Math 8245 Group Theory

Peter Webb

December 14, 2018

Contents

Chapter 1

Fundamental structures in group theory

1.1 Semidirect Products

We already know about direct products. We write $G = K \rtimes Q$ to mean K is a normal subgroup of G, Q is a subgroup of G, and $K \cap Q = 1$, $KQ = G$. We say G is a *semidirect product* of K by Q, and Q is a *complement* of K in G.

In Rotman's book the condition $K \triangleleft G$ is not required for a complement, and the argument for uniqueness implied there there does not work without this condition. I think it is usual to require K to be normal.

Examples 1.1.1.

- Direct products are semidirect products in which both subgroups are normal.
- $S_n = A_n \rtimes C_2$
- $D_{2n} = C_n \rtimes C_2$
- C_4 is not a semidirect product.
- Q_8 is not a semidirect product.
- The crystallographic group of the infinite pattern

$$\begin{array}{ccc} \triangleleft & \triangleleft & \triangleleft \\ \triangleleft & \triangleleft & \triangleleft \\ \triangleleft & \triangleleft & \triangleleft \end{array}$$

is a semidirect product, but the crystallographic group of the infinite pattern

$$\begin{array}{ccccc} \rfloor & & \rfloor & & \rfloor \\ & \rceil & & \rceil & \\ \rfloor & & \rfloor & & \rfloor \end{array}$$

is not a semidirect product. The term *crystallographic group* means the group of *rigid motions* of the plane that preserve the pattern.

- C_6 and S_3 are both semidirect products of C_3 by C_2.

Class Activity. Which of the following have a non-trivial semidirect product decomposition? A_4, A_5, C_4, C_{10}. Is this easy or difficult?

Definition 1.1.2. A homomorphism $\phi : G \to Q$ is a *split epimorphism* if and only if there is a homomorphism $s : Q \to G$ so that $\phi s = 1_Q$.

Theorem 1.1.3 (Rotman 7.20 parts (i) and (iii)). *1. A split epimorphism is an epimorphism.*

2. Let $\phi : G \to Q$ be a group homomorphism. Then ϕ is split epi if and only if $G = K \rtimes Q_1$ where $K = \operatorname{Ker}\phi$, for some subgroup $Q_1 \leq G$ mapped isomorphically to Q by ϕ, if and only if ϕ is surjective and $\operatorname{Ker}\phi$ *has a complement in G.*

Class Activity. The kernel K may have many complements. Find an example where there is more than one.

Corollary 1.1.4. *Let $\phi : G \to Q$ and $s : Q \to G$ be group homomorphisms with $\phi s = 1_Q$. Then $G = \operatorname{Ker}\phi \rtimes sQ$.*

Example 1.1.5. For a short exact sequence of groups $1 \to K \xrightarrow{\theta} G \xrightarrow{\phi} Q \to 1$ to say that ϕ is split epi it is not equivalent to say that θ is split (i.e. has a right inverse). Consider $1 \to C_3 \to S_3 \to C_2 \to 1$. Show that θ is split if and only if $G = K \times Q_1$ for some subgroup $Q_1 \leq G$ mapped isomorphically to Q by ϕ.

Definition 1.1.6. Let $K \lhd G$. Conjugation within G defines a homomorphism $\theta : G \to \operatorname{Aut} K$. Specifically, if $x \in G$ and $a \in K$ then $\theta_x(a) = xax^{-1}$. In general, such a homomorphism θ is an *action* of Q on K.

Definition 1.1.7. The group $\operatorname{Inn} K$ of *inner* automorphisms of K is the group of automorphisms of the form $\alpha(a) = bab^{-1}$ for some fixed $b \in K$.

Class Activity. When $G = S_n$ and $K = A_n$, does G have image in $\operatorname{Inn} A_n$?

When $G = K \rtimes Q$ the restriction of θ to Q gives a mapping $\theta : Q \to \operatorname{Aut} K$. We will say that G *realizes* θ in this situation.

Example 1.1.8. With the two semidirect products $C_6 = C_3 \rtimes C_2$ and $S_3 = C_3 \rtimes C_2$ the two homomorphisms $Q = C_2 \to K = C_3$ are different, realized by the two different semidirect products. The notation $\rtimes$ does not distinguish between them.

Definition 1.1.9. Let Q and K be groups and suppose we are given a homomorphism $\theta : Q \to \operatorname{Aut} K$. We define a group $K \rtimes_\theta Q$ to be $K \times Q$ as a set, and with multiplication $(a, x)(b, y) = (a\theta_x(b), xy)$.

Theorem 1.1.10 (Rotman 7.22)**.** *$K \rtimes_\theta Q$ is a semidirect product that realizes θ. Better: $K \rtimes_\theta Q$ has subgroups $K_1 \cong K$ and $Q_1 \cong Q$ so that it realizes the homomorphism $Q_1 \to Q \xrightarrow{\theta} \operatorname{Aut} K \to \operatorname{Aut} K_1$.*

The construction of $K \rtimes_\theta Q$ could be called the *external* semidirect product and the original definition $G = K \rtimes Q$ the *internal* semidirect product, extending the notion of internal and external direct products.

Theorem 1.1.11 (Rotman 7.23)**.** *If $G = K \rtimes Q$ and $\theta : Q \to \operatorname{Aut} K$ is defined by $\theta_x(a) = xax^{-1}$ then $G \cong K \rtimes_\theta Q$ (via an isomorphism that identifies K with K_1 and Q with Q_1).*

Hence any two semidirect products that realize θ are isomorphic. This resolves the issue that the notation $\rtimes$ does not carry complete information about the semidirect product. On the other hand, it is usual to write just $\rtimes$ instead of $\rtimes_\theta$.

Proof. We define a mapping

$$K \rtimes_\theta Q \to K \rtimes Q$$
$$(a, x) \mapsto ax$$

We check that $(a, x)(b, y) = (a\theta_x(b), xy) \mapsto a\theta_x(b)xy = axbx^{-1}xy = (ax)(by)$. Thus the mapping is a homomorphism. We check that it is bijective. □

Exercise 1.1.12. Let $\theta, \psi : Q \to \operatorname{Aut} K$ be two homomorphisms and let $\beta \in \operatorname{Aut} K$ and $\gamma \in \operatorname{Aut} Q$ be automorphisms. Which, if any, of the following always imply that $K \rtimes_\theta Q \cong K \rtimes_\psi Q$?

1. $\psi = \beta\theta$
2. $\psi = \theta\gamma$
3. $\psi = \beta\theta$ where $\beta \in \operatorname{Inn} K$
4. $\psi = \theta\gamma$ where $\gamma \in \operatorname{Inn} Q$
5. for all $x \in Q$, $\theta(x) = \beta\psi(x)\beta^{-1}$

1.1.1 Small p-groups

Semidirect products are important because many groups that arise in practice can be constructed this way. We describe the non-abelian groups of order p^3 when p is a prime.

Example 1.1.13 (Example 7.15 of Rotman)**.** Consider the groups of the form $(C_p \times C_p) \rtimes C_p$ where p is a prime. First we consider the possible actions of C_p on $C_p \times C_p$. The automorphism group $\operatorname{Aut}(C_p \times C_p) = GL(2, p)$ has size $(p^2-1)(p^2-p) = p(p-1)^2(p+1)$, so that Sylow p-subgroups are copies of C_p and they are all conjugate to the subgroup generated by the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. This means that all non-identity actions of C_p on

$C_p \times C_p$ will give isomorphic semidirect products. We may take a generator of $C_p = \langle c \rangle$ to act on $C_p \times C_p = \langle a, b \rangle$ in additive notation via the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, and in multiplicative notation as ${}^c a = a$, ${}^c b = ab$.

Class Activity. In the last paragraph, is the sentence that ends, ‘will give isomorphic semidirect products’ obvious?

1. $(C_p \times C_p) \rtimes C_p$ is isomorphic to the subgroup of $SL(3,p)$
$$\{ \begin{bmatrix} 1 & u & v \\ 0 & 1 & w \\ 0 & 0 & 1 \end{bmatrix} \bigm| u, v, w \in \mathbb{F}_p \}.$$
Class Activity. Why?

2. $(C_p \times C_p) \rtimes C_p$ has a presentation
$$\langle a, b, c \bigm| a^p = b^p = c^p = [a,b] = [c,a] = 1,\ [c,b] = a \rangle.$$

3. If p is odd then every non-identity element of $(C_2 \times C_p) \rtimes C_p$ has order p. This provides an example of two non-identity groups whose lists of orders of elements are the same, but which are non-isomorphic.

Proof. 2. Let G be the group with the presentation in 2. Then $\langle a, b \rangle \lhd G$ and $|\langle a, b \rangle| \leq p^2$, $|\langle c \rangle| \leq p$ so $|G| \leq p^3$. On the other hand, the semidirect product is an image of G and has order p^3, so the two groups must be isomorphic.

3. From the relation $[c, b] = a$ we deduce $cbc^{-1} = ab$. Every element in this group can be written $a^r b^s c^t$. To simplify the notation slightly, consider an element $a^r b^s c$. By induction, $(a^r b^s c)^d = a^{dr+s+2s+\cdots(d-1)s} b^{ds} c^d$. Putting $d = p$ and using the fact that $1 + 2 + \cdots (p-1)$ is divisible by p, we see that the pth power of this element is 1. □

Class Activity. Where does the argument in proving 3 go wrong when $p = 2$. What is the name of the group $(C_2 \times C_2) \rtimes C_2$, where we have taken $p = 2$? Which of the properties listed above hold when $p = 2$?

Example 1.1.14. Consider groups of the form $C_{p^2} \rtimes C_p$ where p is a prime. Here $\operatorname{Aut} C_{p^2} \cong C_{p(p-1)}$ is cyclic (why?) and any two non-identity actions of $C_p = \langle c \rangle$ on $C_{p^2} = \langle a \rangle$ will give isomorphic groups (why?). We may take c to act on $\langle a \rangle$ as ${}^c a = a^{p+1}$ and now $C_{p^2} \rtimes C_p$ has a presentation
$$\langle a, c \bigm| a^{p^2} = c^p = 1,\ cac^{-1} = a^{1+p} \rangle.$$

Theorem 1.1.15. *1. Let p be an odd prime. Every non-abelian group of order p^3 is isomorphic to one of the two just described.*

2. Every non-abelian group of order 8 is isomorphic to D_8 or Q_8.

Proof. We sketch the proof of 1. There is a theorem as follows:

Theorem 1.1.16 (Rotman 5.46)**.** *Let G be a p-group with a unique subgroup of order p. Then G is cyclic or $p = 2$ and $G \cong Q_{2^n}$ where*

$$Q_{2^n} = \langle x, y \mid x^{2^{n-1}} = y^4 = 1,\ yxy^{-1} = x^{-1},\ x^{2^{n-2}} = y^2 \rangle$$

is the generalized quaternion group of order 2^n.

Assuming this, if $|G| = p^3$ is non-abelian with p odd, choose a non-central subgroup C_p. There is a normal subgroup of order p^2 not containing it. We obtain G as a semidirect product. Now classify the possible semidirect products as the ones we have considered. □

Class Activity. Each sentence in the last paragraph could be discussed.

Exercise 1.1.17. Compute the structure of the center $Z(G)$, the derived subgroup G' and the abelianization G/G' for each of the above groups G.

Example 1.1.18. We describe the semidihedral groups. A theorem states that

$$\operatorname{Aut} C_{2^n} \cong C_{2^{n-2}} \times C_2$$

when $n \geq 3$, but we do not need to know this theorem to see that $C_2 \times C_2$ acts on $C_{2^n} = \langle x \rangle$ as the set of four automorphisms determined by the following:

$$\begin{aligned} x &\mapsto x \\ x &\mapsto x^{-1} \\ x &\mapsto x^{2^{n-1}-1} \\ x &\mapsto x^{2^{n-1}+1} \end{aligned}$$

Letting $C_2 = \langle y \rangle$ act on C_{2^n} as $x \mapsto x^{-1}$ gives the dihedral group $D_{2^{n+1}} = C_{2^n} \rtimes C_2$. Letting $C_2 = \langle y \rangle$ act on C_{2^n} as $x \mapsto x^{2^{n-1}-1}$ gives the semidihedral group

$$SD_{2^{n+1}} = C_{2^n} \rtimes C_2 = \langle x, y \mid x^{2^n} = y^2 = 1,\ yxy = x^{2^{n-1}-1} \rangle.$$

The third group $C_{2^n} \rtimes C_2$ that is not a direct product is less important.

The semidihedral group $SD_{2^{n+1}}$ has three subgroups of order 2^n, and they are copies of C_{2^n}, D_{2^n} and Q_{2^n}. The group $GL(2,3)$ of 2×2 invertible matrices over $\mathbb{F}_3$ of order 48 has SD_{16} as its Sylow 2-subgroup. The three classes of dihedral, semidihedral and generalized quaternion groups share the property that they are the 2-groups of maximal class, as well as being the non-abelian 2-groups of 2-rank at most 2.

1.1.2 Wreath products

We follow Rotman between Theorems 7.24 and 7.27.

Let Q and D be groups and let Q permute a set Ω. We may identify the full direct product $\prod_{\omega\in\Omega} D$ as the set of functions D^Ω from Ω to D, and inside that there is the *restricted* direct product K, consisting of functions that take the value 1 on all except finitely many ω. Given $d \in D$ and $\omega \in \Omega$ let $d_\omega : \Omega \to D$ be the function

$$d_\omega(\psi) = \begin{cases} 1 & \omega \neq \psi \\ d & \omega = \psi \end{cases}$$

Thus K is the subgroup of the direct product generated by the elements d_ω. Now Q acts on D^Ω as $({}^q f)(\psi) = f(q^{-1}\psi)$ and we calculate that ${}^q(d_\omega) = d_{q\omega}$.

Class Activity. Does ${}^q(d_\omega)$ equal $d_{q\omega}$ or $d_{q^{-1}\omega}$?

We see that Q also acts as automorphisms of the restricted direct product K. We define the (permutational) *wreath product* of D and Q to be $D \wr Q := K \rtimes Q$. The subgroup K is called the *base group* of the wreath product. If Ω is not specified we take it to be the regular representation of Q.

If, now, D also acts on a set Λ we can make both Q and K act on the product $\Lambda \times \Omega$ as follows:

$$\begin{aligned} d_\omega(\lambda, \psi) &= \begin{cases} (d\lambda, \psi) & \text{if } \omega = \psi \\ (\lambda, \psi) & \text{if } \omega \neq \psi \end{cases} \\ q(\lambda, \psi) &= (\lambda, \psi). \end{aligned}$$

Picture:

$$\Lambda \times \Omega = \overbrace{\begin{array}{ccc} | & & | \\ \Lambda & \cdots & \Lambda \\ | & & | \end{array}}^{\Omega}$$

Theorem 1.1.19 (Rotman 7.24-26)**.**

1. *The wreath product $D \wr Q$ permutes $\Lambda \times \Omega$ via the above action.*
2. *If D and Q are both faithful in their actions, so is $D \wr Q$ on $\Lambda \times \Omega$.*
3. *If D and Q both act transitively then $D \wr Q$ is transitive on $\Lambda \times \Omega$.*
4. *$T \wr (D \wr Q) \cong (T \wr D) \wr Q$.*

Note that the proof Rotman gives makes the assumption that D and Q act faithfully.

