented groups is *recursively recognizable* if there is an effective method which when applied to an arbitrary finite presentation π determines whether or not $gp(\pi)$ has the property P. More formally, P is *recursively recognizable* if $\{\pi \mid gp(\pi) \in P\}$ is a recursive set of finite presentations.

It turns out that very few interesting properties of groups are recursively recognizable. To formulate the key result we need the following definition.

Definition 6.1 *An abstract property P of finitely presented groups is said to be a* Markov property *if there are two finitely presented groups G_+ and G_- such that*

1. *G_+ has the property P; and*
2. *if G_- is embedded in a finitely presented group H then H does not have property P.*

These groups G_+ and G_- will be called the positive *and* negative witnesses *for the Markov property P respectively.*

It should be emphasized that if P is a Markov property then the negative witness does not have the property P, nor is it embedded in any finitely presented group with property P.

For example the property of being finite is a Markov property. For G_+ one can take $\langle a \mid a^2 = 1\rangle$ which is a finite group. For G_- one can take the group $\langle b, c \mid b^{-1}cb = c^2\rangle$ which is an infinite group and therefore not embedded in any finite group.

Similarly, the property of being abelian is a Markov property. Indeed the two groups chosen as witnesses for the property of being finite will also serve as witnesses for the property of being abelian.

An example of a property which is not a Markov property is the property of being perfect, that is $G/[G,G] \cong 1$. For it is not hard to show (and indeed will follow from the constructions given below) that any finitely presented group can be embedded in a perfect finitely presented group. Hence there can be no negative witness G_- for the property of being perfect.

An abstract property P of finitely presented groups is *hereditary* if H embedded in G and $G \in P$ imply that $H \in P$, that is, the property P is

inherited by finitely presented subgroups. A property of finitely presented groups P is *non-trivial* if it is neither the empty property nor is it enjoyed by all finitely presented groups. Suppose P is a non-trivial, hereditary property of finitely presented groups. Then, since P is non-trivial, there are groups $G_+ \in P$ and $G_- \notin P$. But if G_- is embedded in a finitely presented group H, then $H \notin P$ because P is hereditary. Thus P is a Markov property with witnesses G_+ and G_-. This proves the following:

Lemma 6.8 *If P is a non-trivial hereditary property of finitely presented groups, then P is a Markov property.*

Another useful observation is the following:

Lemma 6.9 *If $\emptyset \neq P_1 \subseteq P_2$ are properties of finitely presented groups and if P_2 is a Markov property, then P_1 is also a Markov property.*

For if G_- is a negative witness for P_2 and if $K \in P_1$, then P_1 is a Markov property with positive and negative witnesses K and G_-.

Recall from the previous section that Higman has constructed a universal finitely presented group, say U. If P is a Markov property with positive and negative witnesses G_+ and G_-, then G_- is embedded in U so $U \notin P$. Moreover, if U is embedded in a finitely presented group H then so is G_- and hence $H \notin P$. Thus P is a Markov property with positive and negative witnesses G_+ and U. Hence U is a negative witness for every Markov property.

The main unsolvability result concerning the recognition of properties of finitely presented groups is the following:

Theorem 6.10 (Adian-Rabin) *If P is a Markov property of finitely presented groups, then P is not recursively recognizable.*

Before indicating a proof of this result, we note the following easy corollaries:

Corollary 6.11 *The following properties of finitely presented groups are not recursively recognizable:*

1. *being the trivial group;*
2. *being finite;*

3. *being abelian;*

4. *being nilpotent;*

5. *being solvable;*

6. *being free;*

7. *being torsion-free;*

8. *being residually finite;*

9. *having a solvable word problem;*

10. *being simple;*

11. *being automatic.*

For each of (1) through (9) is a non-trivial, hereditary property and hence is a Markov property. For (10), it is known that finitely presented, simple groups have solvable word problem and hence, by the above lemma, being simple is a Markov property. Similarly for (11), automatic groups have solvable word problem and so being automatic is a Markov property.

Corollary 6.12 *The isomorphism problem for finitely presented groups is recursively unsolvable.*

For by (1) in the previous corollary there is no algorithm to determine of an arbitrary presentation π whether or not $gp(\pi) \cong 1$.

Proof of the Adian-Rabin Theorem: We are going to give a simple proof of the Adian-Rabin Theorem which is our modification of one given by Gordon. The construction is quite straightforward and variations on the details can be applied to obtain further results. So suppose that P is a Markov property and that G_+ and G_- are witnesses for P. We also have available a finitely presented group U having unsolvable word problem.

