Free groups - basics

B.Sury

Stat-Math Unit
Indian Statistical Institute
Bangalore, India

AIS - May 2010 - IIT Bombay

"Only that thing is free which exists by the necessities of its own nature, and is determined in its actions by itself alone" - Baruch Spinoza

"Everything that is really great and inspiring is created by the individual who can labor in freedom" - Albert Einstein

"Now go we in content
To liberty, and not to banishment."
... W.Shakespeare (As You Like It)

"Freedom is like drink. If you take any at all, you might as well take enough to make you happy for a while" - Finley Peter Dunne

Introduction

Free objects in a category (whatever these animals may be) are the most basic objects in mathematics. A paradigm is the theory of free groups. They arose naturally while studying the geometry of hyperbolic groups but their fundamental role in group theory was recognized by Nielsen (who named them so), Dehn and others. In these lectures, we introduce free groups and their subgroups and study their basic properties. The lectures by Professor Anandavardhanan would treat the more general notions of free products with amalgamation and HNN extensions in detail. However, it is beneficial to look at them already for our purposes as they bear repetition. We shall do so briefly in the course of these lectures.

The notions of free groups, free products, and of free products with amalgamation come naturally from topology. For instance, the fundamental group of the union of two path-connected topological spaces joined at a single point is isomorphic to the so-called free product of the individual fundamental groups.

(More generally) The Seifert-van Kampen theorem asserts that if $X = V \cup W$ is a union of path-connected spaces with $V \cap W$ non-empty and path-connected, and if the homomorphisms $\pi_1(V \cap W) \to \pi_1(V)$ and $\pi_1(V \cap W) \to \pi_1(W)$ induced by inclusions, are injective, then $\pi_1(X)$ is isomorphic to the so-called free product of $\pi_1(V)$ and $\pi_1(W)$ amalgamated along $\pi_1(V \cap W)$.

Many naturally occurring groups can be viewed in terms of these constructions.
For instance, $SL(2, \mathbf{Z})$ *is the free product of* $\mathbf{Z}/4$ *and* $\mathbf{Z}/6$ *amalgamated along a subgroup isomorphic to* $\mathbf{Z}/2$.
The fundamental group of the Klein bottle is isomorphic to the free product of two copies of $\mathbf{Z}$ *amalgamated along* $2\mathbf{Z}$.

The so-called HNN extensions also have topological interpretations. Suppose V and W are open, path-connected subspaces of a path-connected space X and suppose that there is a homeomorphism between V and W inducing isomorphic embeddings of $\pi_1(V)$ and $\pi_1(W)$ in $\pi_1(X)$. One constructs a space Y by attaching the handle $V \times [0, 1]$ to X, identifying $V \times \{0\}$ with V and $V \times \{1\}$ with W. Then, the fundamental group $\pi_1(Y)$ of Y is the HNN extension of $\pi_1(V)$ relative to the isomorphism between its subgroups $\pi_1(V)$ and $\pi_1(W)$.

The HNN extensions give several universal constructions in group theory like 'every countable group can be embedded as a subgroup of a 2-generated countable group'. They even occur concretely; for instance, the proof that the automorphism group of a free group of finite rank > 2 does not admit a faithful matrix representation goes by showing that a certain HNN extension does not.

One final word about free groups is that they abound. In any finitely generated group of matrices, there is a free non-abelian subgroup unless the group has a solvable subgroup of finite index (this is the so-called Tits-alternative which has been generalized to several other groups as well).

1 Free groups

Informally, when we have a sequence $g_1, \cdots, g_n$ of (not necessarily distinct) elements from some group G which satisfy some constraint/relation like $g_1^{a_1} g_2^{a_2} \cdots g_n^{a_n} = 1$ for some non-zero integers a_i, the group is not deemed to be *free*. For instance, in any finite group, there are such constraints. Even in infinite groups, often one comes across such relations; for instance, the matrix $W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ satisfies $W^2 = I$. We shall look for a group where there are 'no nontrivial relations' and call it a free group. This concept seems esoteric and one can imagine it being useful because one may perhaps obtain any group by starting out with a free group and imposing more and more relations on it. The interesting part is that the free group is a concretely occurring – even ubiquitous – object ! For instance, we shall see that the group of all integer 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ having determinant 1 and having a, d odd and b, c even, will turn out to be a free group. If we look at the fundamental group of the symbol ∞, this also turns out to be free (and isomorphic to the above group of matrices !). Without further ado, we turn to the formal definitions now.

Definition
Given a nonempty set S, and a map $\theta : S \to F$ into a group F, the pair (F, θ) is said to be *a free group on* S if, for any function $\alpha : S \to G$ to any group G, there is a unique homomorphism $\widetilde{\alpha} : F \to G$ such that $\alpha = \widetilde{\alpha} \circ \theta$.

If (F, θ) is free on a set S, observe :

(i) θ *must necessarily be one-one.*
(ii) (*F,inclusion*) *is free on the subset Image*(θ).
(iii) *Image*(θ) *generates* F.
Proof : Assume (i),(ii) without loss of generality and so $S \subset F$. Consider the subgroup $< S >$ of F generated by S; this is, by definition, the intersection of all subgroups containing S. Consider the inclusions i, i_1, i_2 of $< S >$ in F, S inside $< S >$ and S inside F. So $i \circ i_1 = i_2$. The unique extension of i_2 to F is clearly the identity map. If $\pi_1 : F \to < S >$ is the unique extension of i_1 to F, then clearly $i \circ \pi_1$ must be the identity map by uniqueness.

To construct a free group on any arbitrary nonempty set X :
Let X' be a set in bijection with X and disjoint from X. Under a fixed bijection, write the element of X' corresponding to an element $x \in X$ as x^{-1}. Define a *nonempty word* in X to mean a formal expression of the form $x_1^{\epsilon_1} \cdots x_r^{\epsilon_r}$ where $x_i \in X, \epsilon_i = \pm 1$ and $r > 0$. Call two such expressions to be the same word if they have the same elements in the same positions. Define the 'product' (in that order) of two nonempty words $x_1^{\epsilon_1} \cdots x_r^{\epsilon_r}$ and $y_1^{\upsilon_1} \cdots y_s^{\upsilon_s}$ to be the juxtaposed word $x_1^{\epsilon_1} \cdots x_r^{\epsilon_r} y_1^{\upsilon_1} \cdots y_s^{\upsilon_s}$. One also denotes the empty word (!) by 1 and defines the product of any word w with 1 in both the ways to be w itself. Call $x_r^{-\epsilon_r} \cdots x_1^{-\epsilon_1}$ to be the inverse of the nonempty word $x_1^{\epsilon_1} \cdots x_r^{\epsilon_r}$; write $(x_1^{\epsilon_1} \cdots x_r^{\epsilon_r})^{-1} = x_r^{-\epsilon_r} \cdots x_1^{-\epsilon_1}$. We also write $1^{-1} = 1$. On the set of all nonempty words together with the empty word 1, one defines the following relation.
$w_1 \sim w_2$ *if* w_2 *is obtained from* w_1 *by a finite sequence of the operations :*
inserting or deleting expressions like xx^{-1} *or* $x^{-1}x$ *for* $x \in X$.
This is an equivalence relation and the set F of equivalence classes $[w]$ will be given the structure of a group now in an obvious manner. Define $[w_1][w_2] = [w_1 w_2]$. It is easy to verify that this gives a well-defined group structure on F (although associativity is a messy check as usual !). Defining the map $\theta : X \to F$ by $x \mapsto [x]$, let us show that (F, θ) is free on X.
Suppose $\alpha : X \to G$ be a map into any group. Extend α to words on X by putting $\alpha(x_1^{\epsilon_1} \cdots x_r^{\epsilon_r}) = \alpha(x_1)^{\epsilon_1} \cdots \alpha(x_r)^{\epsilon_r}$ on nonempty words and $\alpha(1)$ =identity of G. If $w_1 \sim w_2$, then evidently $\alpha(w_1) = \alpha(w_2)$ because a product of the form xx^{-1} or of the form $x^{-1}x$ maps to the identity of G. Thus, α induces a map $\widetilde{\alpha}$ from F to G which is clearly a homomorphism. It is clear that this homomorphism $\widetilde{\alpha}$ on F is unique with the property that $\widetilde{\alpha}[x] = \alpha(x)$ for all $x \in X$; for, the image of X in F generates F.

Reduced words

It is much more convenient to work with (instead of with the equivalence classes as above) so-called reduced words. Call a word w *reduced* if it does not contain any x adjacent to x^{-1} with $x \in X$. In particular, the empty word is reduced. The key observation is :

Claim : *Each equivalence class in F contains a unique reduced word.*

Indeed, start with any word w and, by cancelling off expressions of the forms xx^{-1} and $x^{-1}x$ with $x \in X$, land in a reduced word. So, each equivalence class does contain a reduced word.

To show that two reduced words cannot represent the same element unless they are equal, is clearly the same as showing that no reduced nonempty word represents 1. For, suppose $x_1 \cdots x_r \sim y_1 \cdots y_s$ with both $x_1 \cdots x_r$ and $y_1 \cdots y_s$ reduced. Then, by cancelling off the right ends as much as possible, we may assume $x_r \neq y_s$. Look at the *reduced* (why?) word $x_1 \cdots x_r y_s^{-1} \cdots y_1^{-1}$. Make all the insertions and deletions into the first part $x_1 \cdots x_r$ which makes it $y_1 \cdots y_s$. This can be viewed as making insertions and deletions in the word $x_1 \cdots x_r y_s^{-1} \cdots y_1^{-1}$ which clearly changes it to $y_1 \cdots y_s y_s^{-1} \cdots y_1^{-1} = 1$. Thus, the reduced word $x_1 \cdots x_r y_s^{-1} \cdots y_1^{-1} \sim 1$.

So, suppose $x_{i_1}^{\epsilon_1} \cdots x_{i_n}^{\epsilon_n}$ is a reduced word ~ 1 and $\epsilon_r = \pm 1$ for each r. Consider the following homomorphism from G to $S_{n+1} = Sym\{1, 2, \cdots, n+1\}$. Map x_{i_r} to a permutation which sends r to $r+1$ if $\epsilon_r = 1$ and map x_{i_r} to a permutation which sends $r+1$ to r if $\epsilon_r = -1$. Note that this is possible because there can be a clash only if the following situation arises. Some $x_{i_r} = x_{i_{r-1}}$ and r is sent to $r+1$ because $\epsilon_r = 1$ while it has to be sent to $r-1$ with $\epsilon_{r-1} = -1$. But this is not possible in a reduced word. Send all other $x \in X$ to the iedntity. Hence, we do have at least one well-defined homomorphism from F to S_{n+1} in which the reduced word $x_{i_1}^{\epsilon_1} \cdots x_{i_n}^{\epsilon_n}$ is mapped to the identity permutation. However, this word is mapped to a permutation which cannot be the identity (where does 1 go if $\epsilon_1 = 1$ for instance?). This is the reason for choosing permutation group on $n+1$ symbols. Thus, no reduced nonempty word can give the identity element. Thus, the set of reduced words (including the empty word) are in bijection with equivalence classes of all words.

The above remarks show that the free group on a set X can be thought of as the set of all reduced words. Using this identification and, writing $[w]$ simply as w for simplicity, and writing xx as x^2 etc, we have the so-called :

Normal form :

In a free group F on a subset X, every nontrivial element has a unique expression of the form $x_1^{r_1} \cdots x_k^{r_k}$ with $x_i \in X$, $k \geq 1, x_i \neq x_{i+1}$ for all $1 \leq i < k, r_i \neq 0$.

The advantage of the normal form is that it also characterizes free groups from the above remarks. In other words,
Let G be a group, X be subset. Then, G is free on X if, and only if, every nontrivial element of F has a unique expression of the form $x_1^{r_1} \cdots x_k^{r_k}$ with $k \geq 1; x_i \in X (1 \leq i \leq k); x_i \neq x_{i+1} (1 \leq i < k); r_i \neq 0 \forall i$.

Definition. Let F be free on a subset X and $w \in F$, $w \neq 1$. Consider the unique expression $w = x_1^{r_1} \cdots x_k^{r_k}$ with $k \geq 1; x_i \in X (1 \leq i \leq k); x_i \neq x_{i+1} (1 \leq i < k); r_i \neq 0 \forall i$. Then, the *length of w with respect to F* is defined to be $|r_1| + \cdots + |r_k|$. One also defines the length of the identity element 1 to be 0. This notion of length is very useful (see below for an immediate application and the exercises for other instances).
We mentioned that the associativity of the product law is a messy check; let us do it here working with reduced words. Remember that the product of two reduced words $a_1 a_2 \cdots a_r$ and $b_1 b_2 \cdots b_s$ (in that order) is $a_1 \cdots a_{r-i} b_{i+1} \cdots b_s$ where $a_r b_1 = a_{r-1} b_2 = \cdots = a_{r-i+1} b_i = 1$ and $a_{r-i} b_{i+1} \neq 1$.
Look at reduced words a, b, c. The equality $(ab)c = a(bc)$ is evident if ab and bc are formed without cancellations. In general, we induct on the length of b and, for fixed b, on $l(a) + l(b) + l(c)$. If $b = 1$, then again $(ab)c = ac = a(bc)$.
Assume $l(b) = 1$. So, $b \in X \cup X^{-1}$. Now, if ab involves a cancellation, then a ends in b^{-1}; that is, $a = wb^{-1}$. If bc is formed without cancellations, then clearly $(ab)c = wc$ and $a(bc) = (wb^{-1})(bc) = wc$ as this is the way the two reduced words wb^{-1} and bc are multiplied in $F(X)$. Thus, $(ab)c = a(bc)$ when bc is reduced. Suppose now that bc involves a cancellation; so $c = b^{-1}v$ starts with b^{-1}. Therefore, $(ab)c = wc = w(b^{-1}v)$ while $a(bc) = a(v) = (wb^{-1})v$. Since $l(w) + l(v) < l(a) + l(c)$, the last two terms are equal by induction hypothesis. Hence, the associativity holds when $l(b) = 1$.
Now, suppose the reduced word b has $l(b) = n > 1$ and write $b = b_1 b_2$ with $l(b_1), l(b_2) < n$. Then, by induction hypotheses

$$(ab)c = (a(b_1 b_2))c = ((ab_1)b_2)c = (ab_1)(b_2 c) = a(b_1(b_2 c)) = a((b_1 b_2)c).$$

Thus, the proof of associativity is complete.

The following fact is very easily proved using the universal property.

Lemma.
If $|X_1| = |X_2|$, then $F(X_1) \cong F(X_2)$.

Lemma.
If $F(X_1) \cong F(X_2)$*, then* $|X_1| = |X_2|$.
Proof.
Look at the sets $\mathrm{Hom}(F(X_1), \mathbf{F}_2)$ and $\mathrm{Hom}(F(X_2), \mathbf{F}_2)$ of group homomorphisms to the field $\mathbf{F}_2$ with 2 elements. These sets are vector spaces over $\mathbf{F}_2$ with bases X_1 and X_2 respectively. Fixing an isomorphism $\theta : F(X_1) \to F(X_2)$, we have an isomorphism of $\mathbf{F}_2$-vector spaces from $\mathrm{Hom}(F(X_2), \mathbf{F}_2)$ to $\mathrm{Hom}(F(X_1), \mathbf{F}_2)$ given by $\phi \mapsto \phi \circ \theta$. Thus, their bases must have the same cardinality, which proves that $|X_1| = |X_2|$.

In view of the above lemmata 2 and 3, one may define the *rank* of any free group to be the cardinality of a set X on which they are free. However, the length function does depend on the choice of X.

Examples of free groups occurring naturally

Example 0.
The only free group of rank 1 is, up to isomorphism, the infinite cyclic group.