Proof. 2. Let G be the subgroup of the group of permutations of $\Lambda \times \Omega$ generated by the permutations given by the d_ω and the q as above. The d_ω generate a subgroup isomorphic to the product of the copies of D. The permutations given by the $q \in Q$

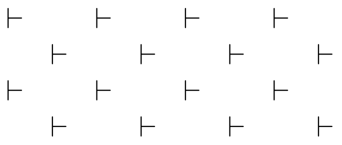
generate a copy of Q. These two subgroups generate G and intersect in 1. We check that the conjugation of Q on $\prod_{\omega\in\Omega} D$ is correct for the wreath product. We calculate

$$\begin{aligned} qd_\omega q^{-1}(\lambda,\psi) &= qd_\omega(\lambda, q^{-1}\psi) \\ &= q\begin{cases}(d\lambda, q^{-1}\psi) & \text{if } \omega = q^{-1}\psi \\ (\lambda, q^{-1}\psi) & \text{if } \omega \neq q^{-1}\psi\end{cases} \\ &= \begin{cases}(d\lambda, \psi) & \text{if } \omega = q^{-1}\psi \\ (\lambda, \psi) & \text{if } \omega \neq q^{-1}\psi\end{cases} \\ &= d_{q\omega}(\lambda,\psi). \end{aligned}$$

This shows that $qd_\omega q^{-1} = d_{q\omega}$ so that the subgroup of G generated by the d_ω is normal. Since the conjugation action is correct for the wreath product we have $G \cong D \wr Q$. □

Class Activity. Is the group $\langle(1,2,3),(4,5,6),(1,4)(2,5)(3,6)\rangle$ isomorphic to either $C_3 \wr C_2$ or $C_2 \wr C_3$?

Example 1.1.20. The group of rigid motions of the pattern



is $\mathbb{Z} \wr C_2$.

Example 1.1.21. Groups of the form $D \wr \mathbb{Z}$ are sometimes called *lamplighter* groups.

Example 1.1.22. In his book Rotman describes a graph whose automorphism group is $C_2 \wr S_5$.

Rotman now presents Theorem 7.27 of Kaloujnine describing the Sylow subgroups of symmetric groups.

Class Activity. How many orbits does the Sylow 3-subgroup of S_{35} have on $\{1,\ldots,35\}$?

1.2 G-sets

We introduce a part of the theory of G-sets, suitable for understanding the approach GAP uses to compute with permutation groups, using stabilizer chains. Rotman's book describes other results about G-sets, such as the Cauchy-Frobenius lemma, often known as 'Burnside's Lemma'.

Let G be a group. A *G-set* is a set Ω with an action of G by permutations. There are right and left G-sets and by an action of G on Ω from the right we mean a mapping $\Omega \times G \to \Omega$ so that $\omega(gh) = (\omega g)h$ and $\omega \cdot 1 = \omega$ hold for all $\omega \in \Omega$ and $g, h \in G$. With the convention that functions are applied from the right, the specification of a right

G-set is equivalent to the specification of a homomorphism $G \to S_\Omega$, the symmetric group on Ω. Similarly a left G-set is equivalent to the specification of a homomorphism $G \to S_\Omega$ provided we adopt the convention that mappings are applied from the left. Because GAP applies mappings from the right, we will work with right G-sets.

For each $\omega \in \Omega$ the set $\omega G = \{\omega g \mid g \in G\}$ is the *orbit* of Ω that contains ω. We say that G acts *transitively* on Ω if there is only one orbit. We put

$$\mathrm{Stab}_G(\omega) = G_\omega = \{g \in G \mid \omega g = \omega\}$$

and this is the *stabilizer* of ω in G. For example:

- if G permutes the set of its subgroups by conjugation then $\mathrm{Stab}_G(H) = N_G(H)$,
- if G permutes the set of its elements by conjugation then $\mathrm{Stab}_G(x) = C_G(x)$,
- if G permutes the right cosets $H\backslash G = \{Hg \mid g \in G\}$ by right multiplication then $\mathrm{Stab}_G(Hg) = H^g = g^{-1}Hg$. This is part 3. of the result below.

A *homomorphism* $f : \Omega \to \Psi$ of G-sets is a mapping with $f(\omega g) = (f(\omega))g$ always, and if this condition is satisfied we say that the mapping f is *equivariant* for the action of G. Such a homomorphism of G-sets is an isomorphism if and only if it is bijective, if and only if there is a G-set homomorphism $f_1 : \Psi \to \Omega$ with $1_\Psi = ff_1$ and $1_\Omega = f_1f$.

Class Activity. Is this obvious?

We probably already know the 'orbit-stabilizer' theorem. Part 2 of the next proposition is a more sophisticated version of this result, applying to infinite G-sets and containing more information.

Proposition 1.2.1. *1. Every G-set Ω has a unique decomposition $\Omega = \bigcup_{i\in I} \Omega_i$ where I is some indexing set and the Ω_i are orbits of Ω.*

2. If Ω is a transitive G-set and $\omega \in \Omega$ then $\Omega \cong \mathrm{Stab}_G(\omega)\backslash G$ as G-sets. Thus if Ω is finite then $|\Omega| = |G : \mathrm{Stab}_G(\omega)|$.

3. When $H \le G$, the stabilizer of the element Hg in the space of right cosets $H\backslash G$ is $H^g = g^{-1}Hg$.

4. If $H, K \le G$, there is a G-set homomorphism $f : H\backslash G \to K\backslash G$ with $f(H) = Kg$ if and only if $H \subseteq K^g$.

5. If $H, K \le G$ then $H\backslash G \cong K\backslash G$ as G-sets if and only if K and H are conjugate subgroups of G.

6. Every equivariant map between transitive G-sets is an epimorphism.

7. $\mathrm{Aut}_{G-\mathrm{set}}(H\backslash G) \cong N_G(H)/H$.

We see from 4. that every homomorphism $H\backslash G \to K\backslash G$ is the composite of a homomorphism $H\backslash G \to K^g\backslash G$ specified by $H \mapsto K^g$ where $H \leq K^g$, followed by an isomorphism $K^g\backslash G \to K\backslash G$ specified by $K^g \mapsto Kg$.

Proof. 2. Given $\omega \in \Omega$, define a mapping $G \to \Omega$ by $g \mapsto \omega g$. This is a map of G-sets. We check that the set of elements of G mapped to ωg is $\mathrm{Stab}_G(\omega)g$, so that there is induced a G-equivariant bijection between the two sets as claimed.

4. We first observe that if $f : \Omega \to \Psi$ is a map of G sets then $\mathrm{Stab}_G(\omega) \subseteq \mathrm{Stab}_G(f(\omega))$. From this, the implication '$\Rightarrow$' follows. Conversely, if $H \subseteq K^g$ we show that the specification $f : H\backslash G \to K\backslash G$ by $f(Hx) = Kgx$ is well defined. This is because if $Hx = Hy$ the $xy^{-1} \in H$ so $xy^{-1} \in K^g$ and $x = g^{-1}kgy$ for some $k \in K$. Thus $Kgx = Kgg^{-1}kgy = Kgy$. The mapping f is G-equivariant, so we have a homomorphism as claimed.

7. The mapping $N_G(H) \to \mathrm{Aut}_{G-\mathrm{set}}(H\backslash G)$ given by $g \mapsto (H \mapsto Hg)$ is a surjective homomorphism of groups. Its kernel is H. □

Let H be a subgroup of a group G. A *right transversal* to H in G is the same thing as a set of right coset representatives for H in G, that is: a set of elements $g_1, \ldots, g_t$ of G so that $G = Hg_1 \cup \cdots \cup Hg_t$.

Proposition 1.2.2. *Let G act transitively on a set Ω and let $\omega \in \Omega$ be an element with stabilizer G_ω. Then elements $\{g_i \mid i \in I\}$ of G form a right transversal to G_ω in G if and only if $\Omega = \{\omega g_i \mid i \in I\}$ and the ωg_i are all distinct.*

Proof. This comes from the isomorphism of G-sets $\Omega \cong G_\omega\backslash G$ under which $\omega g \leftrightarrow G_\omega g$. □

Algorithm 1.2.3. This observation provides a way to compute a transversal for $\mathrm{Stab}_G(\omega)$ in G. Take the generators of G and repeatedly apply them to ω, obtaining various elements of the form $\omega g_{i_1} g_{i_2} \cdots g_{i_r}$ where the g_{i_j} are generators of G. Each time we get an element we have seen previously, we discard it. Eventually we obtain the orbit ωG, and the various elements $g_{i_1} g_{i_2} \cdots g_{i_r}$ are a right transversal to $\mathrm{Stab}_G(\omega)$ in G.

There is an example below with a group of permutations of six points.

The elements of this transversal come expressed as words in the generators of G. It is what GAP does, except that it does the above with the inverses of the generators of G. If an inverse generator g^{-1} sends an already-computed element u to a new element v, the generator g is stored in position v in a list. This means that applying g to v gives u. By repeating this we eventually get back to the first element of the orbit. It is this list of generators that GAP stores in the field ' `transversal`' of a stabilizer chain. Elements of a right transversal are obtained by multiplying the inverses of the generators in reverse sequence.

1.2.1 Stabilizer chains

Computing chains of stabilizers is the most important technique available in computations with permutation groups. The idea of doing this in the context of computational group theory is due to Charles Sims. The following theorem of Schreier allows us to compute generators for stabilizer subgroups and the whole approach is known as the *Schreier-Sims algorithm.*

Theorem 1.2.4 (Schreier)**.** *Let X be a set of generators for a group G, $H \le G$ a subgroup, and T a right transversal for H in G such that the identity element of G represents the coset H. For each $g \in G$ let $\overline{g} \in T$ be such that $H\overline{g} = Hg$. Then*

$$\{tg(\overline{tg})^{-1} \bigm| t \in T, g \in X\}$$

is a set of generators for H.

Note that since $Htg = H\overline{tg}$, the elements $tg(\overline{tg})^{-1}$ lie in H always. Also $\overline{\overline{a}} = \overline{a}$ and $\overline{\overline{a}b} = \overline{ab}$. The generators in the set are called *Schreier generators.* Not only do they generate H but, if the elements of the transversal are expressed as words in the generators of G, then the generators of H are also expressed as words in the generators of G.

Proof. Suppose that $g_1 \cdots g_n \in H$ where the g_i lie in X. Then

$$g_1 \cdots g_n = (g_1 \overline{g_1}^{-1})(\overline{g_1} g_2 \overline{g_1 g_2}^{-1})(\overline{g_1 g_2} g_3 \overline{g_1 g_2 g_3}^{-1}) \cdots (\overline{g_1 \cdots g_{n-1}} g_n)$$

is a product of the Schreier generators. Note that $g_1 \cdots g_n \in H$ so that $\overline{g_1 \cdots g_n} = 1$. □

If G permutes Ω, a *base* for G on Ω is a list of elements $\omega_1, \omega_2, \ldots, \omega_s$ of Ω so that the stabilizer $G_{\omega_1,\omega_2,\ldots,\omega_s}$ equals 1. Here $G_{\omega_1,\omega_2,\ldots,\omega_r}$ is the stabilizer inside the subgroup $G_{\omega_1,\omega_2,\ldots,\omega_{r-1}}$ of ω_r, for each r. Let us write G_r instead of $G_{\omega_1,\omega_2,\ldots,\omega_r}$ and $G_0 = G$. In this situation the chain of subgroups

$$G = G_0 \ge G_1 \ge \cdots \ge G_s = 1$$

is called a *stabilizer chain* (for G, with respect to the given base). We will consider for each r the subset Ω_r of Ω which is defined to be the G_r-orbit containing ω_{r+1}. Thus $\Omega_0 = \omega_1 G$, $\Omega_1 = \omega_2 G_1$ etc. A *strong generating set* for G (with respect to the base) is a set of generators for G which includes generators for each of the subgroups G_r. Thus in a strong generating set, G_r is generated by those generators that happen to fix each of $\omega_1, \ldots, \omega_r$.

Proposition 1.2.5. *Each Ω_i is acted on transitively by G_i. As G_i-sets, $\Omega_i \cong G_{i+1}\backslash G_i$. Hence $|G| = |\Omega_0| \cdots |\Omega_{s-1}|$.*

Proof. We have $\omega_{i+1} \in \Omega_i$ and $\mathrm{Stab}_{G_i}(\omega_{i+1}) = G_{i+1}$. □

Given a set of generators $G = \langle g_1, \ldots, g_d \rangle$ and a subgroup $H \leq G$ a *right Schreier transversal* for H in G is a right transversal with elements expressed as words in the generators, as suggested by the following $1, g_{i_1}, g_{i_1}g_{i_2}, g_{i_3}, \ldots$ so that each initial segment of a word appears (earlier) in the list. Schreier transversals correspond to rooted trees.

Example 1.2.6. Let $G = \langle (1,5)(2,6), (1,3)(4,6), (2,3)(4,5) \rangle$ and write these generators as $a = (1,5)(2,6)$, $b = (1,3)(4,6)$, $c = (2,3)(4,5)$. Find a set of coset representatives for $\mathrm{Stab}_G(1)$.

Solution: We construct a Schreier tree:

$$\begin{tikzcd}
1 \arrow[r, "a"] \arrow[d, "b"'] & 5 \arrow[r, "c"] & 4 \\
3 \arrow[r, "c"] & 2 \arrow[r, "a"] & 6
\end{tikzcd}$$

giving coset representatives $1, a, b, ac, bc, bca$. These form a Schreier transversal: every initial segment of a word is in the transversal. These generators have order 2, and GAP stores their inverses in the list $[1, c, b, c, a, a]$.

Class Activity. Given that the element $abc = (1,4,6,3)(2,5) = x$ lies in G, find the coset representative that represents $\mathrm{Stab}_G(1)x$.

Table of $\overline{tg}$:

1	a	b	ac	bc	bca	
a	1	b	ac	bca	bc	a
b	a	1	bca	bc	ac	b
1	ac	bc	a	b	bca	c

Table of $tg\overline{tg}^{-1}$:

1	a	b	ac	bc	bca	
1	a^2	bab^{-1}	$acac^{-1}a^{-1}$	1	$bca^2c^{-1}b^{-1}$	a
1	aba^{-1}	b^2	$acba^{-1}c^{-1}b^{-1}$	$bcbc^{-1}b^{-1}$	$bcabc^{-1}a^{-1}$	b
c	1	1	ac^2a^{-1}	bc^2b^{-1}	$bcaca^{-1}c^{-1}b^{-1}$	c

Observe that 5 of these entries are necessarily 1. Upon evaluation of these expressions in G the last table becomes the following:

1	a	b	ac	bc	bca	
1	1	$(2,4)(3,5)$	$(2,3)(4,5)$	1	1	a
1	$(2,4)(3,5)$	1	$(2,5)(3,4)$	$(2,3)(4,5)$	$(2,5)(3,4)$	b
$(2,3)(4,5)$	1	1	1	1	$(2,4)(3,5)$	c

We see that, in the stabilizer chain, G_0 acts on Ω of size 6, $G_1 = \langle (2,3)(4,5), (2,4)(3,5) \rangle$ acts on $\{2,3,4,5\}$ of size 4, and $G_{12} = 1$, so that $|G| = 4 \cdot 6 = 24$. The fact that $G_{12} = 1$ we can see by inspection, because the group is so small, but to continue the algorithm properly we go through Schreier's theorem.

Theorem 1.2.7 (Schreier). *Let G have d generators and let $H \leq G$ have finite index. Then H can be generated by $|G : H|(d-1)+1$ elements.*

Proof. Consider the generators $tg(\overline{tg})^{-1}$ for H, and write $n = |G : H|$. The number of edges in the Schreier tree is $n-1$. Each gives an entry 1 in the table of generators. The number of table entries which are not 1 is at most $dn - n + 1 = n(d-1)+1$. □

Write $d(G)$ for the smallest size of a set of generators of G. The last result can be written $d(H) - 1 \leq |G : H|(d(G) - 1)$. When G is a free group it turns out that we always get equality in this bound. We will see this when we come to the section on free groups and, more generally, groups acting on trees. In the example, there were 5 edges in the Schreier tree, and these accounted for the 5 identity elements in the first table.

Algorithm 1.2.8. Given a stabilizer chain with a transversal for each stabilizer group in the next, we can test whether a permutation belongs to a group. If it does, and the transversal elements are words in the generators, we can express the permutation as a word in the generators. This algorithm solves problems such as restoring Rubik's cube to its initial position, given a random permutation of its faces.

