Infinite Groups and Decision Problems

Lectured by J. Button and M. Chiodo

Lent Term 2016

1. Introduction

A **group** is a set G with a binary operation that:
(i) is associative (a **semigroup**),
(ii) also has an identity e (a **monoid**)
(iii) and every element has an inverse.

We have non abelian groups and infinite groups but what about examples of groups that are both non abelian and infinite?

Example 1.1.

(i) If X is a set then we will write $\Sigma(X)$ for the group of all permutations (bijections from X to itself) under composition.

(ii) If F is a field and $M_n(F)$ is the set of $n \times n$ matrices with entries in F then the subset $GL(n, F)$ of invertible matrices is a group under multiplication (and $M_n(F)$ is a semigroup).

Subgroups

Proposition 1.2. A subset $H \subseteq G$ is a **subgroup** of $G \Leftrightarrow H \neq \emptyset$ and for all $a, b \in H$, $ab^{-1} \in H$.

Notation. We will write $H \leqslant G$ for a subgroup H of G and $H < G$ for a proper subgroup, that is $H \neq G$. The trivial subgroup will be written $\{e\}$.

Proposition 1.3.

(i) $L \leqslant H$ and $H \leqslant G \Rightarrow L \leqslant G$.

(ii) If $H_i \leqslant G$ for all $i \in I$ then $\bigcap_{i \in I} H_i \leqslant G$.

For $H_1, H_2 \leqslant G$, $H_1 \cup H_2$ is not a subgroup unless $H_1 \leqslant H_2$ or $H_2 \leqslant H_1$. But...

Proposition 1.4 (Ascending sequence of subgroups). If $H_0 \leqslant H_1 \leqslant \ldots \leqslant G$ (which means $H_n \leqslant G$ and $H_n \leqslant H_{n+1}$ for all n) then $\bigcup_{n=0}^{\infty} H_n \leqslant G$.

Example 1.5. Let the natural numbers $\mathbb{N}$ be $\{0, 1, 2, \ldots\}$ and let $G = \Sigma(\mathbb{N})$. Set H_n to be those permutations of $\mathbb{N}$ that fix everything outside $\{0, \ldots, n\}$. Thus this subgroup H_n of $\Sigma(\mathbb{N})$ is a copy of the symmetric group S_{n+1}.

Then $\cup_{n=0}^{\infty} H_n$ is a subgroup of $\Sigma(\mathbb{N})$ by (1.4) and is proper: the permutations "of finite support" (fixing all but finitely many elements).

Generators

Let $X \subseteq G$.

Definition 1.6. The subgroup $\langle X \rangle$ **generated by** X is $\bigcap H$, where this intersection is over all H with $X \subseteq H \leqslant G$. It is the smallest subgroup of G containing X. We will write $\langle x_1, \ldots, x_n \rangle$, $\langle x_1, x_2, \ldots \rangle$, $\langle X, Y \rangle$, $\langle X_i : i \in I \rangle$, etc as variants on this notation.

Definition 1.7. A group G is **finitely generated** (f.g.) if we have some $n \in \mathbb{N}$ and elements $g_1, \ldots, g_n$ such that $G = \langle g_1, \ldots, g_n \rangle$. (For $n = 0$ we regard $\{g_1, \ldots, g_n\}$ as $\emptyset$ and $\langle \emptyset \rangle = \{e\}$.)

Example 1.8. $G = \langle g \rangle$ means that G is cyclic. Either $G = \mathbb{Z}$ or, if G has order m, we write $G = C_m$. Note that, in either the finite or the infinite case, any subgroup of a cyclic group is cyclic.

Proposition 1.9. If $X \subseteq G$ then the elements of $\langle X \rangle$ are

$$x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}$$

for $k \in \mathbb{N}$ (where we again interpret the $k = 0$ case as equal to e), $n_1, \ldots, n_k \in \mathbb{Z}$, and $x_1, \ldots, x_k \in X$ (but they need not be distinct).

Proof. The set of elements of this form is a subgroup of G by (1.2) and contains X so contains $\langle X \rangle$. However every such expression above must be in $\langle X \rangle$ by closure. □

Thus G finite $\Rightarrow$ G f.g. $\Rightarrow$ G countable.

Warning! If we have $H < G$ for G finite/countable then H is finite/countable but G f.g. does not imply H is f.g. - see (2.20) later.

However f.g. is preserved by homomorphic images (quotients) because $\theta(\langle g_1, \ldots, g_n \rangle) = \langle \theta(g_1), \ldots, \theta(g_n) \rangle$ by (1.9).

Cosets

If $H \leqslant G$ then the **left cosets** of H in G are the sets $gH = \{gh : h \in H\}$ for each $g \in G$.

Proposition 1.10 (Lagrange for infinite groups). The left cosets of H in G form a partition of G and any left coset is in bijection with H.

Note. We have right cosets Hg in bijection with H too. Also there is a bijection from the set of left cosets to the set of right cosets given by gH goes to Hg^{-1} (not to Hg).

Definition 1.11.

(i) The **index** of H in G is the cardinality of the set of left (or right) cosets, written $[G : H]$ if finite.

(ii) A **left (or right) transversal** is a choice of left (or right) coset representatives, with exactly one for each coset.

Normal subgroups

For $g, x \in G$ and $H \leqslant G$ the **conjugate** of x by g is $gxg^{-1} \in G$ and the **conjugate** of H by g is $gHg^{-1} = \{ghg^{-1} : h \in H\} \leqslant G$.

Definition 1.12. The subgroup N is **normal** in G (written $N \trianglelefteq G$) if the following equivalent conditions hold:

(i) $gN = Ng$ for all $g \in G$

(ii) $gNg^{-1} = N$ for all $g \in G$

(iii) $gNg^{-1} \leqslant N$ for all $g \in G$

(iv) N is a union of conjugacy classes of G.

Examples.

(i) $\{e\} \trianglelefteq G$ and $G \trianglelefteq G$

(ii) If G is abelian then $H \trianglelefteq G$ for all subgroups H (but not the converse)

(iii) If $[G : H] = 2$ then $H \trianglelefteq G$.

Proposition 1.13, cf. 1.3. (i) If $N \trianglelefteq G$ and $H \leqslant G$ then $N \cap H \trianglelefteq H$. (But $M \trianglelefteq N$, $N \trianglelefteq G \not\Rightarrow M \trianglelefteq G$.)

(ii) $N_i \trianglelefteq G \ \forall i \in I \Rightarrow \bigcap_{i\in I} N_i \lhd G$.

Proposition 1.14, cf. 1.4. If $N_1 \leqslant N_2 \leqslant \ldots \leqslant G$ with $N_n \trianglelefteq G \ \forall n \in \mathbb{N}$ then $\bigcup_{n=1}^{\infty} N_n \trianglelefteq G$.

Normal closures

Let $X \subseteq G$.

Definition 1.15, cf. 1.6. The **normal closure** $\langle\langle X \rangle\rangle^G$ of X in G is $\bigcap N$ over all N with $X \subseteq N \trianglelefteq G$. It is the smallest normal subgroup of G containing X.

Proposition 1.16, cf. 1.9. If $X \subseteq G$ then the elements of $\langle\langle X \rangle\rangle^G$ are

$$g_1 x_1^{n_1} g_1^{-1} g_2 x_2^{n_2} g_2^{-1} \ldots g_k x_k^{n_k} g_k^{-1}$$

for $k \in \mathbb{N}$, $n_1, \ldots, n_k \in \mathbb{Z}$, $x_1, \ldots, x_k \in X$ and $g_1, \ldots, g_k \in G$ (but not necessarily distinct).

Proof. This collection of elements does form a subgroup containing X and it is normal in G: on conjugation by $g \in G$ one just changes the above expression by writing g in front of and g^{-1} behind every subexpression $g_i x_i^{n_i} g_i^{-1}$. But it is also in $\langle\langle X \rangle\rangle^G$. □

Note. This is a relative notion, in that if $X \subseteq H \leqslant G$ then $\langle\langle X \rangle\rangle^H$ need not equal $\langle\langle X \rangle\rangle^G$; indeed we have $\langle X \rangle \leqslant \langle\langle X \rangle\rangle^H \leqslant \langle\langle X \rangle\rangle^G$ because we have more to conjugate with in G than in H. Hence the ^{G} superscript: for the abstract closure $\langle X \rangle$ one could write $\langle X \rangle^G$ but $\langle X \rangle^H = \langle X \rangle^G$ anyway.

Set products

If $A, B \leqslant G$ then the **set product** is $AB = \{ab : a \in A, b \in B\}$. It is not in general a subgroup of G.

Proposition 1.17.

(i) $AB \leqslant G \Leftrightarrow AB = BA$ and if so then $AB = \langle A, B \rangle$.

(ii) For $N \trianglelefteq G$ and $H \leqslant G$ we have $NH = HN$.

Homomorphisms

A function $\theta : G \to H$ for groups G, H is a **homomorphism** if $\theta(xy) = \theta(x)\theta(y)$ for all $x, y \in G$.

It is an **isomorphism** if θ is bijective (both surjective ('onto') and injective ('1-1')), which is equivalent to the existence of an inverse $\theta^{-1} : H \to G$, (where 'inverse' here means 'two-sided inverse'). In this case, θ^{-1} is unique and is also a homomorphism. We write $G \cong H$, and they are then the same as abstract groups.

Sets and Functions

If X, Y are sets and $f : X \to Y$ is a function then for $U \subseteq X$ the **image** (or pushforward) of U is $f(U) = \{f(x) : x \in U\} \subseteq Y$, and for $V \subseteq Y$ the **inverse image** (or pullback) of V is $f^{-1}(V) = \{x \in X : f(x) \in V\} \subseteq X$.

Lemma 1.18.

(i) $f^{-1}f(U) \supseteq U$ and is equal if f is injective.

(ii) $ff^{-1}(V) \subseteq V$ and is equal if f is surjective.

Theorem 1.19.

If $\theta : G \to H$ is a homomorphism and $A \leqslant G$, $B \leqslant H$, we have $\theta(A) \leqslant H$, $\theta^{-1}(B) \leqslant G$.

If $B \trianglelefteq H$ then $\theta^{-1}(B) \trianglelefteq G$. If θ is surjective then $A \trianglelefteq G \Rightarrow \theta(A) \trianglelefteq H$.

Consequently the **image** $\theta(G)$ of θ is a subgroup of H and the **kernel** $\ker \theta = \theta^{-1}(\{e\})$ is a normal subgroup of G. In fact we have

Corollary 1.20. For $\theta : G \to H$ with $A \leqslant G$, $B \leqslant H$ and $K = \ker\theta$ we have $\theta^{-1}\theta(A) = AK$ and $\theta\theta^{-1}(B) = B \cap \theta(G)$.

Quotients and the Isomorphism Theorems

If $N \trianglelefteq G$ then the set of (left) cosets forms a group under the well-defined multiplication $xN \cdot yN = (xy)N$. We say the group G/N is a **quotient** of G.

Theorem 1.21 (*The* Isomorphism Theorem).

If $\theta : G \to H$ is a homomorphism then $G/\ker\theta \cong \theta(G)$.

What are the subgroups of G/N? If $N \leqslant H \leqslant G$ then $N \trianglelefteq H$ by (1.13) and $H/N \leqslant G/N$.

Theorem 1.22 (Correspondence Theorem).

If $N \trianglelefteq G$ then the subgroups of G/N are exactly those of the form H/N for $N \leqslant H \leqslant G$ and the normal subgroups are exactly those of the form L/N for $N \leqslant L \trianglelefteq G$.

Note: On setting $\pi : G \to G/N$ to be the natural homomorphism, if $H \leqslant G$ then $\pi(H) = HN/N$.

Theorem 1.23 (Product Isomorphism Theorem).

If $H \leqslant G$ and $N \trianglelefteq G$ then $H/(H \cap N) \cong HN/N$.

Theorem 1.24 (Quotient Isomorphism Theorem).

Let $N, L \trianglelefteq G$ with $N \leqslant L$. Then $(G/N)\big/(L/N) \cong G/L$.

Definition 1.25. Given $\theta : G \to H$ and $N \trianglelefteq G$ with $\pi : G \to G/N$ the natural homomorphism, we say that θ **factors through** G/N if there exists $\overline{\theta} : G/N \to H$ with $\theta = \overline{\theta} \circ \pi$.

We must have $N \leqslant \ker\theta$ and this is sufficient by setting $\overline{\theta}(gN) = \theta(g)$.

Actions

We say the group G **acts** on the set X (on the left) if there is a function $\psi : G \times X \to X$ such that

$$\psi(e, x) = x \text{ and } \psi(g_1, \psi(g_2, x)) = \psi(g_1 g_2, x) \text{ for all } g_1, g_2 \in G \text{ and } x \in X.$$

Note for each $g \in G$ the function $x \mapsto \psi(g, x)$ is a permutation of X (put in g^{-1}, g and then g, g^{-1} to get an inverse).

Equivalently there is a homomorphism $\rho : G \to \Sigma(X)$ given by $\rho(g)(x) = \psi(g, x)$.

We say G acts **faithfully** (effectively) if ρ is injective. We can then unambiguously write $g(x)$ for $\rho(g)(x)$; we sometimes do this anyway. We say the action is (fixed point) **free** if $\rho(g)(x) = x \Rightarrow g = e$ (which implies faithful).

For $x \in X$ the **orbit** $\mathrm{Orb}(x) = \{\rho(g)(x) : g \in G\} \subseteq X$ and the **stabiliser** $\mathrm{Stab}(x) = \{g \in G : \rho(g)(x) = x\} \leqslant G$. The orbits form a partition of X and the action is **transitive** if there's one orbit. If $y = \rho(g)(x)$ then $\mathrm{Stab}(y) = g\mathrm{Stab}(x)g^{-1}$ so stabilisers are rarely normal.

Theorem 1.26 (Orbit-Stabiliser for infinite groups). If G acts on X then for $x \in X$ the set of left cosets of $\mathrm{Stab}(x)$ is in bijection with the set $\mathrm{Orb}(x)$.

Example 1.27.

(i) G acts on itself via $\rho(g)(x) = gx$. This is transitive, and also free so $G \leqslant \Sigma(G)$ which is Cayley's Theorem.

(ii) G acts on itself by conjugation: $\rho(g)(x) = gxg^{-1}$. Here $\mathrm{Orb}(x)$ is the **conjugacy class** of x while the stabiliser is the **centraliser** $C_G(x) = \{g \in G : gx = xg\}$ of x (or just $C(x)$ if clear). Note $\langle x \rangle \leqslant C(x)$ but it's not generally abelian.

We also have for $H \leqslant G$ the centraliser $C_G(H) = \{g \in G : gh = hg \text{ for all } h \in H\} = \bigcap_{h \in H} C_G(h)$. We set $C_G(G) = Z(G)$, the **centre** of G, which is abelian and normal.

Moreover G acts on the set of its subgroups by conjugation: $\rho(g)(H) = gHg^{-1} \leqslant G$. Then $\mathrm{Orb}(H)$ is the set of subgroups conjugate to H and the stabiliser is the **normaliser** $N(H) = \{g \in G : gHg^{-1} = H\} \leqslant G$. It is the largest subgroup of G in which H is normal.

Automorphisms

An isomorphism (homomorphism) from G to G is an **automorphism** (**endomorphism**).

Example 1.28. For any $g \in G$, $\alpha_g(x) = gxg^{-1}$ is an automorphism so $H \cong gHg^{-1}$ for all $H \leqslant G$. These are the **inner** automorphisms and form a group $\mathrm{Inn}(G)$ under composition.

We have $\alpha_g = e \Leftrightarrow g \in Z(G)$ so $G/Z(G) \cong \mathrm{Inn}(G)$. Moreover all automorphisms form a group $\mathrm{Aut}(G)$, with $\mathrm{Inn}(G) \trianglelefteq \mathrm{Aut}(G)$ and the quotient is defined to be the **outer** automorphism group $\mathrm{Out}(G)$ of G.

Definition 1.29, cf. 1.12. The subgroup C of G is **characteristic** in G if $\alpha(C) = C$ $\forall \alpha \in \mathrm{Aut}(G)$ (but $\alpha(C) \leqslant C$ $\forall \alpha$ is enough), so $C \trianglelefteq G$.

Proposition 1.30.

(i) A characteristic in B, B characteristic in $C \Rightarrow A$ characteristic in C.

(ii) A characteristic in B and $B \trianglelefteq C \Rightarrow A \trianglelefteq C$.

Direct and Semidirect products

We can form the **direct product** $G_1 \times G_2$ from groups G_1, G_2 via $(x_1, x_2) \cdot (y_1, y_2) = (x_1y_1, x_2y_2)$. This is **external**, we can also do this **internally**.

Proposition 1.31. If $M, N \trianglelefteq G$ with $MN = G$ and $M \cap N = \{e\}$ then $\theta : M \times N \to G$ given by $\theta(m, n) = mn$ is an isomorphism.

Definition 1.32. Given groups G_1, G_2 and a homomorphism $\varphi : G_2 \to \mathrm{Aut}(G_1)$, the **semidirect product** $G_1 \rtimes_\varphi G_2$ is the set of ordered pairs with multiplication

$$(x_1, x_2) \cdot (y_1, y_2) = (x_1(\varphi(x_2)(y_1)), x_2y_2).$$

Example 1.33.

(i) φ the trivial homomorphism gives the direct product.

(ii) Take $\mathbb{Z} = \langle z \rangle$ (written additively) and $C_2 = \{e, c\}$. Then $\mathbb{Z} \rtimes_\varphi C_2$ with $\varphi(c)(z) = -z$ is the **infinite dihedral group**.

Proposition 1.34, cf. 1.31. If $H \leqslant G$ and $N \trianglelefteq G$ with $NH = G$ and $N \cap H = \{e\}$, so $G/N \cong H$ by (1.23), then $\theta : N \rtimes_\varphi H \to G$ given by $\theta(n, h) = nh$ and $\varphi(h)(n) = hnh^{-1} \in N$ is an isomorphism.

So again the internal and external versions are equivalent. There is another point of view:

If $G/N = Q$ with $\pi : G \to G/N$ the natural projection, we call G an **extension** of N by Q (but knowing N and Q does not determine G in general, hence *an* extension). We say that the extension **splits** if $\exists H \leqslant G$ such that $\pi : H \to Q$ is an isomorphism.

Now for $G = NH$ a semidirect product, π restricted to H is an isomorphism as $H \cap \ker \pi = \{e\}$. But a split extension implies $H \cap N = \{e\}$ and $NH = G$ as $\pi(H) = \pi(G)$, so they're the same.

Example 1.35. The **special** linear group $SL(n, \mathbb{C}) = \{A \in GL(n, \mathbb{C}) : \det A = 1\}$ is normal in $GL(n, \mathbb{C})$ with quotient $\mathbb{C} \setminus \{0\}$. Now let $n = 2$ and I_2 be the 2×2 identity matrix. Then $PSL(2, \mathbb{C}) = SL(2, \mathbb{C})/\{\pm I_2\}$. This extension does not split because we have elements g (of order four) in $SL(2, \mathbb{C})$ with $\pi(g)$ of order two in $PSL(2, \mathbb{C})$, but the only element of order 2 in $SL(2, \mathbb{C})$ is $-I_2$ with $\pi(-I_2) = e$.

Abelian groups and Abelianisation

Abelian groups that are not finitely generated can be rather strange. However all is well in the f.g. case:

Theorem 1.36 (Classification of finitely generated abelian groups). If G is a finitely generated abelian group then there are unique integers $r, s \in \mathbb{N}$ and $1 < d_1|d_2|\ldots|d_s$ such that

$$G \cong \mathbb{Z}^r \times C_{d_1} \times \ldots \times C_{d_s}.$$

This can be used to prove that subgroups of f.g. *abelian* groups are themselves f.g.

Now let G be *any* group and take $x, y \in G$. The **commutator** $[x, y] = xyx^{-1}y^{-1} \in G$.

Definition 1.37. The **commutator/derived** subgroup G' of G is $\langle [x, y] : x, y \in G \rangle$.

Proposition 1.38. We have $G' \trianglelefteq G$ with G/G' abelian.

Proof. For α an automorphism of G, we have $\alpha([x, y]) = [\alpha(x), \alpha(y)] \in G'$, so G' is in fact characteristic. Then in G/G', $xyG' = yx([x^{-1}, y^{-1}]G') = yxG'$. □

Corollary 1.39. The quotient group G/G' is the largest abelian quotient of G, meaning that if $N \trianglelefteq G$ with G/N abelian then $G' \leqslant N$, so $G \twoheadrightarrow G/N$ factors through G/G' as in (1.25).

Definition 1.40. For any group G, the **abelianisation** of G is this abelian quotient G/G'.

Thus if the abelianisation of a group G is non trivial/infinite then so is G. Moreover if G is f.g. then (1.36) applies to its abelianisation.

2. Free groups and Free Products

Definition 2.1. Let X (say $\{a, b, \ldots\}$) be a set of symbols (which we will call letters).

A **word** on X is a finite sequence of elements of X. Note: The empty word $\emptyset$ is included too.

We write X^+ for the set of *all* words on X, so formally $X^+ = \cup_{n=0}^{\infty} X^n$ where n is the **word length**.

Now let X^{-1} ($= \{a^{-1}, b^{-1}, \ldots\}$ say) be a set with the same cardinality as but disjoint from X, along with a particular bijection $\iota : X \to X^{-1}$ given by $\iota(a) = a^{-1}$, etc. We think of X^{-1} as **formal inverses** for the elements of X.

Also let $X^* = (X \cup X^{-1})^+$, so that here our letters are either elements of X or their formal inverses.

We define $X^{red} \subseteq X^*$ to be the set of **reduced** words on $X \cup X^{-1}$, that is the words which contain no subword xx^{-1} or $x^{-1}x$ for $x \in X$. Later on this could be called **freely reduced** once other notions of reduced are given.

PreDefinition 1. We want to define the free group on X as the set of reduced words X^{red}, where for $w_1, w_2 \in X^{red}$ we would set $w_1 \cdot w_2$ to be the concatenation of these two words, then cancelling any disallowed subwords that occur.

But we won't define it this way (some books do), as the proof of associativity is fiddly because of the cancelling. However it is clear that the empty word $\emptyset$ is the identity and inverses of words are formed by writing the word backwards and changing all letters to their inverses.

Here is another approach, adopted by some.

PreDefinition 2. Instead define an equivalence relation $\sim$ on the set of all words X^* to be the one *generated* by:
Words v and w are equivalent if they are equal or if one is obtained from the other by a deletion (or indeed insertion) of any subword xx^{-1} or $x^{-1}x$ for $x \in X$.

We then define the free group on X to be the set of equivalence classes $[w]$ in $X^*/\sim$ with multiplication defined by $[w_1] \cdot [w_2] = [w_1 w_2]$ (where $w_1 w_2$ is concatenation). Then associativity follows because $([w_1] \cdot [w_2]) \cdot [w_3]$ equals $[w_1 w_2] \cdot [w_3] = [(w_1 w_2) w_3] = [w_1 (w_2 w_3)]$ which itself equals $[w_1] \cdot ([w_2] \cdot [w_3])$.

So what's the problem?

Actually we will define the free group on X as a *subgroup* of the group of all permutations of the reduced words X^{red} (the "Van der Waerden trick"):

Definition 2.2. The **free group** $F(X)$ on the set $X = \{x_i : i \in I\}$ is the *subgroup* of $\Sigma(X^{red})$ which is generated by the elements $\chi_i : X^{red} \to X^{red}$ for $i \in I$ defined as follows:
For any word $w \in X^{red}$ we have

$$\chi_i(w) = \begin{cases} x_i w & \text{if } w \text{ does } \textit{not} \text{ start with } x_i^{-1} \\ w' & \text{if } w = x_i^{-1} w'. \end{cases}$$

This does have image in X^{red}. Moreover χ_i is a bijection because it has inverse χ_i^{-1} where

$$\chi_i^{-1}(w) = \begin{cases} w' & \text{if } w = x_i w' \\ x_i^{-1} w & \text{if } w \text{ does } \textit{not} \text{ start with } x_i. \end{cases}$$

Proposition 2.3. The map $M : X^{red} \to F(X)$ given by replacing each $x_i^{\pm 1}$ in a given word by the respective $\chi_i^{\pm 1}$, and then multiplying the letters together in $F(X)$ (i.e. composing) is an injection.

Proof. If $M(w_1) = M(w_2)$ for $w_1, w_2 \in X^{red}$ then note $M(w_j)(\emptyset) = w_j$ $(j = 1, 2)$ by repeated use of (2.2) from right to left. □

Corollary 2.4. The map $M : X^{red} \to F(X)$ is bijective, and given any (not necessarily reduced) word $w \in X^*$, if we delete all subwords of the form xx^{-1} and $x^{-1}x$ in any order, we reach a unique $w_{red} \in X^{red}$.

Proof. Extend M to a map $X^* \to F(X)$ using exactly the same definition. This is surjective by (1.9) and, given an unreduced word in X^*, each deletion of subwords xx^{-1} or $x^{-1}x$ reduces the word length of w, so we reach some $w_{red} \in X^{red}$. But deletions do not change the group element in $F(X)$, so $M(w) = M(w_{red})$ and w_{red} is unique by (2.3).□

PreDefinition 1 revisited. As $M : X^{red} \to F(X)$ is a bijection, we can make X^{red} into a group by defining multiplication $\cdot$ as $M(w_1 \cdot w_2) = M(w_1) \circ M(w_2)$ for $w_1, w_2 \in X^*$, where $\circ$ is multiplication/composition in $F(X)$. But, writing $w_1 w_2$ for concatenation, we know that $M(w_1 w_2) = M(w_1) \circ M(w_2)$ so we can cancel in $F(X)$ to conclude that $w_1 \cdot w_2$ is the (unique) reduction of $w_1 w_2$ by (2.4).

Thus we have proved that PreDefinition 1 is valid and equivalent to our definition.

PreDefinition 2 revisited. Similarly we see that for any word $w \in X^*$ there is exactly one reduced word w_{red} in $[w]$ by (2.4). Thus if we have $v, w \in X^*$ with $v \sim v_{red} \in X^{red}$ and $w \sim w_{red} \in X^{red}$ then $[v] \cdot [w]$ would equal $[vw]$ whereas $[v_{red}] \cdot [w_{red}]$ would have to be $[v_{red}w_{red}]$ (where the concatenation $v_{red}w_{red}$ is in X^* but not necessarily in X^{red}).

But $v_{red}w_{red} \sim vw$ because we can first cancel within v and within w and then concatenate, or do this in the opposite order. Thus if $v', w' \in X^*$ with $v \sim v'$ and $w \sim w'$ then $[v'w'] = [v'_{red}w'_{red}] = [v_{red}w_{red}] = [vw]$, so we are well defined and equivalent to PreDefinition 1.

Theorem 2.5 (Universal property of free groups). A free group $F(X)$ has the property that any map $f : X \to H$ (where H is any group) extends uniquely to a homomorphism $f^* : F(X) \to H$ such that $f^*(\chi_i) = f(x_i)$.

Moreover if G is a group with generating set S having this universal property then G is **free on** S, namely there is an isomorphism from G to some free group $F(X)$ which sends S to X. (We might also say that S **freely generates** G.)

Proof. First define $f(x_i^{-1}) = f(x_i)^{-1} \in H$, so f extends uniquely to $X \cup X^{-1}$. Then let $f^*(t_1 \dots t_k) = f(l_1) \dots f(l_k)$, where the element $t_j \in \{\chi_i^{\pm 1} : i \in I\}$ is represented by the letter $l_j \in X \cup X^{-1} = \{x_i^{\pm 1} : i \in I\}$, i.e. $M(l_j) = t_j$.

Then f^* is well-defined by (2.3) and (2.4) as if we have two expressions for the same element of $F(X)$ then they end up at the same reduced word after all the deletions. Moreover f is a homomorphism by definition and any such homomorphism must satisfy the property in the first paragraph.

For the last part, let X be an abstract set such that there is a bijection $b : S \to X$ and define $\theta : G \to F(X)$ by using the hypothesis on G and S to extend b. Then θ is surjective because its image includes the generating set X.