Example 1.
Consider the functions α and β on the set $\mathbb{C} \cup \{\infty\}$ defined by the rules

$$\alpha(x) = x + 2, \quad \beta(x) = \frac{x}{2x+1}.$$

here the symbol ∞ is subject to such formulae as $\frac{1}{0} = \infty$ and $\frac{\infty}{\infty} = 1$. Then α and β are bijections since they have inverses

$$\alpha^{-1}(x) = x - 2, \quad \beta^{-1}(x) = \frac{x}{1-2x}.$$

Thus α and β generate a group F of permutations of $\mathbb{C} \cup \{\infty\}$.
F is free on the set $\{\alpha,\ \beta\}$.
Note that F is nothing but the group generated by $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$.
Proof.
Note that every non-zero power of α maps the interior of the circle $|z| = 1$ to the exterior of the unit circle and a non-zero power of β maps the exterior of the unit circle to the interior with 0 removed. Thus, no nontrivial word can be the identity (why?); lemma 1 shows that F is free on α, β.

Example 2. (Generalization of example 1)
For any complex number z with $|z| \geq 2$, the matrices $\begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}$

generate a free group of rank 2.
It is convenient to postpone the proof to section 2 where the so-called ping-pong lemma is discussed.

Open problem.
Does there exist a rational number α with $0 < |\alpha| < 2$ such that the group generated by $\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix}$ is free ?

We mention in passing that several rational α with $0 < |\alpha| < 2$ are known for which the group generated by $\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix}$ is *NOT* free. The proof interestingly goes via the Brahmagupta-Pell equation.

Example 3. (Not easy to see)
Let p be an odd prime. The subgroup G_p of the symmetric group on $\mathbf{R}$ generated by $x \mapsto x + 1$ and $x \mapsto x^p$ is free of rank 2.

The most basic fact about free groups in abstract group theory is :

Proposition.
Every group is isomorphic to the quotient group of a free group.
Proof.
If G is any group and S is any subset generating it (G itself, for example), consider a set X in bijection with S and consider the free group $F(X)$ on X. If $\theta : X \to S$ is a bijection, then the universal property of $F(X)$ gives a surjective homomorphism from $F(X)$ to G.

Remarks.
(i) In later lectures, a theory of groups acting on graphs would be developed. In particular, they would yield a characterization of free groups as those groups which act freely (that is, without fixed points for nontrivial elements) on a tree.
(ii) In an exercise below, we see that free groups are torsion-free. Actually, any group which is torsion-free and contains a free group of finite index, is free. This is proved by Jean-Pierre Serre using methods from homological algebra.

Exercises.

Q 0. Prove that a group G is free on a subset X if and only if X generates G and no reduced word in $X \cup X^{-1}$ of positive length is the identity.
Q 1. Show that free groups $F(X)$ do not have elements of finite order > 1. In fact, if $a^n = b^n$ for a, b in a free group F, prove that $a = b$.
Q 2. Prove that each element of a free group has at the most finitely many roots; that is, for each $w \in F$, show $\sqrt{w} := \{a \in F : a^n = w \text{ for some } n\}$ is finite.
Q 3. Show that in a free group, two commuting elements a, b must satisfy $a = c^u, b = c^v$ for some element c and some integers u, v. In particular, a free group has nontrivial center if and only if its rank is 1.
Q 4. If a, b are elements in a free group F, and satisfy $a^p b^q = b^q a^p$ for some non-zero integers p, q then prove a, b are integer powers of a common element.
Q 5. If $w \neq 1$ in a free group F, then prove that the centralizer $C(w)$ is an infinite cyclic group.
Q 6. If $w \neq 1$ in a free group, show that w cannot be conjugate to w^{-1}.
Q 7. If N is a normal subgroup of a group G such that G/N is free, then prove that there exists a subgroup H of G satisfying $G = HN$ and $H \cap N = \{1\}$.
Q 8. Let H be a subgroup of infinite index in a free group F. Show that for each subgroup $K \neq \{1\}$ of F, $H \cap K \neq \{1\}$.
Q 9. Let $F(X)$ be free on a set X and fix $x \in X$. Let $s_x : F(X) \to \mathbf{Z}$ be the function which takes any reduced word to the sum of the exponents of the terms which are equal to x. Prove that $s_x(w) = 0$ for some $w \in F$ if and only if, $w \in [F, F]$.
Q 10. (Prochronistically!) Assuming that subgroups of free groups are free, show that the centralizer of a nontrivial element of a free group is infinite cyclic.
Q 11. Prove that if F is free of rank n, then $F/[F, F] \cong \mathbf{Z}^n$.
Q 12. Compute the number of words of length n in a free group of rank d.
Q 13. Prove that if $F = F_0 \supset F_1 \supset F_2 \cdots$ is a chain such that each F_{i+1} is a proper, characteristic subgroup of F_i, then $\cap_i F_i = \{1\}$. In particular, this holds for the derived series.
Q 14. Show that the matrices $Y := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $X := \begin{pmatrix} 1 & e \\ 0 & 1 \end{pmatrix}$ generate a group isomorphic to F_2. Here e is the exponential and one may assume that e is not a root of a nonzero integral polynomial.
Q 15. Prove that in the free group $F(x, y)$ on two generators x, y, the

subgroup H generated by $y^n x y^{-n}; n \in \mathbf{Z}$ is not finitely generated. Hence, show that $[F_2, F_2]$ is not finitely generated.

Solutions.

0.

Consider the unique homomorphism $\theta : F(X) \to G$ extending the inclusion map of X into G. The two conditions are equivalent to surjectivity and injectivity of θ and, therefore, to the freeness of G. Conversely, if G is free on X, then the unique homomorphism $\phi : G \to F(X)$ extending the inclusion of X into $F(X)$ (which exists because G is free on X) is evidently the inverse of θ since X generates G. Thus, ϕ is an isomorphism which means it is a bijection and so the conditions hold.

1.

Let $w = x_1 \cdots x_n$ be any nontrivial reduced word where $x_i \in X \cup X^{-1}$. That is, length of w is $l(w) = n$. Write w^2 as a reduced word

$$w^2 = x_1 \cdots x_n x_1 \cdots x_n = x_1 x_2 \cdots x_{n-r} x_{r+1} x_{r+2} \cdots x_n.$$

Thus, $l(w^2) = l(w) - 2r$ where $0 \leq r \leq n$. We claim that $r < n/2$. If $n = 2k$ and $r \geq k$, then $x_k = x_{k+1}^{-1}$, a contradiction of the hypothesis. Thus, $r < n/2$ when n is even. If $n = 2k + 1$ and $r > k$, then $x_{k+1}^2 = 1$, a contradiction of the fact that no reduced words of positive length can be trivial. Thus, $r \leq n/2$ in either case. Therefore, $w = u^{-1} y u$ where $u^{-1} = x_1 \cdots x_r = x_n^1 \cdots x_{n-r+1}^{-1}, y = x_{r+1} x_{r+2} \cdots x_{n-r}$ is cyclically reduced (that is, first letter $x_{r+1} \neq x_{n-r}^{-1}$, the inverse of the last letter. Note that $y \neq 1$ as $w \neq 1$. Also, $l(y^d) = dl(y)$ for all $d \geq 1$ (this is where the cyclically reducedness comes up). Now $l(w^2) = 2l(w) - 2r = 2n - 2r > 2n - n = n = l(w)$. Further, for any d also, $w^d = u^{-1} y^d y$ implies that $l(w^d) = dl(y) + 2r > (d-1)l(y) + 2r = l(w^{d-1})$. Hence $w^d \neq 1$ for all $d \geq 1$. Thus, free groups are torsion-free.

Let $a^n = b^n$. Write $a = u^{-1} x u, b = v^{-1} y v$ as above with x, y cyclically reduced. Let $l(u) = r, l(v) = s$. Then,

$$l(a^n) = nl(x) + 2r = nl(y) + 2s = l(b^n)$$

$$l(a^{2n}) = 2nl(x) + 2r = 2nl(y) + 2s = l(b^{2n})$$

Thus, we get $l(x) = l(y)$ and so $r = s$. Hence there are no cancellations on each side of $u^{-1} x u = v^{-1} y v$ means $u = v, x = y$ which gives $a = b$.

2.

If $w = 1$, then $\sqrt{w} = \{1\}$. Assume $w \neq 1$. As before, $w = a^n$ means $w = u^{-1}x^n u$ where x is cyclically reduced and $l(a^n) = nl(x) + 2r \geq n$ as $x \neq 1$. Hence, the only n's occurring are those finitely many ones which must satisfy $n \leq l(w)$. For any such n, the previous question shows that there can be at the most one a satisfying $a^n = w$.

3.

We may assume that $a, b \neq 1$. Write $a = x_1 \cdots x_m, b = y_1 \cdots y_n$ with $l(a) = m, l(b) = n$ with $m \leq n$ say. Let $ab = x_1 \cdots x_{m-r} y_{r+1} \cdots y_n$ be the reduced expression (so $l(ab) = m+n-2r$ and $0 \leq r \leq m$). Now, $ab = ba$ implies that their lengths are same and so, the reduced expression of ba is $y_1 \cdots y_{n-r} x_{r+1} \cdots x_m$. Considering the three cases $r = 0, r = m, 0 < r < m$ and applying induction on $l(a) + l(b) = m + n$, we will easily get the result (page 9 of Johnson). It is a nice exercise now to deduce from this that free groups cannot have nontrivial centre unless they are cyclic.

4.

By manipulating the given equation, we may express it as $a^p b^q = b^q a^p$ for some natural numbers p, q. So, $a^p = b^q a^p b^{-q} = (b^q a b^{-q})^p$ which gives $a = b^q a b^{-q}$ by question 1. Once again, then $b^q = a b^q a^{-1} = (aba^{-1})^q$ which gives $b = aba^{-1}$; that is, $ab = ba$. By question 2, we get the result.

5.

Let $u, v \in C(w)$ be nontrivial elements. Then, by question 2, there exist $a, b \in F$ and non-zero integers p, q, r, s such that $u = a^p, w = a^q, v = b^r, w = b^s$. As a^q, b^s commute (they are equal !), question 2* shows that $a = c^l, b = c^m$ for some element c and some non-zero integers l, m. Evidently, $u = c^{lp}$ commutes with $v = c^{mr}$. So $C(w)$ is abelian. It is an infinite group as it contains $< w >$. Let x be a nontrivial element of $C(w)$ of smallest length possible. Let $y \in C(w)$ be arbitrary. As x, y commute as just proved, there exists $z \in F$ and integers u, v such that $x = z^u, y = z^v$. Now, $y = z^v \in C(w)$ means both z, w are integral powers of same element by question 2* and so, they commute; that is, $z \in C(w)$. But $l(x) = l(a^u) > l(a)$ unless $u = \pm 1$. By minimality of choice of x, we must have $u = \pm 1$ and so $y = z^v = x^{\pm v}$. So $C(w) =< x >$.

6.

As we saw, $w = u^{-1}xu$ where x is cyclically reduced (that is, reduced and first and last symbols are not inverses of each other). So, w and w^{-1} are conjugate if and only if x and x^{-1} are. But, a cyclically reduced element y some conjugate cyc^{-1} of which is also cyclically reduced must satisfy the property that cyc^{-1} is a cyclic permutation of y (by induction on $l(c)$). Applying this

to the elements x and x^{-1}, we get $x = ab$ and $x^{-1} = ba$ for some a, b. So, $a^{-1}b^{-1} = x = ab$ which means that $a = a^{-1}, b = b^{-1}$ as the expressions are reduced. But then $a^2 = 1 = b^2$ which means, by 1, that $a = 1 = b$. So $x = 1$ and so $w = 1$, a contradiction. Thus, w cannot be conjugate to w^{-1} if $w \neq 1$.

10.

If $w \neq 1$, $a, b \in C_F(w)$, then $w = x^r = y^s$ where $a \in < x >, b \in < y >$. We have used the fact that two commuting elements are powers of the same element. As $< x, y >$ is free (by assumption), while the relation $x^r = y^s$ is a nontrivial relation. Therefore, $< x, y >$ has rank 1; that is, it is cyclic. Hence there is z such that $x = z^u, y = z^v$. So, $a, b \in < z >$ which means $C_F(w)$ is abelian. Thus, it is infinite cyclic.

14.

We need to show that any nontrivial reduced word in X, Y does not give the identity matrix.

We shall show by induction that any matrix of the form

$$X^{a_1}Y^{b_1}\cdots X^{a_r}Y^{b_r}$$

with a_i, b_i nonzero integers, has $(1,1)$-th entry $e^r \prod_{i=1}^r a_i b_i + p(e)$ and $(2,1)$-th entry $q(e)$ where $p(e), q(e)$ are (possibly constant) polynomials in e with integer coefficients and degrees strictly less than r.

As e does not satisfy any nonzero integer polynomial, the above inductive assertion can easily be seen to prove that nontrivial reduced words in X and Y do not give the identity matrix. Indeed, if a word of the form

$$Y^{b_0}X^{a_1}Y^{b_1}\cdots X^{a_r}Y^{b_r}$$

with $r > 0$, is I, then

$$X^{a_1}Y^{b_1}\cdots X^{a_r}Y^{b_r} = Y^{-b_0}$$

which is impossible since the left side does not have 1 at the $(1,1)$-th place. Similarly, the other words are also dealt with. Now, let us prove the inductive assertion.

Clearly, $X^aY^b = \begin{pmatrix} 1+abe & ae \\ b & 1 \end{pmatrix}$ which shows the assertion holds for $r = 1$.

Suppose $n > 1$ and that

$$X^{a_1}Y^{b_1}\cdots X^{a_n}Y^{b_n}$$

has $(1,1)$-th entry $e^n \prod_{i=1}^n a_i b_i + p(e)$ and $(2,1)$-th entry $q(e)$ where $p(e), q(e)$ are (possibly constant) polynomials in e with integer coefficients and degrees

strictly less than n. Then,

$$X^{a_0}Y^{b_0}X^{a_1}Y^{b_1}\cdots X^{a_n}Y^{b_n}$$
$$= \begin{pmatrix} 1+a_0b_0e & a_0e \\ b_0 & 1 \end{pmatrix} \begin{pmatrix} e^n\prod_{i=1}^n a_ib_i + p(e) & * \\ q(e) & * \end{pmatrix}$$
$$= \begin{pmatrix} e^{n+1}\prod_{i=0}^n a_ib_i + p_1(e) & * \\ q_1(e) & * \end{pmatrix}$$

where p_1, q_1 are integral polynomials in e with degrees less than $n+1$. This proves the inductive assertion.

15.

Clearly, each element of the subgroup H is uniquely expressible as a reduced word in the elements $y^nxy^{-n}; n \in \mathbf{Z}$. Thus, H is not finitely generated.

2 Free abelian groups

A free group F on a set X was defined by the universal property that an abstract map from X to any group G can be extended uniquely as a homomorphism from F to G. If we do the same for abelian groups G, we would arrive at free abelian groups. Equivalently, these groups can also be defined more concretely as follows. An abelian group G is said to be *free abelian on a subset* S if each element of G is a unique (finite) integer linear combination of elements of S. One says that G has rank n if, and only if, $|S| = n$; as before $|S|$ is determined by G. Note that a group G is free abelian of rank n if, and only if, it is isomorphic to $\mathbf{Z}^n$, the set of integral n-tuples under coordinate-wise addition. The key theorem is that a subgroup of a free abelian group is free abelian of rank not exceeding that of the bigger group. From this, one deduces a structure theorem of finitely generated abelian groups. But, a more refined version is as follows :

Theorem (Invariant factor theorem)
If H is a subgroup of a free abelian group G of rank n, then H is free abelian of rank $r \leq n$. Further, there are bases $\{e_1, \cdots, e_n\}$ of G and $\{d_1e_1, \cdots, d_re_r\}$ of H respectively where d_i divides d_{i+1} for $i < r$. The integers d_i are uniquely determined up to sign and are called the invariant factors of H.