Given a permutation π find the coset representative x_1 of the coset $G_1\pi$ by computing the action of π on Ω. We compute $(\omega_1)\pi$. If $\pi \in G$ this must equal $(\omega_1)g$ for some unique g in a right transversal for G_1 in G_0 and so $\pi g^{-1} \in G_1$. In fact, $\pi \in G$ if and only if $(\omega_1)\pi = (\omega_1)g$ for some g in the transversal and $\pi g^{-1} \in G_1$. We now continue to test whether $\pi g^{-1} \in G_1$ by repeating the algorithm.

Example 1.2.9. Continuing the previous example: is $(1,2,3)$ in G? Since $(1,2,3)c^{-1}b^{-1} = (4,5,6) \notin G_1$, the answer is No.

Class Activity. Is $(1,3,5)(2,6,4)$ in G? If it is, write this permutation as a word in the given generators of G.

Algorithm 1.2.10. We give an algorithm for listing the elements of G. We start by listing elements in the subgroups at the small end of the stabilizer chain, at each stage listing them by cosets in the next biggest stabilizer. Thus, if the elements of G_{i+1} have been listed and $t_1, \ldots, t_s$ is a transversal for G_{i+1} in G_i then $G_i = G_{i+1}t_1 \cup \cdots \cup G_{i+1}t_s$. In the example we get

$$[(), (3,5)(2,4), (2,3)(4,5), (3,4)2,5), a, (3,5)(2,4)a, (2,3)(4,5)a, \ldots,$$

starting with the 4 elements of G_1, and continuing with the cosets of G_1 put in the order given by the Schreier transversal. This puts an ordering on the elements of G. GAP orders everything.

Class Activity. Examine the list of elements of some groups, such as S_4 to see the coset structure in the list.

Other algorithms, such as computing generators for a Sylow p-subgroup of a group, or for the normalizer of a subgroup, depend on computing a stabilizer chain. This approach to computation within permutation groups is due to Charles Sims.

1.3 Nilpotent groups

As background material, we probably already know the following results.

Proposition 1.3.1. *Let p be a prime, G a finite p-group, and let $1 \neq H \lhd G$ be a non-identity normal subgroup. Then $1 \neq H \cap Z(G)$.*

Proof. G permutes the elements of H by conjugation with orbits whose lengths are powers of p. Thus the number of orbits of length 1 is divisible by p. The identity element lies in an orbit of length 1, so there is some other element also in an orbit of length 1. □

Corollary 1.3.2. *Let p be a prime and G a finite p-group with $|G| \neq 1$. Then $Z(G) \neq 1$.*

Proof. Take $H = G$ in the previous result. □

Corollary 1.3.3. *Let p be a prime and G a non-abelian group of order p^3. Then $|Z(G)| = p$.*

Proof. If $|Z(G)| = p^2$ then $G/Z(G)$ is cyclic and $G = Z(G)$. □

We define the lower and upper central series of a group G by

$$G = \gamma_1(G) \geq \gamma_2(G) \geq \cdots$$

where $\gamma_{I+1}(G) = [\gamma_i(G), G]$, and $1 = \zeta^0(G) \leq \zeta^1(G) \leq \cdots$ where $\zeta^{i+1}(G)/\zeta^i(G) = Z(G/\zeta^i(G))$.

If G is a finite p-group then $\zeta^c(G) = G$ for some c, called the *class* of G.

Theorem 1.3.4. *For any group G, there is an integer c with $\zeta^c(G) = G$ if and only if $\gamma_{c+1}(G) = 1$. Moreover, in this case, $\gamma_{i+1} \leq \zeta^{c-i}(G)$ for all*

Proof. We prove that if $\zeta^c(G) = G$ then $\gamma_{i+1} \leq \zeta^{c-i}(G)$, by induction on i. It will follow from this that $\gamma_{c+1}(G) = 1$.

Assuming that $\zeta^c(G) = G$ then $\gamma_1 \leq \zeta^c(G)$, and this starts the induction. Assuming it has been proved for i, we have

$$\gamma_{i+2}(G) = [\gamma_{i+1}(G), G] \leq [\zeta^{c-i}(G), G] \leq \zeta^{c-i-1}(G)$$

since $\zeta^{c-i}(G)/\zeta^{c-i-1}(G) = Z(G/\zeta^{c-i-1}(G))$.

The converse is proved similarly. □

Definition 1.3.5. A group with $\zeta^c(G) = G$ for some c is called *nilpotent* (of class c). It is a fact that finite nilpotent groups are direct products of p-groups.

Theorem 1.3.6. *Let G be a finite p-group.*

1. *If $H < G$ is a proper subgroup of G then $H \neq N_G(H)$.*

2. Maimal subgroups of G are normal and have index p.

Proof. 1. We find i so that $\zeta^i(G) \subseteq H$ and $\zeta^{i+1}(G) \nsubseteq H$. Then $H \triangleleft H \cdot \zeta^{i+1}(G)$, which is a larger subgroup than H.

2. If H is a maximal subgroup of G then its normalizer must be G, by part 1. The factor group G/H has no non-trivial subgroups, so is cyclic of order p. □

Definition 1.3.7. We put $\Phi(G) = \bigcap$maximal subgroups of G. This is the *Frattini subgroup* of G. We say that an element $x \in G$ is a *non-generator* of G if it can be omitted from every generating set of elements of G (that contain x).

We continue with Theorems 5.47 – 5.50 of Rotman's book.

Corollary 1.3.8. *The number of maximal subgroups of a finite p group G is $\frac{p^d-1}{p-1}$, where $d = d(G)$ is the minimum size of a set of generators of G.*

Chapter 2

Free constructions with groups

2.1 Construction of free groups

We follow the start of Chapter 11 of Rotman's book, specifically 11.1-11.6. As far as the construction of free groups is concerned, we already know the definition of a free group, have some idea that they consist of reduced words in their generators and inverses, and have seen their role in presentations of groups. The trouble is that we might not know for sure that any free groups (of rank bigger than 1) exist. We describe an algebraic construction of free groups. There is also a topological construction as the fundamental group of a graph. This might be thought to be more immediate, but it relies on first knowing what the fundamental group is, and then we must verify that it satisfies the free property. It is more direct to take an algebraic approach.

Definition 2.1.1. Let X be a subset of a group F. We say that F is *free* on X (or the *free group generated by* X etc.) if and only if for every group G and mapping $f : X \to G$ there exists a unique group homomorphism $\tilde{f} : F \to G$ extending f. We say that X is a *free set of generators* for F, or a *basis* for F. The size of X is called the *rank* of F.

Intuitively, if $X = \{x, y\}$ then $F = F(X)$ is the set of words such as $x^2yxy^{-1}x^{-5}$ etc, and such elements of F expressed as *reduced* words are distinct elements of F if and only if the words look distinct; but this needs to be proved. The free group $F(X)$ may have many bases, such as $\{x, y\}$, $\{x, xy\}$, $\{x^2y, xy\}$ and so on. We already probably know that if F is free with basis X then its abelianization is free abelian with basis the image of X, and since any two bases of a free abelian group must have the same size, any two bases of F must also have the same size. From the same considerations we see that the minimum number of generators $d(F)$ equals the size of X. It is true, but not obvious (and we will not prove it), that any set of generators of $F(X)$, with the same size as X, is a basis of $F(X)$. We also know already that any two free groups on generating sets of the same size are isomorphic (and the possible isomorphism biject with the bijections between the two sets of the same size).

Theorem 2.1.2 (Rotman's Theorem 11.1)**.** *If X is a set, there exists a group F that is free on X.*

To prove this we introduce some terminology. A *word* in X is a string $x_1^{\epsilon_1} \cdots x_n^{\epsilon_n}$ of finite length where $\epsilon_i = \pm 1$ for each i and $x_i \in X$. We write $x^1 = x$. A word is *reduced* if and only if symbols x and x^{-1} are never adjacent. Two reduced words are distinct if and only if they appear to be distinct. It is tempting to try to define F to be the set of reduced words in X and define multiplication of words by juxtaposition, followed by reduction, but the combinatorics of proving that the associative law holds in this fashion are not elegant. For example, consider the product $(y^{-1}x^{-1})(xyz)(z^{-1}y^{-1}x^{-1})$. This equals $y^{-1}x^{-1}$, but if we do the multiplication in one order these terms appear at the left end of the word, and if we do it in the other order they appear at the right end. Somehow the argument that shows associativity must cope with this phenomenon. Instead, we realize $F(X)$ as a set of permutations of the set of reduced words. Permutations automatically satisfy associativity.

Proof. For each $x \in X$ define

$$|x^\epsilon| x_1^{\epsilon_1} \cdots x_n^{\epsilon_n} = \begin{cases} x^\epsilon x_1^{\epsilon_1} \cdots x_n^{\epsilon_n} & \text{if } x^\epsilon \neq x_1^{-\epsilon_1} \\ x_2^{\epsilon_2} \cdots x_n^{\epsilon_n} & \text{otherwise} \end{cases}$$

Then both $|x||x^{-1}|$ and $|x^{-1}||x|$ act as the identity mapping on the set W of reduced words on X. It follows that $|x|$ and $|x^{-1}|$ are inverse permutations of W. Let F_0 be the subgroup of S_W generated by $\{|x| \bigm| x \in X\} = X_0$. Every element of F_0 can be written as a product $|x_1^{\epsilon_1}| \cdots |x_n^{\epsilon_n}|$ where $|x^\epsilon|$ and $|x^{-\epsilon}|$ are never adjacent. This expression is unique, because for any element $\alpha \in F_0$, if $\alpha(1) = x_1^{\epsilon_1} \cdots x_n^{\epsilon_n}$ then $\alpha = |x_1^{\epsilon_1}| \cdots |x_n^{\epsilon_n}|$.

We verify that F_0 is free on the generating set X_0. If G is any group and $f : X_0 \to G$ is a mapping, define $\tilde{f} : F_0 \to G$ by $\tilde{f}(|x_1^{\epsilon_1}| \cdots |x_n^{\epsilon_n}|) = f(|x_1|)^{\epsilon_1} \cdots f(|x_n|)^{\epsilon_n}$. This is well-defined, by the property of uniqueness. We observe, finally, that $\tilde{F}$ is a homomorphism and it is uniquely determined by f. □

Corollary 2.1.3. *Every group is a quotient of a free group.*

As a consequence, every group can be specified by a presentation $G = \langle X \bigm| R \rangle$.

Corollary 2.1.4. *Let X be a subset of a group G. Then G is freely generated by X if and only if each element of G is uniquely expressible as a reduced word in X.*

Proof. We have seen the implication '$\Rightarrow$' in the construction of $F(X)$. Conversely, suppose that each element of G is uniquely expressible as a reduced word in X, and let $F(X)$ be a free group generated by X. Consider the homomorphism $F(X) \to G$ that uniquely extends the identity mapping on X. It is surjective since X generates G. It is injective since each element of $F(X)$ is sent to a different element of G. Thus it is an isomorphism. □

2.2 Coset enumeration

We continue to follow 11.7 and 11.8 of Rotman's book and also the book by Johnson. These describe the Todd-Coxeter algorithm from 1936, implemented on computer from the 1950s onwards. For each relator of length n in a presentation we set up a table with $n+1$ columns; etc.

Theorem 2.2.1. *Suppose that the Todd-Coxeter algorithm for the presentation $G = \langle X \mid R \rangle$ terminates. Then G is finite, and the rows of the coset tables determine a permutation representation of G. This permutation representation is the regular representation.*

Proof. The rows give permutations, by construction. The permutations satisfy the relations in R. The group H generated by these permutations is an image of G. We have constructed a finite transitive permutation representation of G, and it has every other transitive permutation representation of G as an image, since the only restriction on it is that the relations in R hold. The fact that coset collapse may have occurred introduces some delicacy into justifying this last statement. From this it follows by Proposition 1.2.1 part 4 that G is finite and the permutation representation is the regular representation. □

Examples 2.2.2. $G = \langle a, b \mid b^2,\ bab^{-1}a^{-2} \rangle$ has order 6, no extra symbols are needed.

We eliminate c from the 'Fibonacci' group $F(2,3) = \langle a, b, c \mid ab = c,\ bc = a,\ ca = b \rangle$ to get $F(2,3) = \langle a, b \mid baba^{-1},\ abab^{-1} \rangle$. This has order 8 and no extra symbols are needed.

$G = \langle a, b \mid a^2,\ ababa \rangle$ has order 4, no extra symbols are needed.

$G = \langle a, b \mid a^2,\ b^2,\ (ab)^3 \rangle$ has order 6, no extra symbols are needed.

$G = \langle a, b \mid aba^{-1} = b^2,\ bab^{-1} = a^2 \rangle$ has order 1, at least 6 symbols are needed.

$G = \langle a, b, c \mid abc = b,\ bca = c,\ cab = a \rangle$ has order 48 and is not suitable for hand computation.

Coxeter presentations of D_6 and D_8 can be done by hand and need no extra symbols.

Class Activity. Find the order of the group with presentation $\langle x, y \mid xy^2,\ x^2y^3 \rangle$. How many symbols did you introduce in doing the calculation?

Proposition 2.2.3. *The presentation $T_n = \langle x, y \mid x^n y^{n+1},\ x^{n+1} y^{n+2} \rangle$ of the identity group needs at least $2n+1$ symbols to complete the Todd-Coxeter algorithm.*

Proof. Consider the tables for these relators. We can only stop introducing new symbols by completing rows, and the first row to be completed will require $2n+1$ new symbols. □

I believe it is the case, for a finite presentation of a finite group, that there is always a way to choose new symbols so that the algorithm terminates, but I do not know where to find this result in the literature (if it is true). In view of the previous result, the time taken for the algorithm to terminate is not bounded by a function of the group order. Whether or not it is true that there is always a way for the algorithm to terminate, given a presentation of a finite group, it is the case that if an implementation of the algorithm does not terminate after some time, for some presentation, we cannot deduce that the group is infinite. If we do not know that the group presented is finite in advance, it is algorithmically undecidable to determine whether the group is finite or, indeed, if it is the identity group.

2.3 Cayley graphs

These are often introduced as a pictorial way to view the way multiplication works inside a group. We will use them as a tool to characterize free groups and other related groups. Sometimes the Cayley graph of G is taken to have the elements of G as its vertices, but we make the definition more general than this.

Let G act on a set Ω (from the right) and let X be a set of generators of G. The *Cayley graph* $\Gamma(\Omega, X)$ is the graph whose vertex set is Ω and where there is an edge $\omega_1 \to \omega_2$ labeled by $x \in X$ if and only if $\omega_2 = \omega_1 x$.

Example 2.3.1. The Cayley graph of S_3 acting on $\Omega = G$ in the regular representation looks like a triangular prism. When $\Omega = \{1, 2, 3\}$ the graph has three vertices joined in a triangle with some extra edges. Also do the Cayley graph of $C_2 \times C_2$ using two generators.

Example 2.3.2. Picture of the tree for the free group of rank 2.

Proposition 2.3.3. *Let F be a group with generating set X. Then F is a free group freely generated by X if and only if $\Gamma(F, X)$ is a tree.*

Given a set X of generators of a group G we define the *length* of an element $g \in G$ (with respect to X) to be the minimal length $\ell(g)$ of a word in the $x^{\pm 1}$, $x \in X$, whose evaluation in G is g. In the case of a free group with free generators X this length is the length of the reduced word that represents the element g. In this situation, if $x \in X$ then $|\ell(gx^{\pm 1}) - \ell(g)| = 1$.

Proof. Suppose that F is freely generated by X. Each non-identity vertex $x_1^{\epsilon_1} \cdots x_n^{\epsilon_n}$ of $\Gamma(F, X)$ has one adjacent vertex of shorter length, namely $x_1^{\epsilon_1} \cdots x_{n-1}^{\epsilon_{n-1}}$, and all other adjacent vertices are strictly longer. If there were a circuit in the graph, consider a minimal circuit and a vertex in it of maximal length: it must have two distinct adjacent vertices of shorter length. But such adjacent vertices must be equal, which is a contradiction.