Moreover if $\theta(g) = \emptyset$ for $g \in G$ and g is written as a word on $S \cup S^{-1}$ then we can reduce this word to nothing by cancelling in G, because the process of applying θ to each letter is just the replacement of this letter in $S \cup S^{-1}$ with its image in $X \cup X^{-1}$ under the bijection b. Thus g is the identity in G and θ is also injective. □

Thus from now on when given a free group $F(X)$ we can blur the difference between x_i and χ_i above and regard the elements of X, which form a canonical generating set for $F(X)$, as both symbols and maps.

Proposition 2.6. If $F(X)$ and $F(Y)$ are free groups on X and on Y then $F(X) \cong F(Y) \iff$ the sets X and Y are in bijection.

Proof. ($\Leftarrow$) If $f : X \to Y$ is a bijection, then extend to a homomorphism $f^* : F(X) \to F(Y)$ and $(f^{-1})^* : F(X) \to F(Y)$. But $(f^{-1})^* f^*$ fixes X and so is the identity homomorphism (by uniqueness). And the same for $f^*(f^{-1})^*$, so f^* is a bijective homomorphism.

($\Rightarrow$) For $F(X)$ a *free* group and x any element of X we have the **parity homomorphism** $p_x : F(X) \twoheadrightarrow C_2$ given by the parity of the number of occurrences of x and x^{-1} in $w \in F(X)$.

For *any* group G, let $S_G \trianglelefteq G$ be $\langle g^2 : g \in G \rangle$, the subgroup *generated* by the squares. Then the quotient group G/S_G is abelian with all non-identity elements of order 2, and

thus it is a vector space over the field $\mathbb{F}_2$.

Now the images of the elements of X in $F(X)/S_{F(X)}$ span it, as X generates $F(X)$. Moreover they are linearly independent, as if we had distinct elements $x_1, \ldots, x_k$ of X with $x_1 + \ldots + x_k = 0$ in $F(X)/S_{F(X)}$ then $x_1 \ldots x_k$ is a product of squares in $F(X)$, which is a contradiction on applying p_{x_1}.

Thus X has the same cardinality as a basis of this vector space, but the cardinality of a basis is well defined (the dimension). Moreover the same holds for Y because $F(X) \cong F(Y)$ implies that $F(X)/S_{F(X)} \cong F(Y)/S_{F(Y)}$ (both as groups and vector spaces). Thus X and Y are in bijection. □

Consequently we can unambiguously define F_n to be *the* **free group of rank** n where $F_n \cong F(X)$ if and only if $|X| = n$, and also F_∞ for $F(\mathbb{N})$. Then $F_0 = \{e\}$, $F_1 = \mathbb{Z}$, but if $a \neq b \in X$ then $ab \neq ba$ in $F(X)$ by (2.4), so $F(X)$ is non abelian (and of course infinite) whenever $|X| \geqslant 2$.

Corollary 2.7. Every (finitely generated) group is a quotient of a (finitely generated) free group.

Proof. If $H = \langle h_i : i \in I \rangle$, then take $X = \{x_i : i \in I\}$ and use (2.5). □

Definition 2.8. A word w on $X \cup X^{-1}$ is **cyclically reduced** if it is reduced and the first and last letters of w are not mutual inverses.

We can write any reduced word w_0 uniquely as $u|w|u^{-1}$ where u (possibly empty) is reduced, w is cyclically reduced and $|$ means there is no cancellation between the two reduced words on either side.

Proposition 2.9. A free group is torsion free: the only **torsion** (finite order) element is e.

Proof. Given a reduced word $w_0 \neq \emptyset$, write it accordingly as $u|w|u^{-1}$ where $w \neq \emptyset$ and is cyclically reduced. But for $n > 0$, we have $w^n = w|\ldots|w \neq \emptyset$ so $w_0^n = u|w|\ldots|w|u^{-1} \neq \emptyset$. □

Proposition 2.10. If w, w' are cyclically reduced words on $X \cup X^{-1}$ then they are conjugate elements in $F(X)$ iff w' is a cyclic permutation of $w = \ell_1|\ell_2|\ldots|\ell_n$ for $\ell_i \in X \cup X^{-1}$. Namely, $w' = \ell_k|\ell_{k+1}|\ldots|\ell_n|\ell_1|\ldots|\ell_{k-1}$ for some k.

Proof. ($\Rightarrow$) If $w' = cwc^{-1}$ (with c reduced and $c \neq \emptyset$) then as w' is cyclically reduced there exists cancellation in either cw or wc^{-1}, but not both as w is cyclically reduced too. Wlog, say $c = d|\ell_1^{-1}$ and $w = \ell_1|v$ for d (and v) reduced then $w' = dv\ell_1 d^{-1}$, but $v\ell_1 = v|\ell_1$ is a cyclic permutation of w. So either $d = \emptyset$, or repeat with the shorter word d in place of c (and $v|\ell_1$ in place of $w = \ell_1|v$). □

Free Products

Definition 2.11. Let $\{G_\lambda : \lambda \in \Lambda\}$ be an indexed family of groups. A **reduced sequence** in $\{G_\lambda\}$ is a finite sequence $g_1 \ldots g_r$ of elements in the disjoint union $\bigsqcup_{\lambda \in \Lambda} G_\lambda$ such that:

No g_j is equal to the identity in any G_λ and no successive g_j, g_{j+1} are in the same G_λ.

Let $\mathcal{A}$ be the set of all finite sequences in $\bigsqcup G_\lambda$ and let $\mathcal{R}$ be the set of all reduced sequences, again both including $\emptyset$. We say that a reduced sequence is P-**reduced**

(i.e. reduced in the sense of a free *P*roduct) if we need to distinguish between different notions of being reduced.

Definition 2.12, cf. (2.2). The **free product** $*_{\lambda\in\Lambda}G_\lambda$ of the groups $\{G_\lambda : \lambda \in \Lambda\}$ is the subgroup of $\Sigma(\mathcal{R})$ generated by all elements γ_{g_λ} for $\lambda \in \Lambda$ and $g_\lambda \in G_\lambda \setminus \{e_\lambda\}$, where

$$\gamma_{g_\lambda}(g_1 \ldots g_n) = \begin{cases} g_\lambda g_1 \ldots g_n & \text{if } g_1 \notin G_\lambda \\ (g_\lambda g_1) g_2 \ldots g_n & \text{if } g_1 \in G_\lambda \text{ and } g_\lambda g_1 \neq e_\lambda \\ g_2 \ldots g_n & \text{if } g_1 \in G_\lambda \text{ and } g_\lambda g_1 = e_\lambda. \end{cases}$$

Note that this does define a map from $\mathcal{R}$ to $\mathcal{R}$ with inverse given by $\gamma_{g_\lambda}^{-1} = \gamma_{g_\lambda^{-1}}$. We also set γ_{e_λ} to be the identity map.

Proposition 2.13, cf. (2.3). The function $f : \mathcal{A} \to *_{\lambda\in\Lambda}G_\lambda$ given by $f(g_1 \ldots g_n) = \gamma_{g_1} \circ \ldots \circ \gamma_{g_n}$ restricts to a bijection from $\mathcal{R}$ to $*_{\lambda\in\Lambda}G_\lambda$.

Proof. If $g_1, g_2 \in G_\lambda$ then $\gamma_{g_1} \circ \gamma_{g_2} = \gamma_{g_1 g_2}$, so we can gather up neighbouring γs lying in the same group, removing any appearance of an identity element, to get a surjection even when restricted to $\mathcal{R}$.

Also if $g_1 \ldots g_n$ is a reduced sequence then

$$f(g_1 \ldots g_n)(\emptyset) = \gamma_{g_1} \circ \ldots \circ \gamma_{g_n}(\emptyset) = g_1 \ldots g_n.$$

□

Note. Consequently for a given $\lambda \in \Lambda$ the map $\iota_\lambda : G_\lambda \to *_{\lambda\in\Lambda}G_\lambda$ given by $\iota_\lambda(g_\lambda) = \gamma_{g_\lambda}$ is a homomorphism which is injective.

Theorem 2.14, cf. (2.5) (Universal property of free products). For any group H and any collection of homomorphisms $\theta_\lambda : G_\lambda \to H$ for $\lambda \in \Lambda$, there exists a unique homomorphism $\theta : *_{\lambda\in\Lambda}G_\lambda \to H$ such that $\theta_\lambda = \theta \circ \iota_\lambda$ for all λ.

Proof. For $g_1 \ldots g_n \in \mathcal{A}$, we define $\theta(g_1 \ldots g_n) = \theta_{\lambda_1}(g_1) \ldots \theta_{\lambda_n}(g_n)$, where $g_j \in G_{\lambda_j}$.

Similarly we can check this is a well defined homomorphism satisfying the required condition and again it is unique. □

Note. If $F(X)$ is the free group on $X = \{x_i : i \in I\}$ then it is equal to $*_{i\in I}G_i$, where each $G_i = F(\{x_i\}) \cong F_1 \cong \mathbb{Z}$ are infinite cyclic.

We will usually be taking the free product of two groups H_1, H_2 say (the free **factors**), whereupon we will write $H_1 * H_2$ (or equally $H_2 * H_1$) rather than $*_{\lambda\in\Lambda}H_\lambda$ for $\Lambda = \{1, 2\}$. Now by (2.13) we see that this free product is infinite if neither of H_1, H_2 are trivial, whereas $H_1 * \{e\} \cong H_1$ (a trivial free product). Moreover for non-identity elements $h_1 \in H_1$ and $h_2 \in H_2$ we have $(h_1 h_2)^n$ is not the identity for $n > 0$, and $h_1 h_2 \neq h_2 h_1$ so non trivial free products are always non abelian groups.

We also explicitly state here the content of (2.13) when we have a free product of two groups as it will be used later:

Theorem 2.15 (Normal form theorem for free products). For groups H_1, H_2, any element $g \in H_1 * H_2$ can be written uniquely (the **normal form** for g) as $g_1 \ldots g_n$ where $n \geqslant 1$ (or $n = 0$, but only if g is the identity when we write $\emptyset$), each $g_i \in H_1 \cup H_2$ but no g_i is the identity in H_1 or in H_2, and successive g_i, g_{i+1} are taken from alternating factors. □

How do we recognise in practice that a group is a (non trivial) free product?

Suppose X is a topological space and G a subgroup of Homeo(X), the group of all homeomorphisms of X. We say that a subset S of X is a **G-packing** if $g(S) \cap S = \emptyset$ for all $g \in G \setminus \{e\}$.

Theorem 2.16 (Klein's Combination Theorem, 1883). If $G_1, G_2 \leqslant$ Homeo(X), each having a G_i-packing S_i such that $S_1 \cup S_2 = X$ and $S_1 \cap S_2 \neq \emptyset$, then the subgroup G of Homeo(X) given by $G = \langle G_1, G_2 \rangle$ is isomorphic to $G_1 * G_2$.

Proof. Note for $x \in S_1$, $g(x) \notin S_1$ for any $g \in G_1 \setminus \{e\}$ so $g(x) \in S_2$. Of course the same is true on swapping 1 and 2. Now take any point s in $S_1 \cap S_2$ and a non identity element in $G_1 * G_2$, written as a non empty reduced sequence $g_1 \ldots g_n$ with (wlog) $g_n \in G_1$.

Then $g_n(s) \in S_2 \setminus S_1$, $g_{n-1}g_n(s) \in S_1 \setminus S_2$, etc, so $g_1 \ldots g_n(s) \neq s$.

Now we can apply (2.14) to create the homomorphism $\theta : G_1 * G_2 \to \langle G_1, G_2 \rangle$ by setting $H = \langle G_1, G_2 \rangle$ and $\theta_j : G_j \to H$ the subgroup inclusions for $j = 1, 2$. This is clearly surjective as G_1, G_2 are hit.

But the proof of (2.14) says that $\theta(g_1 \ldots g_n)$ is just $g_1 \ldots g_n$, i.e. the reduced sequence is sent to the corresponding composition of homeomorphisms which was seen above to move s and so is not the identity map. Thus θ is injective too. □

Example 2.17.

(i) Let $X = \mathbb{R}^2$ and consider the two reflections a in the line $x = 0$ and b in the line $x = \frac{1}{2}$.

See Picture (1) - "Ping Pong"

Now $G_1 = \langle a \rangle$ and $G_2 = \langle b \rangle$ are both copies of C_2, thus $\langle a, b \rangle = C_2 * C_2$ by (2.16). This is the infinite dihedral group (1.33) as clearly $\langle a, b \rangle = \langle a, ba \rangle$ with $ba(x) = x + 1$ and $a(ba)a^{-1} = (ba)^{-1}$.

(ii) (For those who have seen Möbius Transformations)

Let $f(z) = \dfrac{az + b}{cz + d}$ with $a, b, c, d \in \mathbb{C}$, $ad - bc \neq 0$.

This is a bijection of $\mathbb{C} \cup \{\infty\}$, with inverse $f^{-1}(z) = \dfrac{dz - b}{-cz + a}$.

Now let $f(z) = z + 2$ and $g(z) = \dfrac{z}{2z + 1}$, so that here $G_1 = \langle f \rangle$ and $G_2 = \langle g \rangle$ are both copies of $\mathbb{Z}$. Using Picture (2) and (2.16) with X the upper half plane $\mathcal{U} = \{z \in C : \mathrm{Im}(z) > 0\}$, we get that $\langle f, g \rangle = \mathbb{Z} * \mathbb{Z} = F_2$ is free on $\{f, g\}$.

See Picture (2) - "Upper half plane"

Proposition 2.18. Let $\mathcal{F} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, $\mathcal{G} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \in SL(2, \mathbb{Z})$.

Then $\langle \mathcal{F}, \mathcal{G} \rangle \cong F_2$ is free on $\{\mathcal{F}, \mathcal{G}\}$.

Proof. We can regard the Möbius transformations f, g as elements of $PSL(2, \mathbb{Z}) = SL(2, \mathbb{Z})/\pm I_2$ with $\pi : SL(2, \mathbb{Z}) \to PSL(2, \mathbb{Z})$ the natural quotient map. Now as $\langle f, g \rangle$ is free on $\{f, g\}$ we can use (2.14) to define a homomorphism $\theta : \langle f, g \rangle \to SL(2, \mathbb{Z})$ which sends f, g to $\mathcal{F}, \mathcal{G}$. But $\pi\theta$ is the identity as it fixes f, g so θ is injective and hence $\langle \mathcal{F}, \mathcal{G} \rangle = \theta(\langle f, g \rangle) = \theta(F_2) \cong F_2$. □

We now begin to look at subgroups of free groups, starting with the following useful example which we record as a lemma and which will be seen again later.

Lemma 2.19. Take the group F_2, free on a, b say. Now define elements $a_n = b^n a b^{-n} \in F_2$ and the subset $S = \{a_n : n \in \mathbb{N}\}$.

Then S *freely generates* the subgroup $\langle a_n : n \in \mathbb{N}\rangle \cong F_\infty$ of F_2.

Proof. Let $*_{n\in\mathbb{N}}\langle x_n\rangle$ be the countably infinite free product of the infinite cyclic groups $\langle x_n\rangle$. As mentioned, this is also the group F_∞ and is free on $\{x_n : n \in \mathbb{N}\}$. Now use (2.5) or (2.14) to obtain a homomorphism $\theta : F_\infty \to F_2$ with image $\theta(F_\infty) = \langle S\rangle$ by extending the map or homomorphisms that send(s) x_n to a_n.

So given a P-reduced sequence $g_1 \dots g_k \in F_\infty$ which is non trivial, thus $k \geqslant 1$, we have

$$\theta(g_1 \dots g_k) = a_{n_1}^{p_1} \dots a_{n_k}^{p_k}$$

where for $1 \leqslant i \leqslant k$ we have $g_i = x_{n_i}^{p_i} \in \langle x_{n_i}\rangle \setminus \{e\}$, so $p_1, \dots, p_k \in \mathbb{Z} \setminus \{0\}$ and $n_1, \dots, n_k \in \mathbb{N}$. But successive n_i are distinct because $g_1 \dots g_k$ is P-reduced. However the right hand side is the element

$$b^{n_1} a^{p_1} b^{n_2 - n_1} a^{p_2} \dots b^{n_k - n_{k-1}} a^{p_k} b^{-n_k}$$

which we see is freely reduced in F_2 and not empty, so is not the identity. Thus θ is injective. □

Note. In place of S, one could use any subset T of S with exactly the same proof to conclude that T freely generates the subgroup $\langle T\rangle$ of F_2 and thus $\langle T\rangle \cong F_{|T|}$ by (2.6).

Corollary 2.20. If G is a finitely generated non abelian free group, or even a finitely generated group containing a non abelian free group, then G has a subgroup which cannot be finitely generated.

Proof. By repeated use of (2.19) and the note thereafter, we have

$$F_\infty < F_2 < F_3 < F_4 < \ldots < F_\infty < F_2.$$

Now G being finitely generated means that it and any subgroup of it is countable. Any countable non abelian free group must be one of those listed above, so G contains a subgroup isomorphic to F_∞. But as shown in (2.6) any finitely generated group H must have $H/S(H)$ finite dimensional, whereas $F_\infty/S(F_\infty)$ is infinite dimensional. □

Example: "Schröder Bernstein fails bigtime for groups and homomorphisms, even if the groups are finitely generated!"

From (2.20) we see that there is an injective homomorphism from F_2 to F_3 but also one from F_3 to F_2. However they are not isomorphic groups by (2.6).

So some strange things can happen with subgroups of free groups.... hang on though: all subgroups of free groups we have seen so far were themselves free. Could this always be true?

Yes! This is the **Nielsen Schreier Theorem** for which we shall present a topological proof, once we have covered the required concepts.

3. Topological proof of the Nielsen Schreier Theorem

Handout: **Background in Algebraic Topology**

Definition 3.1. A **graph** Γ (a 1-dim CW complex) is made up of a set V of **vertices** with the discrete topology and a set E of **edges** $\{I_\alpha : \alpha \in A\}$, where each edge I_α is a homeomorphic copy of the compact interval $[0,1]$ with its two endpoints labelled $I_\alpha(0)$ and $I_\alpha(1)$. All of these edge endpoints are identified with points in V as we are also given **attaching maps** $f_0, f_1 : E \to V$ so that Γ is given the quotient topology

$$\Gamma = \bigsqcup_{\alpha\in A} I_\alpha \sqcup V \Big/ I_\alpha(0) = f_0(I_\alpha), I_\alpha(1) = f_1(I_\alpha) \text{ for all } \alpha \in A.$$

Namely we start with the space $U = \bigsqcup_{\alpha\in A} I_\alpha \sqcup V$ with the disjoint union topology and then set Γ to be the quotient topological space formed by identifying each edge endpoint with some vertex according to f_0 and f_1.

For any edge I_α we will write e_α for the "interior" of I_α and call it an **open edge** as it embeds homeomorphically in Γ as a copy of the open interval $(0,1)$. The closure $\overline{e_\alpha}$ in Γ of an open edge e_α is either an embedded homeomorphic copy of I_α when $f_0(I_\alpha) \neq f_1(I_\alpha)$, or is homeomorphic to the circle S^1 when $f_0(I_\alpha) = f_1(I_\alpha)$.

Notes.

1. An arbitrary subset S of Γ is open (closed) in $\Gamma \iff S \cap \overline{e_\alpha}$ open (closed) in $\overline{e_\alpha}$ for all α.

2. Γ is locally path-connected and locally contractible. The definition does not insist that Γ is connected (although here that will be our case of interest) but we do have connected $\iff$ path-connected.

3. Given any vertex v, a basic open neighbourhood of v in Γ can be taken as the union of v with various half open intervals, each lying in the image of I_α for all edges I_α joined to v and each containing the endpoint of this edge which is identified with v.

   Taking these and all open intervals in e_α for each $\alpha \in A$ gives us a collection of basic open sets for the topology on Γ.

A **simple** graph is one where there are no self loops (an edge with its endpoints identified) or multiple edges (distinct edges between the same endpoints). As we are just concerned with the topology rather than the combinatorial structure of our graphs, we note that any graph Γ is homeomorphic to a simple graph (just add various vertices).

A **subgraph** of Γ is a union Δ of edges and vertices such that $e_\alpha \subseteq \Delta \implies \overline{e_\alpha} \subseteq \Delta$.

Proposition 3.2. If $\widetilde{X}$ is a covering space of the connected graph Γ then $\widetilde{X}$ is a graph with vertices and edges the lifts of those in Γ.

Proof. Setting $V(\widetilde{X}) = p^{-1}(V)$ for $p : \widetilde{X} \to \Gamma$ gives us the vertices of $\widetilde{X}$. To obtain the edges, for each $\alpha \in A$ take the natural map from $I_\alpha \cong [0,1]$ to Γ and select any vertex $\widetilde{v}$ that lies above $f_0(I_\alpha)$, whereupon the unique lift of this map gives us an edge in $\widetilde{X}$.

Thus $\widetilde{X}$ is now given the structure of a graph, and the covering space topology on $\widetilde{X}$ agrees with this graph topology on basic open sets, so they are the same. $\square$

If S is a simple graph, an **edge path** $v_0 \dots v_n$ in S is a finite sequence of vertices such that any two consecutive vertices span an edge in S. It is **reduced** if any three consecutive vertices are distinct ("no backtracks"). A **cycle** is a closed ($v_0 = v_n$) edge path.

A **tree** is a connected simple graph with no reduced cycles. Note that a connected subgraph of a tree is also simple and contains no reduced cycles, hence itself is a tree.

Proposition 3.3. Given any connected graph Γ and any vertex v_0 in Γ, there exists a subgraph $\Delta \simeq \{v_0\}$ such that Δ contains all of $V(\Gamma)$.

Proof. Let $\Gamma_0 = \{v_0\} \subseteq \Gamma_1 \subseteq \Gamma_2 \subseteq \dots$ be a sequence of connected subgraphs where Γ_{i+1} is Γ_i together with the edges $\overline{e_\alpha}$ for all $e_\alpha \in \Gamma \setminus \Gamma_i$ having an endpoint in Γ_i.

Then $\bigcup \Gamma_i$ is open (a small neighbourhood of a point in Γ_i lies in Γ_{i+1}) and closed in Γ (by definition of the topology because it is a union of closed edges), so it is all of Γ.

Next define inductively connected subgraphs $\Delta_0 = \Gamma_0 \subseteq \Delta_1 \subseteq \Delta_2 \subseteq \dots$, where Δ_i has the same vertices as Γ_i but where Δ_{i+1} is Δ_i union *one* edge from each $v \in \Gamma_{i+1} \setminus \Gamma_i$ to Δ_i.

Now Δ_{i+1} deformation retracts onto Δ_i in the standard way, so set $\Delta = \bigcup \Delta_i$ which contains $V(\Gamma)$. We now define a deformation retraction from Δ to $\Delta_0 = \{v_0\}$ by performing each homotopy from Δ_{i+1} to Δ_i during the interval $[1/2^{i+1}, 1/2^i]$. We then put all these together to create a map H from $\Delta \times [0,1]$ to Δ which is continuous and such that $H(\cdot, 0)$ is the identity, $H(\cdot, 1)$ is the constant map to v_0 and $H(v_0, t) = v_0$, hence this is our deformation retraction. □

Note. This contractible subgraph Δ cannot contain a reduced cycle, as seen by by taking a vertex $v \in \Delta_n \setminus \Delta_{n-1}$ where n is maximal over all vertices v in this cycle. Thus if Γ is also a simple graph then Δ is a tree.

Corollary 3.4. A tree T is contractible.

Proof. Apply (3.3) to T, thus we obtain a subgraph Δ of T which is contractible. If an edge e in T is spanned by vertices $v \neq w$ but e is not in Δ then take reduced edge paths $v_0 \dots uv$ and $wx \dots v_0$ in Δ. These can be found by taking the minimum integer i with $v \in \Delta_i$ and using the description of Δ in (3.3), then doing the same for w. Note $u \neq w$ and $v \neq x$ as $e \notin \Delta$. So $v_0 \dots vw \dots v_0$ is a reduced cycle in T. ⨳ □

Theorem 3.5. The fundamental group of a connected graph Γ is free.

Proof. Wlog we can assume that Γ is simple and set $T \subseteq \Gamma$ to be Δ in (3.3), so T is a tree, with $\{e_\alpha : \alpha \in A\}$ the edges in $\Gamma \setminus T$, and w_α, w'_α the endpoints of e_α (which are distinct as Γ is simple). For each α we replace Γ by $T \cup e_\alpha$, so that we are assuming for now there is only one edge outside T.

We next take a reduced cycle c_α in Γ from v_0 including $w_\alpha w'_\alpha$ as in (3.4) so that

$$c_\alpha = v_0 u_1 u_2 \dots u_m \; v_n \dots v_1 v_0, \text{ where } u_m = w_\alpha, v_n = w'_\alpha$$

and $v_0 u_1 u_2 \dots w_\alpha$, $w'_\alpha \dots v_1 v_0$ both lie in T with no vertex repeated in either edge path. Thus if there is a repeated vertex in c_α (other than v_0 at the start and finish) then we must have some u_i equal to some v_j, in which case we replace c_α with the reduced cycle $u_i \dots v_j$ and this still contains $w_\alpha w'_\alpha$.

After finitely many of these replacements we end up with the (relabelled) cycle $c_\alpha = v_0 \dots v_0$ which has no repeated vertices, thus topologically this is a loop L, containing

the edge e_α, with $L \cong S^1$ so $\pi_1(L) \cong \mathbb{Z}$.

See Picture (3) - "Seeing the wood for the trees"

But $\Gamma \simeq L$ as the components K of $\overline{\Gamma \setminus L}$ lie in the tree T and so themselves are trees. If $a \neq b \in K \cap L$, then go from a to b in K and then back via L (not via $w_\alpha w'_\alpha$) to get a reduced cycle in T.

Thus by (3.4) we can deformation retract each K (simultaneously) onto the single point in $K \cap L$, giving $\pi_1(\Gamma) \cong \mathbb{Z}$.

Now we return to the original graph Γ with various edges outside T. For each $\alpha \in A$, let m_α be the 'midpoint' of e_α and set $U_\alpha = \left(\Gamma \setminus \bigcup_{a \in A} m_a\right) \cup m_\alpha$.

Then U_α is open in Γ and path-connected (as if we are at a point in e_a then we can move to a vertex) with $U_\alpha \cap U_\beta (\cap U_\gamma) = \Gamma \setminus \bigcup_{a \in A} m_a$, which deformation retracts onto T, so (3.4) implies that all of these sets are simply connected. Also U_α deformation retracts onto $T \cup e_\alpha$ so $\pi_1(U_\alpha) \cong \mathbb{Z}$ by the above, which tells us (by applying Seifert-Van Kampen at any vertex v) that $\pi_1(\Gamma, v) \cong *_{\alpha \in A} \mathbb{Z}$. □

Corollary 3.6. For every free group $F(X)$, there exists a simple graph Γ with $\pi_1(\Gamma) = F(X)$.

Proof. Take a loop ℓ_α for each $x_\alpha \in X$, all joined at a single vertex v, turn it into a triangle and then use (3.5). □

Theorem 3.7 (Nielsen-Schreier Theorem, 1927). A subgroup of a free group is free.

Proof. On being given $H \leqslant F(X)$ we first use (3.6) to get a graph Γ with $\pi_1(\Gamma) \cong F(X)$, then apply the subgroups $\leftrightarrow$ coverings correspondence to obtain a cover $\widetilde{\Gamma} \twoheadrightarrow \Gamma$ with $\pi_1(\widetilde{\Gamma}) = H$, but $\widetilde{\Gamma}$ is also a (connected) graph by (3.2), so H is free by (3.5). □

If X is finite and H has finite index in $F(X)$ then we can easily determine the rank of H.

Theorem 3.8 (Nielsen-Schreier index formula). If H has index i in the free group F_n then H is a free group of rank $i(n-1)+1$.