The proof is carried out by induction on n using the division algorithm as follows. It is clear for $n = 1$. Assume $n > 1$ and that the theorem holds for $m < n$. Corresponding to any basis of G, there is a positive integer

with the property that it is the smallest positive integer that occurs as a coefficient in the expression of elements of H in terms of this basis. This positive integer can depend on the basis and let l_1 be the smallest such with respect to all bases of G. Let $v_1, \cdots, v_n$ be a corresponding basis for G such that $v = l_1 v_1 + \sum_{i=2}^{n} a_i v_i \in H$. Dividing all the a_i by l_1, we have $a_i = q_i l_1 + r_i$ with $0 \leq r_i < l_1$. Evidently, $v = l_1(v_1 + \sum_{i=2}^{n} q_i v_i) + \sum_{i=2}^{n} r_i v_i$ and $v_1 + \sum_{i=2}^{n} q_i v_i, v_2, \cdots, v_n$ is another basis of G. By the minimality of l_1, we must have $r_i = 0$ for all $i \geq 2$. Thus, writing w_1 for $v_1 + \sum_{i=2}^{n} q_i v_i$, $v = l_1 w_1 \in H$. Look at the subset H_0 of H which have coefficients of w_1 to be zero in terms of the basis $w_1, v_2, \cdots, v_n$ of G. Clearly, H_0 is a subgroup of H such that $H_0 \cap \mathbf{Z}v = \{0\}$. Also, if $h \in H$, write $h = b_1 w_1 + \sum_{i=2}^{n} b_i v_i$. Once again, dividing the b_i's by l_1, say, $b_i = m_i l_1 + s_i$ with $0 \leq s_i < l_1$, we have $h - m_1 v = s_1 w_1 + \sum_{i=2}^{n} b_i v_i \in H$. Thus, by the minimality of l_1 we get $s_1 = 0$ i.e., $h - m_1 v \in H_0$. Thus, $H = H_0 \oplus \mathbf{Z}v$. Now, H_0 is contained in the subgroup $G_0 = \sum_{i=2}^{n} \mathbf{Z}v_i$. By induction hypothesis, G_0 has a basis $w_2, \cdots, w_n$ and there exists $r \leq n$ such that H_0 has a basis of the form $d_2 w_2, \cdots, d_r w_r$ with $d_2 | d_3 | \cdots | d_n$. Clearly, therefore, H itself has rank $r \leq n$ and $l_1 w_1, d_2 w_2, \cdots, d_n w_n$ is a basis for H. We have only to show that $l_1 | d_2$. Once again, writing $d_2 = c l_1 + d$ with $0 \leq d < l_1$, we notice $l_1 w_1 + d_2 w_2 = l_1(w_1 + c w_2) + d w_2 \in H$ where $w_1 + c w_2, w_2, \cdots, w_n$ is a basis of G. Thus, minimality of l_1 forces $d = 0$ i.e., $l_1 | d_2$. The proof is complete.

Corollary.
(Structure theorem for finitely generated abelian groups)
A finitely generated abelian group is isomorphic to $\mathbf{Z}^m \times \mathbf{Z}_{d_1} \times \cdots \times \mathbf{Z}_{d_r}$ for some $m \geq 0$ and d_i dividing d_{i+1}. The integer m as well as all the d_i's (up to sign) are uniquely determined.

The existence of bases as in the invariant factor theorem is equivalent to the following statement about matrices :
Lemma.
Given any $A \in M_{m,n}(\mathbf{Z})$ of maximum possible rank, there exist $P \in GL(m, \mathbf{Z})$ and $Q \in GL(n, \mathbf{Z})$ such that PAQ is a matrix whose'diagonal' entries are $d_1, d_2, \cdots$ where $d_i | d_{i+1}$. Furthermore, $\mathrm{GL}(n, \mathbf{Z})$ is generated by elementary matrices $I + E_{ij}$.
Proof
Suppose the matrix statement holds. Let H be a subgroup of a free abelian group G of rank n. Then, H is also free abelian of rank $m \leq n$ (this we are assuming known through other arguments). Let $\alpha : \mathbf{Z}^m \to H$ and $\beta : G \to \mathbf{Z}^n$ be isomor-

phisms. If $i : H \leq G$ denotes the inclusion map, we have the composite map $\beta \circ i \circ \alpha$ corresponds to a matrix $A \in M_{n,m}(\mathbf{Z})$ with respect to the canonical ordered bases of $\mathbf{Z}^m$ and $\mathbf{Z}^n$. The matrix statement gives us $P \in GL(n, \mathbf{Z})$ and $Q \in GL(m, \mathbf{Z})$ such that

$$AQ = P \cdot \begin{pmatrix} d_1 & \cdots & 0 \\ \ddots & \ddots & \ddots \\ 0 & \cdots & d_m \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix}$$

where $d_i | d_{i+1}$.
Hence, the bases

$$\{v_1, \cdots, v_n\} = \{Pe_1, \cdots, Pe_n\}$$

of $\mathbf{Z}^n$ and

$$\{w_1, \cdots, w_m\} = \{Qe_1, \cdots, Qe_m\}$$

of $\mathbf{Z}^m$ are so that

$$\{\beta^{-1}(v_1), \cdots, \beta^{-1}(v_n)\}$$

is a basis for G and $\{\alpha(w_1), \cdots, \alpha(w_m)\}$ is a basis for H.
Now, note that the matrix identity above implies that $AQ(e_i) = P(d_i e_i)$ where e_i on the left side are in $\mathbf{Z}^m$ and those on the right side are in $\mathbf{Z}^n$.
That is, $\beta\alpha(Qe_i) = d_i P(e_i)$.
So, we have $\beta\alpha(w_i) = d_i v_i$, which means that the bases $\{\beta^{-1}(v_1), \cdots, \beta^{-1}(v_n)\}$ of G and $\{\alpha(w_1), \cdots, \alpha(w_m)\}$ of H are as asserted in the invariant factor theorem.
Conversely, let us assume that the invariant factor theorem holds. Consider any $A \in M_{n,m}(\mathbf{Z})$ of rank $\max(m, n)$. Without loss of generality, we shall take $m \leq n$ for, otherwise, we could take the transpose. Now, A defines a homomorphism

$$T_A : \mathbf{Z}^m \to \mathbf{Z}^n \; ; v \mapsto Av.$$

Now the image of T_A is a free abelian group generated by the n vectors $Ae_1, \cdots, Ae_m$.
Since the matrix A has rank m, the vectors $Ae_1, \cdots, Ae_m$ are linearly independent vectors over $\mathbf{Q}$. Therefore, they are linearly independent over $\mathbf{Z}$ also. In other words, Image T_A is free abelian subgroup of $\mathbf{Z}^n$ of rank m.
By the invariant factor theorem, let us choose bases $\{v_1, \cdots, v_n\}$ of $\mathbf{Z}^n$ and $\{d_1 v_1, \cdots, d_m v_m\}$ of Image T_A such that $d_i | d_{+1}$. Call $Aw_i = d_i v_i$ for all $i \leq m$.
Let $P \in GL(n, \mathbf{Z})$ denote the matrix effecting the change of basis from the canonical basis to the v_i's. Similarly, let $Q \in GL(m, \mathbf{Z})$ be the matrix effecting the change of basis from the canonical basis to the w_i's.
Then, $P^{-1}AQ(e_i) = d_i v_i$ for all $i \leq m$. In other words, $P^{-1}AQ$ has the form

asserted.
The above proof of the invariant factor theorem clearly shows the generation of $GL(n, \mathbf{Z})$ by the elementary matrices.

Exercise.
Prove that $SL(n, \mathbf{Z})$ is perfect for $n \geq 3$. Is this true for $n = 2$?

Lemma.
For any $A \in M_{m,n}(\mathbf{Z})$ define $h_i(A)$ to be the GCD of all $i \times i$ minors of A. If A has maximal rank, then for any $P \in GL(m, \mathbf{Z})$ and $Q \in GL(n, \mathbf{Z})$, the numbers $h_i(A) = h_i(PA) = h_i(AQ)$ for all i. The invariant factors of a matrix $A \in M_{m,n}(\mathbf{Z})$ are $h_1(A), \frac{h_2(A)}{h_1(A)}, \frac{h_3(A)}{h_2(A)}, \cdots$ etc.

We know that $GL(n, \mathbf{Z})$ is generated by the matrices of the form $X_{ij} = I + E_{ij}; i \neq j$ and the matrices $diag(\pm 1, \cdots, \pm 1)$. elsewhere. We shall check for each r that

$$h_r(AX_{ij}) = h_r(X_{ij}A)$$

for all $i \neq j \leq n$.
By the previous lemma, we need to consider only A of the 'diagonal' form with non-zero entries $d_1, \cdots, d_m$ with $d_i | d_{i+1}$.
Therefore, it is clear that $h_r(AD) = h_r(DA)$ for $D = diag(\pm 1, \cdots, \pm 1)$.
Now, for such A, we have, if $i > m$ that $AX_{ij} = A$ and, if $i \leq m$, $AX_{ij} = A + A'$ where A' is a matrix whose only nonzero entry is d_i at the (i, j)-th place.
Clearly, $h_r(AX_{ij}) = h_r(A)$.
Similarly, we see also that $h_r(X_{ij}A) = h_i(A)$. Therefore, we have the first assertion.
For the second, we merely note that for 'diagonal' matrices A as above, with $d_i | d_{i+1}$, the numbers $h_i(A) = d_1 \cdots d_i$. Thus, the invariant factors are successive quotients of the h_i's.

Proposition.
The matrices $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ generate $SL(2, \mathbf{Z})$.

Proof
Now $S^{-1}X^{-1}S = Y := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Therefore, *all* the matrices in $SL(2, \mathbf{Z})$ which are of the form $\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}$ are in $< S, X >$.
If $< S, X > \neq SL(2, \mathbf{Z})$, define

$$b_0 = min\{|b| : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}) \setminus < S, X >\}.$$

Note that $b_0 \neq 0$ if $< S, X > \neq SL(2, \mathbf{Z})$. If $g_0 := \begin{pmatrix} a & b_0 \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}) \setminus < S, X >$, then $g_0 S^{-1} X^n$ has $(1,2)$-th entry $a - nb_0$. But, if n is so chosen that $|a - nb_0| < b_0$ (possible when $b_0 \neq 0$), we have $g_0 S^{-1} X^n \in < S, X >$ by the choice of b_0. Thus $g_0 \in < S, X >$, a contradiction of the assumption that $< S, X > \neq SL(2, \mathbf{Z})$. Hence the proof.

A more transparent proof is given in the notes. This is :
We shall prove, equivalently, that S and $A := S^{-1}X^{-1}$ generate $SL(2, \mathbf{Z})$. Note that $AS = Y := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.
Now, start with any $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z})$. We shall show that left and right multiplications by powers of X and Y lead to $\pm I$ by the usual Euclidean division algorithm.
For any integer l, we have $X^l g = \begin{pmatrix} a + lc & b + ld \\ c & d \end{pmatrix}$. This shows us that one can divide a by c and replace a by its residue mod c.
Similarly, one can see that by left multiplication by some Y^l, one can reduce c mod a. Repeating these finitely many times, the division algorithm implies that one of a and c becomes zero; the other has to be ± 1 as the determinant is always 1.
So, g becomes $g_1 = \begin{pmatrix} 0 & \pm 1 \\ \mp 1 & d \end{pmatrix}$ or $g_2 = \begin{pmatrix} \pm 1 & b \\ 0 & \pm 1 \end{pmatrix}$.
Now,

$$g_1 X^{\pm d} = \begin{pmatrix} 0 & \pm 1 \\ \mp 1 & 0 \end{pmatrix} = S^{\mp 1}$$

and $g_2 = X^b$ or $-X^{-b}$.
Since $-I = S^2$, the assertion follows.

Lemma.
A free group of any finite rank is a subgroup of finite index in $SL_2(\mathbb{Z})$.
'Proof'.
We use the following lemma which we already proved in § 1. It shows that F_2 is of finite index in $SL(2, \mathbb{Z})$. Now, F_2 has $\mathbb{Z}^2$ as a quotient and $\mathbb{Z}^2$ clearly quotients of any order. Thus, F_2 contains subgroups of every finite index. Finally, it is a fact called the Nielsen-Schreier Theorem (which we shall prove two sections from now) that a subgroup of index d in F_2 is a free group of rank $1 + d$.

Remark.

If $F(x,y) \to \mathbf{Z}/n\mathbf{Z}$ is the homomorphism $x \mapsto 0, y \mapsto 1$, then its kernel is of index n,

Lemma.
The matrices $g = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ *and* $h = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ *generate a free group.*

As mentioned earlier, more generally $\begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}$ and $h = \begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}$ generate a free group for any complex number z with $|z| \geq 2$.
This, and other such results can be most conveniently proved using the notion of free products and a trick due to Klein known as the ping-pong lemma. Although the general notion of 'free product' will be discussed in detail by Professor Anandavardhanan, we will also briefly introduce it because even for questions on subgroups of free groups etc., this notion is useful. Firstly, we define it in the special case of two groups.

If G_1, G_2 are subgroups of a group G, then G is said to be the *free product* of G_1 and G_2 (written $G = G_1 * G_2$) if, for every group H and homomorphisms $\theta_i : G_i \to H$, there is a unique homomorphism $\theta : G \to H$ so that $\theta|_{G_i} = \theta_i$ for $i = 1, 2$.

The universal property makes it clear what is meant by saying that a group G is the free product of a certain family of subgroups. In section 4, we will show how to construct it after defining presentations. A free group is a free product of copies of cyclic subgroups corresponding to a basis. For the moment, we state and accept without proof (prochronistically !) that

$G = G_1 * G_2$ if and only if G is generated by $G_1 \cup G_2$ and no word of the form $x_1^{u_1} y_1^{v_1} x_2^{u_2} y_2^{v_2} \cdots x_r^{u_r} y_r^{v_r}$ with $u_1 \geq 0, v_r \geq 0$ and other $u_i, v_i > 0$ is trivial for nontrivial elements $x_i \in G_1$ and $y_i \in G_2$.

The Ping-Pong lemma:
Suppose G is a group acting on a set S. Suppose there are two nonempty subsets S_1, S_2 of S with S_2 not included in S_1 and subgroups G_1 and G_2 of G such that G_1 has at least 3 elements and satisfy $g(S_2) \subset S_1 \ \forall\ g \in G_1 \setminus \{1\}$ and $h(S_1) \subset S_2 \ \forall\ h \in G_2 \setminus \{1\}$. Then, the subgroup G_0 of G generated by G_1 and G_2 is isomorphic to the free product of G_1 with G_2.
In the case of g and h, note first that $h = wgw$ where $w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Therefore, if $G_1 =< g >$ and $G_2 =< w >$ and take $S_1 = \{z \in \mathbb{C} : |Re(z)| > 1\}$, $S_2 = \{z \in \mathbb{C} : \ |z| < 1\}$, then the subgroup G_0 of $PSL_2(\mathbb{C})$ generated by G_1

and G_2 is their free product. But then since $h = wgw$, the group generated by g and h is their free product which means it is the free group of rank 2.