Conversely, suppose that $\Gamma(F, X)$ is a tree. Words in the generators and their inverses are in bijection with paths in the tree from the identity to vertices in the tree, and reduced words correspond to reduced paths. Since there is a unique reduced path from the identity to each vertex, each element of F has a unique expression as a reduced word and so F is freely generated by X, by Corollary 2.1.4 □

If $\Omega = G$ is the regular G-set then G acts on $\Gamma(G, X)$ from the *left* as a group of graph automorphisms, with an element $g \in G$ sending an edge $y \to yx$ to $gy \to gyx$.

Lemma 2.3.4. *Let G be a group with generating set X and let H be a subgroup of G. Then*

$$H\backslash\Gamma(G, X) \cong \Gamma(H\backslash G, X).$$

Proof. The vertices of $H\backslash\Gamma(G, X)$ are the orbits Hg of elements $g \in G$, and these are the right cosets of H. There is an edge $Hg \to Hgx$ in $\Gamma(H\backslash G, X)$ for each $x \in X$, and this corresponds to the H orbit of the edge $g \to gx$ of $\Gamma(G, X)$. □

Example 2.3.5. Let G be free of rank 2 on generators x, y and let $H = \langle x \rangle$. Draw $G\backslash\Gamma(G, X)$ and $H\backslash\Gamma(G, X)$.

Proposition 2.3.6. *Let G be a group with generating set X and let H be a subgroup of G. Schreier transversals for H in G biject with maximal rooted trees in the Cayley graph $\Gamma(H\backslash G, X)$, rooted at the subgroup H.*

Proof. This is a question of examining what these things mean. □

2.4 Covering spaces and free groups acting on trees

At this point we can borrow the theory of covering spaces from topology, for which a suitable reference is the book by J.R. Munkres: *Elements of Algebraic Topology*, Addison-Wesley 1984. For our application the topological spaces we study are all graphs (with loops and multiple edges), and so it is possible to apply this theory without the topological terms that apply generally.

Let K and $\tilde{K}$ be topological spaces. We say $p : \tilde{K} \to K$ is a *covering* if and only if each x in K has an open neighborhood U such that $p^{-1}U$ is a disjoint union of open subsets of $\tilde{K}$ each mapped homeomorphically onto U by p.

In the full topological situation, if K is connected and locally path connected, pick any base point x_0 in K. We say $p : \tilde{K} \to K$ is a *universal covering* if and only if $\pi_1(\tilde{K}, \tilde{x}_0) = 1$, where $p(\tilde{x}_0) = x_0$. When K and $\tilde{K}$ are graphs they are automatically locally path connected, and it is equivalent to require that $\tilde{K}$ be a tree.

Topologically, we say that an action of a group G on a space $\tilde{K}$ is *properly discontinuous* if and only if for all $x \in \tilde{K}$ there exists a neighborhood V of x such that $gV \cap V = \emptyset$ for all non-identity $g \in G$. When $\tilde{K}$ is a graph and G is acting by graph automorphisms, it is equivalent to require that the action be *free*, meaning that the stabilizer of every point in the tree is 1. In what follows, when G acts on a graph we will always mean that G acts via graph automorphisms. We will also require that G act *without inversions*, meaning that the setwise stabilizer of each edge fixes that edge pointwise. This condition is implied by a free action (on both vertices and edges).

Proposition 2.4.1. *Let G act freely on a graph Γ. Then the map $\Gamma \to G\backslash\Gamma$ is a covering.*

Corollary 2.4.2. *Let X be a set of generators for a group G. The action of G on the Cayley graph $\Gamma(G, X)$ is free. Hence the quotient map $p : \Gamma(G, X) \to G\backslash\Gamma(G, X)$ is a covering. If G is freely generated by X, it is a universal covering.*

Given a vertex v_0 in a graph Γ we may define the fundamental group $\pi_1(\Gamma, v_0)$ to be the set of equivalence classes of paths in Γ that start and end at v_0. Topologically, equivalence means that the paths are based homotopy equivalent, but in the case of graphs we can express the condition combinatorially: two such paths are equivalent if one can be obtained from the other by inserting and deleting a succession of paths consisting of an edge, followed by the same edge in the opposite direction. Multiplication in this group is determined by concatenation of paths. We follow the convention that the path on the left is the first path followed and the path on the right is the second. There may be the same issue ensuring associativity as with free groups.

Lemma 2.4.3. *If G acts freely on a tree Γ then $\pi_1(G\backslash\Gamma, v_0) \cong G$, where v_0 is a distinguished vertex of $G\backslash\Gamma$.*

Proof. Let $\hat{v}_0$ be a vertex in Γ with $p(\hat{v}_0) = v_0$. We obtain a mapping $G \to \pi_1(G\backslash\Gamma, v_0)$ as follows: for each $g \in G$ there is a unique (shortest) path α_g from $\hat{v}_0$ to $g\hat{v}_0$ and we send g to the homotopy class $[p(\alpha_g)]$. Given a circuit in $(G\backslash\Gamma, v_0)$ based at v_0 it has a unique lift to a path in Γ starting at $\hat{v}_0$. The end point of the path has the form $g\hat{v}_0$ for some g, and depends only on the equivalence class of the path. We send the circuit to g, and this defines a mapping $\pi_1(G\backslash\Gamma, v_0) \to G$. These two mappings are inverse and are group homomorphisms. □

Class Activity. How many times did we use the fact that the action is free in the last proof?

We deduce the following:

Corollary 2.4.4. *Let Y be a graph with a single vertex y_0 and n loops. Then $\pi_1(Y, y_0)$ is a free group of rank n, freely generated by the paths that consist of a single edge.*

The graph in the last result is called a *wedge* of circles or, more florally, a *bouquet* of circles.

Lemma 2.4.5. *Let Y be a connected graph with a distinguished vertex y_0, and let T be a subtree of Y that contains every vertex. Let $\overline{Y}$ be the graph with y_0 as its only vertex, and whose edges are the edges of Y that do not lie in T. Then $\pi_1(Y, y_0) \cong \pi_1(\overline{Y}, y_0)$.*

Proof. For each vertex $y \in Y$ let α_y be the geodesic from y_0 to y in T, and let α_y^{-1} denote the geodesic from y to y_0 in the opposite direction. We define a homomorphism $\pi_1(Y, y_0) \to \pi_1(\overline{Y}, y_0)$ by sending each path to the same path with the edges in T omitted. In the opposite direction we define a homomorphism $\pi_1(\overline{Y}, y_0) \to \pi_1(Y, y_0)$ as follows. If e is an edge of Y not in T, going from vertex x to vertex y, we replace it by the path $\alpha_x e \alpha_y^{-1}$. These homomorphisms are inverse on both sides, and hence are isomorphisms. □

Corollary 2.4.6. *Let Y be a connected graph with a distinguished vertex y_0.*

1. *Then $\pi_1(Y, y_0)$ is a free group of rank $-\tilde{\chi}(Y) = E - V + 1$ where E is the number of edges of Y and V is the number of vertices of Y.*
2. *Let T be a maximal subtree of Y and for each vertex y let α_y be the geodesic from y_0 to y in T. Then the elements $\alpha_v e \alpha_w^{-1}$ ranging over edges e (from v to w) that are not in T freely generate $\pi_1(Y, y_0)$.*

Proof. We know that $\pi_1(Y, y_0) \cong \pi_1(\overline{Y}, y_0)$ is free, and the rank of this free group is the number of edges of Y not in T. This number is $E - V + 1$ and the elements of $\pi_1(Y, y_0)$ that correspond to the single edges in Y_0 are free generators. These elements are the ones listed. □

Theorem 2.4.7. 1. *Let G be a group. Then G is a free group if and only if G can act freely on a tree.*

2. *Subgroups of free groups are free.*

Proof. 1. We have seen that if G is a free group then it does act freely on a tree, namely its Cayley graph. On the other hand, if G is a group that acts freely on a tree Γ then $G \cong \pi_1(G\backslash\Gamma, v_0)$, which is a free group.

2. If $H \leq G$ where G is a free group, then G acts freely on a tree Γ, hence so does H. Therefore H is free, by part 1. □

We now take the last basic result and look more carefully at the generators of the subgroup H. From the theory of groups acting on trees we obtain a result that does not mention trees.

Theorem 2.4.8 (Nielsen-Schreier)**.** *Let H be a subgroup of finite index in a free group F of finite rank $d(F)$. Then $d(H) - 1 = |F : H|(d(F) - 1)$. Furthermore, H is freely generated by the non-identity elements of the form $tx\overline{tx}^{-1}$ where t ranges over the elements of a right Schreier transversal to H in F, x ranges over the generators of F, and $\overline{tx}$ is the transversal element representing the same coset as tx.*

Proof. Let $F = F(X)$ act on its Cayley graph $\Gamma(F, X)$, so that $H\backslash\Gamma(F, X)$ is the Cayley graph $\Gamma(H\backslash G, X)$, by Lemma 2.3.4. By Proposition 2.3.6 a maximal tree in $\Gamma(H\backslash G, X)$ corresponds to a Schreier transversal for H in G. The free generators for $H \cong \pi_1(\Gamma(H\backslash G, X), v_0)$ specified as $\alpha_v e \alpha_w^{-1}$ in Corollary 2.4.6 exactly correspond to elements $tx\overline{tx}^{-1}$, and such an element equals 1 if and only if the edge e lies in T. We have seen before when considering stabilizer chains that the number of these elements is as claimed, and we can now interpret this as an Euler characteristic. If E_H, V_H are the numbers of edges and vertices in the Cayley graph $H\backslash\Gamma(F, X)$ and E_F, V_F are the corresponding numbers for $F\backslash\Gamma(F, X)$ then because these are orbit graphs under a free action we have $E_H = |F : H|E_F$ and $V_H = |F : H|V_F$. Thus the Euler characteristics behave multiplicatively: $\chi(H\backslash\Gamma(F, X)) = |F : H|\chi(\backslash\Gamma(F, X))$ and this gives the subgroup rank formula by Corollary 2.4.6. □

In the above proof, one way to see which elements of G correspond to the paths $\alpha_v e \alpha_w^{-1}$ is to consider their image in $\Gamma(G\backslash G, X)$ which is a bouquet of circles corresponding to the generators of G, but we do not have to do this, because the edges of $\Gamma(H\backslash G, X)$ are labeled with elements of X.

Example 2.4.9. The normal subgroup N of $F(a, b)$ generated by $[a, b], a^2, b^2$ has quotient $C_2 \times C_2$ and the quotient $N\backslash\Gamma(F(a, b)$ is the Cayley graph $\Gamma(C_2 \times C_2, \{\bar{a}, \bar{b}\})$, which looks like (picture). We may take a Schreier transversal $\{1, a, ab, b\}$ corresponding to a maximal tree

$$\begin{tikzcd}
b & ab \\
1 \arrow[u, "b"] \arrow[r, "a"] & a \arrow[u, "b"]
\end{tikzcd}.$$

Table of tg:

1	a	ab	b	
a	a^2	aba	ba	a
b	ab	ab^2	b^2	b

Table of $tg\overline{tg}^{-1}$:

1	a	ab	b	
1	a^2	$abab^{-1}$	$bab^{-1}a^{-1}$	a
1	1	ab^2a^{-1}	b^2	b

The five non-identity elements in the last table freely generate N, which has rank $d(N) = 4(d(F) - 1) + 1 = 5$.

Example 2.4.10. Similar with $S_3 = \langle a, b \mid a^3, b^2, baba \rangle$.

Example 2.4.11. Similar with $Q_8 = \langle a, b \mid a^4, b^4, a^2b^{-2}, bab^3a^{-3} \rangle$.

2.5 $SL(2, \mathbb{Z})$ and the tree on which it acts.

Moebius functions are mappings from the Riemann sphere $\mathbb{C} \cup \{\infty\}$ to itself, of the form

$$f(z) = \frac{az + b}{cz + d}.$$

Proposition 2.5.1. *1. Moebius functions take circles and straight lines to circles and straight lines, preserving angles.*

2. The composition of Moebius functions is given by matrix multiplication.

Proof. 1. When $c = 0$ the effect of f is to scale and translate, which has the claimed effect. When $c \neq 0$ we can write

$$f(z) = \frac{az + b}{cz + d} = \frac{\frac{a}{c}(cz + d) - \frac{ad}{c} + b}{cz + d} = \frac{a}{c} + \frac{b - \frac{ad}{c}}{cz + d}.$$

This has the effect on z of successively scaling by c, adding d, applying $z \to \frac{1}{z}$, scaling by $b - \frac{ad}{c}$, and adding $\frac{a}{c}$. Each of these operations sends circles and straight lines to circles and straight lines. The least obvious case is the mapping $z \to \frac{1}{z}$, and we give a proof in this case. Under this mapping a circle $|z-w| = \rho$, for some real ρ, becomes $|\frac{1}{z} - w| = \rho$. Thus $|1-zw| = \rho|z|$ and $|1-(u+iv)(x+iy)|^2 = (1-ux+vy)^2+(uy+vx)^2 = \rho^2(x^2+y^2)$. This is a quadratic equation in which the coefficients of x and y are the same and the coefficient of xy is zero, so it describes a circle unless the coefficient of x^2 and of y^2 is zero, in which case we get a straight line. A similar argument shows that lines are sent to circles or lines.

To say that Moebius functions preserve angles is to say that they are conformal. This is a consequence of the fact that they are analytic away from $z = -\frac{d}{c}$, using a result from complex analysis. □

In the following we write I_2 for the 2×2 identity matrix.

Proposition 2.5.2. *$SL(2, \mathbb{Z})$ acts on the Riemann sphere $\mathbb{C} \cup \{\infty\}$ by Moebius transformations. The center of $SL(2, \mathbb{Z})$ is generated by $-I_2$ and acts trivially. Thus we obtain an action of $PSL(2, \mathbb{Z}) = SL(2, \mathbb{Z})/\{\pm I_2\}$ which sends circles and straight lines to circles and straight lines, preserving angles between them. It preserves the real axis and also the upper half plane $\{z \in \mathbb{C} \mid \operatorname{Im}(z) > 0\}$. Circles and straight lines orthogonal to the real axis are sent to circles and straight lines orthogonal to the real axis.*

Proposition 2.5.3. *$SL(2, \mathbb{Z})$ is generated by elements*

$$\alpha = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \textit{and} \quad \beta = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}$$

of orders 4 and 6 with square equal to $-I_2$.

Proof. Let

$$\gamma = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Then $\beta = \gamma\alpha$ and $\gamma = \beta\alpha^{-1}$. We show that $SL(2, \mathbb{Z}) = \langle \alpha, \gamma \rangle$ and this will suffice. Notice that

$$\alpha\gamma\alpha^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \in \langle \alpha, \gamma \rangle$$

so that

$$\begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ n & 1 \end{bmatrix} \in \langle \alpha, \gamma \rangle$$

for all integers n. These matrices correspond to integer elementary operations and by left multiplication by these and α we can reduce matrices to I_2. For example,

$$\begin{bmatrix} 3 & 1 \\ 11 & 4 \end{bmatrix} \xrightarrow{\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \xrightarrow{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \xrightarrow{\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

□

Proposition 2.5.4. *Let e be the arc of the unit circle between i and $\omega = \frac{1+i\sqrt{3}}{2}$. Then*

$$\mathrm{Stab}_{SL(2,\mathbb{Z})}(i) = \langle \alpha \rangle = \{\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \pm \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\},$$

$$\mathrm{Stab}_{SL(2,\mathbb{Z})}(\omega) = \langle \beta \rangle = \{\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \pm \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}, \pm \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}\},$$

and

$$\mathrm{Stab}_{SL(2,\mathbb{Z})}(e) = \{\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\}.$$

The set-wise stabilizer of e fixes e pointwise, so it makes no difference in the last equality which stabilizer we consider.