Proof. For a finite graph Γ, define the **Euler characteristic** $\chi(\Gamma)$ (*not* the chromatic number) to be $|V| - |E|$. If $p : \widetilde{\Gamma} \to \Gamma$ has degree i then $\chi(\widetilde{\Gamma}) = i\chi(\Gamma)$ by the proof of (3.2).

So take Γ as in (3.6) with $\pi_1(\Gamma) = F_n$ and $\chi(\Gamma) = 1 - n$. For H of index i in F_n with $\pi_1(\widetilde{\Gamma}) = H$, we have $\chi(\widetilde{\Gamma}) = i(1-n)$.

Now for the subgraph Δ in (3.5), note $\chi(\Delta) = 1$ by the proof of (3.3). Thus the number of edges e_α lying in $\widetilde{\Gamma} \setminus \Delta$ is $-(\chi(\widetilde{\Gamma}) - \chi(\Delta)) = i(n-1)+1$ which is the rank of H by the proof of (3.5). □

[**Aside.** If $N \trianglelefteq F(X)$ then it can also be shown by topological means that if N has infinite index and $N \neq \{e\}$ then N has infinite rank. Indeed this can be extended to free products: if a subgroup N is normal and of infinite index in a non trivial free product then either $N = \{e\}$ or N cannot be finitely generated.]

4. Presentations of groups

On being given any group G and a generating set $S = \{g_i : i \in I\}$ for G, we have by (2.7) that $G \cong F(X)/N$ where X is some set in bijection with S and $N \trianglelefteq F(X)$.

Definition 4.1. A **presentation** $P = \langle X|R\rangle$ for a group G is a set X and a subset R of reduced words on $X \cup X^{-1}$, which are thus elements of $F(X)$, such that $G \cong F(X)/\langle\langle R\rangle\rangle^{F(X)}$.

The elements of X are called **generators**, and the elements of R are **relators**.

Hence every group has a presentation by the comment above. Conversely, on being given some presentation $P = \langle X|R\rangle$ we will write $\overline{P}$ for the group thus defined by this presentation. Furthermore given a word w in $F(X)$, we write $\overline{w} \in \overline{P}$ for the corresponding element of the group $F(X)/\langle\langle R\rangle\rangle^{F(X)}$ under the natural projection map $\pi = \pi_P : F(X) \twoheadrightarrow F(X)/\langle\langle R\rangle\rangle^{F(X)}$. It should be clear that for any group element $g \in \overline{P}$, there will be a reduced word in X^* such that $\overline{w} = g$ in $\overline{P}$. Also we have $\overline{vw} = \overline{v}\,\overline{w}$ for $v, w \in F(X)$.

Moreover if $P = \langle X|R\rangle$ and $w \in F(X)$ then by (1.15) we have that $\overline{w} = e$ in $\overline{P}$ if and only if we have $m \in \mathbb{N}$ such that

$$w = \prod_{i=1}^{m} v_i r_{j_i}^{\pm 1} v_i^{-1} \text{ in } F(X)$$

where $r_{j_i} \in R$ and all of w, v_i, r_{j_i} are reduced words in X^* (even though the expression on the right hand side need not be reduced).

Theorem 4.2 (von Dyck, 1882: "a quotient = more relators"). If $P = \langle X|R\rangle$ is a group presentation with $G = \overline{P}$ then a group H is a quotient of $G \iff$ there is a subset S of $F(X)$ such that $H = \overline{\langle X|R \cup S\rangle}$.

Proof. ($\Leftarrow$) If $\langle X|R\cup S\rangle$ is a presentation for H then we see that the natural projection $\pi : F(X) \twoheadrightarrow F(X)/\langle\langle R\cup S\rangle\rangle^{F(X)} \cong H$ factors through $G = F(X)/\langle\langle R\rangle\rangle^{F(X)}$ as $\langle\langle R\rangle\rangle^{F(X)} \leqslant \langle\langle R\cup S\rangle\rangle^{F(X)}$, so H is a quotient of G.

($\Rightarrow$) Now suppose that $H = G/L$ for $L \trianglelefteq G$ and $G = F(X)/M$, where $M = \langle\langle R\rangle\rangle^{F(X)}$. Take any normal subgroup N of $F(X)$ with $L = NM/M$ (for instance use correspondence) and let S be any subset of N with $\langle\langle S\rangle\rangle^{F(X)} = N$. Note that $\langle\langle R\cup S\rangle\rangle^{F(X)} = \langle\langle N\cup M\rangle\rangle^{F(X)} = NM \trianglelefteq F(X)$. Now by (1.24) we have $G/L = (F(X)/M)/(NM/M) \cong F(X)/NM$ so $H = \overline{\langle X|R\cup S\rangle}$. □

Definition 4.3. A **finite presentation** is a presentation $P = \langle X|R\rangle$ where both $X = \{x_1, \ldots, x_n\}$ and $R = \{r_1, \ldots, r_m\}$ are finite sets.

We write $\langle x_1, \ldots, x_n|r_1, \ldots, r_m\rangle$ for P.

A group G is **finitely presented** (f.p.) if there exists *some* finite presentation $\langle x_1, \ldots, x_n|r_1, \ldots, r_m\rangle$ defining G, i.e. $G = \overline{\langle x_1, \ldots, x_n|r_1, \ldots, r_m\rangle}$ ("finitely many generators and finitely many relators"). If so then G will be generated by $\{\overline{x_1}, \ldots, \overline{x_n}\}$ thus "**f.p. implies f.g.**".

Note. A group G will be defined by many different presentations. Even if G is f.p., it does not mean that all of these will themselves be finite presentations.

[**Aside.** If the finite presentation $\langle X|R\rangle$ defines a group G which is infinite then the kernel $\langle\langle R\rangle\rangle^{F(X)}$ of the natural projection cannot be finitely generated (unless it is trivial) by the previous aside, even though it is "finitely normally generated" in $F(X)$.]

Example 4.4.

0. Finite groups are f.p. (take all elements as generators and the multiplication table for the relators).

1. The free group F_n has a presentation $\langle x_1, \ldots, x_n | \emptyset \rangle$ (we may even write $\langle x_1, \ldots, x_n | \, \rangle$), although it is certainly defined by other more awkward presentations too.

2. Similarly the finite cyclic group C_m has a presentation $\langle x | x^m \rangle$, but for instance $\langle x, y | x^{14}, y^{21}, x^3 = y^4 \rangle$ is a "not immediately obvious" presentation defining C_7. Here $x^3 = y^4$ is a **relation** which stands for the relator $x^3 y^{-4}$ but allows us a bit more variation in writing out a presentation.

How do we know that there exist f.g. groups which are not f.p.? By a famous paper of Bernhard Neumann called "Some remarks on infinite groups". He first establishes the following proposition to deal with different presentations defining the same group.

Proposition 4.5 (B. Neumann, 1937). If a group G is defined by the finite presentation $P_1 = \langle X \,|\, r_1, \ldots, r_m \rangle$ for $X = \{x_1, \ldots, x_n\}$ and also by the presentation $P_2 = \langle Y \,|\, s_1, s_2, \ldots \rangle$ for $Y = \{y_1, \ldots, y_k\}$ with finitely many generators but infinitely many relators then there is l such that $\langle\langle s_1, s_2, \ldots \rangle\rangle^{F(Y)}$ is equal to $\langle\langle s_1, \ldots, s_l \rangle\rangle^{F(Y)}$, so that all but finitely many relators in the second presentation are redundant.

Proof. The elements $\overline{x_1}, \ldots, \overline{x_n}$ generate G, so for each of the elements $\overline{y_i}$ in G we have a (reduced) word $v_i \in F(X)$ such that $\overline{y_i} = \overline{v_i}$ in G. This works vice versa so that we also have $\overline{x_i} = \overline{w_i}$ for $w_i \in F(Y)$. Note that $\overline{r_j} = e$ in G for all $1 \leqslant j \leqslant m$. We now substitute w_j for each appearance of x_j in the word v_i, and in the word r_d, to obtain

$$\overline{y_i} = \overline{v_i \left(w_1(y_1, \ldots, y_k), \ldots, w_n(y_1, \ldots, y_k)\right)} \text{ in } G \text{ for } 1 \leqslant i \leqslant k$$

and

$$e = \overline{r_d \left(w_1(y_1, \ldots, y_k), \ldots, w_n(y_1, \ldots, y_k)\right)} \text{ in } G \text{ for } 1 \leqslant d \leqslant m$$

and we set v_i', $1 \leqslant i \leqslant k$ and r_d', $1 \leqslant d \leqslant m$ to be the words on $Y \cup Y^{-1}$ obtained on the right hand side above by this substitution process.

At this point we set $R = \{r_1, \ldots, r_m\}$, and $N = \langle\langle s_1, s_2, \ldots \rangle\rangle^{F(Y)}$ so that $G = F(X)/\langle\langle R \rangle\rangle^{F(X)} \cong F(Y)/N$. We now will consider the group H defined by the finite presentation

$$\langle y_1, \ldots, y_k \,|\, y_i = v_i' \ (1 \leqslant i \leqslant k),\ r_d' = e \ (1 \leqslant d \leqslant m) \rangle. \tag{1}$$

Then the projection $\pi : F(Y) \twoheadrightarrow F(Y)/N$ factors through H. This is because each relation in the presentation (1) also holds in G from above, thus the normal closure M in $F(Y)$ of the (implied) relators in (1) must lie in N. Hence we obtain a homomorphism $\theta : H \twoheadrightarrow F(Y)/N \cong G$.

Now let us return to the presentation P_1. We also have a homomorphism from $F(X)$ onto H formed by sending each x_i to $\overline{w_i(y_1, \ldots, y_k)}$ and extending via (2.5). This factors through $G = F(X)/\langle\langle R \rangle\rangle^{F(X)}$ because $\overline{r_d(x_1, \ldots, x_n)} = e$ in G, so we obtain another homomorphism $\varphi : G \twoheadrightarrow H$. Then $\theta\varphi(\overline{x_i}) = \overline{x_i}$ with φ surjective, thus the projection $\theta : F(Y)/M \twoheadrightarrow F(Y)/N$ with $M \leqslant N$ is an isomorphism so that $N = M = \langle\langle \text{finite set} \rangle\rangle^{F(Y)}$.

But $N = \bigcup_{i=1}^{\infty} \langle\langle s_1, \ldots, s_i \rangle\rangle^{F(Y)}$ so this sequence of ascending normal subgroups terminates at l say. □

Corollary 4.6. The f.g. group G defined by the presentation

$$\langle a, b \,|\, [a^{2n+1}ba^{-(2n+1)}, b] \text{ for } n = 1, 2, 3, \ldots \rangle$$

is *not* finitely presented.

Proof. Notation: we set $c_n = [a^{2n+1}ba^{-(2n+1)}, b]$.

In the alternating group A_j for odd $j \geqslant 7$, let $\alpha = (12\ldots j)$ and $\beta = (123)$. Then $\alpha^k\beta\alpha^{-k}$ commutes with β if $3 \leqslant k \leqslant j-3$, but not if $k = j-2$.

Now for any integer $l \geqslant 2$, consider the group A_{2l+3}. We have that the group elements $c_1(\alpha,\beta), \ldots, c_{l-1}(\alpha,\beta)$, given by multiplying out $\alpha^{\pm 1}, \beta^{\pm 1}$ in each word c_i, are all equal to e but $c_l(\alpha,\beta) \neq e$. So $c_l \notin \langle\langle c_1, \ldots, c_{l-1}\rangle\rangle^{F_2} \leqslant F_2$ because we have homomorphisms $\theta_l : F_2 \to A_{2l+3}$ given by $a \mapsto \alpha$, $b \mapsto \beta$ where $\theta_l(c_l) \neq e$ but $\theta_l(c_i) = e$ for $1 \leqslant i < l$. Now apply (4.5). □

[**Aside.** This is a simplified version of the example in the B. Neumann paper which is then adapted to show that there are uncountably many finitely generated groups up to isomorphism. Of course there are only countably many finitely presented groups up to isomorphism (as there are only countably many finite presentations) so we could say that "most f.g. groups are not f.p."!]

If we have two groups, each already defined by a presentation, then it would be good if we could use these to obtain a presentation that defines their direct or free product. In fact the free product case is more succinct (sometimes the result below is used to *define* free products but then one would need to show the resulting group is independent of the presentations chosen).

Proposition 4.7. If $G = \overline{\langle X \,|\, R\rangle}$ and $H = \overline{\langle Y \,|\, S\rangle}$ for disjoint sets X and Y then $G * H = \overline{\langle X \cup Y \,|\, R \cup S\rangle}$.

Proof. Let A be the group defined by this presentation $\overline{\langle X \cup Y \,|\, R \cup S\rangle}$. First note that the natural projection from $F(X \cup Y)$ to $G * H$ given by sending each x_i to $\overline{x_i} \in G \leqslant G * H$ and similarly y_j to $\overline{y_j}$ factors through A because all the relators in R and S are the identity in G and in H, hence in $G * H$. This gives us a homomorphism $A \twoheadrightarrow G * H$.

But there is also a homomorphism from G to A (and similarly from H to A) sending $\overline{x_i}$ (or $\overline{y_i}$) back whence it came. So extend these by (2.14) to get the homomorphism $\varphi : G * H \to A$. Now compose with θ to get that $\theta\varphi = \mathrm{id}$ (because it fixes each $\overline{x_i}$ and $\overline{y_j}$ which together generate $G * H$), but φ is surjective so it is an isomorphism. □

Proposition 4.8. Suppose that

$$\begin{aligned} N &= \overline{\langle n_i : i \in I \,|\, r_j : j \in J\rangle}, \\ H &= \overline{\langle h_k : k \in K \,|\, s_l : l \in L\rangle} \end{aligned}$$

and φ is a homomorphism from H to $\mathrm{Aut}(N)$. For each $k \in K$ and $i \in I$, let $w_{k,i} \in F(\{n_i : i \in I\})$ be a (reduced) word such that $\varphi(\overline{h_k})(\overline{n_i}) = \overline{w_{k,i}}$ in N. Then the semidirect product $G = N \rtimes_\varphi H$ is defined by the presentation

$$P = \langle \{n_i : i \in I\} \cup \{h_k : k \in K\} \,|\, r_j\,(j \in J),\, s_l\,(l \in L),\, h_k n_i h_k^{-1} = w_{k,i}\,(i \in I, k \in K)\rangle.$$

Proof. Regard $H = \langle\overline{h_k}\rangle$ and $N = \langle\overline{n_i}\rangle$ as subgroups of G. We define $\theta : \overline{P} \twoheadrightarrow G$ in the usual way: send each n_i to $\overline{n_i} \in G$, each h_k to $\overline{h_k} \in G$, extend to $F(\{n_i\} \cup \{h_k\})$ and then factor through $\overline{P}$ as the given relations in P hold in G too.

Now any element x of $\overline{P}$ is equal to some $\overline{y_x}$ in $\overline{P}$ where y_x is a word on $\{n_i^{\pm 1} : i \in I\} \cup \{h_k^{\pm 1} : k \in K\}$. But for each k and i we have $\overline{h_k}\,\overline{n_i} = \overline{w_{k,i}}\,\overline{h_k}$ in $\overline{P}$, where we know that $w_{k,i}$ is a word in the n_is only. So we can move all occurrences of the last h_k in the word y_x to the end of this word, changing it as we go but so that it is still a word on $\{n_i^{\pm 1} : i \in I\} \cup \{h_k^{\pm 1} : k \in K\}$. We can do this with successive appearances of these h_k, in order from right to left, and eventually we end up with a (reduced) word $u_x v_x$ on $\{n_i^{\pm 1} : i \in I\} \cup \{h_k^{\pm 1} : k \in K\}$ but such that u_x is a word on $\{n_i^{\pm 1} : i \in I\}$ only, v_x on $\{h_k^{\pm 1} : k \in K\}$ only and such that $\overline{u_x}\,\overline{v_x} = \overline{y_x}$ in $\overline{P}$.

Suppose that $\theta(x) = e$ in G, so that in the notation above $\theta(\overline{u_x}\,\overline{v_x}) = \theta(\overline{u_x})\theta(\overline{v_x}) = e$. Then $\theta(\overline{u_x})$ lands in N and $\theta(\overline{v_x})$ lands in H with $N \cap H = \{e\}$ so $\theta(\overline{u_x}) = \theta(\overline{v_x}) = e$. But $\theta(\overline{u_x})$ is $\overline{u_x} \in N$, so u_x must be in the normal closure of the relators r_j and thus $\overline{u_x} = e$ in $\overline{P}$. The same holds for $\overline{v_x}$ so $x = \overline{y_x}$ is the identity and θ is an isomorphism. □

We now have as an immediate corollary from (4.7) and (4.8):

Corollary 4.9. If G and H are both finitely presented then the free product $G * H$ and any semidirect product $G \rtimes H$ or $H \rtimes G$ is also finitely presented.

Example 4.4 continued.

3. The groups $\mathbb{Z}^2$, $\mathbb{Z}^3$ and $\mathbb{Z}^4$ are defined respectively by the presentations $\langle a, b \,|\, [a, b]\rangle$, $\langle a, b, c \,|\, [a, b], [a, c], [b, c]\rangle$ and $\langle a, b, c, d \,|\, [a, b], [a, c], [a, d], [b, c], [b, d], [c, d]\rangle$.

4. The dihedral group D_{2m} of order $2m$ has a presentation $\langle x, y \,|\, x^n, y^2, yxy^{-1} = x^{-1}\rangle$. The infinite dihedral group has a presentation $\langle x, y \,|\, y^2, yxy^{-1} = x^{-1}\rangle$.

A very brief introduction to semigroups

Definition 4.10. A **semigroup** S is a set with an associative binary operation and is **commutative** if $ab = ba$ always. A **subsemigroup** is a subset T of S that is closed under multiplication ($a, b \in T$ implies $ab \in T$).

A **semigroup homomorphism** f is a function from one semigroup to another such that $f(ab) = f(a)f(b)$ always. A **semigroup isomorphism** is a bijective semigroup homomorphism, whereupon the inverse map f^{-1} is also a semigroup homomorphism.

Example 4.11.

1. Any set S with the operation $ab = a$ (not commutative).

2. The natural numbers $\mathbb{N}$ under addition (either with or without 0).

3. All integers at least 26 under addition, but all integers at least -2 under addition is *not* an example.

4. Any subset of integers with the operation $ab = \max(a, b)$.

5. The main example: let X be a set of symbols. Whereas X^+ was the set of all finite words on X including $\emptyset$, let $X^\circ = X^+ \setminus \{\emptyset\}$ be the set of **non empty finite words** on X. Both of these are semigroups under the operation of concatenation and are non commutative if $|X| > 1$.

As there is no cancellation, free semigroups can be defined directly:

Definition 4.12. The **free semigroup** on a set X *is* the semigroup X° (whereas X^+ is the free monoid) and clearly satisfies the universal property: if T is a semigroup and $f : X \to T$ any function then f can be extended uniquely to a semigroup homomorphism from X° to T. Moreover if a semigroup S has a subset Y with this universal property then S is semigroup isomorphic to Y°.

In analogy with free groups, we see that every semigroup is the image under some semigroup homomorphism of a free semigroup. Also if $|X| = |Y|$ then the free semigroups X° and Y° are semigroup isomorphic. Moreover if $|X| < |Y|$ then there cannot be a surjective semigroup homomorphism from X to Y (one could use word length and note that words of length more than 1 cannot map to words of length 1).

However subsemigroups of free semigroups need *not* be free: if $|X| > 1$ then X° is the natural numbers with 0 removed under addition and $S = \{2, 3, 4, \ldots\}$ is a subsemigroup of X° and so is also commutative. Thus S would need to be free on a single element n, but on sending $n \in S$ to $1 \in X^\circ$ how would we extend this to a semigroup homomorphism from S to X°?

Also we do not have a direct generalisation of normal subgroups, but we say a **congruence** on the semigroup S is an equivalence relation $\sim$ (or indeed a partition) which is compatible with the multiplication (if $a \sim c$ and $b \sim d$ then $ab \sim cd$), so that the induced multiplication on the set E of equivalence classes is well defined and turns E into a semigroup, the **quotient** semigroup written $S/\sim$, or here even $S/\equiv$ to emphasise that $\equiv$ is a special type of equivalence relation.

The following is the only example of a congruence we will need and can be thought of as similar to PreDefinition 2 of a free group.

Definition 4.13. As a congruence on a semigroup S is a special type of relation (a relation being just a subset of $S \times S$) and the intersection of congruences on S is also a congruence, we can talk about the congruence **generated by** any $R \subseteq S \times S$ as the intersection of all the congruences containing R.

In particular let X° be the free semigroup on the set X and suppose that R is a relation on X°. If $\equiv$ is the congruence on X° generated by R then we say that $P = \langle X \mid R \rangle$ is a **semigroup presentation** for the quotient semigroup $X^\circ/\equiv$.

We say that P is a **finite semigroup presentation** if both X and R are finite.

Lemma 4.14. Let $P = \langle X \mid R \rangle$ be a semigroup presentation where $R = \{(x_i, y_i) : i \in I\} \subseteq X^\circ \times X^\circ$ and let $\overline{P}$ denote the quotient semigroup $X^\circ/\equiv$ given by the congruence $\equiv$ generated by R.

Then two elements w, w' of X° are equal in $\overline{P}$ if and only there is a finite sequence $w = w_0, w_1, \ldots, w_m = w'$ of words $w_j \in X^\circ$ such that w_{j+1} differs from w_j by the replacement of some subword x of w_j with y of w_{j+1}, where either (x, y) or (y, x) is in R.

Proof. ($\Leftarrow$) If (x, y) or (y, x) is in R, so that $x \equiv y$, then for any $u, v \in X^\circ$ we will have $ux \equiv uy$ and $uxv \equiv uyv$, thus we do have $w_j \equiv w_{j+1}$ and $\equiv$ is an equivalence relation.

($\Rightarrow$) If we define the relation $\equiv_R$ on X° by $w \equiv_R w'$ if there is some sequence from w to w' as above then we see that $\equiv_R$ is an equivalence relation on X°, and even a congruence, containing R. As $\equiv$ is the smallest congruence containing R, we will have $w \equiv w' \Rightarrow w \equiv_R w'$. □

5. HNN extensions and free products with amalgamation

These two constructions, which generalise the idea of a free product, are amongst the most versatile in group theory. They are best thought of as different sides of the same coin, with the amalgamated free product introduced by Schreier in 1926 and the HNN extension in 1949 (named after Graham Higman, Bernhard Neumann and Hanna Neumann).

Definition 5.1. Let G, H be groups containing isomorphic subgroups $A \leqslant G$, $B \leqslant H$ and let $\varphi : A \to B$ be an isomorphism.

The **free product with amalgamation** is the group

$$(G * H)/\langle\langle a = \varphi(a) \text{ for all } a \in A\rangle\rangle^{G*H},$$

written $G *_\varphi H$. Thus we form the free product $G * H$ and identify A in $G * H$ with its image $\varphi(A) = B$ under φ.

Proposition 5.2. If G and H are both f.g. then (regardless of whether A, or equally B, is f.g.) so is $G *_\varphi H$.

If G and H are both f.p. and A (thus B) is f.g. then $G *_\varphi H$ is also f.p. (regardless of whether A is f.p.).

Proof. If G and H are f.g. then so is $G * H$ (by Example Sheet 1?!) and hence any quotient of $G * H$ is too.

Now suppose that G and H are defined by the finite presentations $\langle X \mid R\rangle$ and $\langle Y \mid S\rangle$ so that $\langle X \cup Y \mid R \cup S\rangle$ is a finite presentation defining $G * H$ by (4.7). If $\{a_1, \dots, a_k\}$ generates A then the normal closures in $G*H$ of $\{a(\varphi(a))^{-1} : a \in A\}$ and $\{a_i(\varphi(a_i))^{-1} : 1 \leqslant i \leqslant k\}$ are equal. It is then the case that there are k elements $w_1, \dots, w_k \in F(X \cup Y)$ which can be added to $R \cup S$ so that the quotient of $F(X \cup Y)$ by the normal closure in $F(X \cup Y)$ of $R \cup S$ union these k elements is isomorphic to G_φ.

This is essentially the proof of (4.2 $\Rightarrow$) so we return to the notation there: suppose that $L = \langle\langle l_1, \dots, l_k\rangle\rangle^G$ and $w_1, \dots, w_k$ are elements of $F(X)$ with $\pi(w_i) = l_i$ under the projection π from $F(X)$ to $F(X)/M = G$. Then on setting $S = \{w_1, \dots, w_k\}$ and $\langle\langle S\rangle\rangle^{F(X)} = N$, we have $MN/N = \pi(N) = \langle\langle \pi(S)\rangle\rangle^{\pi(F(X))} = L$ so that $H = \overline{\langle X \mid R \cup S\rangle}$ as before. □

Definition 5.3, cf. (5.1). If G is a group and $A, B \leqslant G$ are isomorphic subgroups of G with $\varphi : A \to B$ some isomorphism between them then the **HNN extension** $G*_\varphi$ is defined by taking some other symbol t (the **stable letter** of the HNN extension), with $\langle t\rangle$ regarded as a copy of of $\mathbb{Z}$, and forming the group

$$(G * \langle t\rangle)/\langle\langle tat^{-1} = \varphi(a)\rangle\rangle^{G*\langle t\rangle}.$$

Thus we take the free product $G * \mathbb{Z}$ and then identify A in $G * \mathbb{Z}$ with its image $\varphi(A) = B$ via conjugation by t.

We call G the **base group** of the HNN extension.

The following has a very similar proof to (5.2) and is omitted.

Proposition 5.4, cf. (5.2). If G is f.g. then so is G_φ (regardless of A).

If G is f.p. and A is f.g. then $G*_\varphi$ is also f.p. (regardless of whether A is f.p.).

This is all very well but how do we know that these constructions define interesting new groups? Any HNN extension has a surjective homomorphism to $\mathbb{Z}$ by sending t to 1 and all of G to 0 (the associated homomorphism of the HNN extension, so that t is an element of infinite order in $G*_\varphi$) but could it be that there is "collapse" within the presentation so that our HNN extension is merely $\mathbb{Z}$ or our amalgamated free product is even trivial?

Actually this does not occur because the the factors G, H in an amalgamated free product, and the base G in an HNN extension, embed naturally as subgroups. This fact is the key to the power of these constructions and is not at all obvious. Here we will prove this first for HNN extensions and then deduce it for amalgamated free products, though one might equally want to do it the other way round. It will come as no surprise that we will utilise group actions on words.

HNN extensions

Definition 5.5. Given the HNN extension $G*_\varphi$ with $\varphi : A \to B$ an isomorphism and stable letter t, an **H-sequence** (from *H*NN) is a finite sequence of the form $g_0 t^{\varepsilon_1} g_1 \dots t^{\varepsilon_n} g_n$ where $n \geqslant 0$, each $g_i \in G$ and each $\varepsilon_i \in \{\pm 1\}$. Of course any H-sequence can also be regarded as an element of $G*_\varphi$ by multiplying out the elements.

Choose right transversals T_A and T_B for A and B in G such that both include e_G. A **normal form** is an H-sequence such that if $\varepsilon_i = +1$ then $g_i \in T_A$, if $\varepsilon_i = -1$ then $g_i \in T_B$, and there is no subsequence te_Gt^{-1} or $t^{-1}e_Gt$. Let $\mathcal{N}$ be the set of all normal forms.

Note first the slightly different treatment of the identity here from the free product case: any g_i is allowed to be the identity e_G in G, but $n \geqslant 0$ means the empty word does not occur. Instead $n = 0$ gives us an element $g_0 \in G$, so taking $n = 0$ and $g_0 = e_G$ gives us the identity.