Proof of the ping-pong lemma.
Let us note first that $G_1 \cap G_2 = \{1\}$. For, if $1 \neq g_1 = g_2 \in G_1 \cap G_2$, then look at some $s_2 \in S_2 \setminus S_1$. Then, for $x_1 \in G_1, x_1 \neq 1, g_1^{-1}$,

$$s_2 = x_1 g_1 g_2^{-1} x_1^{-1}(s_2) \in S_1$$

since x_1^{-1} carries s_2 into an element of S_1 which is, in turn, taken by g_2^{-1} into an element of S_2 which is finally taken by $x_1 g_1$ to an element of S_1. Thus, $s_2 \in S_1$, a contradiction. Thus, $G_1 \cap G_2 = \{1\}$.

Consider any reduced word of the form $w = g_1 h_1 g_2 h_2 \cdots g_r$ where $g_i \in G_1 \setminus \{1\}$ and $h_i \in G_2 \setminus \{1\}$. Note that $w(S_2) \subseteq S_1$. If $w = 1$, then for each $s_2 \in S_2$, we have $s_2 = w(s_2) \in S_1$. Thus, $S_2 \subseteq S_1$, a contradiction. So $w \neq 1$.

Now, if $w = h_1 g_1 \cdots h_r$ is a reduced word, get $x_1 \in G_1$ such that $x_1 \neq 1$. Then, the reduced word $x_1 h_1 g_1 \cdots h_r x^{-1} \neq 1$ by the above argument. So, $w \neq 1$.

If $w = g_1 h_1 \cdots g_r h_r$ is a reduced word, then get $x_1 \in G_1, x_1 \neq 1, g_1^{-1}$. So, $x_1 w x_1^{-1} \neq 1$ by the above argument. Hence, $w \neq 1$.

Similarly, if $w = h_1 g_1 \cdots h_r g_r$ is a reduced word, then $g_r w g_r^{-1}$ is a nontrivial word by the last statement. Hence $w \neq 1$.

Recall the matrices $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ which we showed to be generators for $SL(2, \mathbf{Z})$.
Lemma.
The images s, b of S and SX in $PSL(2, \mathbb{Z}) := SL(2, \mathbb{Z})$ have orders 2 and 3, respectively, and $PSL(2, \mathbb{Z})$ is isomorphic to the free product
$< s > * < b > \cong \mathbb{Z}/2 * \mathbb{Z}/3$.
Remark. This will be useful in computing the abelianization of $SL(2, \mathbb{Z})$.
Proof.
The images of the matrices S, SX correspond to the 'linear fractional transformations' $\alpha : z \mapsto \frac{-1}{z}$, $\beta : z \mapsto \frac{z-1}{z}$ which have orders 2 and 3 respectively. A simple computation shows that $\alpha(P) \subseteq N$, $\beta^{\pm}(N) \subset P$, $\beta^{\pm}(N) \neq P$ where P, N are, respectively, the sets of positive and negative irrational numbers. By the ping-pong lemma, we conclude that α, β generate a free product of $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$.

In the distributed notes, a different (more traditional) proof is given. This goes as follows.
Since $S^2 = -I$ also represents the identity element in $PSL(2, \mathbf{Z})$, the image s of S has order 2 in $PSL(2, \mathbf{Z})$.
Also, the image b of $B := SX = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$ in $PSL(2, \mathbf{Z})$ has order 3 as $(SX)^3 = -I$.
We know that the elements s, b generate the whole group; so, we need only show that no matrix

$$SB^{a_1}SB^{a_2}\cdots SB^{a_r}$$

with each a_i either 1 or 2, can be the matrices $I, -I$.
Since $SB = -X$ and $SB^2 = Y$, it follows that any word in the positive powers of SB and SB^2 is a matrix $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in which a, b, c, d are of the same sign. Therefore, if $b \neq 0$, then the corresponding entry $-b - d$ of SBg and b of SB^2g are non-zero as well. Similarly, if $c \neq 0$, the corresponding entries of SBg and SB^2g are nonzero. Since SB and SB^2 have the property that either the $(1, 2)$-th entry or the $(2, 1)$-th entry is non-zero, any word g in their positive powers has this property; hence g can not be the identity matrix. Therefore, $PSL(2, \mathbf{Z})$ is the free product $< s > * < b >$.

A remark and an exercise.
The commutator subgroup $[F_2, F_2]$ has infinite index in F_2. We shall prove later that :
(i) subgroups of free groups are free and,
(ii) the rank of a normal subgroup of infinite index in F_2 is infinite.

Therefore, there is a subgroup of countably infinite index in $SL(2, \mathbf{Z})$. But, this can be seen explicitly similarly to how we viewed F_2. Indeed, here it is deduced explicitly using the ping-pong lemma.

Example.
Consider an infinite sequence of integers $2 \leq r_1 < r_2 < \cdots$ *where* $r_{i+1} - r_i \geq 3$. *Then, the subgroup of* $SL(2, \mathbf{Z})$ *generated by the matrices* $\begin{pmatrix} -r_i & -1 + r_i^2 \\ 1 & -r_i \end{pmatrix}$ *is free of countably infinite rank.*
We prove that the matrices $T_i = \begin{pmatrix} -r_i & -1 + r_i^2 \\ 1 & -r_i \end{pmatrix}$ generate a free group of infinite rank where $2 \leq r_1 < r_2 < \cdots$ with $r_o - r_{i-1} \geq 3$ for all $i > 1$. We apply the ping-pong lemma.

Consider the discs $K(\gamma) = \{z : |z + d| \leq 1\}$ where $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$. We note that γ maps the interior of $K(\gamma)$ to the exterior of $K(\gamma^{-1})$ and the exterior of $K(\gamma)$ to the interior of $K(\gamma^{-1})$.
Put $T_{-i} = T_i^{-1}$ for convenience. Note that the discs $K(T_i)$ are pairwise disjoint and that (because of the choice of the sequence $\{r_i\}$) there are points P outside all the discs $K(T_i)$ for $i \in \mathbb{Z}$.
Let $S_k \cdots S_1$ be a word in the T_i's. It is easy to see that this word cannot fix P (indeed, $P \notin K(S_1)$ implies $S_1(P)$ is inside $K(S_1^{-1})$ which implies $S_1(P)$ is outside $K(S_2)$ etc.) By Ping-pong, this gives us a free group of countably infinite rank as the matrices T_i generate an infinite cyclic group each.

3 Presentations of groups and Dehn's decision problems

Definition.
A *presentation* of a group G *with generators* X *and relations* R is an isomorphism of G with $F(X)/N$ where R is a subset of $F(X)$ and N is the smallest normal subgroup of $F(X)$ containing R. One usually writes $G =< X|R >$. One calls a group *finitely presentable* if one can choose a presentation $< X|R >$ for G with both X, R finite. One often writes the relations as equations $r = 1$ for all $r \in R$.

Remarks.
The structure theorem of finitely generated abelian groups implies that every such group is automatically finitely presented. In fact, it shows even more - if A is abelian and can be generated by n elements, then it has a presentation of the form $A =< X|R \cup [X, X] >$ where $|X| = n$ and $|R| \leq n$. There is a convenient way to write the relations as a *relation matrix*. Let $X = \{x_1, \cdots, x_n\}, R = \{r_1, \cdots, r_k\}$, and write each r_i as a word in the x's. Gather together the total power m_{ij} of x_j occurring in the expression for r_i, the relation matrix is the $k \times n$ matrix with (i, j)-th entry m_{ij}. Note that the invariant factor theorem gives the structure of the group in terms of the invariant factors of the relation matrix. As a corollary, it follows that if $< X|R >$ is a finitely presentation of a group G such that $|X| > |R|$, then already we have $G/[G, G]$ (so, a fortiori G itself) to be infinite. The fact that any finite group is finitely presentable is not completely obvious but is

proved in the following result.

Lemma.
(i) If N is a normal subgroup of a group G and both $N, G/N$ are finitely generated, then G is finitely generated.
(ii) A subgroup of finite index in a finitely generated group is also finitely generated. Hence, any finite group is finitely presentable.
Proof.
(i) If $x_1, \cdots, x_r \in N$ generate N and $y_1N, \cdots, y_sN$ generate G/N, then clearly $x_1, \cdots, x_r, y_1, \cdots, y_s$ generate G. Indeed, if $g \in G$, then $gN = w(y_1^{\pm}N, \cdots, y_s^{\pm}N)$ which means that $g^{-1}w(y_1^{\pm}, \cdots, y_s^{\pm}) = w_0(x_1^{\pm}, \cdots, x_r^{\pm}) \in N$ for some words w, w_0. Thus, g is a word in the x_i's and the y_j's and their inverses.
(ii) Let $H \leq G$ be of finite index and write $G = \sqcup_{i=1}^{n} Hg_i$ where $g_1 = 1$. Then, for each $g \in G$, there is a corresponding permutation $i \mapsto ig$ of $\{1, 2, \cdots, n\}$ such that $Hg_ig = Hg_{(ig)}$. Here, we have denoted by ig the action of g on i; this is convenient because we have adopted the convention of applying g first in a product gg_1.
Now, we have $g_ig = h(i, g)g_{(ig)}$ for some $h(i, g) \in H$. Let X be a finite set of generators for G. We claim that the elements $h(i, x), x \in X \cup X^{-1}$ generate H.
Let $h \in H$. Write $h = x_1 \cdots x_r$ with $x_i \in X \cup X^{-1}$. Now,

$$h = g_1h = g_1x_1 \cdots x_r = h(1, x_1)g_{(1x_1)}x_2 \cdots x_r$$

$$= h(1, x_1)h((1x_1), x_2)g_{(1x_1x_2)}x_3 \cdots, x_r$$

$$= h(1, x_1)h((1x_1), x_2) \cdots h((1x_1 \cdots x_{r-1}), x_r)g_{(1x_1 \cdots x_r)}$$

$$= h(1, x_1)h((1x_1), x_2) \cdots h((1x_1 \cdots x_{r-1}), x_r)g_{(1h)}.$$

Note that $h = g_1h \in Hg_{(1h)}$ implies that $Hg_{(1h)} = H$; that is, $g_{(1h)} = g_1 = 1$. This proves the assertion that H is finitely generated.

To deduce the assertion that any finite group G is finitely presentable, we write a surjective homomorphism $\theta : F \to G$ from a free group F of finite rank. Since F is finitely generated and Ker θ is of finite index $|G|$ in F, what we proved shows that Ker θ is finitely generated as well. Thus, if R is a finite set of generators for Ker θ and X a finite basis of F, then $G =< X|R >$.

In a later section, we will prove the Nielsen-Schreier theorem asserting that a subgroup of index m in a free group of rank n is also free, of rank $1+m(n-1)$.

This also proves that a subgroup of index m in an n-generated group can be generated by $1 + m(n-1)$ elements.

Examples of presentations

1. $\mathbf{Z} = \langle x \mid \phi \rangle$; $\mathbf{Z}_n = \langle x \mid x^n \rangle$.
2. $\mathbf{Z} \oplus \mathbf{Z} = \langle x,\ y \mid [x,\ y] \rangle$.
3. $\mathbf{Z}^n = \langle x_1,\ \ldots\ ,\ x_n \mid \{[x_i,\ x_j]\ :\ 1 \leq i < j \leq n\} \rangle$
4. $SL(2, \mathbf{Z}) =< x, y | x^2 y^{-3}, x^4 >$.
Furthermore, this implies that the abelianisation of $SL(2, \mathbf{Z})$ is the cyclic group of order 12.
To see this, let us use the result from the previous section which asserts that $PSL(2, \mathbb{Z})$ is isomorphic to the free product $< s > * < b > \cong \mathbb{Z}/2 * \mathbb{Z}/3$ where s, b are the images in $PSL(2, \mathbb{Z})$ of the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\ ,\ X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The presentation for $SL(2, \mathbb{Z})$ follows now by sending S to x and SX to y. Thus,

$$SL(2, \mathbf{Z})_{ab} =< x, y | x^2 y^{-3}, x^4, xyx^{-1}y^{-1} >$$

$$=< e, f | 2e - 3f, ef - fe, 4e > \cong (\mathbf{Z}e \oplus \mathbf{Z}f)/ < 2e - 3f, 4e > .$$

The invariant factors of the subgroup $< 2e - 3f, 4e >$ above are the invariant factors of the matrix $A = \begin{pmatrix} 2 & -3 \\ 4 & 0 \end{pmatrix}$. The latter is computed by computing $h_1(A) = 1 = d_1$ and $h_2(A) = 12 = d_1 d_2$. Clearly, $d_1 = 1$ and $d_2 = 12$. Therefore, the abelianisation of $SL(2, \mathbf{Z})$ is the cyclic group of order 12.
5. $\langle x,\ y \mid x^2 y^3,\ x^3 y^4 \rangle$ is a presentation for the trivial group.
6. The symmetric group $\mathcal{S}_3$ of degree 3 has a presentation

$$\mathcal{S}_3 = \langle r,\ s \mid r^3,\ s^2,\ srsr \rangle.$$

7. Let $D_n,\ n > 1$ be the symmetry group of the regular n-gon P_n. This group is generated by the rotation r with angle $2\pi/n$ and a reflection s in the line through the centre and one of the vertices.

$$D_n = \langle r,\ s \mid r^n,\ s^2,\ (sr)^2 \rangle.$$

8. Let D_∞ be the infinite dihedral group consisting of the motions of $\mathbb{R}$ which map the integers to integers, i.e. the transformations $\mathbb{R} \to \mathbb{R},\ x \mapsto \pm x + k$, with $k \in \mathbf{Z}$.

$$D_\infty = \langle s,\ t \mid s^2,\ stst \rangle.$$

9. $Q = \langle \{x_n\ :\ n \geq 1\} \mid \{x_n = x_{nk}^k\ :\ n,\ k \geq 1\} \rangle$
(Think of $x_n = 1/n!$).
10. Let $n \geq 2,\ X = \{x_1, \ldots, x_{n-1}\}$ and $R = \{x_i^2, (x_i x_{i+1})^3 for\ 1 \leq i \leq n-2,\ [x_i,\ x_j]\ for\ |j-i| > 1\}$. Then, the symmetric group $\mathcal{S}_n = \langle X \mid R \rangle$.
Idea of proof:
Consider $\varphi : X \longrightarrow \mathcal{S}_n,\ x_i \mapsto \sigma_i = (i,\ i+1),\ 1 \leq i \leq n-1$ and observe that $\mathcal{S}_n$ is a homomorphic image of $G = \langle X \mid R \rangle$. Consider $H = \langle x_2,\ \ldots\ ,\ x_n \rangle$ and its cosets

$$K_1 = H,\ K_2 = K_1 x_1,\ K_3 = K_2 x_2, \ldots, K_n = K_{n-1} x_{n-1}$$

and verify that for each i $(1 \leq i \leq n),\ j$ $(1 \leq j \leq n-1)$ we have $K_i x_j = K_l$ for some l $(1 \leq l \leq n)$. Thus H has index at most n and by induction H has order at most $(n-1)!$.
11. The group

$$G = \langle x,\ y \mid x^2,\ y^3,\ (xy)^5 \rangle$$

is the alternating group A_5.
Proof: The elements $\sigma = (12)(34),\ \tau = (135)$ of A_5 satisfy $\sigma^2 = 1,\ \tau^3 = 1,\ (\sigma\tau)^5 = 1$, and hence generate a subgroup of order 30 or 60; since A_5 has no subgroup of order 30 (being simple), therefore $\sigma,\ \tau$ generate A_5.
12. The group with generators $a_1,\ a_2,\ a_3$ and relations

$$a_1^{-1} a_2 a_1 = a_2^2;$$

$$a_2^{-1} a_3 a_2 = a_3^2,$$

$$a_3^{-1} a_1 a_3 = a_1^2$$

is the identity group.
13. (Higman) The group G generated by $a_1,\ a_2,\ a_3,\ a_4$ subject to

$$a_1^{-1} a_2 a_1 = a_2^2;$$

$$a_2^{-1} a_3 a_2 = a_3^2,$$

$$a_3^{-1} a_4 a_3 = a_4^2,$$

$$a_4^{-1}a_1a_4 = a_1^2$$

has no proper normal subgroup of finite index. Indeed, if there is one, say N, then at least one of the elements a_i has nontrivial order n_i in the finite group G/N. Choose p to be the smallest prime dividing some n_i. To fix notation, suppose $p|n_1$. Now $a_4^{-1}a_1a_4 = a_1^2$ gives

$$a_1 = a_4^{-n_4}a_1a_n^{n_4} = a_1^{2^{n_4}}.$$

So $2^{n_4} \equiv 1 \bmod n_1$ (and so mod p also) and so p is odd. Also $n_4 > 1$ also. But $1 < \mathrm{ord}_p(2) < p$ gives $1 < n_4 < p$ which means a prime dividing n_4 would be smaller than p, a contradiction. Thus, $N = G$.
One can also identify this group with a free product with amalgamation in such a way that it is evident that the group is infinite. Since every finitely generated group has at least one maximal normal subgroup, it follows that there exists a finitely generated infinite simple group, namely the quotient group G/N where N is any maximal normal subgroup of G.