Proof. To compute the stabilizer of i: $\frac{ai+b}{ci+d} = i$ if and only if $ai + b = -c + di$, so that $a = d$ and $b = -c$. Thus elements of $SL(2, \mathbb{Z})$ stabilizing i have the form $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ and the determinant is $a^2 + b^2 = 1$. This forces $a = \pm 1$ and $b = 0$ or $b = \pm 1$ and $a = 0$, giving the four matrices in $\langle \alpha \rangle$.

For the stabilizer of ω, note that $\omega^2 = \omega - 1$. If $\frac{a\omega+b}{c\omega+d} = \omega$ then

$$a\omega + b = c\omega^2 + d\omega = c(\omega - 1) + d\omega = (c + d)\omega - c.$$

Thus $a = c + d$ and $b = -c$. The matrix $\begin{bmatrix} c+d & -c \\ c & d \end{bmatrix}$ has determinant $c^2 + cd + d^2 = 1$. If $c \neq 0 \neq d$ then, say, $0 < |c| \leq |d|$ and $1 = c^2 + cd + d^2 \geq c^2 > 0$. Thus $c = 1$, in which case $d = -1$, or $c = -1$, in which case $d = 1$. The other assumption $0 < |d| \leq |c|$ gives the same solutions. Otherwise $c = 0$ implies $d = \pm 1$ and $d = 0$ implies $c = \pm 1$. This gives the 6 matrices of $\langle \beta \rangle$.

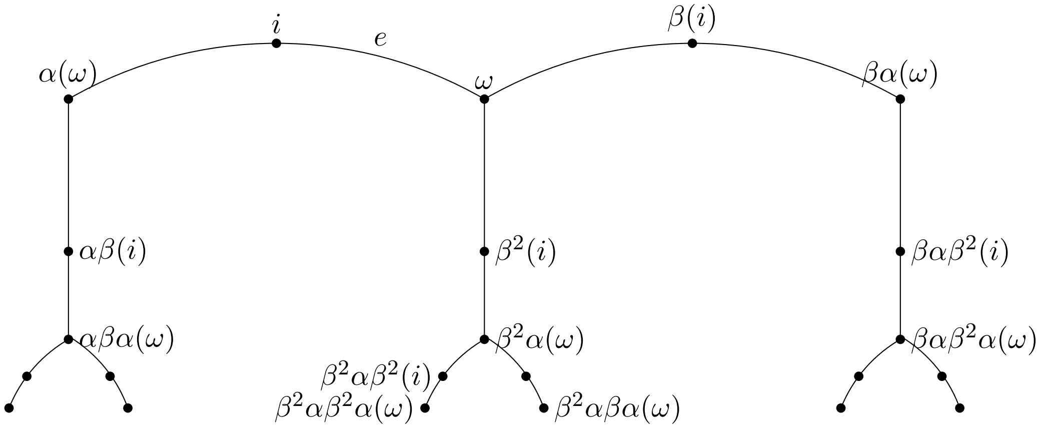
If an element of $SL(2, \mathbb{Z})$ stabilizes the arc e, it must send the unit circle to itself. From the expression of a Moebius transformation as a composite of translations, scalings and an inversion we deduce that the elements in $SL(2, \mathbb{Z})$ that preserve the unit circle are the four matrices in $\mathrm{Stab}_{SL(2,\mathbb{Z})}(i)$. Of these matrices, the only ones that stabilize e are $\pm I_2$, and these matrices stabilize e pointwise. We could also say that because $SL(2, \mathbb{Z})$ preserves the real axis, if an element preserves the unit circle it must preserve $\{\pm 1\}$, and argue from there. □

Let $T = SL(2, \mathbb{Z}) \cdot e$ be the subset of $\mathbb{C}$ that is the union of the arcs ge as g ranges through $SL(2, \mathbb{Z})$. Our goal to show that T has the structure of a tree, with vertices the points gi and $g\omega$, and edges the arcs ge, where $g \in SL(2, \mathbb{Z})$.

Corollary 2.5.5. *The subset T of the upper half plane is connected.*

Proof. The arc e is connected, and G is generated by the stabilizers of its points. □

We next construct some points in T.


i
e
$\beta(i)$
$\alpha(\omega)$
ω
$\beta\alpha(\omega)$
$\alpha\beta(i)$
$\beta^2(i)$
$\beta\alpha\beta^2(i)$
$\alpha\beta\alpha(\omega)$
$\beta^2\alpha(\omega)$
$\beta\alpha\beta^2\alpha(\omega)$
$\beta^2\alpha\beta^2(i)$
$\beta^2\alpha\beta^2\alpha(\omega)$
$\beta^2\alpha\beta\alpha(\omega)$

Some of the points just plotted are $\omega = \frac{1}{2} + i\frac{\sqrt{3}}{2}$, $\beta(i) = 1 + i$, $\beta^2(i) = \frac{1}{2} + i\frac{1}{2}$, $\beta^2\alpha(\omega) = \frac{1}{2} + i\frac{\sqrt{3}}{6}$, $\beta^2\alpha\beta^2(i) = \frac{2}{5} + i\frac{1}{5}$, $\beta^2\alpha\beta^2\alpha(\omega) = \frac{5}{14} + i\frac{\sqrt{3}}{14}$.

A *fundamental domain* for the action of a group on a space is a set of orbit representatives. When the space has some topological or geometric structure we usually expect that the person using the term has in mind some further property, such as that the fundamental domain is connected.

Proposition 2.5.6. *T is a tree whose vertices are the images of i and ω under $SL(2,\mathbb{Z})$, which acts simplicially on this tree, with fundamental domain e. The stabilizer of every vertex is thus a conjugate of* $\mathrm{Stab}_{SL(2,\mathbb{Z})}(i)$ *or* $\mathrm{Stab}_{SL(2,\mathbb{Z})}(\omega)$.

Proof. We construct the union of the images of e in stages. Every point in this union can be written $(\alpha)\beta^{n_1}\alpha\beta^{n_2}\alpha\cdots\beta^{n_t}(\alpha)(x)$ for some x in e and optional α at the ends. We assemble the union according to the number of alternating terms that are powers of α and β. We first join on to e the images $\alpha(e), \beta(e)$ and $\beta^2(e)$. These join on to e at their vertices. Next we glue on $\beta\alpha(e)$, $\beta^2\alpha(e)$, $\alpha\beta(e)$ and $\alpha\beta^2(e)$. These connect at vertices to the vertices of the previous edges. We continue in this way and at every stage edges are joined on at vertices to what has already been constructed, so that T is connected. We can see from this construction that there is an arm of the tree where every vertex begins α, and two other arms where the vertices begin with β, and with β^2. Orientations of edges around each vertex are preserved.

We will get a tree as long as no two edges constructed in this way overlap. Suppose that this were to happen: two edges overlap. They will be labelled by words in α and β applied to e. If those words start the same at the left, we can apply the inverse of the common subword to show that we may assume these words start differently, either beginning with α and β, or with α and β^2, or with β and β^2. By applying a power of β we may assume they begin with α and β^2, putting the place where these edges cross in the region under i. Now among the closest vertices in the α arm to those in the β^2 arm are those of the form $(\alpha\beta)^n(i) = \frac{-n}{1+n^2} + \frac{i}{1+n^2}$, since $\alpha\beta = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}$, so that $(\alpha\beta)^n$ acts as $\begin{bmatrix} 1 & 0 \\ -n & 1 \end{bmatrix}$. These vertices have negative real part, and similarly the corresponding vertices in the β^2 arm have positive real part. This shows that edges do not cross. Therefore the set is a tree. □

Each element of $SL(2, \mathbb{Z})$ is the evaluation of a word

$$(\pm\alpha), \pm\beta^{i_1}, \pm\alpha, \pm\beta^{i_2}, \ldots, (\pm\alpha)$$

with optional first and last elements $(\pm\alpha)$ and where $i_j \in \{1, 2\}$. Because the elements $\pm I_2$ are central this can also be written as the evaluation of a word

$$\pm(\alpha), \beta^{i_1}, \alpha, \beta^{i_2}, \ldots, (\alpha).$$

Define the length of such a word to be the number of its terms.

Proposition 2.5.7. *The edges of the tree T are labelled by the words of the above form in such a way that, for each word w, the edge $\tilde{w}(e)$ is labeled by $\pm w$, where $\tilde{w}$ is the evaluation of w in $SL(2, \mathbb{Z})$. The distance of the edge labeled by w from e is the length of w.*

Proof. This comes from the proof that T is a tree. □

We deduce a normal form theorem for elements of $SL(2, \mathbb{Z})$.

Corollary 2.5.8. *Each element of $SL(2, \mathbb{Z})$ can be uniquely written*

$$\pm(\alpha)\beta^{i_1}\alpha\beta^{i_2}\ldots(\alpha)$$

with the $i_j \in \{1, 2\}$ and the first and last (α) optional.

Proof. This is because if $g \in SL(2, \mathbb{Z})$ can be written as such a word, that word is the label of the edge ge, and different words label different edges. □

We deduce more remarkable things about $SL(2, \mathbb{Z})$.

Corollary 2.5.9. *Every torsion-free subgroup of $SL(2, \mathbb{Z})$ is free.*

Proof. The stabilizer of any point of the tree is a finite group, and so no non-identity element of a torsion-free subgroup can stabilize any point of the tree. Because the action is simplicial it follows that the action of the subgroup is free. The torsion-free subgroup thus acts freely on a tree and so must be a free group. □

We can also make a deduction about finite subgroups of $SL(2, \mathbb{Z})$. We say that a group acting on a tree acts *without inversions* if the stabilizer of every edge stabilizes it pointwise.

Lemma 2.5.10. *Every finite group G acting on a tree without inversions must stabilize a vertex.*

Proof. Let v be any vertex of the tree and consider the finite orbit Gv. The set of geodesics between all pairs of vertices in Gv form a G-stable finite subtree. If this subtree is a single vertex, it is fixed. If it is an edge, it is also fixed. Otherwise it has a vertex of valence at least 2. We can remove a G-orbit of twigs from this tree to get a smaller G-stable subtree, and repeat this process until we are left with an edge or a vertex. □

Corollary 2.5.11. *Every finite subgroup of $SL(2,\mathbb{Z})$ is conjugate to a subgroup of $\langle\alpha\rangle$ or $\langle\beta\rangle$, and hence is cyclic of order 1, 2, 3, 4, or 6.*

Proof. A finite subgroup must be contained in the stabilizer of some vertex, and that stabilizer is a conjugate of $\langle\alpha\rangle$ or $\langle\beta\rangle$. □

The last result could also be proved by considering rational canonical forms of matrices.

Example 2.5.12. The example of $SL(2,\mathbb{Z})$ is spectacular, but we can do something similar with the infinite dihedral group $\mathbb{Z} \rtimes C_2$, where the non-identity element of C_2 acts by inversion on $\mathbb{Z}$. This group can be realized as the group of rigid motions of $\mathbb{R}$ generated by reflection in 0 and reflection in 1, so that it acts on the tree that is $\mathbb{R}$, with the integers as its vertices. We now make similar deductions as for $SL(2,\mathbb{Z})$: that torsion-free subgroups are free, and that finite subgroups are conjugate to subgroups of one of the stabilizer C_2. In this example we did not need the technology of the tree to see this, but it is interesting to see this theory at work.

Definition 2.5.13. If $H = \langle x_i \bigm| r_j\rangle$ and $K = \langle y_k \bigm| s_l\rangle$ the *free product* of H and K is the group $H * K$ with presentation $H * K = \langle x_i, y_k \bigm| r_j, s_l\rangle$. There are group homomorphisms $\theta : H \to H * K$ and $\phi : K \to H * K$ sending symbols x_i and y_j to the same symbols. Every element of $H*K$ can be written as a product $\theta(h_1)\phi(k_1)\cdots\theta(h_n)\phi(k_n)$ for some n, where the h_i lie in H and the k_j lie in K. More generally, suppose J is a subgroup of H that is also isomorphic to a subgroup of K via some injective homomorphism $\alpha : J \to K$. We define the *free product with amalgamation*

$$H *_J K = (H * K)/N = \langle x_i, y_k \bigm| r_j,\, s_l,\, u\phi(u^{-1}), u \in J\rangle$$

where N is the normal subgroup generated by elements $\theta(u)\phi(\alpha(u))^{-1}$ as u ranges through elements of J.

Example 2.5.14. The free product of a free group of rank r with a free group of rank s is a free group of rank $r + s$.

Proposition 2.5.15. *There are isomorphisms $C_4 *_{C_2} C_6 \to SL(2,\mathbb{Z})$ and $C_2 * C_3 \to PSL(2,\mathbb{Z})$.*

Proof. $C_4 *_{C_2} C_6$ is defined by a presentation $\langle x, y \bigm| x^4,\, y^6,\, x^2y^{-3}\rangle$ and there is a surjective homomorphism $C_4 *_{C_2} C_6 \to SL(2,\mathbb{Z})$ sending $x \mapsto \alpha$ and $y \mapsto \beta$ since the relations for x, y are satisfied by $\alpha\beta$. It is surjective since α, β generate $SL(2,\mathbb{Z})$.

It is injective since every element of $C_4 *_{C_2} C_6$ can be written $(x^2)(x)y^{i_1}xy^{i_2}\cdots(x)$ with $i_j \in \{1,2\}$ and distinct such words get sent to distinct elements of $SL(2,\mathbb{Z})$. Finally, $C_2 * C_3$ is obtained by factoring out $\langle x^2\rangle$ from $C_4 *_{C_2} C_6$ and $PSL(2,\mathbb{Z})$ is obtained from $SL(2,\mathbb{Z})$ by factoring out the image $\{\pm I_2\}$. This accounts for the second isomorphism. □

Corollary 2.5.16. *Every element of $C_4 *_{C_2} C_6 = \langle x, y \mid x^4,\, y^6,\, x^2y^{-3}\rangle$ can be uniquely written in the form $(x^2)(x)y^{i_1}xy^{i_2}\cdots(x)$.*

Proposition 2.5.17. *There are surjective group homomorphisms $C_4 *_{C_2} C_6 \to C_{12}$ and $SL(2,\mathbb{Z}) \to C_{12}$ which sends the generators x, y and α, β of these groups to elements of orders 4 and 6 in C_{12}.*

Proposition 2.5.18. *The kernel K of the surjective homomorphism $SL(2,\mathbb{Z}) \to C_{12}$ is the commutator subgroup of $SL(2,\mathbb{Z})$. It acts freely on the tree of $SL(2,\mathbb{Z})$ and hence is a free group. It is free of rank 2. The abelianization of $SL(2,\mathbb{Z})$ is cyclic of order 12.*

Proof. The stabilizers of vertices of the tree are the conjugates of $\langle\alpha\rangle$ and $\langle\beta\rangle$. Under the homomorphism they are mapped to the conjugates of the subgroups C_4 and C_6 of C_{12}, namely those subgroups. Thus all vertex stabilizers inject into C_{12} and hence no non-identity element of K can stabilize any vertex (or edge) of the tree. Thus K acts freely on the tree, and so is free. That K has rank 2 is shown in the exercises. The fact that the abelianization of $SL(2,\mathbb{Z})$ is cyclic of order 12 follows because the same is true of $C_4 *_{C_2} C_6$. □

Just to record the result, we have already seen the following:

Proposition 2.5.19. *Each word of length n in the non-identity elements $\pm\alpha$, $\pm\beta$, $\pm\beta^2$ of $SL(2,\mathbb{Z})$ sends e to an edge distant n from e.*

The proof has already been given, but here it is again in different words.