Note also that any element of $G*_\varphi$ can be expressed as an H-sequence (think of the normal form for the free product $G * \langle t \rangle$ and pad it out with various e_Gs between the relevant powers of t). But as $ta = \varphi(a)t$ and $t^{-1}b = \varphi^{-1}(b)t^{-1}$, we can move a's (resp. b's) to the left of t (resp. t^{-1}) in $G*_\varphi$. This means that *every* H-sequence can be put into normal form without changing the element in $G*_\varphi$ which it represents by working from right to left: for instance if we see a subsequence tg_i where $g_i \notin T_A$ then set $g_i = a_i\gamma_i$ for some $a_i \in A$ and $\gamma_i \in T_A$, so that we can replace $\dots g_{i-1}tg_i \dots = \dots g_{i-1}ta_i\gamma_i \dots$ with $\dots (g_{i-1}\varphi(a))t\gamma_i \dots$ in our sequence.

Theorem 5.6 (Normal form for HNN extensions). Every element in $G*_\varphi$ has a *unique* normal form.

Proof. Define $\rho : G*_\varphi \to \Sigma(\mathcal{N})$ as follows:

1. $\rho(g)(g_0 t^{\varepsilon_1} \dots g_n) = (gg_0)t^{\varepsilon_1} \dots g_n$ for $g \in G$.
2. If $\varepsilon_1 = -1$ and $g_0 \in A$ then $\rho(t)(g_0 t^{-1} \dots g_n) = (\varphi(g_0)g_1)t^{\varepsilon_2} \dots g_n$.

   Otherwise, $\rho(t)(g_0 t^{\varepsilon_1} \dots g_n) = \varphi(a)t\gamma_0 t^{\varepsilon_1} \dots g_n$, where $g_0 = a\gamma_0$ for $\gamma_0 \in T_A$ and $a \in A$.

This has inverse $\rho(t^{-1})(g_0 t g_1 \dots g_n) = (\varphi^{-1}(g_0)g_1)t^{\varepsilon_2} \dots g_n$ if $g_0 \in B$, and otherwise $g_0 t^{\varepsilon_1} \dots g_n \mapsto \varphi^{-1}(b)t^{-1}\eta_0 t^{\varepsilon_1} \dots g_n$ (where $g_0 = b\eta_0$ for $\eta_0 \in T_B$).

Also we can check that $\rho(a) = \rho(t^{-1})\rho(\varphi(a))\rho(t)$ for any $a \in A$, so ρ is a well-defined homomorphism from $G*_\varphi$ to $\Sigma(\mathcal{N})$. Moreover $\rho(g_0 t^{\varepsilon_1} \dots g_n)(e_G) = g_0 t^{\varepsilon_1} \dots g_n$ when $g_0 t^{\varepsilon_1} \dots g_n$ is any normal form and this gives us uniqueness. □

Corollary 5.7 ("Base group embeds", HNN 1949). In any HNN extension $G*_\varphi$ the base group G embeds homomorphically in $G*_\varphi$ by sending $g \in G$ to the H-sequence g.

Proof. If $g \neq e_G$ then the H-sequence g is in normal form, and is not the normal form e_G which is the identity in $G*_\varphi$. □

In fact for many uses we do not need unique representative forms for each element in an HNN extension. We just need an easily verified criterion telling us that our form is *not* the identity.

Definition 5.8. If $G*_\varphi$ is an HNN extension then a **pinch** of an H-sequence $g_0 t^{\varepsilon_1} \dots t^{\varepsilon_n} g_n$ is a subsequence of the form $t g_i t^{-1}$ for $g_i \in A$, or $t^{-1} g_i t$ for $g_i \in B$.

An H-sequence is **reduced** (or H-reduced if we need to distinguish) if it contains no pinch, thus in particular all normal sequences are reduced.

Corollary 5.9 ("No pinch", Britton's Lemma 1963). If the H-sequence $g_0 t^{\varepsilon_1} \dots t^{\varepsilon_n} g_n$ is reduced and $n \geqslant 1$ then it is *not* the identity in $G*_\varphi$; indeed it is not even in G.

Proof. As mentioned above, we can change this reduced sequence into a normal form without altering the element in $G*_\varphi$ which it represents. The key point is that when doing this for a reduced form, no $t^{\pm\varepsilon}$ cancel so we end up with the normal form $g'_0 t^{\varepsilon_1} \dots t^{\varepsilon_n} g'_n$ for $n \geqslant 1$. Moreover if $g_0 t^{\varepsilon_1} \dots t^{\varepsilon_n} g_n = g \in G$ then $g_0 t^{\varepsilon_1} \dots t^{\varepsilon_n} g_n g^{-1}$ is also reduced. □

We present a few examples, which could have been given earlier but we now know that we obtain a range of groups in this way.

Example 5.10.

1. If A (and thus B) is the trivial subgroup then $G*_\varphi$ is the free product $G * \mathbb{Z}$.

2. To go to the other extreme, if $A = G = B$ then the isomorphism $\varphi : A \to B$ becomes the automorphism $\alpha : G \to G$, whereupon G is normal in $G*_\varphi$ which is therefore the semidirect product $G \rtimes_\alpha \mathbb{Z}$ (where we interpret $\alpha : \mathbb{Z} \to \mathrm{Aut}(G)$ as $\alpha(t^n)(g) = \alpha^n(g)$). Thus the identity automorphism gives us here the *direct* product $G \times \mathbb{Z}$. (So an HNN extension is only ever abelian if $G = A = B$ is abelian and φ is the identity automorphism.)

3. But we could even have $A \cong G \cong B$ whilst A and/or B are proper subgroups of G. What's a nice easy group to take which is isomorphic to many of its proper subgroups? Aha! The integers!

   So let $G = \langle x \rangle = \mathbb{Z}$ and take non zero integers m, n to obtain subgroups $A = \langle x^m \rangle$ and $B = \langle x^n \rangle$ both also isomorphic to $\mathbb{Z}$. On taking the obvious isomorphism $\varphi(x^m) = x^n$ between these infinite cyclic subgroups, we have the HNN extension $G*_\varphi$ defined by the presentation $\langle x, t \,|\, t x^m t^{-1} = x^n \rangle$ which gives us the famous **Baumslag-Solitar** groups $BS(m, n)$ from 1962.

4. On taking $G = F_2$ to be free on a, b and $A = \langle a \rangle$, $B = \langle b \rangle$ to be infinite cyclic subgroups formed from each generator, we have the obvious isomorphism $\varphi(a) = b$. Then $G*_\varphi = \langle a, b, t \,|\, t a t^{-1} = b \rangle$ is actually free on t, a. Thus although in an HNN extension G (here $G = \langle a, t a t^{-1} \rangle$) always embeds in $G*_\varphi$ as a *proper* subgroup, because it is in the kernel of the associated homomorphism, the base may nevertheless be *isomorphic* to the extension.

When can a group be expressed as an HNN extension? In fact there is an easy, but not especially instructive, equivalent condition: if and only it has a surjective homomorphism θ

to $\mathbb{Z}$ (if so then the group can be written as a semidirect product $\ker(\theta) \rtimes \mathbb{Z}$ which is an HNN extension with base $\ker(\theta)$ by (5.10) Part 2).

Better would be a criterion for when a subgroup of an HNN extension naturally inherits a description as an HNN extension itself:

Definition 5.11. Let $G*_\varphi$ be an HNN extension with base G, stable letter t and A, B subgroups of G where $\varphi : A \to B$ is an isomorphism. We say that a subgroup H of the base G is a **good** subgroup (with respect to this HNN extension) if $\varphi(H \cap A) = H \cap B$.

Lemma 5.12. If $G*_\varphi$ is as defined in (5.11) and H is a good subgroup of G then the subgroup $\langle H, t\rangle$ of the HNN extension $G*_\varphi$ is itself naturally an HNN extension $H*_\psi$ with base H, stable letter t, and where $H \cap A, H \cap B$ are subgroups of H with ψ, defined as the restriction of φ to $H \cap A$, an isomorphism from $H \cap A$ to $H \cap B$. Moreover $\langle H, t\rangle \cap G = H$.

Proof. The condition on H does mean that ψ is an isomorphism. Form the abstract HNN extension $H*_\psi$ with stable letter s and let $\theta : H*_\psi \to G*_\varphi$ be the homomorphism sending s to t and $h \in H$ to its image h in $G*_\varphi$. This is well defined because if $a \in H \cap A$ then $sas^{-1}(\psi(a))^{-1}$ maps to $tat^{-1}(\varphi(a))^{-1} = e$ as $\psi(a) = \varphi(a) \in B$.

The image of θ is clearly $\langle H, t\rangle$ and if $h_0 s^{\varepsilon_1} \dots s^{\varepsilon_n} h_n$ is an element of $H*_\psi$ represented by a reduced sequence then it maps to $h_0 t^{\varepsilon_1} \dots t^{\varepsilon_n} h_n$ which also contains no pinch. Thus θ is injective and further $h_0 s^{\varepsilon_1} \dots s^{\varepsilon_n} h_n$ does not map into G if $n \geqslant 1$. □

Example 5.13 Obviously the base group G and the trivial subgroup $\{e\}$ of G are always good. But if $A \leqslant H$ then $B \leqslant H$ will be needed too (and vice versa) if H is to be good.

In particular for the Baumslag-Solitar group $BS(2,3)$ with $G = \langle x\rangle$, $A = \langle x^2\rangle$ and $B = \langle x^3\rangle$, we have that A is not good but $H = \langle x^k\rangle$ is good if k is coprime to 2 or 3.

Free products with amalgamation

Here we have similar results in that the factor groups embed and there are equivalent notions of normal and reduced forms. As we will not need unique representatives in this course, we will not here pursue normal forms for amalgamated free products but will prove the other results by reducing to the HNN extension case above.

If $G *_\varphi H$ is a free product with amalgamation where $\varphi : A \to B$ is an isomorphism for $A \leqslant G$ and $B \leqslant H$, we say that a P-reduced (meaning free product reduced) sequence $c_1 \dots c_n \in G * H$ is **A-reduced** (for "amalgamated reduced") if whenever $n > 1$ we have that no c_i is in A or B.

We can turn any P-reduced sequence in $G * H$ into an A-reduced sequence by "absorbing" elements of A or B using φ, without changing the element it represents in $G *_\varphi H$.

For instance if we see the subsequence $\dots g_{i-1} b_i g_{i+1} \dots$ in our P-reduced sequence, where $g_{i-1}, g_{i+1} \in G$ and $b_i \in B$, then we can replace b_i with $\varphi^{-1}(b_i) \in A$ and then group together $(g_{i-1}\varphi^{-1}(b_i)g_{i+1})$ as an element of G, thus reducing the length of the sequence.

Corollary 5.14 ("Factor groups embed", cf. 5.7). If $c_1 \dots c_n$ is A-reduced and $n \geqslant 1$ then the element it represents is not equal to the identity in $G *_\varphi H$. In particular G and H embed in $G *_\varphi H$ by taking $n = 1$.

Proof. We form $F = (G * H * \langle t\rangle)/\langle\langle tat^{-1} = \varphi(a) : a \in A\rangle\rangle^{(G*H)*\langle t\rangle}$, which is an HNN extension with base $G * H$, because A and B are naturally subgroups of $G * H$ and they are isomorphic via φ.

Let $\psi : G *_\varphi H \to F$ be defined by $\psi(g) = tgt^{-1}$ and $\psi(h) = h$. This is a well-defined homomorphism on extension to $G *_\varphi H$, as $a\varphi(a)^{-1}$ maps to $tat^{-1}\varphi(a)^{-1} = e$ in F.

For an element of $G *_\varphi H$ given by a A-reduced sequence, if $n = 1$ and $c_1 \in A$ then $c_1 \overset{\psi}{\mapsto} \varphi(a) \neq e$ (normal form). Otherwise, it maps directly to an H-reduced sequence in the HNN extension F, so we are done by (5.9). □

Multiple HNN extensions

Given any group $G = G_0$, one is at liberty to create a sequence of HNN extensions in turn, by forming $G_1 = G*_\varphi$, $G_2 = G_1*_{\varphi'}$ etc. However an especially useful version of this is when we form multiple HNN extensions of a group G "simultaneously" (though here we will only ever need to do this for finitely many extensions).

Definition 5.15. Suppose that G is a group and we have isomorphisms $\varphi_i : A_i \to B_i$ for $i = 1, \ldots, n$, where all A_i and B_i are subgroups of G. Then the **multiple HNN extension** $G*_{\varphi_1,\ldots,\varphi_n}$ with **base group** G and **stable letters** $t_1, \ldots, t_n$, where all $\langle t_i \rangle$ are regarded as copies of $\mathbb{Z}$, is the group

$$(G * \langle t_1 \rangle * \ldots * \langle t_n \rangle)/\langle\langle t_1 a_1 t_1^{-1} = \varphi_1(a_1) : a_1 \in A_1, \ldots, t_n a_n t_n^{-1} = \varphi_n(a_n) : a_n \in A_n \rangle\rangle.$$

Thus we form the free product G with a copy of the rank n free group F_n and then identify each A_i with its image $\varphi_i(A_i)$ via conjugation by t_i.

Lemma 5.16. The multiple HNN extension $G*_{\varphi_1,\ldots,\varphi_n}$ as defined above can be regarded as the final group $G_n = G_{n-1}*_{\varphi_n}$ obtained by the sequence of (single) HNN extensions $G_i = G_{i-1}*_{\varphi_i}$ where we set the base $G = G_0$ and regard $A_i, B_i \leqslant G_0$ as isomorphic subgroups of G_{i-1}.

Proof. We do this by induction on n. Suppose this is true up to i, so that G_i is equal to

$$(G * F_i)/\langle\langle t_1 a_1 t_1^{-1} = \varphi_1(a_1) : a_1 \in A_1, \ldots, t_i a_i t_i^{-1} = \varphi_i(a_i) : a_i \in A_i \rangle\rangle^{G*F_i}$$

where we write F_i for the free product $\langle t_1 \rangle * \ldots * \langle t_i \rangle$ of infinite cyclic groups generated by the relevant stable letters.

First consider the following setup: suppose G and H are groups and we have a subset C of G and S of $G * H$. Then, as $\langle\langle C, S \rangle\rangle^{G*H} = \langle\langle C \rangle\rangle^{G*H} \langle\langle S \rangle\rangle^{G*H}$, we have that $(G * H)/\langle\langle C, S \rangle\rangle^{G*H}$ is isomorphic to

$$(G * H)/\langle\langle C \rangle\rangle^{G*H} \Big/ \left(\langle\langle C \rangle\rangle^{G*H} \langle\langle S \rangle\rangle^{G*H}\right)/\langle\langle C \rangle\rangle^{G*H}.$$

But the natural homomorphism from $G * H$ to $(G/\langle\langle C \rangle\rangle^G) * H$ factors through $(G * H)/\langle\langle C \rangle\rangle^{G*H}$ to give us an isomorphism from this quotient to $(G/\langle\langle C \rangle\rangle^G)*H$. On further quotienting out by the appropriate normal closure of S on both sides, this isomorphism takes $(\langle\langle S \rangle\rangle^{G*H} \langle\langle C \rangle\rangle^{G*H})/\langle\langle C \rangle\rangle^{G*H}$ to $\langle\langle S \rangle\rangle^{(G/\langle\langle C \rangle\rangle^G)*H}$ so that $(G * H)/\langle\langle C, S \rangle\rangle^{G*H}$ is isomorphic to $((G/\langle\langle C \rangle\rangle^G) * H)/\langle\langle S \rangle\rangle^{(G/\langle\langle C \rangle\rangle^G)*H}$.

Now on replacing G with $G * F_i$, H with $\langle t_{i+1} \rangle$ and letting C be the subset

$$\{t_1 a_1 t_1^{-1} \varphi_1(a_1)^{-1} : a_1 \in A_1, \ldots, t_i a_i t_i^{-1} \varphi_i(a_i)^{-1} : a_i \in A_i\}$$

of $G*F_i$ and S the subset $\{t_{i+1} a_{i+1} t_{i+1}^{-1} \varphi(a_{i+1})^{-1} : a_{i+1} \in A_{i+1}\}$ of $G*F_i$, we conclude that $(G * F_{i+1})/\langle\langle C, S \rangle\rangle^{G*F_{i+1}}$, which is the multiple HNN extension with $i + 1$ stable

letters, is isomorphic to $(((G * F_i)/\langle\langle C\rangle\rangle^{G*F_i}) * \langle t_{i+1}\rangle)/\langle\langle S\rangle\rangle^{((G*F_i)/\langle\langle C\rangle\rangle^{G*F_i})*\langle t_{i+1}\rangle}$. But $(G * F_i)/\langle\langle C\rangle\rangle^{G*F_i}$ is G_i by inductive hypothesis and $(G_i * \langle t_{i+1}\rangle)/\langle\langle S\rangle\rangle^{G_i*\langle t_{i+1}\rangle}$ is the HNN extension G_{i+1} of G_i, so this is isomorphic to the above multiple HNN extension. □

We now have two equivalent results for multiple HNN extensions which have already been established for single HNN extensions.

Definition 5.17. Suppose we have a multiple HNN extension $G*_{\varphi_1,\ldots,\varphi_n}$ with base group G, isomorphisms $\varphi_i : A_i \to B_i$ and stable letters $t_1, \ldots, t_n$. An **MH-sequence** (for multiple HNN extension) is is a finite sequence of the form $g_0 t_{j_1}^{\varepsilon_1} g_1 \ldots t_{j_n}^{\varepsilon_n} g_n$ where $n \geqslant 0$, each $g_i \in G$, each $\varepsilon_i \in \{\pm 1\}$ and each j_i comes from $\{1, \ldots, n\}$ (again we can think of it as an element of $G*_{\varphi_1,\ldots,\varphi_n}$).

A t_i**-pinch** of an MH-sequence is a subsequence of the form $t_i g t_i^{-1}$ for $g \in A_i$, or $t_i^{-1} g t_i$ for $g \in B_i$.

An MH-sequence is **reduced** if it contains no t_i-pinch for any $i \in \{1, \ldots, n\}$. Any element in $G*_{\varphi_1,\ldots,\varphi_n}$ can be represented by a reduced MH-sequence (replace all pinches until there are no more).

Theorem 5.18 (Multiple Britton's Lemma, cf. 5.9). If the MH-sequence $g_0 t_{j_1}^{\varepsilon_1} \ldots t_{j_n}^{\varepsilon_n} g_n$ contains no pinch and $n \geqslant 1$ then it is *not* the identity in $G*_{\varphi_1,\ldots,\varphi_n}$, nor is it even in G.

Proof. We do this by induction on n. Suppose true up to $n-1$ and that $g_0 t_{j_1}^{\varepsilon_1} \ldots t_{j_n}^{\varepsilon_n} g_n$ contains no t_i-pinch, but is in G. Regard $G*_{\varphi_1,\ldots,\varphi_n}$ as the single HNN extension with base G_{n-1} and stable letter t_n, as in (5.16). Then on grouping together as subsequences the successive elements that are not equal to t_n or t_n^{-1}, we have that the resulting H-sequence so formed must contain a pinch $t_n a_n t_n^{-1}$ for $a_n \in A_n$ or $t_n^{-1} b_n t_n$ for $b_n \in B_n$.

But this a_n or b_n was obtained by multiplying out a subsequence of our MH-sequence, thus we regard this subsequence as an MH-sequence itself, but one with no appearances of $t_n^{\pm 1}$ so that it is moreover an MH-sequence with respect to the multiple HNN extension $G_{\varphi_1,\ldots,\varphi_{n-1}}$. Now this clearly contains no t_i-pinches as the original sequence did not, so by induction it cannot lie in G, and thus certainly not in A_n or in B_n. □

We now have, in exact analogy with (5.11) and (5.12):

Definition 5.19. Let $G*_{\varphi_1,\ldots,\varphi_n}$ be an HNN extension with base G, stable letters $t_1, \ldots, t_n$ and A_i, B_i subgroups of G where $\varphi_i : A_i \to B_i$ is an isomorphism. We say that a subgroup H of the base G is a **good** subgroup (with respect to this multiple HNN extension) if $\varphi(H \cap A_i) = H \cap B_i$ for each $i = 1, \ldots, n$.

Lemma 5.20. If $G*_{\varphi_1,\ldots,\varphi_n}$ is as defined in (5.19) and H is a good subgroup of G then the subgroup $\langle H, t_1, \ldots, t_n\rangle$ of the multiple HNN extension $G*_{\varphi_1,\ldots,\varphi_n}$ is itself naturally an HNN extension $H*_{\psi_1,\ldots,\psi_n}$ with base H, stable letters t_i, and where $H \cap A_i, H \cap B_i$ are subgroups of H with ψ_i, the restriction of φ_i to $H \cap A_i$, an isomorphism from $H \cap A_i$ to $H \cap B_i$. Moreover $\langle H, t_1, \ldots, t_n\rangle \cap G = H$.

Proof. Again form the abstract HNN extension $H*_{\psi_1,\ldots,\psi_n}$. Now use (5.18) instead of (5.9). □

6. (Lack of) Finite Index Subgroups

If a subgroup H of a group G has finite index in G, namely a finite number of left (equivalently right) cosets, we will write here $H \leqslant_f G$ (and $H \trianglelefteq_f G$ if H is also normal in G, $H <_f G$ if H is also proper etc), whereas $[G : H]$ will denote the the number of these cosets: the **index** of H in G.

Lemma 6.1. If $[G : H] = k$ then for any $g \in G$, there exists i with $1 \leqslant i \leqslant k$ such that $g^i \in H$.

If $H \trianglelefteq G$ then we can take $i \mid k$, or even $i = k$.

Proof. The left cosets $H, gH, \ldots, g^kH$ cannot all be distinct, so $g^iH = g^jH$ for some $1 \leqslant i < j \leqslant k$, and then $g^{j-i} \in H$ with $1 \leqslant j - i \leqslant k$.

If $H \trianglelefteq G$ then by Lagrange the element gH of G/H has order dividing the order k of the finite quotient group G/H. □

The Regular Representation

Any group G acts on itself by (left) multiplication. Now let H be any subgroup and $\mathcal{L}$ be the set of left cosets of H in G. The **(left) regular representation** ρ of G on $\mathcal{L}$ is the action of G given by $\rho(g)(xH) = gxH$.

Note $\mathrm{Orb}(H) = \mathcal{L}$ and the stabiliser of the point $H \in \mathcal{L}$ is the subgroup $H \leqslant G$.

Lemma 6.2. The kernel of this representation ρ satisfies

$$\ker \rho = \bigcap_{x \in G} xHx^{-1}.$$

Proof. We have that the element $g \in G$ satisfies $xH = gxH$ for all $x \in G \iff x^{-1}gx \in H$ for all $x \in G$. □

Definition 6.3. For a subgroup H of G, the **core** of H in G is $\ker \rho$.

Proposition 6.4. The core of H is normal in G and furthermore is the largest normal subgroup of G that is contained in H.

Proof. The core is normal as it is a kernel. It stabilises the point H so is a subgroup of its stabiliser which *is* H.

Now suppose that $N \trianglelefteq G$ and $N \leqslant H$ then $x^{-1}Nx \leqslant H$ for all $x \in G$. □

Theorem 6.5. If $H \leqslant_f G$ with $[G : H] = n$ then there exists $N \trianglelefteq_f G$ with $N \leqslant H$ and $[G : N] \mid n!$.

Proof. The set $\mathcal{L}$ has n elements, so ρ is a homomorphism from G to S_n. Then $|G/\ker \rho| \cong |\rho(G)|$ which divides $|S_n|$. □

But how do we know that an infinite group has proper finite index subgroups? We don't!

Example 6.6. The group of rationals $\mathbb{Q}$ under addition has no proper finite index subgroups.

Proof. Writing everything additively for the moment, suppose $H <_f \mathbb{Q}$ with index $n > 1$ and $q_0 \in \mathbb{Q} \setminus H$. Now H is of course normal in G, so for each $q \in \mathbb{Q}$ we have $nq \in H$ by (6.1). Thus taking $q = q_0/n$ we have $q_0 = nq \in H$.

But $\mathbb{Q}$ is not finitely generated. What about a finitely generated, or even a finitely presented example?

Theorem 6.7 Part 1 (Higman, 1951). The group G defined by the presentation

$$\langle a_1, a_2, a_3, a_4 \mid a_1 a_2 a_1^{-1} = a_2^2,\ a_2 a_3 a_2^{-1} = a_3^2,\ a_3 a_4 a_3^{-1} = a_4^2,\ a_4 a_1 a_4^{-1} = a_1^2 \rangle$$

has no proper finite index subgroups.

Proof. If $H <_f G$ then (6.5) gives a non-trivial finite quotient G/N. Note that for $n > 1$ and a prime p dividing $2^n - 1$, the least prime factor of n is less than p.

Take r the order of 2 mod p, then $r \mid n$ and $r \mid p-1$ (by Fermat's Little Theorem). Now say n_i is the order of the image of $\overline{a_i}$ in G/N. Then $\overline{a_1}^n\, \overline{a_2}\, \overline{a_1}^{-n} = \overline{a_2}^{2^n}$, so $n_2 \mid 2^{n_1} - 1$, and so on.

Let p be the smallest prime dividing $n_1 n_2 n_3 n_4$, with (wlog) $p \mid n_2$. Then n_1 has a smaller prime factor. This is a contradiction unless $n_1 n_2 n_3 n_4 = 1$. □

Theorem 6.7 Part 2 (Higman, same paper). This group G is infinite.

Proof. The group H defined by the presentation $\langle x, y \mid yxy^{-1} = x^2 \rangle$ is an HNN extension in which we see that x and y have infinite order, by (5.7). Let H' be an isomorphic copy of H with presentation $\langle x', y' \mid y'x'y'^{-1} = x'^2 \rangle$. We form $H *_\varphi H'$, where $\varphi(\overline{x}) = \overline{y'}$ is infinite and will be defined by the presentation (via $x = y'$, $z = x'$)

$$\langle x, y, z \mid yxy^{-1} = x^2, xzx^{-1} = z^2 \rangle.$$

Note that $\overline{y}, \overline{z}$ freely generate F_2 by (5.14) as a freely reduced word $w(y, x')$ with powers gathered becomes an A-reduced sequence with respect to this free product with amalgamation $H *_\varphi H'$.

Now take four copies of H, say H_i, each defined by the presentation $\langle a_i, b_i \mid b_i a_i b_i^{-1} = a_i^2 \rangle$ for $i = 1, 2, 3, 4$. Form

$$K = H_1 *_\varphi H_2 = \langle a_1 (= b_2), b_1, a_2 \mid b_1 a_1 b_1^{-1} = a_1^2,\ a_1 a_2 a_1^{-1} = a_2^2 \rangle,$$

which is infinite with $\langle \overline{b_1}, \overline{a_2} \rangle$ free as before, and $L = H_3 *_\varphi H_4$ (send $1 \mapsto 3$, $2 \mapsto 4$).

Finally, make $K *_\theta L$ for $\theta(\overline{b_1}) = \overline{a_4}$, $\theta(\overline{a_2}) = \overline{b_3}$. This is G. □

This page was left intentionally blank.

7. Computability theory

Turing machines

To explore what is *incomputable*, we first need a robust definition of what it means for a problem to be *computable*. There are many (equivalent) ways to do this; we present one of them here, first introduced by Turing[1] in 1937.

Definition 7.1. A **Turing machine** (abbreviated to TM) is a finite object which consists of the following:

1. A finite **alphabet** $S = \{s_0, \ldots, s_m\}$.
2. A finite set of **states** $Q = \{q_0, \ldots, q_n\}$ (we distinguish q_1 as the **initial state** and q_0 as the **halting state**.
3. Two formal symbols L, R, different from any symbols in S or Q.
4. A finite set of **instructions**, which are quadruples in $Q \times S \times (S \cup \{L, R\}) \times Q$, one for each ordered pair $(q_i, s_j) \in Q \times S$, each of one of the following three forms:
   a) (q_i, s_j, s_k, q_l)
   b) (q_i, s_j, L, q_l)
   c) (q_i, s_j, R, q_l)

   such that each starting pair $(q_i, s_j, *, *)$ occurs precisely once.
5. An **eye** which can read one alphabet symbol at a time.
6. A variable internal state q which takes values in Q.