Proposition (Philip Hall).
Let N be a normal subgroup of a group G such that both N and G/N are finitely presentable. Then, G is finitely presentable. Moreover, the numbers of generators and of relations for G can be bounded in terms of those for N and those for G/N.
Proof.
Suppose N has generators $x_1, \cdots, x_m$ and relations $r_1 = \cdots = r_k = 1$ where r_i's are words in the x_j's. Let $y_1N, \cdots, y_nN$ generate G/N and let $s_1 = \cdots = s_l = 1_{G/N}$ be relations where s_i's are words in the y_j's. Clearly, G is generated by $x_1, \cdots, x_m, y_1, \cdots, y_n$. Further, these generators obviously satisfy equalities of the following kind :

$$r_1(x) = \cdots = r_k(x) = 1, s_1(y) = t_1(x), \cdots, s_l(y) = t_l(x),$$

$$y_j^{-1}x_iy_j = u_{ij}(x), y_jx_iy_j^{-1} = v_{ij}(x) \ \forall \ i \leq m, j \leq n.$$

Consider the group $\tilde{G}$ which is presented by these generators and relations; to distinguish G from $\tilde{G}$, we will denote the generators of $\tilde{G}$ by $\tilde{x_i}, \tilde{y_j}$ and relations similarly with $\tilde{r_i}$'s etc. Now, there is a surjective homomorphism θ from $\tilde{G}$ to G which carries the $\tilde{x_i}$'s to x_i's etc. Look at the kernel Ker θ. Restricted to the subgroup $\tilde{N} :=< \tilde{x_1}, \cdots, \tilde{x_m} >$ of $\tilde{G}$, the homomorphism θ is an isomorphism as all relations in N are consequences of the relations $r_i(x) = 1$. Thus, $\tilde{N} \cap$ Ker$\theta = \{1\}$. Further, $\tilde{N}$ is a normal subgroup of $\tilde{G}$ as $\tilde{y_j}\tilde{x_i}\tilde{y_j}^{-1}, \tilde{y_j}^{-1}\tilde{x_i}\tilde{y_j} \in \tilde{N}$

by construction. So θ induces a surjective homomorphism from $\tilde{G}/\tilde{N}$ to G/N which is an isomorphism as all relations in G/N are consequences of the relations $y_j N = 1_{G/N}$. Therefore, Ker $\theta = \{1\}$. So, $G \cong \tilde{G}$ and is finitely presented. Note also there are bounds for numbers of generators and of relations for G in terms of those for N and those for G/N.

Dehn's famous decision problems from 1912 (still unsolved in general). These are :

The word problem.
Given a presentation $< X|R >$, is there an algorithm which decides whether two given element w_1, w_2 of $F(X)$ give the same element of $< X|R >$?
There are examples of finitely presented groups with unsolvable word problem.

The conjugacy problem.
Given a presentation $< X|R >$, is there an algorithm to decide if given elements of $F(X)$ are conjugate in $< X|R >$?
A solution of the conjugacy problem also gives a solution of the word problem. The word problem is easier because there exist groups which have solvable word problem but unsolvable conjugacy problem. For instance, every presentation with a single defining relation has a solvable word problem but it is not known whether any such presentation has a solvable conjugacy problem.

The isomorphism problem.
Is there an algorithm to decide whether two given finite presentations give isomorphic groups ?
This is the hardest of the three. Even special cases like an algorithm to decide whether a given finite presentation gives a trivial group does not exist. However, there are so-called Tietze transformations which allow one to go from one finite presentation to any other finite presentation for the same group !

4 Connections with topology - amalgams etc.

The notions of free groups, free products, and of free products with amalgamation come naturally from topology. For instance, the fundamental group of the union of two path-connected topological spaces joined at a single point is isomorphic to the free product of the individual fundamental groups.

The Seifert-van Kampen theorem asserts that if $X = V \cup W$ is a union of path-connected spaces with $V \cap W$ non-empty and path-connected, and if the homomorphisms $\pi_1(V \cap W) \to \pi_1(V)$ and $\pi_1(V \cap W) \to \pi_1(W)$ induced by inclusions, are injective, then $\pi_1(X)$ is isomorphic to the free product of $\pi_1(V)$ and $\pi_1(W)$ amalgamated along $\pi_1(V \cap W)$.
These notions have found many group-theoretical applications. Recall that:
If $G_i, i \in I$ are groups, then a group G along with injective homomorphisms $\phi_i : G_i \to G$ is said to be their free product if, for every group H and homomorphisms $\theta_i : G_i \to H$, there is a unique homomorphism $\phi : G \to H$ so that $\phi \circ \phi_i = \theta_i$ for all $i \in I$.
In other words, G has the universal repelling property with respect to homomorphisms from G_i's to groups.
To construct G, one starts with a presentation $< X_i|R_i >$ of each G_i and takes $< X|R >$ as a presentation of G where X is the disjoint union of the X_i's and R is the union of the R_i's. The homomorphisms $\phi : G_i \to G$ are, therefore, simply inclusions. The uniqueness of such a free product G up to isomorphism follows from the uniqueness property of ϕ above.
One writes $G = *_{i \in I} G_i$. If I is a finite set, say, $I = \{1, 2, \cdots, n\}$, then it is customary to write $G = G_1 * G_2 * \cdots * G_n$.
For example,
The free group of rank r is the free product $\mathbf{Z} * \cdots * \mathbf{Z}$ of r copies of $\mathbf{Z}$.
More generally, a free group $F(X)$ on a set X is the free product $*_{x \in X} < x >$.
The group $PSL(2, \mathbf{Z})$ is the free product of a cyclic group of order 2 and a cyclic group of order 3.
Recall that if A is a group, $G_i, i \in I$ is a family of groups and $\alpha_i : A \to G_i$ $(i \in I)$ are injective homomorphisms, then a group G is said to be the free product of G_i's amalgamated along A, if there are homomorphisms $\phi_i : G_i \to G$ satisfying $\phi_i \circ \alpha_i = \phi_j \circ \alpha_j$ for all $i, j \in I$ such that the following universal property holds: for every group H and homomorphisms $\theta_i : G_i \to H$ with $\theta_i \circ \alpha_i = \theta_j \circ \alpha_j$ for all $i, j \in I$, there is a unique homomorphism $\theta : G \to H$ with $\phi \circ \phi_i = \theta_i$.
One denotes G by $*_A G_i$ if there is no confusion as to what the maps α_i are. Sometimes, the maps α_i are taken to be not necessarily injective and still the above definition can be carried out. Note that, if α_i are trivial, then $*_A G_i = *G_i$, the free product.
The construction is as follows. If $G_i =< X_i|R_i >$, $i \in I$, then

$$G :=< \sqcup_{i \in I} X_i| \cup R_i \cup \cup_{i,j} R_{ij} >$$

where $R_{ij} = \{\alpha_i(a)\alpha_j(a)^{-1}; a \in A\} >$.
The uniqueness of G up to isomorphism is clear once again by the uniqueness of θ.

For instance, $SL(2,\mathbf{Z}) = \mathbf{Z}/4 *_{\mathbf{Z}/2} \mathbf{Z}/6$.
The fundamental group of the Klein bottle is isomorphic to $\mathbf{Z}*_{2\mathbf{Z}}\mathbf{Z}$. A free product with amalgamation could be the trivial group even if the groups $\alpha_i(A)$ are not. For example, let $\alpha_1 : \mathbf{Z} \to PSL(2,Q)$ be an injective homomorphism and let $\alpha_2 : \mathbf{Z} \to \mathbf{Z}/2$ be the natural homomorphism. Then, $G_1 *_{\mathbf{Z}} G_2 = \{1\}$.
Finally, we recall the notion of HNN extensions named after G.Higman, B.H.Neumann & H.Neumann. The construction is akin to adjoining elements to fields to get field extensions.
Let $G =< X|R >$ be a group and let A be a subgroup. For an injective homomorphism $\phi : A \to G$, the HNN extension of G with respect to ϕ is the group $G^* =< X \cup \{t\}|R \cup \{tat^{-1}\phi(a)^{-1}\} >$.
It is a fact that G^* is independent of the presentation of G chosen and that G embeds naturally into G^*. It can be shown that, given two elements a, b of equal order in a group G, this construction enables one to embed the group G into a bigger group in which a, b are conjugate. The HNN construction also finds a natural topological interpretation.
For, suppose V and W are open, path-connected subspaces of a path-connected space X and suppose that there is a homeomorphism between V and W inducing isomorphic embeddings of $\pi_1(V)$ and $\pi_1(W)$ in $\pi_1(X)$. One constructs a space Y by attaching the handle $V \times [0,1]$ to X, identifying $V \times \{0\}$ with V and $V \times \{1\}$ with W. Then, the fundamental group $\pi_1(Y)$ of Y is the HNN extension of $\pi_1(V)$ relative to the isomorphism between its subgroups $\pi_1(V)$ and $\pi_1(W)$.
In the lectures of Professor Anandavardhanan, more details about free products with amalgamation and about HNN extensions will be proved.

5 Subgroups of free groups

We describe the methods of Nielsen and of Schreier (1901-1929) which tell us that subgroups of free groups are free; they also give explicit generators. The main aim of this section is to prove the following result which is a version of the Riemann-Hurwitz formula for a covering space :

Nielsen-Schreier Theorem
If G is a subgroup of a free group F, then G is a free group. Moreover, if G has finite index m in F, then $m(\text{rank } F - 1) = \text{rank } G - 1$ (that is, the rank of G is precisely $1 + m(n-1)$, where n is the rank of F which may be infinite).

Remark.
There is a direct proof of this theorem verifying the universality property. This has recently been written down by Ribes and others using wreath products.

Corollary of the N-S theorem.
(i) If $H \leq G$ is a subgroup of index m in an n-generated group G, then H is $1 - m + mn$-generated.
(ii) Commuting elements a, b in a free group are powers of a common element.
Proof.
(i) Start with a surjection $\pi : F_n \to G$ and look at $\pi^{-1}(H)$.
(ii) The subgroup $< a, b >$ is abelian and free and hence cyclic.

Definitions and remarks on Nielsen's method.
Let F be free with a basis X. For $w_1, w_2 \in F$, there exist a, b, c such that $w_1 = ab^{-1}, w_2 = bc$ and ac is a reduced expression for $w_1 w_2$. So

$$l(w_1 w_2) = l(w_1) + l(w_2) - 2l(b) \leq l(w_1) + l(w_2).$$

For a product of more than two words, it becomes more complicated and Nielsen's idea is to isolate a subset of words where the numbers of cancellations are limited and to show that simple transformations (akin to elementary transformations of integral matrices) reduce words to these special words. Nielsen's method is now vastly generalized and goes under the name of 'cancellation theory'.

For a finite subset U of G, define *elementary Nielsen transformations* on $U = \{u_1, \cdots, u_n\}$ as :

(T0) delete some u_i when $u_i = 1$,
(i') replace some u_i by u_i^{-1}, and
(ij) replace some u_i by $u_i u_j$ for some $i \neq j$.
Of course, it is unsaid here but understood that the other u_k's are unchanged.

A *Nielsen transformation* is a finite sequence of the elementary Nielsen transformations. For instance, for $U = \{u_1, u_2, 1\}$, the transformation $(12)(2')(T0)$ (read from left to right) takes U to the set $\{u_1 u_2^{-1}, u_2^{-1}\}$.
It is easy to see that the group generated by U is unchanged after we apply Nielsen transformations (as it is clearly so for each elementary Nielsen transformation) and the Nielsen transformations form a group (as each has an inverse).

Define a subset U of F to be *Nielsen-reduced* if, $\forall w_1, w_2, w_3 \in U \cup U^{-1}$:

(N0) $w_i \neq 1$,
(N1) either $w_1 = w_2^{-1}$ or $l(w_1 w_2) \geq \max(l(w_1), l(w_2))$,
(N2) either $w_1 = w_2^{-1}$ or $w_3 = w_2^{-1}$ or $l(w_1 w_2 w_3) > l(w_1) - l(w_2) + l(w_3)$.

For example, any set of the form $\{w, w^{-1}\}$ with $w \neq 1$ is reduced.
Another example is, if $x, y \in X$, then $\{y, x^{-1}yx, x^{-2}yx^2, \cdots, \}$ is Nielsen-reduced.

It is also convenient to fix a well-ordering on $X \cup X^{-1}$. This gives lexicographic ordering '$<$' on the reduced words.
For example, if $X = \{x, y\}$ and the ordering of $X \cup X^{-1}$ is $x < y < x^{-1} < y^{-1}$. Then, the first few elements of $F(X)$ are :

$$1 < x < y < x^{-1} < y^{-1} < x^2 < xy < xy^{-1} < yx < y^2 < yx^{-1}$$

$$< x^{-1}y < x^{-2} < x^{-1}y^{-1} < yx < y^2 < yx^{-1} < \cdots$$

Define the *left half* $L(w)$ of a word w to be the initial (left) part of length $[\frac{l(w)+1}{2}]$. Define a well-ordering of the pairs (w, w^{-1}) (which we will simply denote as $w_1 \prec w_2$) if either $\min(L(w_1), L(w_1^{-1})) < \min(L(w_2), L(w_2^{-1}))$ or they are equal and $\max(L(w_1), L(w_1^{-1})) < \max(L(w_2), L(w_2^{-1}))$.

The main property which we will need is :

Proposition.
Let $U = \{u_1, \cdots, u_n\}$ be any finite subset of F. Then, there exists a Nielsen transformation which changes U into a Nielsen-reduced set V. In fact, there is a polynomial time algorithm to do this.
Proof.
If U does not satisfy (N1), then (after a permutation of $U^{\pm 1}$), $l(u_i u_j) < l(u_i)$ for some $u_i u_j \neq 1$. Clearly, in a free group $l(u^2) \geq l(u)$; thus, $i \neq j$. By using the Nielsen transformation (ij), one may reduce the sum $\sum l(u_i)$. Inductively, assuming that this sum has been chosen to be the minimum possible, then we can assume that U does satisfy (N1). Also, after the transformation (T0), we may suppose (N0) holds. Finally, consider $w_1, = x, w_2 = y, w_3 = z$ with $xy \neq 1 \neq yz$. By (N1), $l(xy) \geq l(x), l(yz) \geq l(z)$. Thus, the part of y which cancels in the formation of xy is as well as the part of y that cancels while forming yz are both of lengths at the most the half of $l(y)$.
Thus, $x = ap^{-1}, y = pbq^{-1}, z = qc$ are all reduced, and so are $xy = abq^{-1}$

and $yz = pbc$.
If $b \neq 1$, then $xyz = abc$ is reduced and therefore,

$$l(xyz) = l(x) - l(y) + l(z) + l(b) > l(x) - l(y) + l(z)$$

which means (N2) already holds.
If $b = 1$, then $x = ap^{-1}, y = pq^{-1}, z = qc$ and indeed (N2) does not hold as $l(p) = l(q) \leq l(x)/2, l(z)/2$ and $p \neq q$.