Using these three items of initial data, we construct a recursive family of finite presentations $\{\pi_w \mid w \in U\}$ indexed by the words of U so that if $w =_U 1$ then $gp(\pi_w) \cong G_+$ while if $w \neq_U 1$ then G_- is embedded in U. Thus $gp(\pi_w) \in P$ if and only if $w =_U 1$. Since U has unsolvable word problem, it follows that P is not recursively recognizable.

The family $\{\pi_w \mid w \in U\}$ is rather like a collection of buildings constructed from playing cards standing on edge. Such a building can be rather unstable so that if an essential card is removed (corresponding to $w =_U 1$) then the entire structure will collapse. The main technical result needed is the following.

Lemma 6.13 (Main Technical Lemma) *Let K be a group given by a presentation on a finite or countably infinite set of generators, say*

$$K = \langle x_1, x_2, \ldots \mid r_1 = 1, r_2 = 1, \ldots \rangle.$$

For any word w in the given generators of K, let L_w be the group with presentation obtained from the given one for K by adding three new generators a, b, c together with defining relations

$$\begin{aligned} a^{-1}ba &= c^{-1}b^{-1}cbc && (6.1)\\ a^{-2}b^{-1}aba^2 &= c^{-2}b^{-1}cbc^2 && (6.2)\\ a^{-3}[w,b]a^3 &= c^{-3}bc^3 && (6.3)\\ a^{-(3+i)}x_iba^{(3+i)} &= c^{-(3+i)}bc^{(3+i)} \quad i = 1, 2, \ldots && (6.4) \end{aligned}$$

where $[w, b]$ is the commutator of w and b. Then

1. *if $w \neq_K 1$ then K is embedded in L_w by the inclusion map on generators;*

2. *the normal closure of w in L_w is all of L_w; in particular, if $w =_K 1$ then $L_w \cong 1$, the trivial group;*

3. *L_w is generated by the two elements b and ca^{-1}.*

If the given presentation of K is finite, then the specified presentation of L_w is also finite.

Proof: Suppose first that $w \neq_K 1$. In the free group $\langle b, c \mid \rangle$ on generators b and c consider the subgroup C generated by b together with the right hand sides of the equations (1) through (4). It is easy to check that the indicated elements are a set of free generators for C since in forming the product of two powers of these elements or their inverses some of the conjugating symbols will remain uncancelled and the middle portions will be unaffected.

Similarly, in the ordinary free product $K * \langle a, b \mid \rangle$ of K with the free group on generators a and b consider the subgroup A generated by b together with the left hand sides of the equations (1) through (4). Using the assumption that $w \neq_K 1$ it is again easy to check that the indicated elements are a set of free generators for A.

Thus assuming $w \neq_K 1$, the indicated presentation for L_w together with the equation identifying the symbol b in each the two factors is the natural presentation for the free product with amalgamation

$$\begin{array}{rcl} (K * \langle a, b \mid \rangle) & * & \langle b, c \mid \rangle. \\ A & = & C \end{array}$$

So if $w \neq_K 1$, then K is embedded in L_w establishing the first claim.

Now let N_w denote the normal closure of w in L_w. Clearly $[w, b] \in N_w$ so by equation (3), $b \in N_w$. But equations (1) and (2) ensure that a, b, c are all conjugate and so a, b, c all belong to N_w. Finally, since each of the system of equations (4) can be solved to express x_i in terms of a, b, c, it follows that $x_i \in N_w$ for $i = 1, 2, \ldots$. Thus each of the generators of L_w belongs to N_w and so $L_w = N_w$. This verifies the second assertion.

Finally, let M be the subgroup of L_w generated by b and ca^{-1}. Equation (1) can be rewritten as $b(ca^{-1})b(ca^{-1})^{-1}b^{-1} = c$ so that $c \in M$. But then from $ca^{-1} \in M$ it follows that $a \in M$. Finally from the system of equations (4) which can be solved for the x_i in terms of a, b, c it follows that $x_i \in M$ for $i = 1, 2, \ldots$ and so $M = L_w$. (For later use we note that neither equation (2) nor equation (3) was used in the proof of the final assertion). This completes the proof of the lemma.

Using this technical lemma it is easy to complete the proof of the Adian-Rabin Theorem. We are given the three finitely presented groups U, G_+ and G_- which can be assumed presented on disjoint alphabets as follows:

$$U = \langle y_1, \ldots, y_k \mid r_1 = 1, \ldots, r_\rho = 1 \rangle$$

$$G_- = \langle s_1, \ldots, s_m \mid u_1 = 1, \ldots, u_\sigma = 1 \rangle$$

$$G_+ = \langle t_1, \ldots, t_n \mid v_1 = 1, \ldots, v_\tau = 1 \rangle$$

Let $K = U * G_-$ the ordinary free product of U and G_- presented as the union of the presentations of its factors. Since U has unsolvable word problem, K also has unsolvable word problem. Also both U and G_- are embedded in K by the inclusion map on generators. For any word w in the generators of

U (these are also generators of K) form the presentation L_w as in the Main Technical Lemma. Finally we form the ordinary free product $L_w * G_+$.