Definition 7.2. We define the **action** of a Turing machine on words in its alphabet by the following:

a) Take a word $w = s_{i_1} \ldots s_{i_k} \in S^+$.

b) Write w on a tape, with one symbol per box. That is, $\overline{|s_{i_1}|s_{i_2}|\ldots|s_{i_k}|}$.

c) Place the eye over the leftmost square of the tape.

d) Set the internal state to $q = q_1$.

e) Look for the quadruple $c = (q, s_j, *, *)$ (where s_j is the symbol under the eye), and implement c according to the following convention:

i) A quadruple of type (q_i, s_j, s_k, q_l) means the machine replaces the read symbol s_j with the symbol s_k, and changes its internal state to $q = q_l$.

ii) A quadruple of type (q_i, s_j, L, q_l) means the eye moves one square to the left, and the machine changes its internal state to $q = q_l$.

iii) A quadruple of type (q_i, s_j, R, q_l) means the eye moves one square to the right, and the machine changes its internal state to $q = q_l$.

Note: If the eye ever moves past the end of the tape, additional squares with symbol s_0 are added as needed, hence the tape is always finite (for this reason we often interpret s_0 as a blank square).

f) If the internal state q is not q_0, then the machine repeats the process from step e) , with the new internal state and the eye at the new position.

g) If the internal state q is q_0, then the machine halts and outputs the (now modified) tape.

Definition 7.3. If, on input of a word w, the Turing machine T eventually halts in finitely-many steps (that is, eventually reaches its halting state q_0), then we say $T(w)$ **halts**, and write $T(w)\downarrow$. We write $T(w)\downarrow = v$ (with $v \in S^+$) to denote that $T(w)$ halts with v written on the tape when it does. If T never reaches its halting state on input w, then we say $T(w)$ **does not halt**, and write $T(w)\uparrow$.

[1] A. Turing, *On computable numbers, with an application to the entscheidungsproblem*, Proc. London Math. Soc. (2) **42**, 230–265 (1937).

Church-Turing thesis

It has been proposed by Church and Turing that the following idea, though inherently unprovable because of it's lack of formalisation, is essentially true (and we paraphrase somewhat here):

Any finite written description of a deterministic step-by-step computation is equivalent to some Turing machine. Moreover, there is a construction that, given such a finite description, will give us an explicit Turing machine T that carries out the original computation.

This is colloquially referred to as the **Church-Turing thesis**. We will make very frequent use of this idea, as at several places in our proofs we will describe, in words, an algorithmic process. This is often referred to as 'Proof by the Church-Turing thesis'. We do not suggest that you formalise all these proofs, as such actions are usually as unenlightening as they are time consuming. We will refrain from explicitly stating that we are appealing to the Church-Turing thesis in proofs, but these instances should be fairly self-evident when we make statements like 'Begin an enumeration of .....', or 'Run the following infinite collection of infinite enumerations as a diagonal process via interleaving....'.

Computable and incomputable sets

Definition 7.4. The **halting set** of a Turing machine T with alphabet S, denoted $\Omega(T)$, is the set of all words $w \in S^+$ on which $T(w)$ halts. That is,

$$\Omega(T) := \{w \in S^+ \mid T(w)\downarrow\}$$

Given a finite alphabet S, a set $A \subseteq S^+$ is said to be **recursively enumerable**, abbreviated r.e. (or **computably enumerable**) if $A = \Omega(T)$ for some Turing machine on alphabet S. $A \subseteq S^+$ is said to be **recursive** (or **computable**) if both A and $S^+ \setminus A$ are recursively enumerable.

If we consider all Turing machines on alphabet $S = \{0, 1\}$, then we can interpret S^+ as $\mathbb{N}$ (via binary expansion). So these Turing machines can be seen to take as input integers. Moreover, by reading the binary string on output tapes, these machines can be seen to output integers. Thus, each Turing machine T on alphabet $S = \{0, 1\}$ defines a unique partial function $f : \mathbb{N} \to \mathbb{N}$, with domain of definition $\Omega(T)$ (interpreted as binary representations of integers). This idea extends to any alphabet $S = \{0, 1, \ldots, n-1\}$; the inputs and outputs of T are thus interpreted as integers in base n.

Given that Turing machines are finite objects, we can make a 'nice' numbering of them (for convenience, we restrict ourselves to the Turing machines with alphabet $S = \{0, 1\}$). We start by writing down all machines with two states ($|Q| = 2$) and ordering their quadruples lexicographically, then all machines with 3 states ($|Q| = 3$) and ordering their quadruples lexicographically, and so on (whenever we have k states, we take the set of states to be the symbols $\{q_0, q_1, \ldots, q_{k-1}\}$, with that ordering). Thus we can construct an algorithmic numbering of all Turing machines on alphabet $S = \{0, 1\}$, which we write as $T_1, T_2, \ldots$ (we fix this numbering from now on). This numbering is completely algorithmic, in the sense that:
1. Given a Turing machine T on alphabet $\{0, 1\}$, we can find n such that $T = T_n$.
2. Given n, we can construct the Turing machine T_n.

Definition 7.5. We say a set $X \subseteq \mathbb{N}$ is **recursively enumerable**, abbreviated r.e. (or **computably enumerable**) if there exists n such that $X = \Omega(T_n)$; X is said to be **recursive** (or **computable**) if there exists m, n such that $X = \Omega(T_n)$ and $\mathbb{N} \setminus X = \Omega(T_m)$.

So what does it really mean for $X \subseteq \mathbb{N}$ to be recursively enumerable? The basic idea is that there exists a process which outputs integers in a sequence $a_1, a_2, \ldots$ such that:

1. ONLY elements of X appear in the list.
2. ALL elements of X eventually appear in the list (possibly with repetition).

How do we get this list? As X is recursively enumerable, there exists n such that $X = \Omega(T_n)$. So take one copy of T_n, input 1, and do 'one step' of the computation of $T_n(1)$ (that is, carry out the instructions of one quadruple). Now take another copy of T_n, do one step of $T_n(2)$, and one *more* step of $T_n(1)$. Then take a new copy of T_n, do one step of $T_n(3)$, one more step of $T_n(2)$, and one more step of $T_n(1)$. This is a very long diagonal process. Each time one of the $T_n(k)$ halts, we add k to our sequence. As all these steps happen in a given order, the sequence will have a total ordering (the order on which the machines halted).

The much stronger notion of being computable comes from this 'listing' process. If $X \subseteq \mathbb{N}$ is computable, then we can construct two processes, one which outputs a sequence $a_1, a_2, \ldots$ for X, and another which outputs a sequence $b_1, b_2, \ldots$ for $\mathbb{N} \setminus X$. We can use this to compute membership in X as follows: given an integer k, simply look for k down the list $a_1, a_2, \ldots$, as well as down the list $b_1, b_2, \ldots$ (do this as a diagonal process: check a_1, then b_1, then a_2, then b_2, ...). Since these lists are disjoint, and their union is $\mathbb{N}$, we know that k will appear in precisely one list, which will tell us if $k \in X$ or $k \notin X$.

Note that being recursive is equivalent to having a characteristic function which is **computable**: there is an algorithm which takes as input an integer n and always eventually halts, outputting yes or no depending on whether n lies in the set or not. Being r.e. is equivalent to having a characteristic function which is **partially computable**: there is an algorithm which takes as input an integer n, which eventually halts with output yes when n lies in the set, but does not halt if n does not lie in the set.

The following lemma, though elementary, is crucial in showing the existence of a *universal* Turing machine.

Definition 7.6. Define **Cantor's pairing function** $\langle ., . \rangle : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by

$$\langle x, y \rangle = \frac{1}{2}(x + y)(x + y + 1) + y$$

Then this is a bijection from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$, which we can extend inductively to define $\langle x_1, \ldots, x_n \rangle := \langle \langle x_1, \ldots, x_{n-1} \rangle, x_n \rangle : \mathbb{N}^n \to \mathbb{N}$. We note that this function, and all its extensions, are computable in the sense of the preceding paragraph.

Proof. Let $x + y = z$. Then we need only make the following elementary observations:

1. $\langle z, 0 \rangle \leqslant \langle x, y \rangle \leqslant \langle 0, z \rangle$
2. $\langle x, y + 1 \rangle = \langle x + 1, y \rangle + 1$
3. $\langle 0, z \rangle + 1 = \langle z + 1, 0 \rangle$

The above imply that a totally ordered listing of $\mathbb{N}$ is given by
$\langle 0, 0 \rangle, \langle 1, 0 \rangle, \langle 0, 1 \rangle, \langle 2, 0 \rangle, \langle 1, 1 \rangle, \langle 0, 2 \rangle, \ldots$

Most 'well-defined' problems in mathematics can be reduced to computing membership in a subset of $\mathbb{N}$, via use of Cantor's pairing function. For example, take the set $X := \{\langle x, y \rangle \in \mathbb{N} \mid x \text{ divides } y\}$. Then the problem 'does x divide y?' can be interpreted as 'is $\langle x, y \rangle$ in X?'

Theorem 7.7 (Turing, 1937). There exists a **universal Turing machine**. That is, a Turing machine which can simulate the action of every Turing machine.

Proof. We define an algorithm as follows: on input of an integer $x = \langle m, n \rangle$, we (begin to) compute $T_m(n)$. Since this is a verbal description of an algorithm, we can conclude that there exists some Turing machine T such that $T(x) = T_m(n)$ (that is, if $T_m(n) \downarrow$ with output k then $T(x) \downarrow$ with output k, and if $T_m(n) \uparrow$ then $T(x) \uparrow$).

There is a very important set of integers, which forms the basis for most incomputable problems in mathematics.

Definition 7.8. The Halting Set, denoted $\mathbb{K}$, is given by

$$\mathbb{K} := \{n \in \mathbb{N} \mid T_n(n) \downarrow\}$$

This 'diagonal' set seems obscure, but has the following useful properties:

Theorem 7.9. The halting set $\mathbb{K}$ satisfies the following:
1. $\mathbb{K}$ is recursively enumerable.
2. $\mathbb{K}$ is not recursive.

Proof. Clearly, $\mathbb{K}$ is recursively enumerable; in the same way that we defined a universal Turing machine $T(x) := T_m(n)$ (where $x = \langle m, n\rangle$), we can define a 'restricted' universal Turing machine that computes *one* entry of each Turing machine, by $T'(n) := T_n(n)$. Again, this is a verbal description of an algorithm, so by the Church-Turing thesis we can indeed construct a Turing machine which performs this computation. Thus $\mathbb{K} = \Omega(T')$, and so $\mathbb{K}$ is r.e.

Now, suppose that $\mathbb{K}$ was recursive. Then there exists some m such that $\Omega(T_m) = \mathbb{N}\setminus\mathbb{K}$. Now look at the index m; we want to know if m is in $\mathbb{K}$ or not. So we have:

$$\begin{aligned} m \in \mathbb{K} &\Leftrightarrow T_m(m) \downarrow \\ &\Leftrightarrow m \in \Omega(T_m) \\ &\Leftrightarrow m \in \mathbb{N} \setminus \mathbb{K} \\ &\Leftrightarrow m \notin \mathbb{K} \end{aligned}$$

This is a contradiction, so m cannot possibly exist. That is, there is no m such that $\Omega(T_m) = \mathbb{N} \setminus \mathbb{K}$, and so $\mathbb{N} \setminus \mathbb{K}$ is not recursively enumerable.

So $\mathbb{K}$ is a set which is *described* by a Turing machine, but for which membership cannot be *computed* by any Turing machine. That is, deciding membership in $\mathbb{K}$ is our first provably *incomputable* problem!

Note that, in general, if we have a recursively enumerable set of Turing machines $X = \{T_{i_1}, T_{i_2}, \ldots\}$ (that is, the set $\{i_1, i_2, \ldots\}$ is r.e.) which take as input values in $\mathbb{N}$, then we can construct a (quasi-universal) Turing machine T which simulates the action of all the machines in X. This is given by $T(\langle m, n\rangle) := T_{i_m}(n)$, and if $|X| < m$ then $T(\langle m, n\rangle) \uparrow$. Note that we do not, *a priori*, need to know the size of X; we just start enumerating these machines $T_{i_1}, T_{i_2}, \ldots$, and if we ever get to T_{i_m} then we input n into T_{i_m} to (try and) compute $T_{i_m}(n)$.

Example 7.10. Let $\{T_i\}_{i\in I}$ be a collection of Turing machines. Let $X_i := \Omega(T_i)$ be the r.e. set of integers defined by T_i, for each $i \in I$. Then
1. If I is finite then $\bigcap_{i\in I} X_i$ is r.e.
2. If I is r.e. then $\bigcup_{i\in I} X_i$ is r.e.

There are many more results from computability theory which, though interesting, we do not need to cover to complete this course. A good and easily-available reference is the set of online course notes by Frank Stephan[2].

[2] *Recursion Theory*, Online course notes from 2012. See http://www.comp.nus.edu.sg/∼fstephan/recursiontheory-pstopdf.pdf

8. The word problem

Recursive presentations

Definition 8.1. If X is a finite set, and $R \subseteq F(X)$ is an r.e. subset of $F(X)$ (the halting set of a Turing machine on $X \cup X^{-1}$), then we say $\langle X|R\rangle$ is a **recursive presentation**; groups with such a presentation are called **recursively presented**.

We will often consider the 'data' of a recursive presentation to be
1. The finite generating set X, and
2. A Turing machine T whose halting set is R; $\Omega(T) = R$.

This way, a recursive presentation can be described with only finite data. This is important later on when dealing with algorithmic properties of recursively presented groups.

Observe that being finitely presented implies being recursively presented, as all finite sets are r.e. Moreover, one can have a recursive presentation of a finitely presented group. This is a very important distinction; having a recursive presentation of a finitely presented group does not always allow us to algorithmically extract a finite presentation from it.

The reason we write 'recursive presentation' rather that 'recursively enumerable presentation' is because if $P = \langle X|R\rangle$ is a recursive presentation of a group, then $\overline{P}$ has another recursive presentation where the relating set is recursive.

Lemma 8.2. Let $P = \langle X|R\rangle$ be a recursive presentation of a group. Then there exists a recursive presentation $P' = \langle X'|R'\rangle$ with R' recursive such that $\overline{P'} \cong \overline{P}$.

Proof. First, if R is finite, then it is recursive, so just take the original presentation P.
So let R be an infinite r.e. set, with enumeration $r_1, r_2, \ldots$. Take $X' := X \cup \{t\}$ and $R' := \{t^i r_i \mid r_i \in R\} \cup \{t\}$. It is clear that $\overline{P'} \cong \overline{P}$, as the generator $\overline{t}$ is trivial in $\overline{P'}$, and the other relators are then unchanged at the group level. To see that the set R' is recursive in $F(X \cup \{t\})$, take any word $w \in F(X \cup \{t\})$. If this is not of the form $t^i z$ for some $z \in F(X)$, then $w \notin R'$. Otherwise w is of the form $t^i z$ with $z \in F(X)$. So start enumerating the elements of R. Eventually, as R is infinite, we will reach the word r_i; then $r_i = z$ (in $F(X)$) iff $w \in R'$. This process always halts, so R' is recursive.

Note that this proof is merely existential. We cannot necessarily construct (a Turing machine enumerating) R' from (a Turing machine enumerating) an arbitrary r.e. set $R \subseteq F(X)$, as we would need *a priori* knowledge of whether R was finite or infinite.

The word problem for groups

Definition 8.3. We define the **word problem** (abbreviated to WP) for a recursive presentation $P = \langle X|R\rangle$ of a group as follows:

$$\text{Given two words } u, v \in F(X), \text{ is } \overline{u} = \overline{v} \text{ in } \overline{P}?$$

This can be rephrased in the following equivalent way:

$$\text{Given a word } w \in F(X), \text{ is } \overline{w} = e \text{ in } \overline{P}?$$

Groups which have a recursive presentation for which there exists an algorithm that, on input of a pair of words $u, v \in F(X)$, decides if $\overline{u} = \overline{v}$ in $\overline{P}$, are said to have **solvable word problem** (or **soluble word problem**, or **decidable word problem**), abbreviated to SWP. If there is no such presentation for which such an algorithm exists, then we say the group has **unsolvable word problem** (or **insoluble word problem**, or **undecidable word problem**), abbreviated to IWP.

So saying G has SWP is equivalent to saying that, for some recurisve presentation P of G, the set of trivial words in P is recursive.

Lemma 8.4. Let $P = \langle X|R\rangle$ be a recursive presentation of a group. Then the set of trivial words $\{w \in F(X) \mid \overline{w} = e \text{ in } \overline{P}\}$ is r.e. This is **uniform** over all recursive presentations. i.e., there is *one* Turing machine which takes as input pairs (P, w) where P is a recursive presentation and w is a word on the generators of P, and halts iff $\overline{w} = e$ in $\overline{P}$.

Proof. This follows from (1.16) and the comment before (4.2), since we can form a recursive enumeration of all words of the form $u_1 r_1^{\pm 1} u_1^{-1} \dots u_n r_n^{\pm 1} u_n^{-1}$, as well as their free reductions. Then w will appear in this enumeration iff $\overline{w} = e$ in $\overline{P}$.

Corollary 8.5. Let G be a recursively presentated group. Then G has solvable word problem iff it has a recursive presentation $P = \langle X|R\rangle$ whose set of non-trivial words $\{w \in F(X) \mid \overline{w} \neq e \text{ in } \overline{P}\}$ is r.e.

Lemma 8.6. Suppose P_1 and P_2 are recursive presentations defining groups, with $\overline{P_2} \leqslant \overline{P_1}$. Suppose, moreover, that there is an algorithm to solve the word problem in P_1. Then there is an algorithm to solve the word problem in P_2.

Proof. Let $P_i = \langle X_i|R_i\rangle$, $i = 1, 2$. As $\overline{P_2} \leqslant \overline{P_1}$, there is some map $\varphi : X_2 \to F(X_1)$ which extends to a map $\varphi' : F(X_2) \to F(X_1)$ which in turn extends to an injective homomorphism $\overline{\varphi'} : \overline{P_2} \to \overline{P_1}$. So, given a word $w \in F(X_2)$, evaluate the word $\varphi'(w)$, and then use the solution to the word problem in P_1 to compute if $\overline{\varphi'(w)} = e$ in $\overline{P_1}$.

Corollary 8.7. Suppose P_1 and P_2 are recursive presentations defining the same group. Suppose, moreover, that there is an algorithm to solve the word problem in P_1. Then there is an algorithm to solve the word problem in P_2.

So having solvable word problem is independent of the recursive presentation we are considering. Note that this is *not* the case for infinitely generated groups: the presentation $\langle x_i \ \forall i \in \mathbb{N} \mid -\rangle$ for F_∞ has solvable word problem, but the presentation $\langle x_i \ \forall i \in \mathbb{N} \mid x_i = e \ \forall i \in \mathbb{K}\rangle$ for F_∞ does not.

Definition 8.8. We define the **subgroup membership problem** (abbreviated to MP) for a recursive presentation $P = \langle X|R\rangle$ of a group and a finite set of words $\{w_1, \dots, w_m\} \subseteq F(X)$ as follows:

$$\text{Given a word } v \in F(X), \text{ is } \overline{v} \in \langle \overline{w_1}, \dots, \overline{w_m}\rangle \text{ in } \overline{P}?$$

If there exists an algorithm that, on input of a word $v \in F(X)$, decides if $\overline{v} \in \langle \overline{w_1}, \dots, \overline{w_m}\rangle$ in $\overline{P}$, then we say that $\langle \overline{w_1}, \dots, \overline{w_m}\rangle$ has **solvable subgroup membership problem** in $\overline{P}$.

The word problem for semigroups

Definition 8.9. We define the **word problem** (WP) for a recursive presentation $P = \langle X|R\rangle$ of a semigroup as follows:

$$\text{Given two words } u, v \in X^\circ, \text{ is } \overline{u} = \overline{v} \text{ in } \overline{P}?$$

Semigroups which have a presentation for which there exists an algorithm that, on input of a pair of words $u, v \in X^\circ$, decides if $\overline{u} = \overline{v}$ in $\overline{P}$, are said to have **solvable word problem** (or **soluble word problem**, or **decidable word problem**), abbreviated to SWP. If there is no such presentation for which such an algorithm exists, then we say the semigroup has **unsolvable word problem** (or **insoluble word problem**, **undecidable word problem**), abbreviated to IWP.

A semigroup with unsolvable word problem

Turing machines are not just confined to mathematical logic. We can fully 'realise' the action of a Turing machine in algebraic structures. The following construction is originally due to Markov and Post (independently), and was the first example of a finitely presented semigroup with unsolvable word problem. In essence, they took the construction of a Turing machine and encoded it into a finite presentation of a semigroup in quite a natural way. This construction mimics the inner workings of the Turing machine in the semigroup, without the need for much extra peripheral structure. It is interesting to compare this with the construction we will see later of a finite presentation of a group with unsolvable word problem, which takes a different machine encoding approach (modular machines, which we shall introduce in the next section), but for which the action of the machine is buried much more deeply within the algebraic structure.

To get an algebraically clearer description of the inner working of a Turing machine, we introduce the notion of an *instantaneous description*.

Definition 8.10. Given a Turing machine T with alphabet $S = \{s_0, \ldots, s_m\}$ and states $Q = \{q_0, \ldots, q_n\}$, we call a word $w \in (S \cup Q)^+$ an **instantaneous description** of T if w takes the form

$$w = s_{i_1} \ldots s_{i_k} q_j s_{i_{k+1}} \ldots s_{i_l}$$

where we may have $k = 0$ (i.e., q_j can be the leftmost letter of w), but we insist that $l > k$ (i.e., q_j cannot be the rightmost letter of w).

The instantaneous description $s_{i_1} \ldots s_{i_k} q_j s_{i_{k+1}} \ldots s_{i_l}$ simply corresponds to "The Turing machine T, in state q_j, reading letter $k{+}1$ of the l-letter word $s_{i_1} \ldots s_{i_l}$". We then see that we can act on this instantaneous description by choosing the appropriate quadruple $(q_j, s_{i_{k+1}}, *, *)$ from T, to get a new instantaneous description.

As a notational convention, given a semigroup presentation $P = \langle X | R \rangle$, we call an element $(x, y) \in R$ a **semigroup relator** (or just **relator**), and often write this as $x = y$.

Definition 8.11. Let T be a Turing machine with alphabet $S = \{s_0, \ldots, s_m\}$, states $Q = \{q_0, \ldots, q_n\}$, and halting state q_0. We define the **associated semigroup** $\overline{\Gamma(T)}$ to be the semigroup presented by

$$\Gamma(T) := \langle q, h, s_0, \ldots . s_m, q_0, \ldots, q_n \mid R(T) \rangle$$

where the relators in $R(T)$ are, for all $i, l \in \{1, \ldots, n\}$, all $j, k \in \{0, \ldots, n\}$, and all $\beta \in \{0, \ldots, m\}$:

$$\begin{aligned}
q_i s_j = q_l s_k \quad &\text{if} \quad q_i s_j s_k q_l \in T \\
q_i s_j s_\beta = s_j q_l s_\beta \quad &\text{if} \quad q_i s_j R q_l \in T \\
q_i s_j h = s_j q_l s_0 h \quad &\text{if} \quad q_i s_j R q_l \in T \\
s_\beta q_i s_j = q_l s_\beta s_j \quad &\text{if} \quad q_i s_j L q_l \in T \\
h q_i s_j = h q_l s_0 s_j \quad &\text{if} \quad q_i s_j L q_l \in T \\
q_0 s_\beta &= q_0 \\
s_\beta q_0 h &= q_0 h \\
h q_0 h &= q
\end{aligned}$$

To give some sort of explanation: the first 5 rows of relators in the definition above precisely mimic the action of the Turing machine (7.2) as it computes a given input, by virtue of the fact that their conditions are all quadruples from the Turing machine. The final three rows of relators ensure that, when we reach the halting state q_0, everything collapses down to q. The symbol h can be viewed as an 'end-marker'.

Theorem 8.12 (Markov-Post, 1947). Let T be a Turing machine, and let $\Gamma(T)$ be the associated semigroup presentation, as in (8.11). If $w \in S^+$, then

$$w \in \Omega(T) \text{ if and only if } \overline{hq_1wh} = \overline{q} \text{ in } \overline{\Gamma(T)}$$

Proof. First, if $w \in \Omega(T)$ then there is a sequence of instantaneous descriptions $q_1w = w_0, w_1, \ldots, w_n = \alpha q_0 \beta$, such that $\alpha, \beta \in S^+$, and each successive pair w_i, w_{i+1} differ by the action of one quadruple from T (so we mirror the action of T on w by the instantaneous descriptions). But by the first 5 relator types of $\Gamma(T)$, if w_i, w_{i+1} differ by the action of one quadruple from T, then $\overline{hw_ih} = \overline{hw_{i+1}h}$ by (4.14). So we get $\overline{hq_1wh} = \overline{h\alpha q_0 \beta h}$, and then $\overline{h\alpha q_0 \beta h} = \overline{q}$ by repeated application of the last 3 relator types of $\Gamma(T)$, again using (4.14).

For the other direction, notice that by (4.14) if $\overline{hq_1wh} = \overline{q}$ in $\overline{\Gamma(T)}$ then we have a sequence of words $hq_1wh = w_0, w_1, \ldots, w_n = q$, all defining the same element in $\overline{\Gamma(T)}$, each differing by application of one relator. Take a sequence of shortest possible length N. Since the only way to reach q is via the relator $hq_0h = q$, then we must have $w_{N-1} = hq_0h$ (by minimality of N). The only way to reach a word of the form hq_0h (not from q) is to have words of the form $h\alpha q_0 \beta h$ in the sequence, where $\alpha, \beta \in S^+$. Let w_M be the first word with q_0 appearing. Again, by the minimality of N, we have that q_0 appears in all words $w_M, w_{M+1}, \ldots, w_{N-1}$. Thus the sequence $w_0, \ldots, w_M$ is obtained by repeated application of relators of the first 5 types in $\Gamma(T)$. Seeing as the first word in the sequence (w_0) is an instantaneous description (buffered either side by h), and all the relators of the first 5 types preserve instantaneous descriptions, then *all* the w_i's are instantaneous descriptions (buffered either side by h), for $i \leqslant M$.

Note that each relator has corresponding 'time' direction in the Turing machine T. For example, the relator $q_is_j = q_ls_k$ (if $q_is_js_kq_l \in T$) is 'forwards in time' if we replace $q_is_j \to q_ls_k$, and 'backwards in time' if we replace $q_ls_k \to q_is_j$. So each adjacent pair w_i, w_{i+1} corresponds to either a 'forwards in time' or 'backwards in time' application of a relator. Obviously, the pair w_{M-1}, w_M corresponds to a 'forwards in time' relator, as this is the only way to introduce the symbol q_0. Assume that not all pairs are 'forwards in time'. As the sequence ends with a 'forwards in time' pair w_{M-1}, w_M, there must be some $1 \leqslant j \leqslant M-1$ such that w_{j-1}, w_j is backwards in time, but w_j, w_{j+1} is forwards in time. But then, by the deterministic nature of a Turing machine, we must have that $w_{j-1} = w_{j+1}$ as words, and so our sequence was not of minimal length. This is a contradiction, so all pairs w_i, w_{i+1} corresponds to a 'forwards in time' application of a relator. This implies that, on input of the word w, the Turing machine eventually reaches the internal state q_0, following these 'forwards in time' pairs w_i by application of quadruples from T. So $T(w)$ halts.

So we have proved that the action of T is *completely* simulated within the finitely presented semigroup $\overline{\Gamma(T)}$, and moreover we can algorithmically construct $\Gamma(T)$ from T as given in (8.12). Later, we will show that there is an analogous construction which simulates the action of a Turing machine within a finitely presented *group*, and then use it to prove the Higman embedding theorem.

Theorem 8.13. There is a finitely presented semigroup with unsolvable word problem.