Example.
Look at $F = F(x, y)$ with the order $1 < x < y < x^{-1} < y^{-1} < \cdots$ etc. Consider the subgroup H of all elements of even length. That is, it is the kernel of the homomorphism from F to $\mathbf{Z}/2\mathbf{Z}$ which sends a word to the sum of the exponents mod 2. So, a set of generators is $U = \{x^2, xy, xy^{-1}, yx, y^2, y^{-1}x\}$. At the first step, we apply $(35)(6')(46)(6')(1')(61)(1')$ to U and arrive at

$$U_1 = \{x^2, xy, xy, y^2, y^2, y^{-1}x^{-1}\}$$

Applying $(2')(32)(2')(4')(54)(4')(62)$ to U_1 and arrive at

$$U_2 = \{x^2, xy, 1, y^2, 1, 1\}$$

Applying T_0 to $3, 5, 6$ we arrive at $U_3 = \{x^2, xy, y^2\}$. This satisfies (N0),(N1) but not (N2) because $l(w_3 w_2^{-1} w_1) = 2$. So, we apply $(2')(32)(2')(3')$ to U_3 and arrive finally at $V = \{x^2, xy, xy^{-1}\}$ which is Nielsen-reduced. From the following proposition, it follows that H is free.

Proposition.
Let U be any (could be infinite also) subset of F satisfying (N0),(N1) and (N2). Then, for each $u \in U^{-1}$, there exist words $a(u), m(u)$ with the 'middle part' $m(u) \neq 1$ such that $u = a(u)m(u)a(u^{-1})^{-1}$ is a reduced expression and such that if $w = u_1 u_2 \cdots u_n$ with $n \geq 0, u_i \in U \cup U^{-1}$, $u_i u_{i+1} \neq 1$, then

$$l(w) = l(u_1 \cdots u_{i-1} a(u_i)) + l(m(u_i)) + l(a(u_i^{-1})^{-1} u_{i+1} \cdots u_n).$$

In particular, $l(w) \geq n$. In even more particular, $< U >$ is free.
Proof.
For $u \in U^{\pm 1}$, look at the longest initial part $a(u)$ that cancels in some product $vu \neq 1$ with $v \in U^{\pm 1}$. The hypothesis (N2) on U implies that the initial part $a(u)$ and the final part $a(u^{-1})^{-1}$ do not exhaust u. Thus, we have

$u = a(u)m(u)a(u^{-1})^{-1}$, a reduced expression for some $m(u) \neq 1$.
Now, consider any $w = u_1 u_2 \cdots u_n$ with $n \geq 0, u_i \in U \cup U^{-1}$, $u_i u_{i+1} \neq 1$. After cancellations between adjacent u_i's, the non-reduced word $u_1 \cdots u_n$ leads to a word $w' = m_1' m_2' \cdots m_n'$ where m_i' is a middle part of u_i containing $m(u_i)$. Thus, $m_i' \neq 1$ and since there is no cancellation between m_i' and m_{i+1}', the word w' is reduced. So, it is the reduced form of w and the terms $m(u_1), \cdots, m(u_n)$ remain uncancelled in w'. Thus, the assertion $l(w) = l(u_1 \cdots u_{i-1} a(u_i)) + l(m(u_i)) + l(a(u_i^{-1})^{-1} u_{i+1} \cdots u_n)$ on length follows. Thus, $l(w) \geq n$ and, therefore, $< U >$ is free.

Remarks.
There is a small point to be made here. Nielsen reduced sets might contain two words of the form w, w^{-1} and hence, may not be bases of $< U >$. However, another set of Nielsen transformations would reduce such a set to a basis.

Corollary (Nielsen subgroup theorem).
A finitely generated subgroup of a free group must be free.
Proof.
If U is any finite subset of F, the proposition previous to the one above shows that one can change U to a Nielsen-reduced set V by Nielsen transformations; thus $< U >=< V >$ which last group is free with basis V, by the above proposition.

Corollary.
A free group of finite rank n cannot be generated by fewer than n elements. Further, any generating set of cardinality n is a basis.
Proof.
Note that when we Nielsen-transform a finite set, the cardinality never exceeds the earlier one. Let Y be a subset of F with $\leq n$ elements for which $< Y >= F_n$. After a Nielsen transformation of Y, we get a basis V of F with $|V| \leq |Y| \leq n$. Therefore, we must have $|V| = |Y|$.

Exercise. Let F be free on $\{x, y\}$. Put $a_n = x^{-n} y x^n$ for all natural numbers n. Prove that no reduced word in $a_1, \cdots, a_n$ (for any n) is trivial. Deduce that the subgroup F_0 of F generated by $a_n, n \in \mathbf{N}$ is free of countably infinite rank.

Definitions and remarks - Schreier's method.
Let F be free on a set X and denote by $l(\cdot)$, the usual length function of F

with respect to X. For any subgroup G of F, a *(right) transversal* is a set containing exactly one element from each right coset of G in F.
A *Schreier transversal* is a right transversal S which contains all the initial parts of all its elements; that is, if $x_1 \cdots x_n \in S$ with $l(x_1 \cdots x_n) = n$, then $x_1 \cdots x_r \in S$ for all $0 \leq r \leq n$. Note that inclusion of the case $r = 0$ means that the empty word (the identity element) always belongs to a Schreier transversal. In order to construct Schreier transversals, let us introduce a well-ordering on F (that is, so that each subset has a least element) as follows. Choose any well-ordering of $X^{\pm 1}$ and order $F(X)$ by length and then lexicographically on words of equal length. That is, for different reduced words, we have $x_1 \cdots x_r < y_1 \cdots y_s$ if either $r < s$ or if $r = s$ and $x_k < y_k$ in $X^{\pm 1}$ where $k = min\{i : x_i \neq y_i\}$. Note that 1 is the least element of any subset of F which contains 1.

Lemma.
Every subgroup G of F has a Schreier transversal. Indeed, the set S consisting of the least elements in each right coset gives a Schreier transversal.
Proof.
Consider S as above. Suppose, if possible, that $x_1 \cdots x_n \in S$ but $x_1 \cdots x_{n-1} \notin S$. Note that $x_1 \cdots x_n$ is a reduced word. Let $y \in Gx_1 \cdots x_{n-1}$ be such that $y < x_1 \cdots x_{n-1}$. Consider yx_n. If $l(y) < n - 1$, then clearly $l(yx_n) < n = l(x_1 \cdots x_n)$ and so $yx_n < x_1 \cdots x_n$ which contradicts the choice of $x_1 \cdots x_n$ as an element of S. So, $l(y) = n - 1$ and let us write $y = x_1 \cdots x_{r-1} y_r \cdots y_{n-1}$ in reduced form where $y_r < x_r$. Here r could be 1. If $y_{n-1} = x_n^{-1}$, then $l(yx_n) = n - 2$ which gives again $yx_n < x_1 \cdots x_n$, again an impossibility. So $y_{n-1} \neq x_n^{-1}$, and the expression $yx_n = x_1 \cdots x_{r-1} y_r \cdots y_{n-1} x_n$ is in reduced form which again shows that $yx_n < x_1 \cdots x_n$, a contradiction.

We should keep in mind that there could be other Schreier transversals too which are different from the one constructed in the lemma.

Lemma.
Let $G \leq F(X)$ and consider the Schreier transversal constructed in the above lemma. For $w \in F$, write $\bar{w}$ for the element of $Gw \cap S$. Then, $S_G := \{sx\overline{sx}^{-1} : s \in S, x \in X^{\pm}\}$ generates G.
Proof.
Note that the map $w \mapsto \bar{w}$ from F to S satisfies :
(i) $\overline{\bar{w}} = \bar{w}$,
(ii) $\bar{w} = w$ if and only if $w \in S$,

(iii) $\overline{\bar{s}xx^{-1}} = s$ for all $s \in S, x \in X^{\pm}$.
The last property follows because $Gsx = G\overline{sx}$ and so, $Gs = G\overline{sx}x^{-1}$.
Evidently, $S_G \subseteq G$ as $Gsx = G\overline{sx}$. Conversely, consider any element $g \in G$ and write $g = x_1 \cdots x_n$ in reduced form. Look at the n elements of S; $s_1 = 1, s_2 = \overline{s_1 x_1}, \cdots, s_{n+1} = \overline{s_n x_n}$. Then $t_i := s_i x_i s_{i+1}^{-1} = s_i x_i \overline{s_i x_i}^{-1} \in S_G \subseteq G$ for $1 \leq i \geq n$. Therefore,

$$t_1 \cdots t_n = s_1 x_1 x_2 \cdots x_n s_{n+1}^{-1} = g s_{n+1}^{-1}.$$

As all other terms are in G, we get $s_{n+1} \in G$. On the other hand, $S \cap G = \{1\}$ which means $s_{n+1} = 1$; that is, $g = t_1 \cdots t_n \in S_G$. Hence $G =< S_G >$.

From S_G, to get a basis of G (and thereby show G is free), we proceed as follows. Look at the subset T of the generating set S_G defined as $T := \{sx\overline{sx}^{-1} : s \in S, x \in X, sx \notin S\}$, and the sets $T^{-1} := \{t^{-1} : t \in T\}, T' := \{sx\overline{sx}^{-1} : s \in S, x \in X^{-1}, sx \notin S\}$.

Lemma. $T' = T^{-1}$.
Proof.
For $s \in S, x \in X^{\pm}$, we noticed that $s = \overline{\bar{s}xx^{-1}}$.
Now, $sx \notin S$ if and only if $sx \neq \bar{s}x$ if and only if $s \neq \bar{s}xx^{-1}$ if and only if $\overline{\bar{s}xx^{-1}} \neq \bar{s}xx^{-1}$ if and only if $\bar{s}xx^{-1} \notin S$.
Now, let $x \in X$. So, the above shows that $T^{-1} \subseteq T'$.
If $x \in X^{-1}$, the above discussion shows that $(T')^{-1} \subseteq T$. Therefore, we have $T^{-1} = T'$.

Key observation of Schreier :
Note that $T \cup T' = S \setminus \{1\}$. If $w = sx\overline{sx}^{-1}, w' = ty\overline{ty}^{-1} \in T \cup T'$, then in the reduced expression of ww', neither x nor y can cancel unless $ww' = 1$. This is how the Schreier property arose. More precisely, :

Lemma.
Let $w = sx\overline{sx}^{-1}, w' = ty\overline{ty}^{-1} \in T \cup T' (= S_G \setminus \{1\})$. Then, $x\bar{s}\bar{x}^{-1}ty$ has a reduced expression of the form xzy for some $z \in F$ unless $t = \bar{s}\bar{x}, xy = 1, s = \bar{t}\bar{y}$.
Proof.
Write reduced expressions $\bar{s}\bar{x} = x_1 \cdots x_m$ and $t = y_1 \cdots y_n$ with $x_i, y_j \in X^{\pm}$. Note that if $y_n y = 1$, then $ty = y_1 \cdots y_{n-1} \in S$ (as $t \in S$) and we would have $w' = 1$, a contradiction. Therefore, $y_n \neq y^{-1}$. In the same way, $x_m \neq x$ for, otherwise $\bar{s}\bar{x}x^{-1} = x_1 \cdots x_m x^{-1} = x_1 \cdots x_{m-1} \in S$ which means $\overline{\bar{s}\bar{x}x^{-1}} = \bar{s}\bar{x}x^{-1}$. But, we already observed in (iii) of an earlier lemma that the left

hand side $\overline{\bar{s}\bar{x}x^{-1}}$ is always equal to s. Therefore, if $x_m = x$, we would have $\bar{s}\bar{x}x^{-1} = s$ which would give $w = 1$, a contradiction. Hence, we have $x_m \neq x$ as well as $y_n y \neq 1$. The upshot is that therefore, $x\bar{s}\bar{x}^{-1}$ begins with x and ty ends with y. Thus, the only thing left is to check for any cancellation in the product $\bar{s}\bar{x}^{-1}t$. Now, the reduced form of $\bar{s}\bar{x}^{-1}ty$ ends with y because, if not, $ty = y_1 \cdots y_n y$ would be an initial part of $\bar{s}\bar{x}$ and would thus be in S which contradicts the assumption that $w' \in T \cup T'$. Similarly, the reduced form of $x\bar{s}\bar{x}^{-1}t$ has x at the beginning for, if not, $(x\bar{s}\bar{x}^{-1})^{-1} = \bar{s}\bar{x}x^{-1}$ would be an initial part of t and would thus be in S - this is impossible because then $sx = \overline{\bar{s}\bar{x}x^{-1}}x = (\bar{s}\bar{x}x^{-1})x = \bar{s}\bar{x}$ would be in S, a contradiction to the assumption that $w \in T \cup T'$. Therefore, nether x nor y can cancel while obtaining the reduced form of $x\bar{s}\bar{x}^{-1}ty$ unless they come together to cancel each other. This happens if and only if $\bar{s}\bar{x}^{-1}t = 1 = xy$. Of course, these two statements also imply $\bar{t}\bar{y} = \overline{\bar{s}\bar{x}x^{-1}} = s$. This proves the lemma.

Proof of Nielsen-Schreier theorem.
As we know S_G generates G and hence so does T because $T \cup T^{-1} = S_G \setminus \{1\}$. We claim that T is a basis of G. Write $w = w_1 \cdots w_n$, $w_i = s_i x_i \overline{s_i x_i}^{-1}$, a *reduced word in* $T \cup T^{-1} = S_G \setminus \{1\}$. So, elements of the form $b = sx\bar{s}\bar{x}^{-1}, b^{-1} = ty\bar{t}\bar{y}^{-1}$ satisfying $t = \bar{s}\bar{x}, xy = 1, s = \bar{t}\bar{y}$ cannot occur in adjacent places in w. By the above lemma, we have that the reduced form of w as a word in $X^{\pm}$ contains the middle letters x_i's as separate letters. Thus, $l(w) \geq n$. As $n \geq 1$, we have $w \neq 1$; so G is free on T.

To compute the rank of the subgroup, we need to compute $|T|$; that is, the number of nontrivial elements of the form $sx\overline{sx}^{-1}$. So, let us compute the number of such elements which become trivial. Note that such an element is trivial if $sx = \overline{sx}$. Now, let P_l (respectively P'_l) denote the number of $\overline{s}$ of length l which end in some $x \in X$ (respectively, $x \in X^{-1}$). Put $P_0 = 1, P'_0 = 0$. Now, if $sx = \overline{sx}$ and s has length l, then $l(sx) = l \pm 1$. For a given l, and arbitrary $x \in X$, there are P'_l of our elements s for which $l(sx) = l - 1$ (because if s ends in x^{-1}, then $\overline{sx}$ is an initial segment of s and so $sx\overline{sx}^{-1} = 1$).

Conversely, if some element of S has length $l+1$ and ends in $x \in X$, then it is of the form $\bar{s}x$ where $\overline{s}$ is its initial segment of length l; and then $\overline{s}x\overline{sx}^{-1} = 1$. The number of such elements is P_{l+1}. The total number of elements of the

form $\overline{s}x\overline{sx}^{-1}$ which reduce to the identity is, therefore,

$$\sum_{l\geq 0}(P'_l + P_{l+1}) = \sum_{l\geq 0}(P'_l + P_l) - P_0 = \#S - 1 = m - 1.$$

Therefore, the rank of the subgroup G is $mn-(m-1)$. The proof is complete.