Proof. We proceed by induction on n. When $n = 0$ the result is true. Also when $n = 1$ the word sends e to an adjacent edge, joined to e at the vertex stabilized by the group element. In general, supposing that the path in the tree from e to $h_1g_2h_2\ldots(e)$ has length $n-1$ and that it starts by passing through the vertex stabilized by h_1, we see that the path in the tree from e to $g_1h_1g_2h_2\ldots(e)$ has length n and that it starts by passing through the vertex stabilized by g_1. The argument is similar whether the left term at the left of the word is in C_2 or C_3. This completes the induction step. □

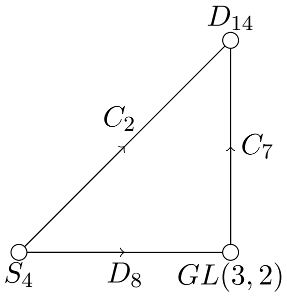
2.5.1 Free products with amalgamation, HNN-extensions and Bass-Serre theory

We have seen in a particular case that if a group acts on a tree then many things may be deduced about that group, including a presentation. Bass-Serre theory shows that every group acting on a tree has a presentation as an iterated free products with amalgamation or HNN-extensions, as well as describing the subgroup and conjugacy structure of such groups. It also shows that a group with such a presentation can act on a tree without inversions in such a way that the stabilizers in the action are the groups that appear in the presentation. The first part if this is useful in describing the structure of groups that interest us for different reasons, and that we know act on a tree. The second part has to do with considering groups given by presentations, that we might not otherwise be interested in.

Many consequences, such as the injectivity of the factors into a free product with amalgamation and the description of subgroups of a free product (the 'Kurosh subgroup theorem') were known prior to the approach of Bass and Serre, proved by arguments using manipulation of elements, at a time when the subject was known as 'combinatorial group theory'. An achievement of Bass-Serre theory was the use of group actions on trees to prove these, and other, results. These days, this theory is considered a part of what is known as 'geometric group theory', although this term arose later as a result of contributions of Gromov.

Definition 2.5.20. A *graph of groups* $(\mathcal{G}, Y)$ is a connected directed graph Y with vertex set V and edge set E, together with a function $\mathcal{G}$ that assigns to each vertex $v \in V$ a group $\mathcal{G}(v)$, and to each directed edge $e \in E$ with origin $o(e)$ and terminus $t(e)$ a subgroup $\mathcal{G}(e) \leq \mathcal{G}(o(e))$ and an injective homomorphism $\phi_e : \mathcal{G}(e) \to \mathcal{G}(t(e))$. The $\mathcal{G}(v)$ are called the *vertex groups* and the $\mathcal{G}(e)$ are called the *edge groups*.

Example 2.5.21. A graph of groups may be described by a diagram such as



where an edge $\overset{H}{\circ}\overset{J}{\longrightarrow}\overset{K}{\circ}$ means there is a subgroup $J \leq H$ and a specified injective homomorphism $\phi : J \to K$.

Suppose we are given an action of a group G on a connected graph Γ, without inversions. We can construct a graph of groups as follows. We choose first a fundamental domain Y (a set of orbit representatives) of a special kind for the action of G on Γ.

Some authors call this set of orbit representatives a *fundamental transversal.* It consists of a subtree Y_0 of Γ together with some other edges e for which $o(e) \in Y_0$. For each such edge e there is an element $g_e \in G$ and a vertex $v = \bar{t}(e)$ in Y so that $t(e) := g_e\bar{t}(e)$. On edges $e \in Y_0$ we take $\bar{t}(e) = t(e)$ and $g_e = 1$. We now define the graph $\overline{Y}$ to have the same vertices and edges as Y, but with the modified terminus function $\bar{t}$. It is isomorphic to the quotient graph $G\backslash\Gamma$. We define $\mathcal{G}(v) = G_v$ and $\mathcal{G}(e) = G_e$ for vertices and edges in $\overline{Y}$. For edges $e \in Y$ the homomorphism ϕ_e is $\phi_e(x) = g_e^{-1}xg_e$. Thus when $e \in Y_0$ this homomorphism is the inclusion of subgroups. Note that

$$G_e \leq G_{t(e)} = G_{g_e\bar{t}(e)} = g_eG_{\bar{t}(e)}g_e^{-1}$$

so that $\phi_e(x) \in G_{\bar{t}(e)}$. We now have graph of groups $(\mathcal{G}, \overline{Y})$.

Example 2.5.22. Starting with $SL(2, \mathbb{Z})$ we constructed a tree on which $SL(2, \mathbb{Z})$ acts. A fundamental transversal for this action consists of the edge $\overset{i}{\circ}\xrightarrow{e}\overset{\omega}{\circ}$ in the complex plane, and in this case it equals the subtree Y_0. The corresponding graph of groups is

$$\overset{\langle\alpha\rangle}{\circ}\xrightarrow[\{\pm I_2\}]{}\overset{\langle\beta\rangle}{\circ}$$

Examples 2.5.23. It is known that other groups of this kind act on trees giving graphs of groups as follows. For $SL_2(\mathbb{Z}[1/p])$ the graph of groups is

$$\overset{SL_2(\mathbb{Z})}{\circ}\xrightarrow[\Gamma_0(p)]{}\overset{SL_2(\mathbb{Z})}{\circ}$$

where $\Gamma_0(p)$ consists of the matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $c \equiv 0 \pmod p$.

When k is a field the group $GL_2(k[X])$ acts on a tree giving graph of groups

$$\overset{GL_2(k)}{\circ}\xrightarrow[B(k)]{}\overset{B(k[X])}{\circ}$$

where $B(R)$ is the group of invertible upper-triangular matrices with entries in R. These examples are described in Serre's book, 'Trees'.

Implicit in the above construction is that we can always find a fundamental domain of the kind specified.

Proposition 2.5.24. *Let G act on a connected graph Γ without inversions. There exists a fundamental transversal Y for the action, consisting of a subtree Y_0, and some other edges whose starting vertices lie in Y_0.*

Proof. Consider the set of subtrees of Γ for which all vertices and edges lie in different orbits. By Zorn's lemma there is a maximal such subtree Y_0, and we argue that it must contain a complete set of representatives of orbits of vertices of Γ. Now for each remaining orbit of edges of Γ we can find a representative whose starting vertex lies in Y_0. □

Given a graph of groups we can go in the opposite direction and construct a group that acts on a graph (in fact, a tree). We construct the action on the graph later, and first construct the group by means of a presentation.

Definition 2.5.25. Let $(\mathcal{G}, Y)$ be a graph of groups. Choose a maximal subtree $Y_0 \subseteq Y$. We define the *fundamental group of the graph of groups* (with respect to Y_0) to be the group $\pi(\mathcal{G}, Y, Y_0)$ given by the presentation with

- generators: all $\mathcal{G}(v)$ where $v \in V$ runs through the vertices of Y, and also a set $\{t_e \mid e \in E\}$;
- relations: defining relations for all the vertex groups $\mathcal{G}(v)$; $t_e^{-1} g t_e = \phi_e(g)$ for all $e \in E$, $g \in \mathcal{G}(e)$; $t_e = 1$ for all edges e in Y_0.

It is a lemma in the theory that changing the choice of maximal subtree Y_0 does not change the isomorphism type of the fundamental group.

Example 2.5.26. When $(\mathcal{G}, Y) = \overset{H}{\circ}\overset{J}{\longrightarrow}\overset{K}{\circ}$ the fundamental group is the free product with amalgamation $H *_J K$.

Example 2.5.27. When $(\mathcal{G}, Y)$ consists of n loops all fastened at a single vertex, with all vertex and edge groups the identity, the fundamental group is a free group F_n of rank n. In the other direction, for the action of F_n on its Cayley graph we could take a fundamental transversal Y to consist of the single vertex 1, together with the n edges emanating from it labeled by the generators of F_n (without the vertices at their other ends). The corresponding graph of groups $(\mathcal{G}, \overline{Y})$ is now a single vertex with n loops attached, with all stabilizer groups 1.

Definition 2.5.28. When a graph of groups consists of a single vertex and a single edge (which is a loop) the corresponding fundamental group is called an *HNN-extension*, named for G. Higman, B.H. Neumann and H. Neumann who were authors of a paper that appeared in 1949 where these groups were considered. Specifically, if H is a subgroup of $G = \langle x_i \mid r_j \rangle$ and we have an injective homomorphism $\phi : H \to G$ the HNN-extension is

$$G *_{H,\phi} = \langle x_i, t \mid r_j,\ \phi(h) = t^{-1} h t \text{ for all } h \in H \rangle.$$

Higman, Neumann and Neumann proved that the canonical homomorphism $G \to G *_{H,\phi}$ is injective, thus resolving positively the following question: if H and K are isomorphic subgroups of G, does there exist a group L, containing G as a subgroup, in which H and K are conjugate? They iterated the construction to produce, for instance, infinite simple groups in which every pair of non-identity elements are conjugate.

We remark that both free products with amalgamation and HNN-extension may be defined as groups having certain universal properties, which we leave the reader to formulate. The property is of the kind that there is a unique homomorphism to each group that contains a realization of the amalgamation information contained in the corresponding graph of groups.

Assuming that vertex groups embed into free products with amalgamation and HNN-extensions, we may see that, given a graph of groups, the fundamental group of the graph of groups is obtained by performing these two constructions successively. The free products with amalgamation may be performed with the vertex and edge groups in the subtree Y_0, and the remaining edges then give rise to HNN-extensions.

Example 2.5.29. Consider the graph of groups with a single vertex and a single edge, both labeled by the same infinite cyclic group $\mathbb{Z}$. Let the homomorphism $\phi : \mathbb{Z} \to \mathbb{Z}$ by multiplication by some integer n. Then $\mathbb{Z} *_{\mathbb{Z},\phi} \cong \mathbb{Z}[\frac{1}{n}] \rtimes \langle t \rangle$ where conjugation by t is multiplication by n on the subgroup $\mathbb{Z}[\frac{1}{n}]$.

Lemma 2.5.30. *Let G act on a graph Γ without inversions with corresponding graph of groups $(\mathcal{G}, \overline{Y})$. If Γ is connected then G is generated by the vertex groups $\mathcal{G}(v) = G_v$ and the elements g_e corresponding to the vertices and edges of $\overline{Y}$.*

Proof. First proof in the simplified case that Y is a tree. Let $H = \langle G_v \bigm| v \in Y \rangle$ be the subgroup generated by the vertex groups. Then we claim that $HY \cap (G - H)Y = \emptyset$. This is because if $hv_1 = gv_2$ with $v_1, v_2 \in Y$ (and thinking that $h \in H$ and $g \in G - H$) then $v_1 = v_2$ and $g^{-1}h \in \operatorname{Stab}(v_1)$, so that $g \in H$. Now if $H \neq G$ we see that X is the union of two subgraphs that do not intersect, a contradiction.

General proof: Step 1: We show that the edges in Γ with an end point v in Y have the form $v \xrightarrow{ge} gg_e w$ where $g \in G_v$, and where $v \xrightarrow{e} g_e w$ is an edge in Y, with the vertex $w = \bar{t}(e)$ in Y; or $v \xleftarrow{hg_e^{-1}e} hg_e^{-1} w$ where $h \in G_v$ and $g_e v \xleftarrow{e} w$ is an edge of Y. To see this, an edge starting at v has the form ge for some $e \in Y$ and the other end of e is $g_e w$ for some vertex $w \in Y$. The start of e must be v because $v = o(ge) = go(e)$ and v is in the same orbit as $o(e)$, and in Y, so $v = o(e)$. Therefore $v = gv$ and $g \in G_v$. An edge ending at v has the form ge for some $e \in Y$ and now $v = gt(e) = gg_e\bar{t}(e)$. Since $v, \bar{t}(e)$ belong to Y and lie in the same G-orbit we have $v = \bar{t}(e)$ and $h = gg_e \in G_v$. Thus $g = hg_e^{-1}$.

Step 2: Let H be the subgroup of G generated by the vertex groups G_v and the elements g_e. We show that $HY = \Gamma$. To see this, observe that all edges with a vertex in Y, together with their vertices, lie in HY by Step 1. From this we next deduce that all edges of Γ incident with HY lie in HY, as do their vertices. This is because if an edge e has vertex hv with v a vertex of Y and $h \in H$, then $h^{-1}e$ is incident with $v \in Y$, so that $h^{-1}e \in HY$. Thus $e \in hHY = HY$. Finally we deduce that $HY = \Gamma$ since Γ is connected.

Step 3: To show that $H = G$, let $g \in G$ and let v be any vertex of Y. Then

$gv = hv'$ for some $h \in H$ and $v' \in Y$. Since Y is a set of orbit representatives, $v = v'$, and $h^{-1}g \in G_v$. Thus $g \in hG_v \subseteq H$. □

Corollary 2.5.31. *Let G act on a connected graph Γ without inversions with corresponding graph of groups $(\mathcal{G}, \overline{Y})$. There is a surjective homomorphism $\pi(\mathcal{G}, \overline{Y}, Y_0) \to G$ extending the identity maps on the vertex groups G_v and sending t_e to g_e for each edge e in Y.*

Proof. There is a homomorphism extending the identity maps on the vertex groups because the relations that define the fundamental group are also satisfied in G. It is surjective by the lemma. □

Theorem 2.5.32. *Let G act on a connected graph Γ without inversions with corresponding graph of groups $(\mathcal{G}, \overline{Y})$. The homomorphism $\pi(\mathcal{G}, \overline{Y}, Y_0) \to G$ is an isomorphism if Γ is a tree.*

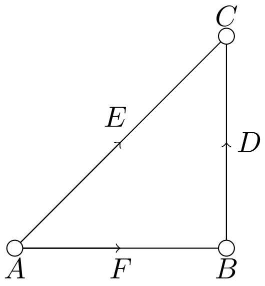
Proof. The argument follows the analogous argument for $SL(2, \mathbb{Z})$ but is notationally more complicated. We show that the homomorphism $\pi(\mathcal{G}, \overline{Y}, Y_0) \to G$ is injective.

As a first step, we claim that we can write any element of $\pi(\mathcal{G}, \overline{Y}, Y_0)$ in the form

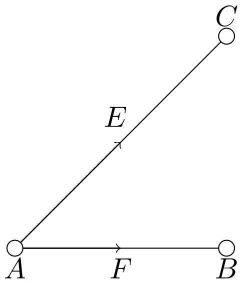
$$x_0 t_{e_1}^{\pm 1} x_1 t_{e_2}^{\pm 1} \cdots t_{e_n}^{\pm 1} x_n$$

where, for each i, $x_i \in G_{v_i}$ and there are edges in $\overline{Y}$ of the form $v_{i-1} \xrightarrow{e_i} v_i$ if $t_{e_i}^{+1}$ appears, and $v_{i-1} \xleftarrow{e_i} v_i$ if $t_{e_i}^{-1}$ appears, and furthermore $v_n = v_0$. Do this by writing any element as a product of stabilizer elements and the t_e, and then insert $1 \in G_v$ or $t_e = 1$ for $e \in Y_0$, along paths in Y_0 between suitably chosen vertices to get the result. Example:

$\overline{Y} =$



with $Y_0 =$



With suggestive notation we have

$$x_B t_D^{-1} x_C = x_B t_F^{-1} 1_A t_E 1_C t_D^{-1} 1_B t_F^{-1} 1_A t_E x_C t_E^{-1} 1_A t_F 1_B$$

of length 7.

Class Activity. What is the length of the similar shortest expression for

$$x_C t_F x'_C t_D x_A?$$

Every such element determines a path in Γ starting at v_0 as follows:

- $x_0 t_{e_1}$ determines $v_0 \xrightarrow{x_0 e_1} x_0 g_{e_1} v_1$
- $x_0 t_{e_1}^{-1}$ determines $v_0 \xleftarrow{x_0 g_{e_1}^{-1} e_1} x_0 g_{e_1}^{-1} v_1$
- $x_1 t_{e_2}$ determines $v_1 \xrightarrow{x_1 e_2} x_1 g_{e_2} v_2$
- $x_1 t_{e_2}^{-1}$ determines $v_1 \xleftarrow{x_1 g_{e_2}^{-1} e_2} x_1 g_{e_2}^{-1} v_2$
- etc.