Proof. Let T_n be the Turing machine from (7.8) with halting set $\Omega(T_n) = \mathbb{K}$, and form $\Gamma(T_n)$ as in (8.11). Assume that $\overline{\Gamma(T)}$ has SWP; then there is an algorithm which takes any pair of words in the generators of $\Gamma(T)$ and decides if they give the same element in $\overline{\Gamma(T)}$. But now, given any $w \in S^+$ with S being the alphabet of T, we can use the algorithm for the word problem in $\overline{\Gamma(T)}$ to compute whether $\overline{hq_1wh} = \overline{q}$, and thus whether $w \in \Omega(T_n)$ by (8.12). But $\Omega(T_n) = \mathbb{K}$, which by (7.9) is not recursive.

9. Modular machines

We define modular machines as an alternate way of mechanical computing. We will show that they can simulate Turing machines in a very natural way.

Definition 9.1. A **modular machine** $\mathcal{M}$ consists of an integer $m > 1$ and a finite set of quadruples each of the form (a, b, c, R) or (a, b, c, L), where $m > a \geqslant 0$ and $m > b \geqslant 0$ and $m^2 > c \geqslant 0$. We require that, for each such pair (a, b), there is at most one quadruple of $\mathcal{M}$ of the form $(a, b, *, *)$.

A **modular machine configuration** is an ordered pair $(\alpha, \beta) \in \mathbb{N}^2$. We write $(\alpha, \beta) \underset{\mathcal{M}}{\longrightarrow} (\alpha_1, \beta_1)$, called a **computational step** of $\mathcal{M}$, if $\alpha = um + a$ and $\beta = vm + b$ (with $0 \leqslant a, b < m$) and there exists c such that either:

1. $(a, b, c, R) \in \mathcal{M}$ and $\alpha_1 = um^2 + c$ and $\beta_1 = v$, or
2. $(a, b, c, L) \in \mathcal{M}$ and $\alpha_1 = u$ and $\beta_1 = vm^2 + c$.

Note that the action of $\mathcal{M}$ on (α, β) depends only on the class of (α, β) modulo m. This is why we call $\mathcal{M}$ a *modular* machine.

We write $(\alpha, \beta) \underset{\mathcal{M}}{\overset{*}{\longrightarrow}} (\alpha', \beta')$ if there exists a finite sequence

$$(\alpha, \beta) = (\alpha_1, \beta_1) \underset{\mathcal{M}}{\longrightarrow} (\alpha_2, \beta_2) \underset{\mathcal{M}}{\longrightarrow} \dots \underset{\mathcal{M}}{\longrightarrow} (\alpha_n, \beta_n) = (\alpha', \beta')$$

Such a sequence is called a **computation** of $\mathcal{M}$.

If, for $\alpha = um + a$, $\beta = vm + b$ $(0 \leqslant a, b < m)$, no quadruple of $\mathcal{M}$ begins with (a, b), then we say (α, β) is **terminal**. If $(0, 0)$ is terminal in $\mathcal{M}$, then we define the set

$$H_0(\mathcal{M}) := \{(\alpha, \beta) \mid (\alpha, \beta) \underset{\mathcal{M}}{\overset{*}{\longrightarrow}} (0, 0)\}$$

Simulating Turing machines using modular machines

We will now see that modular machines can completely simulate the action of Turing machines. First, it helps to slightly re-define Turing machines to have *quintuples* (rather than quadruples), all of the form (q_i, s_j, s_k, q_l, Z), where $Z \in \{L, R\}$.

These act on tapes in almost the same way as quadruple-Turing machines. The machine, in state q_i, reading letter s_j, re-writes the letter to s_k, and then moves left or right 1 square (depending on whether $Z = L$ or R). The rest of the functionality of a quintuple-Turing machine is identical to that of a quadruple-Turing machine. Obviously a 'moving' quadruple from our original definition in (7.1) of the form (q_i, s_j, L, q_l) (or (q_i, s_j, R, q_l)) can be re-written in this quintuple form as (q_i, s_j, s_j, q_l, L) (or (q_i, s_j, s_j, q_l, R)); the two act identically on instantaneous descriptions.

To make the two definitions computationally equivalent, we replace each 'non-moving' quadruple (q_i, s_j, s_k, q_l) with the quintuple $(q_i, s_j, s_k, q_{ijl}, R)$ where q_{ijl} is a new auxiliary state, together with quintuples (q_{ijl}, x, x, q_l, L) for all alphabet symbols $x \in S$. That is, we 'write on the tape and move right, and then immediately move left again while leaving the tape unchanged'. Furthermore, by keeping the same alphabet, we have fully constructed a quintuple-Turing machine from a quadruple-Turing machine.

It is then clear that if T is a quadruple-Turing machine, and T' is its associated quintuple-Turing machine as defined above, then the two are computationally equivalent in the sense that if $w \in S^+$ (S being the alphabet of both T and T'), then

$$T(w) \downarrow = v \;\Leftrightarrow\; T'(w) \downarrow = v$$

We may now consider all our Turing machines to be in quintuple form; if we are given one in quadruple form, we convert it (algorithmically, as above) into quintuple form.

We now describe how to convert a quintuple-Turing machine into an equivalent modular machine. So, take a quintuple-Turing machine T with alphabet S and states Q. Set $m = |S| + |Q| + 1$. Now re-write the symbols S as integers $\{0, \ldots, n\}$, and the states Q as integers $\{n+1, \ldots, m-1\}$. We now define, from the quintuples of T, an associated modular machine $\mathcal{M}$ with modulus m.

First, we describe how to interpret an instantaneous description of a Turing machine as a pair of integers. We do this as follows:
Suppose $b_k \cdots b_1 b_0 q a c_0 c_1 \cdots c_l$ (which we shall call C) is an instantaneous description of T. In our re-writing convention above, we can consider b_i's, c_i's, a to all lie in $\{0, \ldots, n\}$, and q to lie in $\{n+1, \ldots, m-1\}$ (recall that we have re-written the symbols and states of T as integers between 0 and $m-1$).

We set $u := \Sigma_{i=0}^{k} b_i m^i$, $v := \Sigma_{i=0}^{l} c_i m^i$, and then to the instantaneous description C we associate *two* modular machine configurations. These are
1. $(um + a, vm + q)$, called the **left associate** of C, and
2. $(um + q, vm + a)$, called the **right associate** of C.

Now we define the modular machine $\mathcal{M}$ associated to T to have modulus m as above, and for each quintuple (q_i, s_j, s_k, q_l, D) in T (where $D \in \{L, R\}$), except those of the form $(q_0, *, *, *, *)$, we include *both* of the following two quadruples in $\mathcal{M}$:
1. $(q_i, s_j, s_k m + q_l, D)$, and
2. $(s_j, q_i, s_k m + q_l, D)$.

Notice that $\mathcal{M}$ always takes associates to associates, but might flip left $\leftrightarrow$ right.

We will now illustrate how the modular machine $\mathcal{M}$ mimics the action of T. Take a word $w \in S^+$ of T, and compute the left associate of $q_0 w$. Now, suppose we have *any* general instantaneous description C as above, converted to either its left or right associate.

First, suppose T has the quintuple (q, a, a', q', R). Then this quintuple acting on C would yield the instantaneous description $b_k \cdots b_1 b_0 a' q' c_0 c_1 \cdots c_l$. Had we been considering the left (resp. right) associate of C, which is $(um + a, vm + q)$ (resp. $(um + q, vm + a)$), then the *one* quadruple from $\mathcal{M}$ that we could apply here would be $(a, q, a'm + q', R)$ (resp. $(q, a, a'm + q', R)$). Thus, in either case, this one computational step of $\mathcal{M}$ would yield the right associate $(um^2 + a'm + q', v) = \left(\left(\Sigma_{i=0}^{k} b_i m^{i+1} + a'\right)m + q' \ , \ \left(\Sigma_{i=1}^{l} c_i m^{i-1}\right)m + c_0 \right)$. This corresponds to the instantaneous description $b_k \cdots b_1 b_0 a' q' c_0 c_1 \cdots c_l$, as expected.
If instead T has the quintuple (q, a, a', q', L), then this quintuple acting on C would yield the instantaneous description $b_k \cdots b_1 q' b_0 a' c_0 c_1 \cdots c_l$. Had we been considering the left (resp. right) associate of C, which is $(um + a, vm + q)$ (resp. $(um + q, vm + a)$), then the *one* quadruple from $\mathcal{M}$ that we could apply here would be $(a, q, a'm + q', L)$ (resp. $(q, a, a'm + q', L)$). Thus, in either case, this one computational step of $\mathcal{M}$ would yield the left associate $(u, vm^2 + a'm + q') = \left(\left(\Sigma_{i=1}^{k} b_i m^{i-1}\right)m + b_0 \ , \ \left(\Sigma_{i=0}^{l} c_i m^{i+1} + a'\right)m + q' \right)$. This corresponds to the instantaneous description $b_k \cdots b_1 q' b_0 a' c_0 c_1 \cdots c_l$, as expected.
Finally, note that $(am + q_0, bm + s)$ and $(am + s, bm + q_0)$ are terminal in $\mathcal{M}$ for all $s \in S$ and all $a, b \geqslant 0$, and are the *only* associates which are terminal. These correspond to associates of instantaneous descriptions containing the halting state q_0. When we reach such an associate, $\mathcal{M}$ 'terminates', and we convert the associate back to an instantaneous description giving us the output word of $T(w)$.

We summarise the discussion above with the following theorem.

Theorem 9.2. Given any Turing machine T in quadruple form, we can construct from it a computationally equivalent Turing machine T' in quintuple form. Given any Turing machine T' in quintuple form, we can construct from it a modular machine $\mathcal{M}$ which simulates the action of T'.

We need to take this one step further, and, for a quadruple-Turing machine T, relate its halting set $\Omega(T)$ to the set $H_0(\mathcal{M})$ where $\mathcal{M}$ is its associated modular machine. We do this as follows: after having formed the quintuple-Turing machine T' as described in (9.2), we make a new quintuple-Turing machine T'' by introducing a dummy symbol h into the alphabet, and two dummy states q_L, q_R. When the machine would otherwise enter the halting state q_0, make it instead enter q_L. Then, add some extra instructions so that once it enters q_L it keeps scanning left and re-writing all the squares with s_0. When it moves past the left-end of the tape it adds a new square with h on it. Then it reads that h, re-writes it with s_0 and goes into state q_R and repeats the 's_0-writing' process; this time to the right. When it moves past the right-end of the tape, it adds a new square with h on it. Then it reads h, re-writes it with s_0, and enters the original halting state q_0. To do this formally, the extra instructions needed are $(q_L, s, s_0, q_L, L)\ \forall s \in S$, (q_L, h, s_0, q_R, R), $(q_R, s, s_0, q_R, R)\ \forall s \in S$, (q_R, h, s_0, q_0, L), noting that when we move past the end of the tape when in state q_L or q_R, we add a new square with h on it (and not s_0). The point of doing this transformation is that whenever T'' halts, it outputs a tape consisting entirely of s_0's (i.e., it 'wipes the tape' just before halting).

If we construct from this modified quintuple-Turing machine T'' an equivalent modular machine $\mathcal{M}$ as per (9.2) *but where we insist that* s_0 *is assigned the integer* 0 *in the re-labeling and, in a slight deviation from convention, we also assign* 0 *to* q_0 (one can go back and check that this does not ruin the intended function of $\mathcal{M}$), then we can conclude that $T(w)\downarrow$ iff the left associate (α, β) of $q_1 w$ in $\mathcal{M}$ satisfies $(\alpha, \beta) \xrightarrow[\mathcal{M}]{*} (0, 0)$. That is:

Theorem 9.3 Let T be a quadruple-Turing machine. Then from it we can construct a modular machine $\mathcal{M}$ such that, if $w \in S^+$ is a word on the alphabet of T, and (α, β) is the left associate of $q_1 w$ in $\mathcal{M}$, then

$$T(w)\downarrow \ \Leftrightarrow\ (\alpha, \beta) \in H_0(\mathcal{M})$$

The exposition leading to (9.2) contained several important new ideas and concepts. In contrast, the exposition leading to (9.3) can be described as the analogue in computability theory of *abstract nonsense*; there is nothing deep in (9.3), and in the literature one would just explain (9.3) with the argument 'we can clear the tape before halting', with no further justification. As such, only the statement of (9.3) will be examinable, not its proof.

10. Groups with unsolvable word problem

There is an excellent survey by Charles F. Miller III on the word problem in groups, and indeed many other group-theoretic decision problems. This resource is freely available[3].

A recursively presented group with unsolvable word problem

Theorem 10.1. There exists a recursively presented group with IWP.

Proof. Form the recursive presentation

$$Q = \langle a, b, c, d \mid b^n a b^{-n} = d^n c d^{-n}\ \forall\ n \in \mathbb{K} \rangle$$

This is an amalgamated product $\overline{\langle a, b | - \rangle} *_\varphi \overline{\langle c, d | - \rangle}$ over subgroups $A := \langle \overline{b^n a b^{-n}}\ \forall\ n \in \mathbb{K} \rangle$ and $B := \langle \overline{d^n c d^{-n}}\ \forall\ n \in \mathbb{K} \rangle$, both isomorphic to F_∞ by (2.19) as $\mathbb{K}$ is infinite. The

[3] *Decision problems for groups-survey and reflections.* Algorithms and classification in combinatorial group theory (Berkeley, CA, 1989), Math. Sci. Res. Inst. Publ., **23**, Springer, New York, 1–59 (1992). See http://www.ms.unimelb.edu.au/ cfm/papers/paperpdfs/msri_survey.all.pdf

isomorphism $\varphi : A \to B$ is then given by extending the map $\overline{b^n a b^{-n}} \mapsto \overline{d^n c d^{-n}}$ $\forall\, n \in \mathbb{K}$. Then, by the reduced form theorem for amalgamated products (5.14), we have that:

$$\overline{b^n a b^{-n}(d^n c d^{-n})^{-1}} = e \text{ in } \overline{Q} \;\Leftrightarrow\; n \in \mathbb{K}$$

So an algorithm solving the word problem for Q will give an algorithm which decides membership in $\mathbb{K}$, which is impossible by (7.9).

A finitely presented group with unsolvable word problem

We now show that there is a finitely presented group with unsolvable word problem. We will actually show a much stronger result: that we can take any modular machine, and from it define a finitely presented group which simulates the action of this machine. A quick application of the halting set $\mathbb{K}$ then gives us a finitely presented group with unsolvable word problem. In the next section we will use this finitely presented group that simulates a modular machine to show the Higman embedding theorem: *every recursively presented group embeds in some finitely presented group.*

Our approach will be to outline the entire construction first, and state (without proof) each of the intermediate results that we need. After this initial construction, we prove all the steps that are not obvious (these are denoted with a * in the construction below).

Construction 10.2. The following is a construction for simulating a modular machine within a finitely presented group:

1. Define the group $K := \mathbb{Z} * (\mathbb{Z} \times \mathbb{Z})$ with presentation $\langle x, y, t \mid [x, y] = e \rangle$.
2. Define, for all $(r, s) \in \mathbb{Z}^2$, the word $t(r, s) := y^s x^r t x^{-r} y^{-s}$.
3. Define the subgroup $T := \langle \{\overline{t(r,s)}\}_{(r,s)\in\mathbb{Z}^2} \rangle \leqslant K$.
4. *Observe that T is free with basis $\{\overline{t(r,s)}\}_{(r,s)\in\mathbb{Z}^2}$.
5. For $M > a \geqslant 0$, $N > b \geqslant 0$, define

$$T_{a,b}^{M,N} := \langle \{\overline{t(\alpha,\beta)} \mid \alpha \equiv a \mod M,\ \beta \equiv b \mod N\} \rangle \leqslant T \leqslant K$$

$$K_{a,b}^{M,N} := \langle \overline{t(a,b)}, \overline{x}^M, \overline{y}^N \rangle \leqslant K$$

6. *Observe that $T_{a,b}^{M,N} \cong T$ via extension of the map $\overline{t(uM + a, vN + b)} \mapsto \overline{t(u, v)}$ $\forall u, v \in \mathbb{Z}$.
7. *Observe that $K_{a,b}^{M,N} \cong K$ via extension of $\overline{t(a,b)} \mapsto \overline{t}$, $\overline{x}^M \mapsto \overline{x}$, $\overline{y}^N \mapsto \overline{y}$.
8. *Observe that $T \cap K_{a,b}^{M,N} = T_{a,b}^{M,N}$ in K.
9. Let $\mathcal{M} = \{(a_i, b_i, c_i, R) \mid i \in I\} \cup \{(a_j, b_j, c_j, L) \mid j \in J\}$ be a modular machine with modulus m, in which $(0, 0)$ is terminal.
10. Define, for each $i \in I$, the map $\phi_i : K_{a_i,b_i}^{m,m} \to K_{c_i,0}^{m^2,1}$ via extension of the map $\overline{t(a_i, b_i)} \mapsto \overline{t(c_i, 0)}$, $\overline{x}^m \mapsto \overline{x}^{m^2}$, $\overline{y}^m \mapsto y$.
11. Define, for each $j \in I$, the map $\varphi_j : K_{a_j,b_j}^{m,m} \to K_{0,c_j}^{1,m^2}$ via extension of the map $\overline{t(a_j, b_j)} \mapsto \overline{t(0, c_j)}$, $\overline{x}^m \mapsto \overline{x}$, $\overline{y}^m \mapsto \overline{y}^{m^2}$.

12. *Observe that ϕ_i and φ_j are isomorphisms for every $i \in I$, $j \in J$.

13. Define the following HNN extension with stable letters $\{r_i\}_{i\in I}$, $\{l_j\}_{j\in J}$.

$$K_{\mathcal{M}} := K*_{\{\phi_i\}_{i\in I},\{\varphi_j\}_{j\in J}}$$

14. *Observe that $K_{\mathcal{M}}$ is finitely presented.

15. Define the subgroup $T' := \langle T,\ \overline{r_i}\ \forall i \in I,\ \overline{l_j}\ \forall j \in J\rangle \leqslant K_{\mathcal{M}}$.

16. *Observe that T is a good subgroup of K with respect to the HNN extension $K_{\mathcal{M}}$, and thus $T = T' \cap K$.

17. Define $T_{\mathcal{M}} := \langle\{\overline{t(\alpha,\beta)} \mid (\alpha,\beta) \in H_0(\mathcal{M})\}\rangle \leqslant T \leqslant K$.

18. Define $T'_{\mathcal{M}} := \langle T_{\mathcal{M}},\ \overline{r_i}\ \forall i \in I,\ \overline{l_j}\ \forall j \in J\rangle \leqslant K_{\mathcal{M}}$.

19. Define $\langle\bar{t}\rangle' := \langle\bar{t},\ \overline{r_i}\ \forall i \in I,\ \overline{l_j}\ \forall j \in J\rangle \leqslant K_{\mathcal{M}}$.

20. *Observe that $T_{\mathcal{M}}$ is a good subgroup of K with respect to the HNN extension $K_{\mathcal{M}}$, and thus $T_{\mathcal{M}} = T'_{\mathcal{M}} \cap K$.

21. *Observe that $T'_{\mathcal{M}} = \langle\bar{t}\rangle'$ in $K_{\mathcal{M}}$.

22. *Observe that $T_{\mathcal{M}} = \langle\bar{t}\rangle' \cap K$.

23. *Observe that $\overline{t(\alpha,\beta)} \in \langle\bar{t}\rangle'$ iff $(\alpha,\beta) \in H_0(\mathcal{M})$.

24. Define the HNN extension $G_{\mathcal{M}}$ with presentation

$$\langle K_{\mathcal{M}}; k \mid khk^{-1} = h\ \forall h \in \langle t\rangle'\rangle$$

25. *Observe that $G_{\mathcal{M}}$ is finitely presented.

26. *Observe that $\overline{kt(\alpha,\beta)k^{-1}} = \overline{t(\alpha,\beta)}$ in $G_{\mathcal{M}}$ iff $(\alpha,\beta) \in H_0(\mathcal{M})$.

Before we proceed, we require the following lemma.

Lemma 10.3. Let S be a free basis for a free group G. Let $P, Q \subseteq S$ be subsets. Then
1. $\langle P\rangle$ is free with basis P.
2. For any $s \in S$, we have $s \in \langle P\rangle$ iff $s \in P$.
3. $\langle P\rangle \cap \langle Q\rangle = \langle P \cap Q\rangle$.

Proof.
1. Apply the normal form theorem for free products (2.15), as done in (2.19).
2. If $s \neq P$ but $s \in \langle P\rangle$ then S is not a free basis, as otherwise we could express s as a product $s_1 \cdots s_k$ of other elements of S and their inverses, and so s and $s_1 \cdots s_k$ would be two ways of writing the same element of G, contradicting the unique normal form for free products (2.15).
3. If $g \in \langle P \cap Q\rangle$ the g can be written as a product of elements which lie both in P and in Q, and so $g \in \langle P\rangle \cap \langle Q\rangle$. Conversely, if $g \in \langle P\rangle \cap \langle Q\rangle$, then g can be written as a product $p_1 \cdots p_k$ of elements of P and their inverses, as well as a product $q_1 \cdots q_l$ of elements of Q and their inverses. But by the normal form theorem for free products (2.15), as both $P, Q \subseteq S$, we must have that these two (freely reduced) words are the same, and so $g \in \langle P \cap Q\rangle$.

We now prove all the nontrivial steps in (10.2), to show that the construction is valid. We will suppress the use of over lines $\overline{w}$ in the proofs, and just use the underlying words as group elements. This is to save on heavy notation; the meaning should be clear in all cases.

Step 4: T is free with basis $\{\overline{t(r,s)}\}_{(r,s)\in\mathbb{Z}^2}$.
By definition, $\{\overline{t(r,s)}\}_{(r,s)\in\mathbb{Z}^2}$ generates T. By applying the normal form theorem for free products (2.15), in the same way that we did in (2.19), we have that these elements *freely* generate T, and thus T is free.

Step 6: $T^{M,N}_{a,b} \cong T$ via extension of the map sending $\overline{t(uM+a,vN+b)} \mapsto \overline{t(u,v)}$ $\forall u,v\in\mathbb{Z}^2$.
By step 4, we have that $\{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}$ is a free basis for T. Thus any subset $S \subseteq \{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}$ is a free basis for the subgroup $\langle S\rangle$ it generates (10.3). In particular, as $\{t(\alpha,\beta) \mid \alpha \equiv a \mod M, \beta \equiv b \mod N\} \subseteq \{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}$, and both sets have the same cardinality (via the map sending $t(uM+a,vN+b)\mapsto t(u,v)$), then they freely generate isomorphic groups with isomorphism given via the map extending $t(uM+a,vN+b)\mapsto t(u,v)$ $\forall u,v\in\mathbb{Z}$; see (2.6) and (10.3).

Step 7: $K^{M,N}_{a,b} \cong K$ via extension of the map $\overline{t(a,b)}\mapsto \bar{t}$, $\overline{x}^M \mapsto \overline{x}$, $\overline{y}^N \mapsto \overline{y}$.
First, observe that $K^{M,N}_{a,b}$ is conjugate to $K^{M,N}_{0,0} = \langle t, x^M, y^N\rangle$, where $t(0,0)=t$ (conjugate by $x^{-a}y^{-b}$; this commutes with both x,y, thus $x^{-a}y^{-b}t(a,b)y^bx^b = t$). Thus the two are isomorphic; see (1.28). But $\langle t,x^M,y^N\rangle$ is generated by $\langle t\rangle$ and $\langle x^M,y^N\rangle$ which lie in different free factors of K. So $\langle t,x^M,y^N\rangle \cong \langle t\rangle * \langle x^M,y^N\rangle \cong \mathbb{Z}*(\mathbb{Z}\times\mathbb{Z})$; see Q5 of example sheet 1.

Step 8: $T\cap K^{M,N}_{a,b} = T^{M,N}_{a,b}$ in K.
For $\supseteq$, note that $t(uM+a,vN+b) = y^{vN}x^{uM}t(a,b)x^{-uM}y^{-vN} \in K^{M,N}_{a,b}$.
For $\subseteq$, note that $x^Mt(\alpha,\beta) = t(\alpha+M,\beta)x^M$ and $y^Nt(\alpha,\beta)=t(\alpha,\beta+N)y^N$, hence any element of $K^{M,N}_{a,b}$ is of the form $gx^{uM}y^{vN}$ where $g\in T^{M,N}_{a,b}$ and $u,v\in\mathbb{Z}$. If this element is in T, then $u=v=0$ (hence it is in $T^{M,N}_{a,b}$), by the following argument: If $gx^{uM}y^{vN}\in T$ then $x^{uM}y^{vN}\in T$ as $g\in T^{M,N}_{a,b} \leqslant T$. But as $x^{uM}y^{vN}\in T$ which is free with basis $\{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}$ (by step 4), we have that $x^{uM}y^{vN} = t(c_1,d_1)^{\varepsilon_1}\dots t(c_l,d_l)^{\varepsilon_l}$ for $\varepsilon_i\in\{\pm1\}$, and thus $e = y^{-vN}x^{-uM}t(c_1,d_1)^{\varepsilon_1}\dots t(c_l,d_l)^{\varepsilon_l}$. But by the normal form theorem for free products (2.15), as K is a free product, then the word $y^{-vN}x^{-uM}t(c_1,d_1)^{\varepsilon_1}\dots t(c_l,d_l)^{\varepsilon_l}$ can only be trivial if $t(c_1,d_1)^{\varepsilon_1}\dots t(c_l,d_l)^{\varepsilon_l}$ is trivial (or there would be a non-cancelling occurrence of t). Thus $e = y^{-vN}x^{-uM}$, and so $v=u=0$ as x,y generate the $\mathbb{Z}\times\mathbb{Z}$ free factor of K.

Step 12: ϕ_i and φ_j are isomorphisms for every $i\in I$, $j\in J$.
This follows immediately from step 7.

Step 14: $K_{\mathcal{M}}$ is finitely presented.
Note that K is finitely presented, and $K_{\mathcal{M}}$ is a finite tower of HNN extensions of K, each over finitely generated subgroups. It then follows from the comment after (5.4) that $K_{\mathcal{M}}$ is finitely presented.

Step 16: T is a good subgroup of K with respect to the HNN extension $K_{\mathcal{M}}$, and thus $T = T'\cap K$.
To verify T is good, we need to show $\phi_i(T\cap K^{m,m}_{a_i,b_i}) = T\cap K^{m^2,1}_{c_i,0}$ for all $i\in I$, and $\varphi_j(T\cap K^{m,m}_{a_j,b_j}) = T\cap K^{1,m^2}_{0,c_j}$ for all $j\in J$. From step 8, we have $T\cap K^{m,m}_{a_i,b_i} = T^{m,m}_{a_i,b_i}$ and $T\cap K^{m^2,1}_{c_i,0} = T^{m^2,1}_{c_i,0}$. But $\phi_i(t(um+a_i,vm+b_i)) = t(um^2+c_i,v)$, so $\phi_i(T^{m,m}_{a_i,b_i}) = T^{m^2,1}_{c_i,0}$ as ϕ_i is an extension of a map sending a free basis of $T^{m,m}_{a_i,b_i}$ to a free basis of $T^{m^2,1}_{c_i,0}$. So we're done for ϕ_i, and an identical argument works for φ_j. So T is good, and thus by (5.20) $T = T'\cap K$.