Corollary. $[F_n, F_n]$ has infinite rank.
Proof. Write $\{x_1, \cdots, x_n\}$ for a basis of F_n. Clearly every coset contains a unique $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ where $\alpha_i \in \mathbb{Z}$. The set S of these clearly form a Schreier transversal. But, for $k \leq n$, the elements $\overline{g} = x_1^{\alpha_1} \cdots x_k^{\alpha_k}$ satisfy $\overline{g}x_i\overline{gx_i}^{-1} = 1$ if and only if $k \leq i$. Thus, S is infinite.

Aliter for rank computation.
Here is another way of proving the assertion about the rank. Note that the elements of T indexed by the pairs (s, x) in $S\times X$ with $sx \notin S$ are all distinct by the previous lemma. Thus, rank $G = |S|$ rank $F - d$ where $d = |\{(s, x, t) \in S \times X \times S : sx = t\}|$. If we show that $d = m - 1$, then we would have rank $G = m$ rank $F - (m - 1)$. Note that the assertion on rank holds when rank F is infinite. Let us assume then that rank $F < \infty$. Consider the (Cayley) graph $\mathcal{T}$ with m vertices labelled by elements of S where there is a directed edge from s to sx for each $x \in X$. By the Schreier property of S, each vertex is connected to 1 by a path; therefore, the graph is connected. Moreover, there are no circuits because F is free on X. Hence the graph is a tree on $m - 1$ edges (by Euler's formula for the Euler characteristic). Finally, since the edges of the graph are in bijection with the set $\{(s, x, t) \in S \times X \times S : sx = t\}$, we have $m - 1 = |\{(s, x, t) \in S \times X \times S : sx = t\}| = d$.

6 Rewriting process

A method of Reidemeister and Schreier allows us to show that the subgroup of finite index in a finitely presented group is itself finitely presented. The process allows us to find a presentation also ! Let us state the relevant theorem; the proof is easy and we do not discuss it now (it is in the notes).

Reidemeister-Schreier Theorem
Let $G =< X|R >$ be a finite presentation. Let H be a subgroup of finite index m in G. Then, one has a presentation $< T|\tilde{S} >$ for H with $|T| = 1 - m + m|X|, |\tilde{S}| = m|R|$.

Remark. In fact, we will see in the next section an algorithm which starts with a presentation $< X|R >$ of G, and with a set of words in $X^{\pm}$ which generate H and yields the index of H, a Schreier transversal for H in G and the permutation representation of G on the right cosets of H - all in finitely many steps.

Proof of theorem.
We have already discussed how to get T. Let us recall this briefly. Let $\theta : F(X) \to G$ be the canonical homomorphism and let $K := \theta^{-1}(H)$. Now, start with the Schreier transversal S for K in $F(X)$ and look at $T = \{sx\bar{sx}^{-1} : s \in S, x \in X, sx \notin S\})$. Note that this way we got a basis for K and its image under θ yields generators for H also thus. Consider the subset $\tilde{R} := \{srs^{-1} : s \in S, r \in R\}$ of K. Essentially, the construction of relations for H consists in rewriting $\tilde{R}$ nicely as follows. Consider the normal closure $\bar{R}$ of R in $F(X)$; it lies in K. It is generated by the elements wrw^{-1} for $w \in F(X), r \in R$. Write $w = ks$ with $k \in K, s \in S$. Hence, $\bar{R}$ is generated by $ksrs^{-1}k^{-1}$ which means that $\tilde{R}$ is generated by the conjugates in K of the elements of $\tilde{R}$. That is, $\bar{R}$ is the normal closure of $\tilde{R}$ in K. Now, if we write the elements of $\tilde{R}$ as words in $T \cup T^{-1}$ and call the set of words as $\tilde{S}$, the group $K/\bar{R}$ has the presentation $< T|\tilde{S} >$. But clearly $K/\bar{R} \cong H$.

Example
Consider the free group $F = \langle x,\ y \mid \phi \rangle$. The homomorphism $\varphi : F \longrightarrow \mathbf{Z}_n = \{0, 1, 2, \ldots, n-1\},\ n \geq 2$ defined by $x \mapsto 0, y \mapsto 1$ and let U be the kernel of φ. As coset representatives we take $1, y, y^2, \ldots, y^{n-1}$; clearly this system satisfies the Schreier property. Then the Reidemeister-Schreier generators for U are the non-trivial elements of the following set:

$$\{y^i.x.(\overline{y^i x})^{-1},\ y^i.y(\overline{y^{i+1}})^{-1} \mid i = 0, 1, 2, \ldots, n-1\}$$

We obtain the non-trivial elements

$$y^n,\ x_i = y^i.x.y^{-(i+1)}$$

for $i = 0, \ldots, n-2$ and $x_{n-1} = y^{n-1}.x$. The rank of this group is $n+1$. Note that this shows that :
The free group of rank 2 contains a free group of rank n as a subgroup of index n − 1.

7 Coxeter-Todd-Moser algorithm

Let $< X|R >$ be a finitely presentation of a group G. Given a subgroup H of finite index in G, we will discuss an algorithm which yields in finitely many steps : (i) the index, (ii) a Schreier transversal, (iii) the permutation representation of G on the cosets. In order to explain the algorithm clearly, we start with the case $H = \{1\}$; that is, we assume $< X|R >$ is a finite presentation of a finite group and determine : (i) the order of G, (ii) a Schreier transversal of $\bar{R}$ in $F(X)$, (iii) a faithful permutation representation of G.

Description of algorithm.
For each relation $r = x_1 \cdots x_n$ as a reduced in $x_i \in X^{\pm}$, we draw a rectangular table with $n+1$ columns and an unspecified number of rows. In fact, if the procedure stops after r rows, then the order of G turns out to be r. In the table with $n+1$ columns, write the symbols $x_1, \cdots, x_n$ on top of the n vertical lines separating the columns. Enter the symbol 1 in the first and last places of the first row of each table. The other places are as yet empty. In an adjacent place to one of these two 1's, write a 2. For instance, if 2 is written in the place to the immediate right of the leftmost 1, then define $1x_1 = 2$. Record this on a list to be continued. If $x_n = x_1^{-1}$, then it means that $2x_n = 2x_1^{-1} = 1$ as defined; so, fill in a 2 to the left of the rightmost 1 in case $x_n = x_1^{-1}$. Otherwise, leave that place empty. Having filled in 2's and 1's in the first row according to the definition $1x_1 = 2$, there may or may not be any empty place in that row. Since we introduced the symbol 2, start a second row and put 2 in the leftmost and rightmost places. Continue defining as "$2x_1 = 3$ and 3 in the 2nd place of 2nd row" or as "$3x_n = 2$ and 3 in the second-last place". Record this definition. Continuing in this manner, we stop when the rows are complete. What might happen is that we filled a place with a symbol and that forces an identity. That is, suppose on a row, there are three places with i, empty slot and k from left to right. If a symbol j is entered in between because of a definition of the form $ix_r = j$, then we have the 'bonus' result that $jx_s = k$ where x_s is on the vertical line separating the columns with j and k. We record this also in the same list as the definitions although we put the definitions on the first column of the list and the bonuses on the second column of the list. Actually, we don't draw the vertical lines between columns because we wish to put some 'dashes' in between the symbols $1, 2$ etc. When we define something like $1x_1 = 2$, we

put $1-2$ with a single 'dash'. If we have a bonus, say $jx_s = k$, we put $j = k$ with two 'dashes'. If we fill a place with a symbol j to the right of a symbol i because of some earlier bonus, we write $i \equiv j$ with three 'dashes'. If j has been filled because of an earlier definition, we do not put any 'dash'. The monitor list will yield the permutation representation of G on the set of right cosets. The Schreier transversal comes from the definition column.
We must bear in mind that we need to fill in all the tables simultaneously. Let us illustrate this with a simple example first where there is only one relation (and therefore, only one table).

Example : $G =< x|x^5 >$
The only relation is $x^5 = 1$ and so we make a table with 6 columns with x written on the top of where each of the 5 vertical lines separating the columns would be. Start with 1's at the 1st and the 6th place of the first row. Let us put 2 at the 2nd place; that is, we define $1x = 2$ which we write as $1 - 2$. So, the 3rd,4th and 5th places of the row are empty as yet. Start a second row with 2 at the extremities. So, on the 5th place of this row, we must put a 1 and write no 'dash' in between. Continuing in this manner, we have the following table :

1 - 2 - 3 - 4 - 5 = 1
2 3 4 5 ≡ 1 2
3 4 5 ≡ 1 2 3
4 5 ≡ 1 2 3 4
5 ≡ 1 2 3 4 5

The list recorded is :
Definitions :
1x=2
2x=3
3x=4
4x=5
————-
Bonuses :
5x=1
————
The monitor list is :
1x=2,2x=3,3x=4,4x=5,5x=1.

Thus $O(G) = 5$. The monitor list yields the permutation representation of G given by $x \mapsto (12345)$. The Schreier transversal comes from the definition column of the recorded list; the j-th element comes from the first one. In our example, we get $1, 1x, 1x^2, 1x^3, 1x^4$.

Example : $G =< a, b | baba^{-1}, abab^{-1} >$
This is the so-called Fibonacci group $F(2, 3)$. Here, we have two tables as follows.

Table corresponding to $baba^{-1}$:

1-2-3-4=1
2-6-8=3 2
3 4 6=7-3
45264
58125
67586
71457
83718

Table corresponding to $abab^{-1}$:

14-5=21
234=62
37=143
467=54
5268=5
68376
75817
8=1238

Definitions :
1b=2,2a=3,3b=4,4b=5,2b=6,3a=7,6a=8
Bonuses :
1a=4,5a=2,4a=6,6b=7,7b=1,7a=5,8b=3,5b=8,8a=1
Note that $|F(2, 3)| = 8$ which was not easy to guess by inspection.
A schreier transversal is $\{1, b, ba, bab, bab^2, b^2, ba^2, b^2a\}$ and the permutation representation (right regular representation) $G \to S_8$ is described by $a \mapsto (1, 4, 6, 8)(2, 3, 7, 5)$ and $b \mapsto (1, 2, 6, 7)(3, 4, 5, 8)$.

8 More properties of free groups

Lemma.
(i) A free group F is residually-p for any prime p.
(ii) If G is a finitely generated, residually finite group, then every surjective homomorphism from G to itself is an isomorphism.
Proof.
(i) Let p be any prime. Let $g \in F$ be a nontrivial element and suppose $g = x_1^{a_1} \cdots x_n^{a_n}$ where x_i are basis elements (not necessarily distinct), each $a_i \neq 0$ and $x_i \neq x_{i+1}$.
Let d be an integer bigger than the power of p dividing the product $a_1 \cdots a_n$.
We consider the (finite) p-group P consisting of all $(n+1) \times (n+1)$ upper triangular matrices with entries from $\mathbf{Z}/p^d$ and all diagonal entries 1.
Of course, the matrices $I + E_{ij}$ for $i < j$ generate P where E_{ij} has only the one nonzero entry 1 at the (i,j)-th place.
Of course, to define a homomorphism from F to P, we need to specify only the images of the basis elements but we also need for our purpose to ensure that the element g maps to a nontrivial matrix.
Therefore, it is natural to gather together, for each basis element x of F, all the i's such that $x_i = x$ and define

$$\theta : F \to P \ ;$$

$$x \mapsto I \quad if \quad x_i \neq x \ \ \forall \ \ i \leq n \ ,$$
$$x \mapsto \prod_{i:x_i=x} (I + E_{i,i+1}).$$

In the last expression, the factors on the right side commute since consecutive x_i and x_{i+1} are unequal and $E_{i,i+1}E_{j,j+1} = 0$ unless $i + 1 = j$.
Therefore, for a basis element x occurring in g,

$$\theta(x) = I + \sum_{i:x_i=x} E_{i,i+1}.$$

Let us check whether $\theta(g) = I$ is possible. Note that

$$\theta(g) = \theta(x_1)^{a_1} \cdots \theta(x_n)^{a_n}$$

$$= (I + a_1 \sum_{i:x_i=x_1} E_{i,i+1}) \cdots (I + a_n \sum_{i:x_i=x_n} E_{i,i+1}).$$

We see that $E_{1,2}, E_{2,3}, \cdots, E_{n,n+1}$ occur in that order and their coefficients are precisely the integers $a_1, \cdots, a_n$.
Therefore, since $E_{i,j}E_{j,k} = E_{i,k}$, the coefficient of $E_{1,n+1}$ is the product $a_1 \cdots a_n$ which is not zero in $\mathbf{Z}/p^d$ by the choice of d. Hence $\theta(g) \neq I$ and so, F is residually p.
(ii) Let G be any finitely generated, residually finite group. Suppose $\theta : G \to G$ be a surjective homomorphism which has a nontrivial element g in the kernel.
Let N be a normal subgroup of finite index in G such that $g \notin N$.
Since G is finitely generated, there are only finitely many different homomorphisms $\theta_1, \cdots, \theta_n$ from G to G/N. Note that the composite map $\alpha = \beta \circ \pi : G \to G$, where $\pi : G \to G/Ker(\theta)$ is the natural map and $\beta : G/Ker(\theta) \to G$ is induced by θ, is such that α maps g maps to the identity.
Also, since $\theta_i \circ \alpha : G \to G/N; i \leq n$ are distinct, these are just the θ_i in some order. This means that every homomorphism from G to G/N maps g to the identity. This contradicts the fact that the natural homomorphism maps g to a nontrivial element. Therefore, θ must be an automorphism.

We have a very interesting fact about normal subgroups of free groups of finite rank. We shall prove it using a theorem due to Burns and Marshall Hall which we shall not prove.

Proposition.
If a finitely generated subgroup H of a free group F contains a non-trivial normal subgroup of F, then H has finite index in F. In particular, a non-trivial normal subgroup of a free group of finite rank is finitely generated if and only if it is of finite index.

Theorem (Burns, M.Hall).
Let F be a free group, A a finite subset and H a finitely generated subgroup disjoint from A. Then, H is a free factor of a group G which is of finite index in F and disjoint from A.

Proof of proposition.
Let $1 \neq N \subseteq H$, N normal in F. By the above theorem, there exists a group of the form $G = H * K$ which is of finite index in F. Suppose H has infinite index in F. Then $K \neq 1$ (because $H \neq G$ because G has finite index in F). Write $G = F(X)$ where $X = X_H \sqcup X_K$ with $H =< X_H >, K =< X_K >$. If $1 \neq x \in N, 1 \neq y \in K$, then $y^{-1}xy \in N \subseteq H$. However, $y^{-1}xy$ is not a word in the generators X_H. Therefore, H must have finite index in F.

Remarks on Aut F

We mention some interesting facts about the automorphism group of a free group mostly without proofs (excepting the lemma below). The reader may refer to [1] for proofs of these facts as well as for a proof of the proposition above. In what follows, F is free of finite rank $n \geq 2$.

1. The group Aut F is complete; that is, the center of Aut F has trivial center and all its automorphisms are inner.

2. The natural homomorphism $F \to F/[F.F] \cong \mathbf{Z}^n$ induces an epimorphism Aut $F \to GL_n(\mathbf{Z})$. If $n = 2$, the kernel is precisely the subgroup of inner automorphisms of F (which is, of course, isomorphic to F).

3. The group Aut F is finitely presented; a nice presentation due to McCool has generators similar to the elemenrary matrices (so-called Whitehead transformations).

4. The finite subgroups of Aut F map isomorphically onto a subgroup of $GL_n(\mathbf{Z})$.

Lemma.