We splice these, multiplying the second edge by $x_0 g_{e_1}^{\pm 1}$ etc, to get, for example,

$$v_0 \xrightarrow{x_0 e_1} x_0 g_{e_1} v_1 \xrightarrow{x_0 g_{e_1} x_1 e_2} x_0 g_{e_1} x_1 g_{e_2} v_2$$

in case we have g^{+1} both times. The path ends with

$$\to x_0 g_{e_1}^{\pm 1} x_1 g_{e_2}^{\pm 1} \cdots g_{e_n}^{\pm 1} v_n$$

and $v_n = v_0$, noticeing that $x_n v_n = v_n$.

We are now in a position to show that the homomorphism $\pi(\mathcal{G}, \overline{Y}, Y_0) \to G$ is injective. Let $x_0 \cdots x_n$ be an element of $\pi(\mathcal{G}, \overline{Y}, Y_0)$ that is sent to 1. We show that it equals 1 by induction on n. When $n = 0$, $x_0 \in G_{v_0}$ and the composite $G_{v_0} \to \pi(\mathcal{G}, \overline{Y}, Y_0) \to G$ is the identity on G_{v_0} so the map $\pi(\mathcal{G}, \overline{Y}, Y_0) \to G$ is injective on the subgroup generated by G_{v_0}. Now suppose $n > 0$ and the result is true for smaller values. Let $x = x_0 \cdots x_n = 1$ in G. Then the path in Γ starting at v_0 finishes at $x x_n^{-1} v_n = x v_0 = v_0$, since $x_n \in G_{v_n}$. Because Γ is a tree, the path has an extreme vertex $y x_{i-1} g_{e_i}^{\pm 1} v_i$ and here $e_i = e_{i+1}$ and $v_{i-1} = v_{i+1}$. The path around this vertex looks like (in the case $g_{e_i}^{+1}$)

$$y v_{i-1} \xrightarrow{y x_{i-1} e_i} y x_{i-1} g_{e_i} v_i \xleftarrow{y x_{i-1} g_{e_i} x_i g_{e_i}^{-1} e_i} y x_{i-1} g_{e_i} x_i g_{e_i}^{-1} v_{i+1}.$$

Now $y x_{i-1} e_i = y x_{i-1} g_{e_i} x_i g_{e_i}^{-1} e_i$ so $e_i = g_{e_i} x_i g_{e_i}^{-1} e_i$, and $h = g_{e_i} x_i g_{e_i}^{-1} \in G_{e_i}$. Since the relation $x_i = t_{e_i}^{-1} h t_{e_i}$ holds in $\pi(\mathcal{G}, \overline{Y}, Y_0)$, we have $t_{e_i} x_i t_{e_i}^{-1} \in G_{v_{i-1}}$ in $\pi(\mathcal{G}, \overline{Y}, Y_0)$, because $G_{e_i} \subseteq G_{v_{i-1}}$. Thus $x_{i-1} t_{e_i} x_i t_{e_i}^{-1} x_{i+1} = k \in G_{v_{i-1}}$ and x is equal in $\pi(\mathcal{G}, \overline{Y}, Y_0)$ to a word of length $n - 2$ which maps to 1. By induction it is 1 in $\pi(\mathcal{G}, \overline{Y}, Y_0)$, so $x = 1$ in $\pi(\mathcal{G}, \overline{Y}, Y_0)$. The argument when the extreme edges have opposite orientation is similar. □

We comment that the rather complicated looking pair of extreme edges in the above proof comes from applying group elements to the more basic edges

$$v_{i-1} \xrightarrow{x_{i-1} e_i} x_{i-1} g_{e_i} v_i \quad \text{and} \quad v_i \xleftarrow{x_i g_{e_i}^{-1} e_i} x_i g_{e_i}^{-1} v_{i+1}.$$

Class Activity. Is the argument when $n = 0$ legitimate in the above proof? Isn't there something to be shown in establishing that the stabilizer groups inject into a free product with amalgamation or an HNN extension? Is there something different here?

Theorem 2.5.33 (Main theorem of Bass-Serre theory)**.** *Let $(\mathcal{G}, Y)$ be a graph of groups. The following are equivalent for a group G.*

1. *G can act simplicially on a tree without inversions, with associated graph of groups $(\mathcal{G}, Y)$.*
2. *$G \cong \pi(\mathcal{G}, Y, Y_0)$ is isomorphic to the fundamental group of the graph of groups.*

Proof. We have shown that 1. implies 2. The reverse implication may be established by constructing the graph on which $\pi(\mathcal{G}, Y, Y_0)$ acts either as the coset graph of $\pi(\mathcal{G}, Y, Y_0)$ with respect to the subgroups $\mathcal{G}(v)$ and $\mathcal{G}(e)$, or as the largest graph on which this group acts with the given stabilizer information. There is the task of proving that the action of $\pi(\mathcal{G}, Y, Y_0)$ on this graph produces the original graph of groups, and also that the graph is a tree. The latter could be done by considering equivalence classes of words in the generators, written in normal form. Such normal forms are words of the kind we considered already, corresponding to a paths in Y starting and finishing at a given vertex. An equivalence relation is put on these words. We define a permutation action of $\pi(\mathcal{G}, Y, Y_0)$ on the set of these words in the same way that we did in constructing free groups. There is also an approach using homological algebra. We also need to know that the graph is connected, which comes from a lemma: if G acts on a topological space such that there is a connected subset containing representatives of every orbit, and so that G is generated by the stabilizers of these representatives, then the space is connected. □

Example 2.5.34. Any element of a free product with amalgamation $A *_C B$ can be written in form $\cdots a_i b_i a_{i+1} b_{i+1} \cdots$ with $a_i \in A$ and $b_j \in B$, starting and ending with an a or b. Such an expression is equivalent to one of the form $\cdots (a_i c_i)(c_i^{-1} b_i) a_{i+1} b_{i+1} \cdots$. Up to such equivalence the expression is unique. Unique expressions can also be formulated by using coset representatives of C in A and B.

Various corollaries follow, such as the next theorem. Such corollaries were typically known prior to the approach with terminology such as, 'fundamental group of a graph of groups'.

Theorem 2.5.35 (Kurosh subgroup theorem)**.** *Let H be a subgroup of a free product $G = *_{v \in V} \mathcal{G}(v)$. Then H is a free product $F * (*H_{g,v})$ where F is a free group and $H_{g,v} = H \cap g\mathcal{G}(v)g^{-1}$ as g ranges over representatives of the double cosets $H \backslash G / \mathcal{G}(v)$ and $v \in V$.*

Proof. Because G acts on a tree with vertex stabilizers of the form $g\mathcal{G}(v)g^{-1}$ with $g \in G$ and trivial edge stabilizers, so also H acts on the same tree. The vertex stabilizers in H have the form $H \cap g\mathcal{G}(v)g^{-1}$ and again the edge stabilizers are trivial. Each G-orbit of vertices $Gv \cong G/\mathcal{G}(v)$ gives a vertex stabilizer of the form $H \cap g\mathcal{G}(v)g^{-1}$ for each

H-orbit $H\backslash G/\mathcal{G}(v)$ with representative g. Now the form of the presentation of H as the fundamental group of a graph of groups implies the result. Notice that, in this presentation, there are relations $t_e^{-1}1t_e = \phi_e(1)$, but such relations always hold, and can be removed from the presentation, showing that the group presented is the free product of the group generated by these t_e, with the group generated by the vertex stabilizers. □

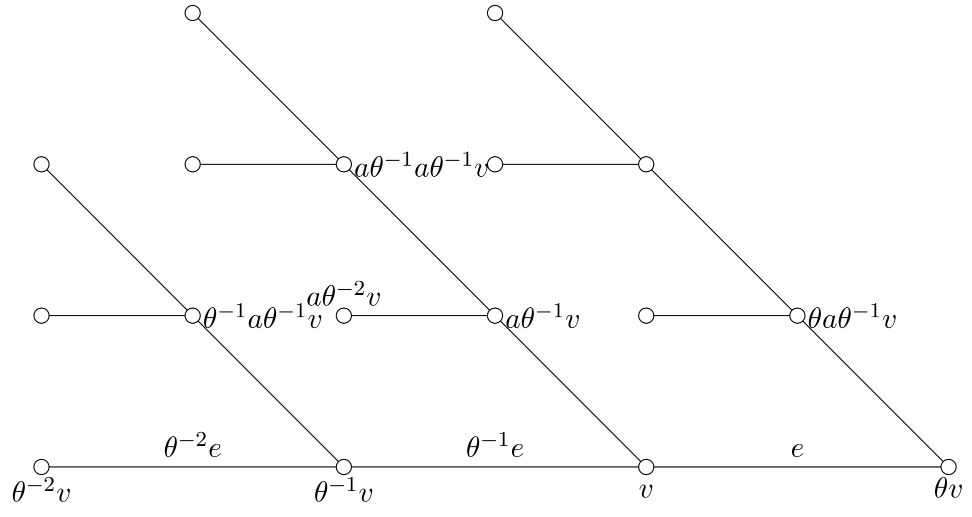
Example 2.5.36. We construct the tree on which

$$G = H*_{K,\phi} = \langle a, \theta \bigm| \theta^{-1}a\theta = a^2 \rangle$$

acts, giving a graph of groups with one vertex and one edge, with infinite cyclic vertex and edge stabilizers $H = K = \langle a \rangle$. The homomorphism $\phi : K \to H$ is the squaring map: $\phi(a^r) = a^{2r}$. We can see from the presentation that $G \cong \mathbb{Z}[\frac{1}{2}] \rtimes \langle \theta \rangle$ where conjugation by θ acts on the additive group $\mathbb{Z}[\frac{1}{2}]$ as division by 2.

To construct the tree Γ we start with the fundamental transversal $Y = v \xrightarrow{e}$ and construct the largest tree on which G acts with Y as a set of orbit representatives. In Γ there is an edge $v \xrightarrow{e} \theta(v)$. The edges starting at v have the form $v \xrightarrow{a^r e} a^r\theta(v)$, possibly with identifications among such edges, to be determined. In the group G we have the relation $a\theta = \theta a^2$, so that $a^r\theta = \theta a^{2r}$ for all integers r. This means that the terminal vertex of the edge $a^r e$ is $a^r\theta v = \theta a^{2r} v = \theta v$, which is the terminal vertex of e, so that all the edges starting at v must be identified. Thus there is only one edge starting at v. Since there is a single orbit of vertices, the same is true of every vertex of Γ.

The edges terminating at v have the form $a^r\theta^{-1}(v) \xrightarrow{a^r\theta^{-1}(e)} v$ with $r \in \mathbb{Z}$. Because of the relation $a^{2r}\theta^{-1}v = \theta^{-1}a^r v = \theta^{-1}v$, there are at most two edges terminating at v, with labels $\theta^{-1}(v) \xrightarrow{\theta^{-1}(e)} v$ and $a\theta^{-1}(v) \xrightarrow{a\theta^{-1}(e)} v$. The number of edges terminating at each vertex of Γ is the same, so is either 1 or 2 everywhere. From this we construct the following tree on which G acts, with all edges directed from left to right:



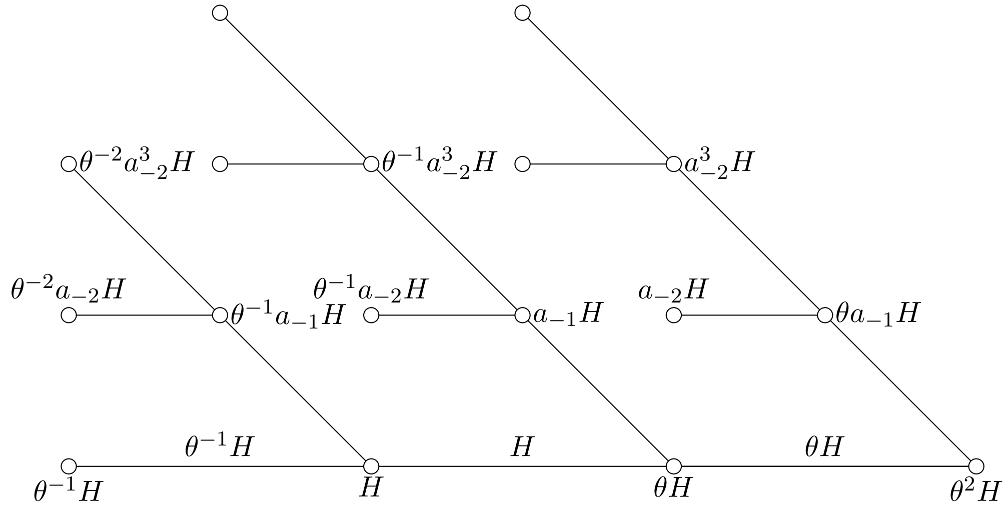
The fact that G acts on this tree has very much to do with the fact that the relation $\theta^{-1}a\theta = a^2$ holds. The element θ acts by moving the tree one place from left to right, and a acts by fixing e, and by permuting the adjacent vertices to v as $\theta^{-1}v \leftrightarrow a\theta^{-1}v$ and $\theta^{-2}v \to a\theta^{-2}v \to \theta^{-1}a\theta^{-1}v \to a\theta^{-1}a\theta^{-1}v \to \theta^{-2}v$.

Example 2.5.37. The tree for

$$G = H{*}_{K,\phi} = \langle a, \theta \mid \theta^{-1}a\theta = a^2\rangle$$

can also be constructed as a coset graph. The vertex set of such a graph is the disjoint union of copies of G/G_v as v ranges over vertices in Y. The edge set is the disjoint union of copies of G/G_e. The end vertices of an edge gG_e are $o(gG_e) = gG_{o(e)}$ and $t(gG_e) = gt_eG_{\bar{t}(e)}$. We may verify that this defines a G-graph.

In this case we have $H = G_v = \langle a\rangle$. Let us define $a_i = \theta^{-i}a\theta^i$ for each $i \in \mathbb{Z}$. We know from the presentation that these conjugates of a generate a normal subgroup U of G, isomorphic to $\mathbb{Z}[\frac{1}{2}]$ when written additively, and that G is the semidirect product $U \rtimes \langle\theta\rangle$. The elements $a_0, a_{-1}, a_{-2}, a^3_{-2}, a_{-3}, a^3_{-3}, a^5_{-3}, a^7_{-3}, \ldots$ for a set of coset representatives for H in U, and the elements $\theta^r a_i^s$ with $r \in \mathbb{Z}$ and the a_i^s as above are a set of coset representatives of H in G. They index both the vertices and edges of the coset graph. An edge gH has starting vertex $o(gH) = gH$, so that each vertex has one edge coming out of it. The terminal vertex $t(gH) = g\theta H = g\theta H\theta^{-1}\theta = g\langle a_{-1}\rangle\theta$. Therefore $t(g_1H) = t(g_2H)$ if and only if $g_1\langle a_{-1}\rangle\theta = g_2\langle a_{-1}\rangle\theta$, if and only if $g_2 = g_1a^m_{-1}$ for some m. There are two cosets of H in each coset of $\langle a_{-1}\rangle$, so each vertex has two edges going in to it: $gH \to g\theta H \leftarrow ga_{-1}H$. For instance, the edges terminating at $\theta a_i^sH = \theta a_i^s\theta^{-1}\theta H = a_{i-1}^s\theta H = a_{i-1}^s a_{-1}\theta H$ are $a_{i-1}sH$ and $a_{i-1}sa_{-1}H$. The tree can be labeled as follows.