Step 20: $T_{\mathcal{M}}$ is a good subgroup of K with respect to the HNN extension $K_{\mathcal{M}}$, and thus $T_{\mathcal{M}} = T'_{\mathcal{M}} \cap K$.
To verify $T_{\mathcal{M}}$ is good, we need to show $\phi_i(T_{\mathcal{M}} \cap K^{m,m}_{a_i,b_i}) = T_{\mathcal{M}} \cap K^{m^2,1}_{c_i,0}$ for all $i \in I$, and $\varphi_j(T_{\mathcal{M}} \cap K^{m,m}_{a_j,b_j}) = T_{\mathcal{M}} \cap K^{1,m^2}_{0,c_j}$ for all $j \in J$. Note that if $\phi_i(t(\alpha,\beta)) = t(\alpha_1,\beta_1)$ or $\varphi_j(t(\alpha,\beta)) = t(\alpha_1,\beta_1)$ then $(\alpha,\beta) \underset{\mathcal{M}}{\longrightarrow} (\alpha_1,\beta_1)$ by definition of ϕ_i, φ_j. Also, $t(\alpha,\beta) \in T_{\mathcal{M}}$ iff $(\alpha,\beta) \in H_0(\mathcal{M})$ by (10.3) since T is free on $\{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}$ (step 4) and $T_{\mathcal{M}}$ is free on a subset of this set. Thus $t(\alpha,\beta) \in T_{\mathcal{M}} \Leftrightarrow (\alpha,\beta) \in H_0(\mathcal{M}) \Leftrightarrow (\alpha_1,\beta_1) \in H_0(\mathcal{M}) \Leftrightarrow t(\alpha_1,\beta_1) \in T_{\mathcal{M}}$. Now observe that $T_{\mathcal{M}} \cap K^{m,m}_{a_i,b_i} = T_{\mathcal{M}} \cap T \cap K^{m,m}_{a_i,b_i}$ as $T_{\mathcal{M}} \leqslant T$, and thus $T_{\mathcal{M}} \cap T \cap K^{m,m}_{a_i,b_i} = T_{\mathcal{M}} \cap T^{m,m}_{a_i,b_i}$ (step 8). By step 4 and (10.3), we see that $T_{\mathcal{M}} \cap T^{m,m}_{a_i,b_i} = \langle \{t(\alpha,\beta) \in T_{\mathcal{M}} \mid \alpha \equiv a_i \mod m,\ \beta \equiv b_i \mod m\}\rangle$ (free on the intersection of their free bases, both of which are subsets of a free basis for T). A similar argument gives that $T_{\mathcal{M}} \cap T^{m^2,1}_{c_i,0} = \langle \{t(\alpha,\beta) \in T_{\mathcal{M}} \mid \alpha \equiv c_i \mod m^2,\ \beta \equiv 0 \mod 1\}\rangle$. By the first part of the argument, if $\phi_i(t(\alpha,\beta)) = t(\alpha',\beta')$ then $t(\alpha,\beta) \in T_{\mathcal{M}} \Leftrightarrow t(\alpha',\beta') \in T_{\mathcal{M}}$. From this it follows that $\phi_i(T_{\mathcal{M}} \cap T^{m,m}_{a_i,b_i}) = T_{\mathcal{M}} \cap T^{m^2,1}_{c_i,0}$ (If $t(\alpha,\beta) \in T_{\mathcal{M}} \cap T^{m,m}_{a_i,b_i}$ then $\phi_i(t(\alpha,\beta)) \in T_{\mathcal{M}}$ as $t(\alpha,\beta) \in T_{\mathcal{M}}$, and $\phi_i(t(\alpha,\beta)) \in T^{m^2,1}_{c_i,0}$ as $t(\alpha,\beta) \in T^{m,m}_{a_i,b_i}$. Thus $\phi_i(t(\alpha,\beta)) \in T_{\mathcal{M}} \cap T^{m^2,1}_{c_i,0}$, and so $\phi_i(T_{\mathcal{M}} \cap T^{m,m}_{a_i,b_i}) \subseteq T_{\mathcal{M}} \cap T^{m^2,1}_{c_i,0}$. Conversely, if $t(\alpha,\beta) \in T_{\mathcal{M}} \cap T^{m^2,1}_{c_i,0}$ then $t(\alpha,\beta) \in T^{m^2,1}_{c_i,0}$, so by step 16 there exists $t(\alpha',\beta') \in T^{m,m}_{a_i,b_i}$ such that $\phi_i(t(\alpha',\beta')) = t(\alpha,\beta)$. But then $t(\alpha',\beta') \in T_{\mathcal{M}}$ as $\phi_i(t(\alpha',\beta')) \in T_{\mathcal{M}}$, so $t(\alpha',\beta') \in T_{\mathcal{M}} \cap T^{m,m}_{a_i,b_i}$, and thus $\phi_i(T_{\mathcal{M}} \cap T^{m,m}_{a_i,b_i}) \supseteq T_{\mathcal{M}} \cap T^{m^2,1}_{c_i,0}$). So we're done for ϕ_i, and an identical argument works for φ_j. Thus $T_{\mathcal{M}}$ is good, and so by (5.20), $T_{\mathcal{M}} = T'_{\mathcal{M}} \cap K$.

Step 21: $T'_{\mathcal{M}} = \langle \bar{t} \rangle'$ in $K_{\mathcal{M}}$.
Clearly $T'_{\mathcal{M}} \supseteq \langle t \rangle'$ in $K_{\mathcal{M}}$, as $t = t(0,0) \in T_{\mathcal{M}} \leqslant T'_{\mathcal{M}}$ and all r_i, l_j lie in $T'_{\mathcal{M}}$.
To prove the $\subseteq$ direction, it suffices to show that $t(\alpha,\beta) \in \langle t \rangle'$ for all $(\alpha,\beta) \in H_0(\mathcal{M})$. We do this by induction on the length of the computation which takes (α,β) to $(0,0)$. Clearly, for $(\alpha,\beta) = (0,0)$ (computation of length 0) we have $t(0,0) = t \in \langle t \rangle'$.
Now assume that $(\alpha,\beta) \underset{\mathcal{M}}{\longrightarrow} (\alpha_1,\beta_1)$ via the modular machine computation (a_i, b_i, c_i, R), and so $\alpha = um + a_i$, $\beta = vm + b_i$, and $\alpha_1 = um^2 + c_i$, $\beta_1 = v$. Then we have

$$\begin{aligned} t(\alpha,\beta) &= y^{\beta} x^{\alpha} t x^{-\alpha} y^{-\beta} \\ &= y^{vm+b_i} x^{um+a_i} t x^{-um-a_i} y^{-vm-b_i} \\ &= y^{vm} x^{um} y^{b_i} x^{a_i} t x^{-a_i} y^{-b_i} x^{-um} y^{-vm} \\ &= y^{vm} x^{um} t(a_i,b_i) x^{-um} y^{-vm} \end{aligned}$$

So we thus have

$$\begin{aligned} r_i t(\alpha,\beta) r_i^{-1} &= \phi_i(t(\alpha,\beta)) \\ &= \phi_i(y^{vm} x^{um} t(a_i,b_i) x^{-um} y^{-vm}) \\ &= y^{v} x^{um^2} t(c_i,0) x^{-um^2} y^{-v} \\ &= y^{v} x^{um^2+c_i} t x^{-um^2-c_i} y^{-v} \\ &= t(um^2 + c_i, v) \\ &= t(\alpha_1,\beta_1) \end{aligned}$$

So if $(\alpha,\beta) \in H_0(\mathcal{M})$, then $(\alpha_1,\beta_1) \in H_0(\mathcal{M})$ by a shorter computation, and so by induction $t(\alpha_1,\beta_1) \in \langle t \rangle'$. Thus $t(\alpha,\beta) = r_i^{-1} t(\alpha_1,\beta_1) r_i \in \langle t \rangle'$ as both r_i and $t(\alpha_1,\beta_1)$ are in $\langle t \rangle'$.
If $(\alpha,\beta) \underset{\mathcal{M}}{\longrightarrow} (\alpha_1,\beta_1)$ via the modular machine computation (a_j, b_j, c_j, L), then the proof is practically identical. So we are done.

Step 22: $T_{\mathcal{M}} = \langle \bar{t} \rangle' \cap K$.
This is immediate from steps 20 and 21.

Step 23: $\overline{t(\alpha,\beta)} \in \langle \bar{t} \rangle'$ iff $(\alpha,\beta) \in H_0(\mathcal{M})$.
If $(\alpha,\beta) \in H_0(\mathcal{M})$ then $t(\alpha,\beta) \in T_{\mathcal{M}} = \langle t \rangle' \cap K$ by step 22. Conversely, it is clear that $t(\alpha,\beta) \in K$, and thus if $t(\alpha,\beta) \in \langle t \rangle'$ then $t(\alpha,\beta) \in \langle t \rangle' \cap K = T_{\mathcal{M}}$ (step 22) and thus $(\alpha,\beta) \in H_0(\mathcal{M})$ as $T_{\mathcal{M}}$ is free on $\{t(\alpha,\beta) \mid (\alpha,\beta) \in H_0(\mathcal{M})\}$ which is a subset of the free basis $\{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}$ for T; see 10.3.

Step 25: $G_{\mathcal{M}}$ is finitely presented.
Note that $K_{\mathcal{M}}$ is finitely presented by step 14, and $G_{\mathcal{M}}$ is an HNN extension of $K_{\mathcal{M}}$ over a finitely generated subgroup $\langle t \rangle'$. It then follows from the comment after (5.4) that $G_{\mathcal{M}}$ is finitely presented.

Step 26: $\overline{k^{-1}t(\alpha,\beta)k} = \overline{t(\alpha,\beta)}$ in $G_{\mathcal{M}}$ iff $(\alpha,\beta) \in H_0(\mathcal{M})$.
By Britton's lemma (5.18), for $g \in K_{\mathcal{M}}$ we have $kgk^{-1} = g \Leftrightarrow g \in \langle t \rangle'$. In particular, $kt(\alpha,\beta)k^{-1} = t(\alpha,\beta) \Leftrightarrow t(\alpha,\beta) \in \langle t \rangle' \Leftrightarrow (\alpha,\beta) \in H_0(\mathcal{M})$ (using step 23).

This concludes the proof that all the steps in the construction are valid. A useful consequence of the construction (10.2) is that we can simulate $H_0(\mathcal{M})$ of *any* modular machine, within a finitely group; step 26 of (10.2) (via the subgroup membership problem: $\overline{t(\alpha,\beta)} \in \langle \bar{t} \rangle'$ iff $(\alpha,\beta) \in H_0(\mathcal{M})$). We use this when proving Higman's embedding theorem.

Theorem 10.4 (Boone-Britton-Novikov). Determining membership in $H_0(\mathcal{M})$ can be reduced to the word problem for $K_{\mathcal{M}}$. *A fortiori*, if $\mathcal{M}$ is taken with $H_0(\mathcal{M})$ nonrecursive via (7.9) and (9.3), then $K_{\mathcal{M}}$ is a finitely presented group with IWP.

There are some slightly more straightforward examples of groups with IWP, for example Borisov's construction[4] of one with 5 generators and 12 relators. Our construction above would yield a very complicated finite presentation.

11. The Higman embedding theorem

The following standard result is immediate from (8.4), and we leave the proof as an exercise on example sheet 3.

Theorem 11.1. Let G be a finitely presented group, and $H \leqslant G$ a finitely generated subgroup. Then H is recursively presented.

Remarkably, recursively presented groups are *precisely* the set of finitely generated subgroups of finitely presented groups. Graham Higman proved this as his famous embedding theorem of 1961:

Theorem 11.2 (Higman 1961). Each recursively presented group embeds into some finitely presented group.

For the moment, we will show that for *each* recursively presented group H there is *some* corresponding finitely presented group G for which $H \hookrightarrow G$. Later, we will see that there is *one* finitely presented group into which *all* recursively presented groups embed[5].

Our approach will be to outline the entire construction first, and state (without proof) each of the intermediate results that we need. After this initial construction, we prove all the steps that are not obvious (these are denoted with a * in the construction below).

[4] V. Borisov, *Simple examples of groups with unsolvable word problem*, Math. Zametki **6**, 521–532 (1969); English transl., Math. Notes **6**, 768–775 (1969).
[5] If you don't already have a headache, then this will give you one.

In a conventional flip[6], we take our Turing machines to start on the *rightmost* letter of an input word and read right to left, rather than the leftmost and read left to right. Functionally, these are the same. Furthermore, as per the paragraph preceding (9.3), we take q_0 as our initial state (as well as our halting state). This is so we can use the result of (9.3).

Construction 11.3. The following gives a construction for embedding a recursively presented group into a finitely presented group.

1. Let $C = \langle c_1, \ldots, c_n \mid S \rangle$ be a recursive presentation of a group, where S is the halting set of the Turing machine T.

2. Re-write every word in $\{c_1, \ldots, c_n\}^*$ as being in the free monoid on $\{c_1, \ldots, c_{2n}\}$, where $c_{n+i} = c_i^{-1}$.

3. *Observe that we can uniformly construct a *new* Turing machine T' whose halting set S' is *all* the trivial words in the group, when written in the free monoid on $\{c_1, \ldots, c_{2n}\}$. We make T' have an extra symbol s_0, different to $\{c_1, \ldots, c_{2n}\}$.

4. Take the corresponding modular machine $\mathcal{M}$ for T' (9.3), with modulus m. We make sure to assign s_0 to 0, and c_i to i for all $1 \leqslant i \leqslant 2n$.

5. *Observe that $m > 2n$, by construction of the modular machine $\mathcal{M}$.

6. To each word $w = c_{i_k} c_{i_{k-1}} \cdots c_{i_1}$ we associate an m-ary representation $\alpha = \Sigma_{j=1}^{k} i_j m^j$, as per (9.3).

7. Define $I := \{\alpha \in \mathbb{N} \mid \alpha \text{ represents a word}\}$. That is, $\alpha = \Sigma_{j=1}^{k} \beta_j m^j$, $1 \leqslant \beta_j \leqslant 2n$.

8. For $\alpha \in I$, define $w_\alpha(c)$ to be the word formed from α.

9. *Observe that $w_{\alpha m+i}(c) = w_\alpha(c) c_i$ for all $1 \leqslant i \leqslant 2n$, $\alpha \in I$.

10. For $\alpha \in I$, write $w_\alpha(b)$, $w_\alpha(bc)$ for the words obtained from $w_\alpha(c)$ by replacing c_i with b_i and $b_i c_i$ respectively (where the b_i's are a new set of symbols).

11. *Observe that, for all $\alpha \in I$, we have that $w_\alpha(c) \in S'$ iff $(\alpha, 0) \in H_0(M)$.

12. Define the group $K_{\mathcal{M}}$ from step 13 in (10.2).

13. Define $U = \{t, r_i \forall i, l_j \forall j\}$ (t, along with all the stable letters of $K_{\mathcal{M}}$).

14. Define $t_\alpha := t(\alpha, 0)$.

15. Define the free product

$$H_1 := K_{\mathcal{M}} * (\overline{C} \times \overline{\langle b_1, \ldots, b_n \mid - \rangle}) * \overline{\langle d | - \rangle}.$$

16. Let $b_{n+i} = b_i^{-1}$ for $1 \leqslant i \leqslant n$.

17. *Observe that $\{\overline{t_\alpha} | \alpha \in I\}$ and $\{\overline{t_\alpha w_\alpha(b) d} | \alpha \in I\}$ are each a free basis for the subgroup they generate in H_1, and thus generate isomorphic subgroups with isomorphism ψ given via extension of the map $\overline{t_\alpha} \mapsto \overline{t_\alpha w_\alpha(b) d}$ $\forall \alpha \in I$.

18. With ψ as in step 17, define the following HNN extension with stable letter p:

$$H_2 := H_1 *_\psi$$

[6]Because the original source of this construction, Cohen's text, presents it as so.

19. Define the subgroup

$$A := \langle \overline{t}, \overline{x}, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n), \overline{p} \rangle \leqslant H_2$$

20. *Observe that A is an HNN extension of the free group F with free basis $\{\overline{t}, \overline{x}, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n)\}$, with stable letter p sending $pt_\alpha p^{-1} = t_\alpha w_\alpha(b) d \ \forall \alpha \in I$.

21. For $1 \leqslant i \leqslant 2n$, define the subgroup

$$A_i := \langle \overline{t_i}, \overline{x}^m, \overline{b_i d}, \overline{b_j} (1 \leqslant j \leqslant n), \overline{p} \rangle \leqslant H_2$$

22. *Observe that $\langle \overline{t_i}, \overline{x}^m \rangle \cap \langle \overline{t_\alpha} : \alpha \in I \rangle = \langle \overline{t_\beta} : \beta \in I \text{ with } \beta \equiv i \mod m \rangle$.

23. *Observe that $\langle \overline{t_i}, \overline{x}^m, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n) \rangle \cap \langle \overline{t_\alpha} : \alpha \in I \rangle = \langle \overline{t_\beta} : \beta \in I \text{ with } \beta \equiv i \mod m \rangle$.

24. *Observe that $\langle \overline{t_i}, \overline{x}^m, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n) \rangle \cap \langle \overline{t_\alpha w_\alpha(b) d} : \alpha \in I \rangle = \langle \overline{t_\beta w_\beta(b) d} : \beta \in I \text{ with } \beta \equiv i \mod m \rangle$.

25. *Observe that, for all $1 \leqslant i \leqslant 2n$, A_i is an HNN extension of the free group on basis $\{\overline{t_i}, \overline{x}^m, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n)\}$, with stable letter p sending $pt_\alpha p^{-1} = t_\alpha w_\alpha(b) d \ \forall \alpha \in I$ with $\alpha \equiv i \mod m$.

26. *Observe that $A \cong A_i$ for all $1 \leqslant i \leqslant 2n$, via ψ_i obtained by extension of the map $\overline{t} \mapsto \overline{t_i}$, $\overline{x} \mapsto \overline{x}^m$, $\overline{d} \mapsto \overline{b_i d}$, $\overline{b_j} \mapsto \overline{b_j} \ \forall 1 \leqslant j \leqslant n$, $\overline{p} \mapsto \overline{p}$.

27. Define the subgroup

$$A_+ := \langle \overline{U}, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n), \overline{p} \rangle \leqslant H_2$$

28. Define the subgroup

$$A_- := \langle \overline{U}, \overline{d}, \overline{b_j c_j} (1 \leqslant j \leqslant n), \overline{p} \rangle \leqslant H_2$$

29. *Observe that A_+ is an HNN extension with base group $\langle \overline{U}, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n) \rangle$ (the free product of $\langle \overline{U} \rangle$ and the free group with basis $\{\overline{d}, \overline{b_j} (1 \leqslant j \leqslant n)\}$), with stable letter p and HNN relations $pt_\alpha p^{-1} = t_\alpha w_\alpha(b) d \ \forall \alpha \in I$ with $\alpha \equiv i \mod m$.

30. *Observe that $A_+ \cong A_-$, via ψ_+ which is obtained by extension of the map $\overline{u} \mapsto \overline{u} \ \forall \overline{u} \in \overline{U}$, $\overline{d} \mapsto \overline{d}$, $\overline{b_j} \mapsto \overline{b_j c_j} \ \forall 1 \leqslant j \leqslant n$, $\overline{p} \mapsto \overline{p}$.

31. With the isomorphisms defined in steps 26 and 30, define the following HNN extension with stable letters $a_1, \ldots, a_{2n}, k$:

$$H_3 := H_2 *_{\psi_1, \ldots, \psi_{2n}, \psi_+}$$

32. *Observe H_3 is finitely presented, and $\overline{C} \hookrightarrow H_3$.

We now prove all the nontrivial steps of Construction 11.2.

Step 3: We can uniformly construct a *new* Turing machine T' whose halting set is *all* the trivial words in the group, when written in the monoid on $\{c_1, \ldots, c_{2n}\}$.
First, using (8.4), as C is a recursive presentation, we can (uniformly) construct a new Turing machine T'' which enumerates *all* the trivial words in C. Note that T'' has input alphabet $\{c_1, \ldots, c_n, c_1^{-1}, \ldots, c_n^{-1}\}$ (as formal symbols). Now simply substitute the symbols $\{c_1^{-1}, \ldots, c_n^{-1}\}$ for $\{c_{n+1}, \ldots, c_{2n}\}$ in T'', and add a new symbol s_0 to be used as 'blank'. Call

this new machine T'.

Step 5: $m > 2n$, by construction of the modular machine $\mathcal{M}$.
This follows immediately from (9.2) and (9.3), where we assign s_0 to 0, and c_i to i for all $1 \leqslant i \leqslant 2n$.

Step 9: $w_{\alpha m+i}(c) = w_\alpha(c)c_i$ for all $1 \leqslant i \leqslant 2n$, $\alpha \in I$.
This becomes obvious when we write out the words in full.

Step 11: For all $\alpha \in I$, we have that $w_\alpha(c) \in S$ iff $(\alpha, 0) \in H_0(M)$.
Note that $w_\alpha(c) \in S'$ iff T' halts on input of a tape with $w_\alpha(c)$ written on it (T' starts reading from the rightmost letter of $w_\alpha(c)$, in start state q_0 by our strange convention from (9.3)). The left associate of the instantaneous description $w_\alpha(c)q_0$ is $(\alpha, 0)$. So by the construction of $\mathcal{M}$ in (9.2) and (9.3), $w_\alpha(c) \in S'$ iff $\mathcal{M}$, on input $(\alpha, 0)$, eventually reaches $(0, 0)$; that is, iff $(\alpha, 0) \in H_0(M)$.

Step 17: $\{\overline{t_\alpha} | \alpha \in I\}$ and $\{\overline{t_\alpha w_\alpha(b)d} | \alpha \in I\}$ are each a free basis for the subgroup they generate in H_1, and thus generate isomorphic subgroups with isomorphism ψ given via extension of the map $\overline{t_\alpha} \mapsto \overline{t_\alpha w_\alpha(b)d}$ $\forall \alpha \in I$..
We showed in step 4 of (10.2) that $\{t_\alpha | \alpha \in I\}$ is a free basis. By the normal form theorem for free products (2.15), it follows that $\{t_\alpha w_\alpha(b)d | \alpha \in I\}$ is also a free basis, as $w_\alpha(b)d$ is from a different free factor to t_α in H_1. The two sets generate isomorphic (free) groups by (2.6), as they both have the same cardinality $|I|$.

Step 20: A is an HNN extension of the free group F with basis $\{\overline{t}, \overline{x}, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n)\}$, with stable letter p sending $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d$ $\forall \alpha \in I$.
Clearly, F is free on the given basis, as it is the union of free basis elements in distinct factor groups of the free product H_1. To use the good subgroup theorem (5.20), we need to show that $\psi(\langle t_\alpha : \alpha \in I\rangle \cap F) = \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle \cap F$. But recall that $t_\alpha := t(\alpha, 0) = x^\alpha t x^{-\alpha}$. Thus it is clear that $\langle t_\alpha : \alpha \in I\rangle \subseteq F$, and similarly $\langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle \subseteq F$. Hence $\langle t_\alpha : \alpha \in I\rangle \cap F = \langle t_\alpha : \alpha \in I\rangle$ and $\langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle \cap F = \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle$, and so $\psi(\langle t_\alpha : \alpha \in I\rangle \cap F) = \psi(\langle t_\alpha : \alpha \in I\rangle) = \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle = \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle \cap F$. The rest now follows by the good subgroup theorem (5.20).

Step 22: $\langle \overline{t_i}, \overline{x}^m\rangle \cap \langle \overline{t_\alpha} : \alpha \in I\rangle = \langle \overline{t_\beta} : \beta \in I \text{ with } \beta \equiv i \mod m\rangle$.
From step 8 of (10.2) we have that $T \cap K^{M,N}_{a,b} = T^{M,N}_{a,b}$ in K. Written out, this is $\langle \{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}\rangle \cap \langle t(a,b), x^M, y^N\rangle = \langle \{t(\alpha,\beta) \mid \alpha \equiv a \mod M,\ \beta \equiv b \mod N\}\rangle$. In particular, we have $\langle \{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}\rangle \cap \langle t(i,0), x^m, y^0\rangle = \langle \{t(\alpha,\beta) \mid \alpha \equiv i \mod m,\ \beta \equiv 0 \mod 0\}\rangle = \langle \{t(\alpha,0) \mid \alpha \equiv i \mod m\}\rangle = \langle \{t_\alpha \mid \alpha \equiv i \mod m\}\rangle$. So now we have shown that $\langle t_i, x^m\rangle \cap \langle \{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}\rangle = \langle t_\alpha : \alpha \equiv i \mod m\rangle$. Intersecting both sides with $\langle t_\alpha : \alpha \in I\rangle$ gives $\langle t_i, x^m\rangle \cap \langle \{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle t_\alpha : \alpha \equiv i \mod m\rangle \cap \langle t_\alpha : \alpha \in I\rangle$. Observing that $\{t_\alpha : \alpha \in I\}$ is a subset of a free basis for T, we have that $\langle \{t(r,s)\}_{(r,s)\in\mathbb{Z}^2}\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle t_\alpha : \alpha \in I\rangle$ by (10.3), and $\langle t_\beta : \beta \equiv i \mod m\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle t_\alpha : \alpha \in I \text{ with } \alpha \equiv i \mod m\rangle$ by (10.3).

Step 23: $\langle \overline{t_i}, \overline{x}^m, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n)\rangle \cap \langle \overline{t_\alpha} : \alpha \in I\rangle = \langle \overline{t_\beta} : \beta \in I \text{ with } \beta \equiv i \mod m\rangle$.
This follows since $\langle t_\alpha : \alpha \in I\rangle$ is contained in the free factor $K_\mathcal{M}$ of H_1, and so $\langle t_i, x^m, d, b_j (1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle t_i, x^m\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle t_\beta : \beta \in I \text{ with } \beta \equiv i \mod m\rangle$ (the final part coming from step 22).

Step 24: $\langle \overline{t_i}, \overline{x}^m, \overline{d}, \overline{b_j} (1 \leqslant j \leqslant n)\rangle \cap \langle \overline{t_\alpha w_\alpha(b)d} : \alpha \in I\rangle = \langle \overline{t_\beta w_\beta(b)d} : \beta \in I \text{ with } \beta \equiv i \mod m\rangle$.
This is similar to step 23. It is immediate that the containment $\supseteq$ holds. To show $\subseteq$, observe that $\langle t_i, x^m, d, b_j (1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle$ is contained in H_1. Now project onto the $K_\mathcal{M}$ factor of H_1; call this map $f : H_1 \twoheadrightarrow K_\mathcal{M}$. But now observe that $f(\langle t_i, x^m, d, b_j (1 \leqslant$

$j \leqslant n)\rangle \cap \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle) = \langle t_i, x^m\rangle \cap \langle t_\alpha : \alpha \in I\rangle$, which by step 23 is equal to $\langle t_\beta : \beta \in I \text{ with } \beta \equiv i \mod m\rangle$. Thus, by 10.3, as the t_α's are a subset of a free basis of T, the only values of α we can have are $\alpha \equiv i \mod m$. So $\langle t_i, x^m, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle = \langle t_i, x^m, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha w_\alpha(b)d : \alpha \in I \text{ with } \alpha \equiv i \mod m\rangle$. But it is then immediate that $\langle t_\alpha w_\alpha(b)d : \alpha \in I \text{ with } \alpha \equiv i \mod m\rangle \subseteq \langle t_i, x^m, d, b_j(1 \leqslant j \leqslant n)\rangle$, and so this intersection is simply $\langle t_\alpha w_\alpha(b)d : \alpha \in I \text{ with } \alpha \equiv i \mod m\rangle$.

Step 25: A_i is an HNN extension of the free group on basis $\{\overline{t}, \overline{x}^m, \overline{d}, \overline{b_j}(1 \leqslant j \leqslant n)\}$, with stable letter p sending $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d \ \forall \alpha \in I$ with $\alpha \equiv i \mod m$.
We need to show, for each i, that $\psi(\langle t_i, x^m, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha : \alpha \in I\rangle) = \langle t_i, x^m, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle$. By steps 23 and 24, this equates to showing that $\psi(\langle t_\beta : \beta \in I \text{ with } \beta \equiv i \mod m\rangle) = \langle t_\beta w_\alpha(b)d : \beta \in I \text{ with } \beta \equiv i \mod m\rangle$, which is immediate from the definition of ψ.