A presentation π_w for these groups $L_w * G_+$ can be obtained by simply writing down all of the above generators together with all of the above defining equations. Such a presentation is defined for any word w in U whether or not $w \neq_U 1$. But it follows from the lemma that if $w \neq_U 1$ then the group G_- is embedded in $gp(\pi_w) = L_w * G_+$ and so $gp(\pi_w) \notin P$ by the definition of a Markov property. On the other hand, if $w =_U 1$ then by the lemma $L_w \cong 1$ and so $gp(\pi_w) \cong G_+$ and hence $gp(\pi_w) \in P$.

Thus we have shown that the recursive collection of presentations

$$\{\pi_w \mid w \text{ a word in } U\}$$

has the property that $gp(\pi_w) \in P$ if and only if $w =_U 1$. Since U has unsolvable word problem, it follows that P is not recursively recognizable. This completes the proof of the Adian-Rabin Theorem.

Bibliography

[1] M. Hall Jr., "The theory of groups", Macmillan, New York, 1959.

[2] D. L. Johnson, "Presentation of groups", LMS Lecture Notes **22**, Cambridge University Press, Cambridge 1976.

[3] D. L. Johnson, "Topics in the theory of group presentations", LMS Lecture Notes **42**, Cambridge University Press, Cambridge 1980.

[4] R. C. Lyndon and P. E. Schupp, "Combinatorial Group Theory", Springer-Verlag, Berlin-Heidleberg-New York,1977.

[5] W. Magnus, A. Karrass and D. Solitar, "Combinatorial Group Theory", Wiley, New York, 1966 (also corrected Dover reprint 1976)

[6] W. S. Massey, "Algebraic Topology: an introduction", Graduate Texts in Matematics **56**, Springer-Verlag, Berlin-Heidleberg-New York, 1977.

[7] C. F. Miller III, *Decision problems for groups - survey and reflections*, pp 1-59 in "Algorithms and Classification in Combinatorial Group Theory", MSRI Publications No. 23, edited by G. Baumslag and C. F. Miller III, Springer-Verlag, 1992.

[8] J. J. Rotman, "An introduction to A;lgebraic Topology", Springer Graduate Texts in Mathematics **119**, Springer-Verlag, Berlin-Heidleberg-New York, 1988.

[9] J. J. Rotman, "An introduction to the theory of groups" (4th edition), Springer Graduate Texts in Mathematics **148**, Springer-Verlag, Berlin-Heidleberg-New York, 1995.

[10] P. Scott and C. T. C. Wall, *Topological methods in group theory*, in "Homological Group Theory", pp 137-203 in LMS Lecture Notes **36**, edited by C. T. C. Wall, Cambridge University Press, 1979.

[11] J-P Serre, "Trees", Springer-Verlag, Berlin-Heidleberg-New York,1980.

Articles

[12] G. Baumslag, *Wreath products and finitely presented groups*, Math. Zeit. **75** (1960/1961) 22–28.

[13] G. Baumslag, *A remark on generalized free products*, Proc. Amer. Math. Soc. **13** (1962) 53–54.

[14] G. Baumslag, *A finitely presented metabelian group with a free abelian derived group of infinite rank*, Proc. Amer. Math. Soc. **35** (1972), 61–62.

[15] G. Baumslag and J. E. Roseblade, *Subgroups of direct products of free groups*, J. London Math. Soc. (2) **30** (1984), 44–52.

[16] R. Bieri and R. Strebel, *Almost finitely presented soluble groups* Comment. Math. Helv., 53 **2** (1978) 258-278.

[17] M. Hall Jr., *Subgroups of finite index in free groups*, Canad. J. Math. **1** (1949), 187–190.

[18] G. Higman, *Subgroups of finitely presented groups*, Proc. Royal Soc. London Ser. A **262** (1961), 455-475.

[19] A. Karrass and D. Solitar, *Subgroups of HNN groups and groups with one defining relation*, Canad. J. Math. **23** (1971) 627-643.

[20] C. F. Miller III, *Subgroups of a direct product with a free group*, Quarterly Journal of Maths **53** (2002) 503-506

[21] J. Stallings, *The topology of finite graphs*, Inventiones Math. **71** (1983), 551-565.