(i) Let G be a finitely generated, residually finite group. Then Aut G is also residually finite.

(ii) Let G be a finitely generated and virtually, a residually p-group for some prime p. Then Aut G is also virtually a residually p-group.

Proof.

(i) Let $\theta \in Aut(G)$ be a nontrivial element; suppose $\theta(g) \neq g$ for some $g \in G$. Then, the element $x := \theta(g)g^{-1} \neq 1$.

Since G is residually finite, there exists $H \leq G$ of finite index such that $x \notin H$.

As G is finitely generated, it has only finitely many subgroups of any given finite index. In particular, the set of subgroups $\{\sigma(H) : \sigma \in Aut(G)\}$ is a finite set since each of them has index $[G : H]$. Letting

$$C := \bigcap_{\sigma \in Aut(G)} \sigma(H)$$

it follows that C is a characteristic subgroup of finite index in G such that $x \notin C$.

Consider the homomorphism $Aut(G) \to Aut(G/C)$.

We claim that the finite quotient group Aut (G/C) of Aut (G), is such that θ maps to a nontrivial element of it. This would prove that Aut (G) is residually finite.

Now, if θ were to map to the identity, this precisely means that $\theta(h)h^{-1} \in C$ for every $h \in G$. As this does not hold for $h = x$, we have the assertion.
(ii) As G is virtually residually-p, there is a subgroup G' of finite index in G which is residually-p. By taking the intersection of all the finitely many subgroups $\sigma(G')$ as σ runs through Aut G, we get a characteristic subgroup G_0 of finite index in G which is residually-p. If $\mathrm{Aut}(G_0)$ is virtually residually-p, so is $\mathrm{Aut}(G)$ as seen by pulling back via the restriction homomorphism from $\mathrm{Aut}(G)$ to $\mathrm{Aut}(G_0)$. Without loss of generality, we, therefore, assume that G itself is residually-p. Let H be any characteristic subgroup of p-power index in G. Consider the p-group $P = G/H$. Now, $P/\Phi(P) \cong \oplus_1^r \mathbf{Z}/p$ where $\Phi(P)$ denotes the Frattini subgroup of P. Moreover, by problem 58, the number r of copies of $\mathbf{Z}/p$ is bounded independently of H; indeed, $r \leq n$, where n is the minimal number of generators of G. Now, by the previous problem,

$$Ker\ Aut(P) \to Aut(P/\Phi(P)))$$

is a p-group. Note that $\mathrm{Aut}(P/\Phi(P)) \cong GL(r, \mathbf{Z}/p)$, and that there are only finitely many homomorphisms from $\mathrm{Aut}(G)$ to $GL(n, \mathbf{Z}/p)$, since there are only finitely many subgroups of index bounded by the order of $GL(n, \mathbf{Z}/p)$ in the finitely generated group Aut G. Call $\mathcal{A}$ to be the intersection of the kernels of all the homomorphisms from $\mathrm{Aut}(G)$ to $GL(n, \mathbf{Z}/p)$. We claim that $\mathcal{A}$ is residually-p. Let $\sigma \in \mathcal{A}, \sigma \neq id$. So, there is g such that $\sigma(g)g^{-1} \neq id$. Let H be a characteristic subgroup of p-power index in G such that $\sigma(g)g^{-1} \notin H$; then $\sigma \notin \mathrm{Ker}(\mathrm{Aut}(G) \to \mathrm{Aut}(G/H))$. Call $P = G/H$. Consider, now, the composite

$$\mathcal{A} \hookrightarrow Aut(G) \to Aut(P) \to Aut(P/\Phi(P)) \hookrightarrow GL(n, \mathbf{Z}/p)$$

By the choice of $\mathcal{A}$, the image goes into $N := Ker(Aut(P) \to Aut(P/\Phi(P)))$, a p-group. Since the image of σ is nontrivial in $\mathrm{Aut}(P), Ker(\mathcal{A} \to N)$ is normal, of p-power index in $\mathcal{A}$, and does not contain σ. This completes the proof.

Remarks.
For any group G, one defines the *stable commutator length* of any element $g \in [G, G]$ to be $\lim_{n\to\infty} l(g^n)/n$ (if it exists) where $l(x)$ for any element $x \in [G, G]$ is the smallest number r such that x can be written as a product of r commutators. Note that $l(g^n) \leq nl(g)$ for any $g \in [G, G]$. Very recently (Journal of the AMS), it has been proved that in a free group F, the stable

commutator length of any element of the commutator subgroup is a rational number.

9 Examples of free products and amalgams

Example 1.
The free product of $\mathbf{Z}/2\mathbf{Z}$ and and $\mathbf{Z}/2\mathbf{Z}$ is isomorphic to the infinite dihedral group D_∞.

$$\mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/2\mathbf{Z} \simeq D_\infty := \{x, y | x^2 = 1, xyx^{-1} = y^{-1}\}.$$

Example 2.
With respect to the canonical maps from $\mathbf{Z}$ to $\mathbf{Z}/2\mathbf{Z}$ and $\mathbf{Z}/3\mathbf{Z}$ we get

$$\mathbf{Z}/2\mathbf{Z} *_{\mathbf{Z}} \mathbf{Z}/3\mathbf{Z} = (0).$$

Example 3.
Consider any injective map from $\mathbf{Z}$ to $\mathrm{PSL}_2(\mathbf{Q})$ and the canonical map from $\mathbf{Z}$ to $\mathbf{Z}/2\mathbf{Z}$ then we have

$$\mathrm{PSL}_2(\mathbf{Q}) *_{\mathbf{Z}} \mathbf{Z}/2\mathbf{Z} = (0).$$

Example 3.
This example realizes $\mathrm{PSL}_2(\mathbf{Z})$ and $\mathrm{SL}_2(\mathbf{Z})$ as amalgams. In fact, amalgamated groups are characterized as groups acting on trees with certain special properties.

$$\mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/3\mathbf{Z} \simeq \mathrm{PSL}_2(\mathbf{Z})$$
$$\mathbf{Z}/4\mathbf{Z} *_{\mathbf{Z}/2} \mathbf{Z}/6\mathbf{Z} \simeq \mathrm{SL}_2(\mathbf{Z})$$

Example [Nagao].
Let K be a field and let $K[X]$ be the polynomial ring in one variable X with coefficients in K. Let $G = \mathrm{GL}_2$ and let B be the standard Borel subgroup consisting of upper triangular matrices in G. Then

$$G(K[X]) = G(K) *_{B(K)} B(K[X]).$$

Example (Ihara).
Let F be a non-Archimedean local field. Let $\mathbf{P}$ be the maximal ideal of the ring of integers $\mathcal{O}$ of F. Let $G = \mathrm{SL}_2(F)$. Let $K = \mathrm{SL}_2(\mathcal{O})$ and let I be the subgroup of elements of K which are upper triangular modulo $\mathbf{P}$. Then

$$G = K *_I K.$$

Rational version of Ihara's example.
For a prime number p let $\Gamma_0(p)$ be the subgroup of elements of $\mathrm{SL}_2(\mathbf{Z})$ which are upper triangular modulo p. Let $\mathbf{Z}[1/p]$ be the subring of $\mathbf{Q}$ containing all rational numbers whose denominators is some power of p. Then

$$\mathrm{SL}_2(\mathbf{Z}[1/p]) = \mathrm{SL}_2(\mathbf{Z}) *_{\Gamma_0(p)} \mathrm{SL}_2(\mathbf{Z}).$$

Theorem (Margulis-Tits).
The group $\mathrm{SL}_3(\mathbf{Z})$ is not an amalgam of the form $G_1 *_A G_2$ for any three groups G_1, G_2 and A such that $G_1 \neq A \neq G_2$.
This is actually true in a very general setting. Let F be a number field and let S be a finite set of primes of F. Let $\mathcal{O}(S)$ denote the ring of S-integers of F. If G is a simple Chevalley group of F-rank at least 2 then the group $G(\mathcal{O}(S))$ is not an amalgam.

Exercises on amalgams

Exercise 1.
Show that

$$\mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/2\mathbf{Z} \simeq D_\infty := \{x, y : x^2 = 1, xy = y^{-1}x\}.$$

Exercise 2.
Let $(m, n) = 1$. Then, with respect to the canonical homomorphisms from $\mathbf{Z}$ to $\mathbf{Z}/n\mathbf{Z}$ and $\mathbf{Z}/m\mathbf{Z}$ show that

$$\mathbf{Z}/n\mathbf{Z} *_{\mathbf{Z}} \mathbf{Z}/m\mathbf{Z} = (0).$$

Exercise 3.
Let G be a simple group which admits $\mathbf{Z}$ as a subgroup. Then with respect to the canonical homomorphism from $\mathbf{Z}$ to $\mathbf{Z}/n\mathbf{Z}$ show that

$$G *_{\mathbf{Z}} \mathbf{Z}/n\mathbf{Z} = (0).$$

Exercise 4.
Determine all finite order elements of $PSL(2, \mathbf{Z})$. Give an example of a subgroup of $PSL(2, \mathbf{Z})$ of index 6. Is it free?
Exercise 5.
Let H be a subgroup of $G = G_1 *_A G_2$. Assume that $A \cdot H = G$. Let $B = A \cap H$ and let $H_i = G_i \cap H$ for $i = 1, 2$. Show that H is generated by H_1 and H_2 and can be identified with $H_1 *_B H_2$. Use this to deduce from Ihara's example the rational version.

10 Some applications to abstract group theory

Proposition.
Every countable group can be embedded as a subgroup of a group which can be generated by two elements of infinite order.
Proof.
Consider $G =< X|R >$ where X is countable and the group $F = G* < a, b >$. The subgroup of $< a, b >$ generated by $\{b^{-n}ab^n : n \geq 0\}$ is free as it is Nielsen-reduced. Writing $X = \{x_1, x_2, \cdots, \cdots, \}$, we claim that the subgroup of F generated by $\{x_n a^{-n} b a^n : n \geq 0\}$ is free (here $x_0 = 1$). Indeed, the projection of F onto $< a, b >$ which sends x_i's to 1 sends the elements $x_n a^{-n} b a^n$ to a basis; hence the original is a basis too. Now, G can be embedded in the HNN extension $< F, t|t^{-1}at = b, t^{-1}b^{-n}ab^n t = x_n a^{-n} b a^n; n \geq 1 >$ of F. Note that this HNN extension is generated by the two elements t, a.

Proposition.
Let $G = *_A G_i$. Then, any element of G of finite order can be conjugated inside one of the G_i. In particular, if all the G_i's are torsion-free then so is G.
Proof.
Let $g \in G = *_A G_i$; write $g = g_1 \dots g_n$. Let $l(g) = n$ be the length of g. If

$l(g) \leq 1$ then $g \in G_i$ for some i. If $l(g) \geq 2$ we say g is *cyclically reduced* if $i_1 \neq i_n$.
We now show inductively that that any g is conjugate to either an element of some G_i or to a cyclically reduced element. Assume that $l(g) = n \geq 2$ and that we have shown this for all elements of length at most $n-1$. Suppose g is not cyclically reduced then $i_1 = i_n$ and so conjugating by g by g_1^{-1} we get $g = g_1 \dots g_n \sim g_2 \dots g_{n-1}(g_n g_1)$ and the length of $g_2 \dots g_{n-1}(g_n g_1)$ is at most $n-1$.
Now take any $g \in G$ which is of finite order. Since all the G_i are torsion free, we get that no conjugate of g is in any G_i. We may replace g by a conjugate and assume that it is cyclically reduced. We leave it to the reader to the check that in this case, for any $r \geq 1$ we have that the length of g^r is rn and so g cannot have been an element of finite order unless $n = 0$, i.e., $g = 1$.
Proposition.
If G_1 and G_2 are two finite groups then their free product $G_1 * G_2$ contains a free subgroup of index $o(G_1)o(G_2)$. In particular, the free product of two finite groups admits a faithful finite-dimensional representation.
Proof.
Consider the direct product $G_1 \times G_2$ of G_1 and G_2. The inclusion maps from the G_i into $G_1 \times G_2$ gives a canonical homomorphism from the free product $G_1 * G_2$ to $G_1 \times G_2$. Clearly this map is surjective. Let K be the kernel of this homomorphism.
Let S be the set of commutators in $G_1 * G_2$ given by

$$S = \{xyx^{-1}y^{-1} : x \in G_1, y \in G_2\}.$$

Let N be the subgroup of $G_1 * G_2$ generated by S. Note that N is normal in $G_1 * G_2$ because, if $x \in G_1, y \in G_2$ and $g_1 \in G_1$, then

$$g_1[x,y]g_1^{-1} = [g_1 x, y][g_1, y]^{-1} \in N.$$

Using the universal definitions of direct product and free product it is easy to see that $N = K$. It suffices now to prove that S is a free subset of $G_1 * G_2$. To this end, it suffices to show that for any sequence $s_1, \dots, s_n \in S$ with $s_i = a_i b_i a_i^{-1} b_i^{-1}$ and any sequence $\epsilon_1, \dots, \epsilon_n \in \{\pm 1\}$ (with the condition that if $\epsilon_k = -\epsilon_{k+1}$ then $s_k \neq s_{k+1},$) the element $g = s_1^{\epsilon_1} \cdots s_n^{\epsilon_n}$ is not the identity element. In fact we will show that

$$l(s_1^{\epsilon_1} \cdots s_n^{\epsilon_n}) \geq n + 3.$$

If $\epsilon_n = 1$ (resp. $\epsilon_n = -1$) then g ends with $a_n^{-1}b_n^{-1}$ (resp. $b_n^{-1}a_n^{-1}$).
This can be seen using induction. Without loss of generality assume that $\epsilon_{n-1} = 1$ (the argument for the case $\epsilon_{n-1} = -1$ is similar.)
If $n = 1$ then there is nothing to prove. Let $n \geq 2$.
If $\epsilon_n = 1$ then we may write g as

$$g = t_1 \dots t_p a_{n-1}^{-1} b_{n-1}^{-1} a_n b_n a_n^{-1} b_n^{-1}$$

with $p \geq n$ by induction hypothesis (because $p+2 \geq (n-1)+3$ by induction hypothesis. Hence $l(g) = (p+2)+4 \geq n+6 > n+3$ and g ends with $a_n^{-1}b_n^{-1}$.
If $\epsilon_n = -1$ then we may write g as

$$g = t_1 \dots t_p a_{n-1}^{-1} (b_{n-1}^{-1} b_n) a_n b_n^{-1} a_n^{-1}$$

with $p \geq n$ by induction hypothesis. Now if $b_{n-1} \neq b_n$ then $l(g) = p + 5 \geq n + 5 > n + 3$. If $b_{n-1} = b_n$ then $a_{n-1} \neq a_n$ since $s_{n-1} \neq s_n$. Now, then

$$g = t_1 \dots t_p (a_{n-1}^{-1} a_n) b_n^{-1} a_n^{-1}$$

which gives $l(g) = p + 3 \geq n + 3$ and in either of these two cases g ends with $a_n^{-1}b_n^{-1}$.

Proposition.
Every countable group G can be embedded in a group K which has the property that all the elements of the same order are conjugate.
Proof.
Consider the set $\{(a_i, b_i) : i \geq 1\}$ of all ordered pairs of elements of G which have the same order. Then, the HNN extension $G_1 :=< G, t_n (n \geq 1) | t_n a_n t_n^{-1} = b_n (n \geq 1) >$ has the property that all elements of the same order in G are conjugate in G_1. In this manner, we get G_2 from G_1, G_3 from G_2 etc. and then the group $G_0 := \cup_{n \geq 1} G_i$ satisfies the assertion made.

References

[1] R.C.Lyndon, P.E.Schupp, *Combinatorial group theory*, Springer-Verlag 1977.