Step 26: $A \cong A_i$ for all $1 \leqslant i \leqslant 2n$, via the map ψ_i which is an extension of the map sending $\overline{t} \mapsto \overline{t}$, $\overline{x} \mapsto \overline{x}^m$, $\overline{d} \mapsto \overline{b_i d}$, $\overline{b_j} \mapsto \overline{b_j}$ $\forall 1 \leqslant j \leqslant n$, $\overline{p} \mapsto \overline{p}$.
Clearly ψ_i is surjective, as it is an extension of a map from a generating set to a generating set (of the same cardinality). To see that ψ_i is a homomorphism, we need to verify that it preserves relations. The only relations of A are $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d$ for all $\alpha \in I$. But

$$\begin{aligned}
\psi_i(pt_\alpha p^{-1}) &= p\psi_i(x^\alpha t x^{-\alpha})p^{-1} \\
&= px^{\alpha m}x^i t x^{-i}x^{-\alpha m}p^{-1} \\
&= pt_{\alpha m+i}p^{-1} \\
&= t_{\alpha m+i}w_{\alpha m+i}(b)d \\
&= t_{\alpha m+i}w_{\alpha m}(b)b_i d \ \text{ (by step 9)}
\end{aligned}$$

and similarly $\psi_i(t_\alpha w_\alpha(b)d) = \psi_i(x^\alpha t x^{-\alpha})w_\alpha(b)b_i d = t_{\alpha m+i}w_\alpha(b)b_i d$.
To see that ψ_i is injective, take a nontrivial element $g \in A$ with $\psi_i(g) = e$. Since A is an HNN extension, then g has a reduced form; a freely reduced word w on $\{t, x, d, b_j(1 \leqslant j \leqslant n), p\}$ representing g which contains no p-pinch. But now consider the image word $\psi_i(w)$. If w contains no occurrence of p, then neither does $\psi_i(w)$; as w is freely reduced then so is $\psi_i(w)$ which is a contradiction as $\psi_i(w)$ is trivial in the base group $\langle\{t_i, x^m, d, b_j(1 \leqslant j \leqslant n)\}\rangle$ which is free on the given generators. If instead w contains an occurrence of p, then so does $\psi_i(w)$, which must thus contain a p-pinch as it is also freely reduced. But then the p-pinch in $\psi_i(w)$ can be 'pulled back' to a p-pinch in w in A as follows: Take w and evaluate ψ_i on the occurrences of $p^\pm$ only. So we are considering a word $\psi_i(v_1)p^{\varepsilon_1}\psi_i(v_2)\dots p^{\varepsilon_l}\psi_i(v_{l+1})$. So the pinch is of the form (without loss of generality) $p\psi_i(v_j)p^{-1}$ with $\psi_i(v_j) \in \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle$. Thus $v_j \in \langle t_\alpha : \alpha \in I\rangle$ (as we can consider p-pinches in terms of the HNN extension H_2). Thus $pv_j p^{-1}$ is a subword of w, and is a p-pinch in A; a contradiction.

Step 29: A_+ is an HNN extension with base group $\langle \overline{U}, \overline{d}, \overline{b_j}(1 \leqslant j \leqslant n)\rangle$ (the free product of $\langle\overline{U}\rangle$ and the free group with basis $\{\overline{d}, \overline{b_j}(1 \leqslant j \leqslant n)\}$), with stable letter p.
First, observe that $\langle t\rangle' := \langle U\rangle$, from step 19 of (10.2). So, using step 22 of (10.2), we have that $\langle U\rangle \cap K = \langle t\rangle' \cap K = T_{\mathcal{M}} = \langle t(\alpha, \beta) : (\alpha, \beta) \in H_0(\mathcal{M})\rangle$. As $\langle t_\alpha : \alpha \in I\rangle \leqslant K$, it follows that $\langle U\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle U\rangle \cap K \cap \langle t_\alpha : \alpha \in I\rangle = \langle t(\alpha, \beta) : (\alpha, \beta) \in H_0(\mathcal{M})\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle t_\alpha : (\alpha, 0) \in H_0(\mathcal{M})\rangle$ (we are dealing with subsets of free bases, which are again free bases by 10.3). We then apply this to obtain the following 2 facts:
1. By a very similar argument to that used in step 23, we have that $\langle U, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha : \alpha \in I\rangle = \langle t_\alpha : (\alpha, 0) \in H_0(\mathcal{M})\rangle$.
2. By a very similar argument to that used in step 24, we have that $\langle U, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle = \langle t_\alpha w_\alpha(b)d : (\alpha, 0) \in H_0(\mathcal{M})\rangle$.
Thus, by the definition of ψ from step 17, we have that $\psi(\langle U, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha : \alpha \in I\rangle) = \psi(\langle t_\alpha : (\alpha, 0) \in H_0(\mathcal{M})\rangle) = \langle t_\alpha w_\alpha(b)d : (\alpha, 0) \in H_0(\mathcal{M})\rangle = \langle U, d, b_j(1 \leqslant j \leqslant n)\rangle \cap \langle t_\alpha w_\alpha(b)d : \alpha \in I\rangle$. So $\langle U, d, b_j(1 \leqslant j \leqslant n)\rangle = A_+$ is a good subgroup of H_2, and so by the

good subgroup theorem (5.20), A_+ is an HNN extension with base group $\langle U, d, b_j (1 \leqslant j \leqslant n)\rangle$ (the free product of $\langle U\rangle$ and the free group with basis $\{d, b_j (1 \leqslant j \leqslant n)\}$), with stable letter p, and with relations $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d \ \forall \alpha \in I$ with $(\alpha, 0) \in H_0(\mathcal{M})$.

Step 30: $A_+ \cong A_-$ via ψ_+, which is obtained by extension of the map $\overline{u} \mapsto \overline{u} \ \forall \overline{u} \in \overline{U}$, $\overline{d} \mapsto \overline{d}$, $\overline{b_j} \mapsto \overline{b_j c_j} \ \forall 1 \leqslant j \leqslant n$, $\overline{p} \mapsto \overline{p}$.
First, take the map $f : H_2 \to H_2$ where f sends each c_j to e, and maps each other generator of H_2 to itself. Then $f(A_-) = A_+$, and so this induces a homomorphism $\psi_- : A_- \to A_+$ which is actually a surjection since f (and thus ψ_-) send the given generators of A_- to the given generators of A_+ ($U \mapsto U$, $d \mapsto d$, $b_j c_j \mapsto b_j$, $p \mapsto p$). Now define the map $\psi_+ : A_+ \to A_-$ to be the 'reverse' of ψ_-, by extending the map $U \mapsto U$, $d \mapsto d$, $b_j \mapsto b_j c_j$, $p \mapsto p$. We show ψ_+ is a homomorphism; in doing so we will have proved that ψ_+ is an isomorphism as then ψ_+ and ψ_+ will be mutual inverses.
So, to show ψ_+ is a homomorphism, we must show that $\psi_+(pt_\alpha p^{-1}) = \psi_+(t_\alpha w_\alpha(b)d)$ for all $\alpha \in I$ with $(\alpha, 0) \in H_0(\mathcal{M})$. First, observe that $\psi_+(pt_\alpha p^{-1}) = pt_\alpha p^{-1}$ by definition of ψ_+. Next, observe that $\psi_+(t_\alpha w_\alpha(b)d) = t_\alpha w_\alpha(bc)d$, again by definition of ψ_+. Now, since b_j and c_l commute for all j, l, we have $w_\alpha(bc) = w_\alpha(c)w_\alpha(c)$. But, by step 11, we have that $w_\alpha(c) = e$ in C whenever $(\alpha, 0) \in H_0(\mathcal{M})$. Thus $\psi_+(pt_\alpha p^{-1}) = pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d = t_\alpha w_\alpha(bc)d = \psi_+(t_\alpha w_\alpha(b)d)$ for all $\alpha \in I$ with $(\alpha, 0) \in H_0(\mathcal{M})$, and so ψ_+ is a homomorphism.

Step 32: H_3 is finitely presented, and $\overline{C} \hookrightarrow H_3$.
First note that H_3 is formed by starting with a free product H_1 of $K_\mathcal{M}$, $C \times F_n$, and $\mathbb{Z}$. Then we HNN extend 2 times, forming H_2 and H_3, with finitely many stable letters each. Thus H_3 is finitely generated, and contains an embedded copy of C.
The relations of H_3 consist of the following collections of words:
1. The finitely many relations of $K_\mathcal{M}$, the relations of the form $b_i c_j = c_j b_i$ for $1 \leqslant i, j \leqslant n$, the relation $ptp^{-1} = td$, and the finitely many relations involving the stable letters of the HNN extension H_3.
2. The relations $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d$ for all $\alpha \in I$ with $\alpha \neq 0$ (from the HNN extension H_2).
3. The set S' of all trivial words from our original group C. That is, $w_\alpha(c) = e$ for all $\alpha \in I$ with $(\alpha, 0) \in H_0(\mathcal{M})$ (by step 11).
We now proceed to show that all the relations of type 3 are consequences of those of types 1 and 2, and then to show that all the relations of type 2 are consequences of those of type 1. First, by step 23 of (10.2), we see that if $(\alpha, 0) \in H_0(\mathcal{M})$ then t_α $(= t(\alpha, 0))$ lies in $\langle t\rangle'$, and thus can be written as a word over U. Moreover, this relationship comes as a consequence of the relations of $K_\mathcal{M}$ (type 1). Then, as a consequence of the relations involving k in H_3 (type 1), we have $kt_\alpha k^{-1} = t_\alpha$ if $(\alpha, 0) \in H_0(\mathcal{M})$ (as t_α can be written as a word over U precisely when $(\alpha, 0) \in H_0(\mathcal{M})$). We have, as a consequence of the relations of type 2, the relations $kw_\alpha(b)k^{-1} = w_\alpha(bc)$ for all $\alpha \in I$ (as we have $kb_j k^{-1} = b_j c_j$ for all j; type 1 relations, from the HNN extension H_3). Moreover, taking $\alpha \in I$ with $(\alpha, 0) \in H_0(\mathcal{M})$, we have that $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d$ (type 2 relation). Observing that $kpk^{-1} = p$ and $kdk^{-1} = d$ are type 1 relations, and that we already have $kt_\alpha k^{-1} = t_\alpha$ as a consequence of relations of type 1, we see that, as a consequence of type 1 relations,

$$\begin{aligned} t_\alpha w_\alpha(b)d &= pt_\alpha p^{-1} \\ &= kpt_\alpha p^{-1}k^{-1} \\ &= kt_\alpha w_\alpha(b)dk^{-1} \\ &= t_\alpha k w_\alpha(b)k^{-1}d \\ &= t_\alpha w_\alpha(bc)d \\ &= t_\alpha w_\alpha(b)w_\alpha(c)d \end{aligned}$$

Thus $w_\alpha(c) = e$ is a consequence of the relations of types 1 and 2, for all $\alpha \in I$ with $(\alpha, 0) \in H_0(\mathcal{M})$. So all relations of type 3 are consequences of relations of types 1 and 2.
It remains to show that all the relations of type 2 are consequences of those of type 1. To

begin, let $w_\alpha(a)$ be the word $w_\alpha(b)$ where we substitute $b_j \mapsto a_j$ $(1 \leqslant j \leqslant 2n)$. We induct on the length of $w^R_\alpha(a)$, the **reverse** of the word $w_\alpha(a)$, to show that $w^R_\alpha(a)tw^R_\alpha(a)^{-1} = t_\alpha$ and $w^R_\alpha(a)dw^R_\alpha(a)^{-1} = w_\alpha(b)d$ are consequences of the type 1 relations. For $|w_\alpha(a)| = 1$ we have $w^R_\alpha(a) = a_i$ (so $\alpha = i$), some $1 \leqslant i \leqslant 2n$. Thus:

$$\begin{aligned} w^R_\alpha(a)tw^R_\alpha(a)^{-1} &= a_i t a_i^{-1} = t_i = t_\alpha \\ w^R_\alpha(a)dw^R_\alpha(a)^{-1} &= a_i d a_i^{-1} = b_i d = w_\alpha(b)d \end{aligned}$$

are already type 1 relations (for $\alpha \in I$ with $\alpha \neq 0$).
Now, for the inductive step, we observe that a word of length 'one more that $w^R_\alpha(a)$' will be of the form $w^R_{\alpha m+i}(a)$, some $1 \leqslant i \leqslant 2n$ (step 9). Thus,

$$\begin{aligned} w^R_{\alpha m+i}(a)tw^R_{\alpha m+i}(a)^{-1} &= a_i w^R_\alpha(a)tw^R_\alpha(a)^{-1}a_i^{-1} \quad \text{(by step 5)} \\ &= a_i t_\alpha a_i^{-1} \quad \text{(a consequence of type 1 relations, by induction)} \\ &= a_i x^\alpha t x^{-\alpha} a_i^{-1} \\ &= x^{\alpha m} a_i t a_i^{-1} x^{-\alpha m} \quad \text{(as } a_i x a_i^{-1} = x^m\text{; a type 1 relation)} \\ &= x^{\alpha m} t_i x^{-\alpha m} \quad \text{(from above; a type 1 relation)} \\ &= t_{\alpha m+i} \end{aligned}$$

are all consequences of type 1 relations. Similarly,

$$\begin{aligned} w^R_{\alpha m+i}(a)dw^R_{\alpha m+i}(a)^{-1} &= a_i w^R_\alpha(a)dw^R_\alpha(a)^{-1}a_i^{-1} \quad \text{(by step 5)} \\ &= a_i w_\alpha(b)da_i^{-1} \quad \text{(a consequence of type 1 relations, by induction)} \\ &= a_i w_\alpha(b)a_i^{-1}b_i d \quad \text{(as } a_i d a_i^{-1} = b_i d\text{; a type 1 relation)} \\ &= w_\alpha(b)b_i d \quad \text{(as } a_i b_j a_i^{-1} = b_j \ \forall i,j\text{; a type 1 relation)} \\ &= w_{\alpha m+i}(b)d \quad \text{(by step 9)} \end{aligned}$$

are also all consequences of type 1 relations.
Now, it remains to show that the type 2 relations $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d$ for all $\alpha \in I$ with $\alpha \neq 0$ are consequences of the type 1 relations. We have $ptp^{-1} = td$ is a type 1 relation. Conjugating both sides by $w^R_\alpha(a)$, we consider the relation $w^R_\alpha(a)ptp^{-1}w^R_\alpha(a)^{-1} = w^R_\alpha(a)tdw^R_\alpha(a)^{-1}$; a consequence of a type 1 relation. Then

$$\begin{aligned} w^R_\alpha(a)ptp^{-1}w^R_\alpha(a)^{-1} &= pw^R_\alpha(a)tw^R_\alpha(a)^{-1}p^{-1} \quad \text{(as } a_i p a_i^{-1} = p \ \forall 1 \leqslant i \leqslant 2n\text{)} \\ &= pt_\alpha p^{-1} \end{aligned}$$

using only consequences of type 1 relations. Similarly, we get

$$\begin{aligned} w^R_\alpha(a)tdw^R_\alpha(a)^{-1} &= w^R_\alpha(a)tw^R_\alpha(a)^{-1}w^R_\alpha(a)dw^R_\alpha(a)^{-1} \\ &= t_\alpha w_\alpha(b)d \end{aligned}$$

using only consequences of relations of type 1. Thus, as $ptp^{-1} = t$ is a type 1 relation, we have that $pt_\alpha p^{-1} = t_\alpha w_\alpha(b)d$ for all $\alpha \in I$ with $\alpha \neq 0$, as consequences of type 1 relations.
So we can discard all relations of H_3 of types 2 and 3, and are left with the (finite set of) type 1 relations. So H_3 is finitely presented.

This concludes the proof that all the steps in the construction are valid. By combining (11.1) and (11.2), we now state the complete classification of finitely generated subgroups of finitely presented groups.

Theorem 11.3. Let H be a finitely generated group. Then H embeds into some finitely presented group iff H is recursively presented.

12. Other decision problems in group theory

Now that we have *one* algorithmically undecidable problem relating to finitely presented groups, we can use some simple algebraic tricks to propagate this out and show that *many* other problems are algorithmically undecidable.

The Adian-Rabin construction

Several decision problems (like the word problem) are solvable in particular classes of finitely presented groups. For example, the word problem is solvable in finitely presented free groups, and finitely presented abelian groups. However, to use such algorithms, we need to *know* that we have a presentation of a group which is free, abelian, etc. So, can we recognise these properties? Well, no. And to show that we can't, we use a very clever result by Adian and Rabin[7].

Theorem 12.1 (Adian-Rabin). There is a construction that, on input of a presentation $P = \langle X|R\rangle$ with X countable, and a word $w \in F(X)$, produces a new presentation $P(w)$ and an explicit homomorphism $\phi : \overline{P} \to \overline{P(w)}$ such that:
1. If $\overline{w} \neq e$ in $\overline{P}$, then ϕ is an embedding, and hence $\overline{P(w)} \neq \{e\}$.
2. If $\overline{w} = e$ in $\overline{P}$, then $\overline{P(w)} \cong \{e\}$.
3. $\overline{P(w)}$ can be generated by 2 elements.
4. If X, R are both r.e. sets, then $\overline{P(w)}$ is recursively presented.
5. If X, R are both finite sets, then $\overline{P(w)}$ is finitely presented.
6. In cases 4. and 5. above, the construction is algorithmic; we can obtain such a recursive/finite presentation directly from P and w.

Proof. Take a presentation $P = \langle X|R\rangle = \langle x_1, x_2, \ldots |R\rangle$ and form the free product $\overline{P} * F_2$ with presentation $Q := P * \langle a, b|-\rangle$. Now take another copy of F_2 with presentation $S := \langle c, d|-\rangle$.

Form the subgroup $A \leqslant \overline{P} * F_2$ by

$$A := \langle \overline{b}, \overline{aba^{-1}}, \overline{a^2bab^{-1}a^{-2}}, \overline{a^3[w,b]a^{-3}}, \overline{a^{(3+i)}x_iba^{-(3+i)}}\ \forall x_i \in X\rangle$$

If $\overline{w} \neq e$ in $\overline{P}$ then, by applying the normal form theorem for free products (2.15) in the same way as we did in (2.19), A is free on its given generating set, and so $A \cong F_\infty$ if $|X| = \infty$ or $A \cong F_{n+4}$ if $|X| = n < \infty$.
Similarly, form the subgroup $B \leqslant F_2$ by

$$B := \langle \overline{d}, \overline{cdcd^{-1}c^{-1}}, \overline{c^2dcd^{-1}c^{-2}}, \overline{c^3dc^{-3}}, \overline{c^{(3+i)}dc^{-(3+i)}}\ \forall x_i \in X\rangle$$

Just like in the case of A, we have that B is free on its given generating set, and so $B \cong F_\infty$ if $|X| = \infty$ or $B \cong F_{n+4}$ if $|X| = n < \infty$. Thus $A \cong B$.

Take the map $\varphi : A \to B$ given by extending the following bijection between generating sets (this is an isomorphism when $\overline{w} \neq e$ in $\overline{P}$):

$$\begin{aligned}
\overline{b} &\mapsto \overline{d} \\
\overline{aba^{-1}} &\mapsto \overline{cdcd^{-1}c^{-1}} \\
\overline{a^2bab^{-1}a^{-2}} &\mapsto \overline{c^2dcd^{-1}c^{-2}} \\
\overline{a^3[w,b]a^{-3}} &\mapsto \overline{c^3dc^{-3}} \\
\overline{a^{(3+i)}x_iba^{-(3+i)}} &\mapsto \overline{c^{(3+i)}dc^{-(3+i)}}\ \forall x_i \in X
\end{aligned}$$

[7] Done independently, and in a *much* more complicated way than what we present here.

Now form the amalgamated product $(\overline{P} * \overline{\langle a, b|-\rangle}) *_\varphi (\overline{\langle c, d|-\rangle})$. This has presentation $P(w)$ given by (after replacing d with b since $\varphi(\overline{b}) = \overline{d}$):
Generators: $X \cup \{a, b, c\}$.
Relations: R along with

$$\begin{aligned} aba^{-1} &= cbcb^{-1}c^{-1} \\ a^2bab^{-1}a^{-2} &= c^2bcb^{-1}c^{-2} \\ a^3[w, b]a^{-3} &= c^3bc^{-3} \\ a^{(3+i)}x_iba^{-(3+i)} &= c^{(3+i)}bc^{-(3+i)} \ \forall x_i \in X \end{aligned}$$

We now prove all the properties of $P(w)$ as stated in the theorem.

When $\overline{w} \neq e$ in $\overline{P}$, φ is an isomorphism, and so $\overline{P(w)}$ is an amalgamated product. So take ϕ to be the natural embedding $\overline{P} \hookrightarrow (\overline{P} * \overline{\langle a, b|-\rangle}) *_\varphi (\overline{\langle c, d|-\rangle}) = \overline{P(w)}$ by (5.14). Thus 1. is proved.

If $\overline{w} = e$ in $\overline{P}$, then A is not free on its generating set, so $\overline{P(w)}$ is not an amalgamated product. Moreover, one can see from the relators of $P(w)$ that, in the case $\overline{w} = e$ in $\overline{P}$, we have that $\overline{P(w)} \cong \{e\}$ (first see that $\overline{b} = e$, then $\overline{c} = e$, then $\overline{a} = e$, then $\overline{x_i} = e$ for all $x_i \in X$). Thus 2. is proved.

We only need $\overline{b}$ and $\overline{a^{-1}c}$ to generate $\overline{P(w)}$; this follows immediately from the relations of $P(w)$ as we have $c = b^{-1}(a^{-1}c)^{-1}b(a^{-1}c)b$, and then $a = c(a^{-1}c)^{-1}$, and then $x_i = a^{-(3+i)}c^{(3+i)}bc^{-(3+i)}a^{(3+i)}$. Thus 3. is proved

Suppose both X and R are r.e. The relations of $P(w)$ are R (which is r.e.), along with the finite set $\{aba^{-1} = cbcb^{-1}c^{-1}, a^2bab^{-1}a^{-2} = c^2bcb^{-1}c^{-2}, a^3[w, b]a^{-3} = c^3bc^{-3}\}$, and the infinite set $\{a^{(3+i)}x_iba^{-(3+i)} = c^{(3+i)}bc^{-(3+i)}\}_{x_i \in X}$ (which is r.e. since X is r.e.). By (7.10), as the relations of $P(w)$ are the union of finitely many r.e. sets, then they form an r.e. set. We add a generator t, and add the relation $t = a^{-1}c$. Using 3., we can (algorithmically) re-write the r.e. set of relations of $P(w)$ using just b and t, as these generate the group. We then discard all other generators apart from b and t. This re-written presentation has only 2 generators. Thus $\overline{P(w)}$ is recursively presented, and so 4. is proved.

Similarly, if X, R are finite, then the re-writing done to prove 4. gives a 2-generator finite presentation for $\overline{P(w)}$. Thus 5. is proved.

In the proofs of cases 4. and 5., the constructions and re-writings given are clearly algorithmic. Thus 6. is proved.

The following consequence, originally proved by Higman, Neumann and Neumann in 1949 but now a corollary of (12.1), is a predecessor to the Higman embedding theorem. Their original proof was the first use of HNN extensions in mathematics.

Theorem 12.2 (Higman, Neumann, Neumann, 1949). Every countable group C embeds into some 2-generator countable group G_C; if C is recursively (resp. finitely) presented, then G_C is also.

Proof. Take a presentation P for C, and form the free product presentation $Q := P * \langle t|-\rangle$ of the group $C * \mathbb{Z}$. Now form the presentation $Q(t)$ as per (12.1); the theorem now follows immediately as $\overline{t} \neq e$ in $\overline{Q}$.

Applications

The first way that we apply (12.1) is to show that the **triviality problem**, of recognising if a finite presentation P defines the trivial group or not, is unsolvable.

Theorem 12.3. There is no algorithm that, on input of a finite presentation P, decides if $\overline{P} \cong \{e\}$ or not.

Proof. Take a finite presentation $P = \langle X|R\rangle$ of a finitely presented group with IWP (10.4); noting that the proof there allows us to construct an *explicit* such presentation. Now, for each word $w \in F(X)$, we can construct the finite presentation $P(w)$ as per (12.1). Now simply observe that $\overline{w} = e$ in $\overline{P}$ iff $\overline{P(w)} \cong \{e\}$, so an algorithm to determine if an arbitrary finite presentation defined the trivial group or not would allow us to solve the word problem for $\overline{P}$; a contradiction.

This immediately implies the following (known as the **isomorphism problem** for finitely presented groups).

Corollary 12.4. There is no algorithm that, on input of two finite presentations P_1, P_2, decides if $\overline{P_1} \cong \overline{P_2}$ or not.

So we can't tell if a finite presentation P gives the trivial group or not, and we can't tell if two finite presentations P_1, P_2 give the same group or not. But can we recognise other properties?

Definition 12.5. A property of groups ρ is an **algebraic property** if it is invariant under isomorphism.

Definition 12.6. We call an algebraic property ρ of finitely presented groups a **Markov property** if there exist two finitely presented groups G_+, G_- such that:
1. G_+ has the property ρ.
2. G_- does not have the property ρ, nor does it embed in any finitely presented group with the property ρ.

We can tie this in with the Adian-Rabin construction (12.1) to show that many properties of finitely presented groups are algorithmically unrecognisable.

Theorem 12.7 (Markov). It is impossible to algorithmically recognise any Markov property amongst finitely presented groups. That is, for any fixed Markov property ρ, there is no algorithm that, on input of a finite presentation P, determines if $\overline{P}$ has that property or not.

Proof. Let ρ be a Markov property, with G_+, G_- as above (with finite presentations P_+, P_- respectively). Let $Q = \langle X|R\rangle$ be a finite presentation of a group with unsolvable word problem from (10.4). Given any word $w \in F(X)$, we can form the finite presentation $P_+ * \big((P_- * Q)(w)\big)$ as per (12.1). Then we have that:
1. If $\overline{w} = e$ in $\overline{Q}$, then $\overline{w} = e$ in $\overline{P_- * Q}$, and so $\overline{(P_- * Q)(w)} \cong \{e\}$. That is, the Adian-Rabin construction applied to $P_- * Q$ with word w gives the trivial group. Hence $\overline{P_+ * \big((P_- * Q)(w)\big)} \cong \overline{P}_+$, and so has property ρ.
2. If $\overline{w} \neq e$ in $\overline{Q}$, then $\overline{w} \neq e$ in $\overline{P_- * Q}$, and so $\overline{P}_- \hookrightarrow \overline{P_- * Q} \hookrightarrow \overline{(P_- * Q)(w)} \hookrightarrow \overline{P_+ * \big((P_- * Q)(w)\big)}$. So $\overline{P_+ * \big((P_- * Q)(w)\big)}$ does not have property ρ, as $\overline{P}_-$ embeds into it.
So in summary, $\overline{P_+ * \big((P_- * Q)(w)\big)}$ has property ρ if and only if $\overline{w} = e$ in $\overline{Q}$. But $\overline{Q}$ is a group with IWP, so the problem of recognising ρ groups amongst finitely presented groups is algorithmically unsolvable.

Lemma 12.8. The following is a (non-exhaustive) list of Markov properties, and hence all are algorithmically unrecognisable amongst finitely presented groups.
Being trivial.
Being finite.
Being free.
Being abelian.
Being cyclic.
Being torsion free.
Being torsion.
Having solvable word problem.

We now state the following remarkable consequence of the HNN theorem (12.2) and Higman's embedding theorem (11.2).

Theorem 12.9. There is a **universal finitely presented group**. That is, a finitely presented group which contains an embedded copy of every finitely presented group.

Proof. Fix a countably infinite alphabet $X = \{x_1, x_2, \ldots\}$. Now, for each n, we can construct all finite presentations of 'length n' (i.e., have precisely n letters appearing as generators/relations), and we can order these lexicographically (take the ordering on X and then read across the presentation from left to right as if it was a string of letters only; adopt the convention $x_i < x_i^{-1} < x_{i+1}$). Thus, by ordering all the presentations of length 1, then of length 2, and so on, we have an enumeration of finite presentations $P_1, P_2, \ldots$. Observe that, from j, we can algorithmically construct the presentation P_j. Moreover, every finitely presented group is isomorphic to the group given by one such presentation (perhaps many).
So now form the presentation $P = P_1 * P_2 * \ldots$. As we can construct P_j from j, then we have that P has an r.e. set of generators, and an r.e. set of relations. So by (12.2), we have that $\overline{P}$ embeds into a recursively presented group G. Finally, we use (11.2) to embed G into a finitely presented group H. Now, as every finitely presented group embeds into $\overline{P}$, then they all embed into H also.