The action of a is to stabilize H, its descendents and the vertices a_iH. It has orbits such as $\theta^{-1}H \leftrightarrow \theta^{-1}a_{-1}H$ and

$$\theta^{-2}H \to \theta^{-2}a_{-2}H \to \theta^{-2}a_{-1}H \to \theta^{-2}a^3_{-2}H \to \theta^{-2}H.$$

Chapter 3

Finite Simple Groups

The Classification of Finite Simple Groups is as follows:

Theorem 3.0.1. *Every finite simple group is cyclic of prime order, an alternating group, a finite simple group of Lie type, or one of the twenty-six sporadic finite simple groups.*

A very good account of this background can be found in the book by D. Gorenstein, R. Lyons and R. Solomon, *The Classification of the Finite Simple Groups,* AMS Math. Surveys and Monographs 40.1 (1994). This text is the first of a long series whose goal is to present a complete proof of the classification theorem.

Finite groups of Lie type may be approached in different ways. The approach of R. Steinberg that emerged from work in the 1950s through to 1968 provides a uniform way to construct these groups as fixed points of endomorphisms of linear algebraic groups over the algebraic closure of a finite field. Prior to this, Chevalley in 1955 had shown that there are analogues over finite fields of simple complex Lie groups, and the groups constructed by this means are known as Chevalley groups. They include the classical groups: linear, symplectic, orthogonal, and unitary. Simple groups are obtained by taking matrices of determinant 1 and factoring out the center of the group.

In some sense, most non-cyclic finite simple groups have the form $PSL(2, q)$ where q is a prime power. The orders of small groups $PSL(n, q)$ are listed in the following table, of order at most 7920.

$q\backslash n$	2	3	4	5
2	6	168	20160	9999360
3	12	5616	6065280	
4	60	20160		
5	60	372000		
7	168	1876896		
8	504			
9	360			
11	660			
13	1092			
17	2448			
19	3420			
16	4080			
23	6072			
25	7800			

Apart from $PSL(2,2)$ and $PSL(2,3)$, the groups listed in the table are simple. There are isomorphisms $PSL(2,4) \cong PSL(2,5)$ and $PSL(2,7) \cong PSL(2,3)$. The two groups listed of order 20160 are not isomorphic. Aside from these groups, the remaining simple groups of order at most 7920 are: A_7 (2520), $PSU(3,3)$ (6048) and the Mathieu group M_{11} (7920). The alternating group A_6 is isomorphic to $PSL(2,9)$ and A_8 is isomorphic to $PSL(4,2)$.

3.1 Multiply transitive permutation groups

We follow results 9.5, 9.7, 9.8, 9.9 from Rotman's book, together with the surrounding definitions.

Examples 3.1.1. S_n and A_n are sharply n and $n-2$-transitive on $\{1,\ldots,n\}$. M_{12} and M_{11} are sharply 5 and 4-transitive on 12 and 11 points. For each prime $p \geq 3$, $C_p \rtimes C_{p-1}$ is sharply 2-transitive on a set of size p. This group is the affine group acting on an affine line over $\mathbb{F}_p$. $PSL(3,4)$ of order $21\cdot 20\cdot 48 = 20160$ acts 2-transitively on the projective plane $\mathbb{P}^2(4)$ over the field $\mathbb{F}_4$ with four elements. This space has $\frac{4^3-1}{4-1} = 21$ points

We continue with 9.19, 9.23, 9.24, 9.25, 9.26 of Rotman.

3.2 The Mathieu groups

The five Mathieu groups were the first sporadic simple groups to be discovered and appeared as multiply transitive permutation groups in papers that appeared in 1860 and 1861. They are associated with extraordinary Steiner systems that require remarkable numerical coincidence to be true for them to exist. We continue with 9.51 - 9.59 of Rotman's book.

3.3 Block designs and Steiner systems

Definition 3.3.1. A $t-(v,k,\lambda)$ *design* consists of

- a finite set S of size $|S|=v$,
- a collection $\mathcal{D}$ of subsets of S of size k,

such that every subset of S of size t is contained in exactly λ members of $\mathcal{D}$. We assume $S\neq\emptyset\neq\mathcal{D}$, and $v\geq k\geq t$. The members of $\mathcal{D}$ are called *blocks*.

Example 3.3.2. When every subset of S of size k is a block we get a $t-(v,k,\lambda)$-design for each $t\leq k$. These designs are called *complete*, and among them are the *trivial* designs, which have parameters $v-(v,v,1)$. The only known designs with $t>5$ are complete.

Definition 3.3.3. When $\lambda=1$ a design is called a *Steiner system*. Instead of writing the parameters as $t-(v,k,1)$ they are often written $S(t,k,v)$.

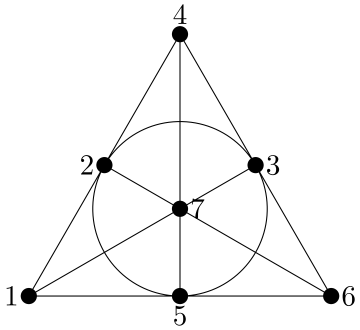
Designs with $\lambda>1$ arose around 1935 in the course of the statistical analysis of experiments. Steiner systems have an older heritage. In the case $t=3$, they appear in Prize question 1733 of the Lady's and Gentlemen's Diary, posed by W.S.B. Woolhouse. The posed problem was solved by Thomas Kirkman in 1847, and in 1850 he posed another problem in the same journal, known as Kirkman's schoolgirl problem. Steiner considered these systems in 1853.

Until 1975, only two non-trivial Steiner systems were known with $t=5$, and only two with $t=4$, associated to the Mathieu groups M_{24},M_{12}, and M_{23},M_{11}. According to Wikipedia, it is now known that there exist Steiner systems $S(4,5,n)$ when n is 11, 23, 35, 47, 71, 83, 107, 131, 167 and 243.

Example 3.3.4. Affine n-space $\mathbb{A}^n(q)$ over $\mathbb{F}_q$ consists of the set $\mathbb{F}_q^n$ with the translates of vector subspaces of dimension s as the affine s-subspaces. Taking $\mathcal{D}$ to be this collection of affine s-dimensional subspace we obtain a $2-(q^n,q^s,\lambda)$-design. When $s=1$ then $\lambda=1$; when $s=2$ then $\lambda=\frac{q^{n-1}-1}{q-1}$. To see this, we may assume that one of the two points is 0 in $\mathbb{F}_q^n$ and then planes containing the second point biject with lines in the quotient space by that vector.

Projective n-space $\mathbb{P}^n$ is the set of lines in $\mathbb{F}_q^{n+1}$. The set of projective s-subspaces is a $2-(\frac{q^{n+1}-1}{q-1},\frac{q^{s+1}-1}{q-1},\lambda)$ design. When $s=2$ then $\lambda=\frac{q^{n-1}-1}{q-1}$ again. When $s=1$ then $\lambda=1$ and we have a $2-(\frac{q^{n+1}-1}{q-1},q+1,1)$, or $S(2,q+1,\frac{q^{n+1}-1}{q-1})$, Steiner system. Taking $n=2$ we have a $2-(q^2+q+1,q+1,1)$, or $S(2,q+1,q^2+q+1)$, Steiner system called a *projective plane* and the blocks are called *lines*.

For example: in $\mathbb{A}^2(2)$ every subset of size 2 of the set of 4 points is a line. $\mathbb{P}^2(2)$ is the projective plane with seven points $\{1, \ldots, 7\}$ and seven lines as follows:

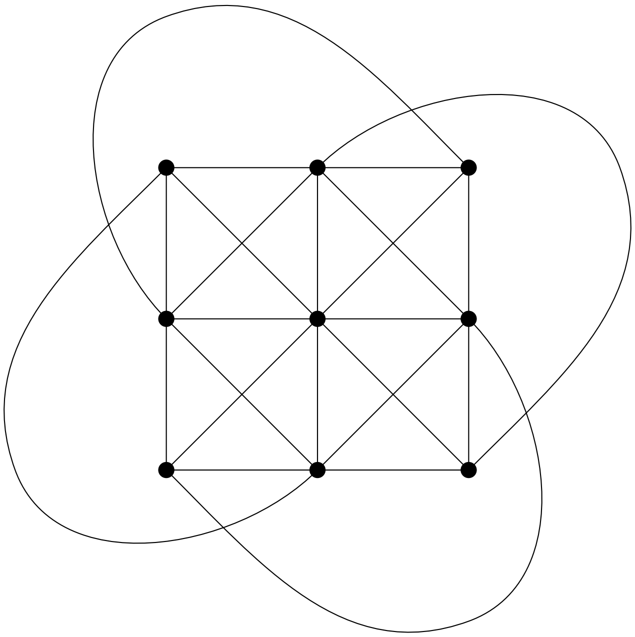


The automorphism group of this configuration is $SL(3, 2) = GL(3, 2) = PSL(3, 2)$. The set of lines (or blocks) is

$$\{\{1, 2, 4\}, \{2, 3, 5\}, \{3, 4, 6\}, \{4, 5, 7\}, \{5, 6, 1\}, \{6, 7, 2\}, \{7, 1, 3\}\}$$

which is obtained from one of them by applying the 7-cycle $(1, 2, 3, 4, 5, 6, 7)$ in $GL(3, 2)$ (a *Singer cycle*).

$\mathbb{A}^2(3)$ looks like the following:



Example 3.3.5. Suppose we wish to compare 13 varieties of corn that will be grown in standardized plots. If only two varieties can be grown in each plot we would need

$\binom{13}{2} = 78$ plots. Suppose instead that four varieties can be grown in each plot. How many plots are needed? It can be done with 13 plots, and no fewer. To see 'no fewer': 6 comparisons are made in each plot; there are 78 comparisons to be made altogether; hence at least $\frac{78}{6} = 13$ plots are needed. It can be done with 13: use $\mathbb{P}^2(3)$. It helps the statistics if each pair of varieties is compared the same number of times in the experiment.

Class Activity. Suppose we have 9 varieties of corn and 3 varieties can be grown in each plot. How many plots are needed? Answers: A9, B10, C11, D12, E None of the above

Example 3.3.6. 13 people enter a Scrabble tournament. How many games must be played (4 people per game) so that each player plays against every other player in some game?

Example 3.3.7. The game Set is played with a deck of 81 cards representing the points in 4-dimensional affine space over $\mathbb{F}_3$. A *set* is an affine line. Each pair of cards determines a unique set, so we have a Steiner system $S(2, 3, 81)$.

Definition 3.3.8. In a $t-(v, k, \lambda)$ design, let λ_i denote the number of blocks containing some fixed set I of i points, where $0 \leq i \leq t$. Thus $\lambda_t = \lambda$. The number λ_0 is the total number of blocks and is often denoted b. Also the number λ_1 of blocks containing a single given point is often denoted r. We will see that these numbers do not depend on the particular subset I that was chosen.

Proposition 3.3.9. *In a $t - (v, k, \lambda)$ design we have, for each i with $0 \leq i \leq t$, the equality*

$$\lambda_i \binom{k-i}{t-i} = \binom{v-i}{t-i} \lambda.$$

It follows that λ_i is independent of the set I that was chosen. Furthermore, the design is also a $i - (v, k, \lambda_i)$ design for each $i \leq t$.

Proof. We count the set of pairs

$$\{(A, B) \mid I \subseteq A \subseteq B, \ A \text{ is a } t\text{-subset}, \ B \text{ is a block}\}$$

in two ways. The number of ways of choosing A so that $I \subseteq A$ is $\binom{v-i}{t-i}$, and with A fixed there are $\lambda_t = \lambda$ possible B. On the other hand, if we choose B first with $I \subseteq B$ there are λ_i possibilities, and then the number of ways of choosing A is $\binom{k-i}{t-i}$. □

Example 3.3.10. When we were considering varieties of corn we had $v = 13$, $k = 4$ and $t = 2$. We computed the number of blocks to be $b = \lambda_0 = \lambda_2 \binom{13}{2} / \binom{4}{2}$.

Corollary 3.3.11. *In a $t - (v, k, \lambda)$ design we have*

$$\begin{aligned} \lambda_0 k &= v\lambda_1, \\ \lambda_1(k-1) &= (v-1)\lambda_2, \\ \lambda_2(k-2) &= (v-2)\lambda_3, \\ \lambda_3(k-3) &= (v-3)\lambda_4, \end{aligned}$$

and so on. Thus

- *in any 1-design, $bk = vr$,*
- *in any 2-design, $r(k-1) = (v-1)\lambda$*

and so $(v-1)\lambda$ must be divisible by $(k-1)$ for such a design to exist. In general we have

$$\lambda_i = \frac{(v-i)(v-i-1)\cdots(v-j+1)}{(k-i)(k-i-1)\cdots(k-j+1)}\lambda_j$$

if $i \leq j$.

Definition 3.3.12. Given a design, there are other designs that may be constructed from it, including a dual design and a residual design which we do not define here. Given a design $\mathcal{D}$ on a set S and a point $x \in S$ we define the *contracted* or *derived* design at x to be the design $\mathcal{D}_x$ on $S_x := S - \{x\}$ where $\mathcal{D}_x := \{B - \{x\} \mid x \in B \in \mathcal{D}\}$. If $\mathcal{D}$ is a $t - (v, k, \lambda)$ design with numbers of blocks λ_i, $0 \leq i \leq t$, then $(S_x, \mathcal{D}_x)$ is a $(t-1) - (v-1, k-1, \lambda)$-design with new numbers of blocks $\lambda_i^{\text{new}} = \lambda_{i+1}^{\text{old}}$ with $0 \leq i \leq t-1$.

Definition 3.3.13. If $\overline{\mathcal{D}}$ is a design on $\bar{S} := S \cup \{\infty\}$ for which $\overline{\mathcal{D}}_\infty = \mathcal{D}$ we say that it is an *extension* of the design $(S, \mathcal{D})$.

Class Activity. Draw pictures of the contractions of $\mathbb{P}^2(2)$ and of $\mathbb{A}^2(3)$.

Example 3.3.14. In a Steiner system $S(5, 8, 24)$ (otherwise known as a $5 - (24, 8, 1)$ design) the numbers $(\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ are $(759, 253, 77, 21, 5, 1)$. The successive contracted designs have parameters $S(4, 7, 23), S(3, 6, 22), S(2, 5, 21)$. We can compute these numbers from the equations

$$\begin{aligned}
4\lambda_4 &= 20\lambda_5 \\
5\lambda_3 &= 21\lambda_4 \\
6\lambda_2 &= 22\lambda_3 \\
7\lambda_1 &= 23\lambda_2 \\
8\lambda_0 &= 24\lambda_1
\end{aligned}$$

starting with $\lambda_5 = 1$. The fact that these equations have integer solutions is a strong divisibility constraint.

Proposition 3.3.15. *If a $t - (v, k, \lambda)$ design with λ_0 blocks is extendable then $(k+1)$ divides $\lambda_0(v+1)$.*

Proof. The extension is a $(t+1) - (v+1, k+1, \lambda)$ design and $\lambda_1^{\text{new}} = \lambda_0^{\text{old}}$. We now apply the equation $\lambda_0^{\text{new}}(k+1) = \lambda_1^{\text{new}}(v+1)$. □

Corollary 3.3.16. *If $\mathbb{P}^2(q)$ is extendable then $q = 2$ or 4.*

Proof. The projective plane $\mathbb{P}^2(q)$ is a $2-(q^2+q+1, q+1, 1)$ design with $\lambda_0 = q^2+q+1$ blocks. If it can be extended then $q+2 \mid (q^2+q+1)(q^2+q+2)$. Since this product equals $(q^2-1+q+2)(q^2+q+2)$ it is equivalent to ask whether $q+2$ divides $q^2(q^2-1)$. Assume that this division does occur.

If $q = p^r$ with p odd, then p does not divide $q+2$, and so $q+2$ divides $q^2-1 = (q+1)(q-1)$. Since no factor of $q+2$ can divide $q+1$, we must have $q+2$ dividing $q-1$, an impossibility. Thus $\mathbb{P}^2(q)$ is not extendable when q is odd.

If $q = 2^r$, and $r > 1$, then $q+2 = 2^r+2$ is divisible by 2, but by no higher power of 2, and so if 2^r+2 divides $2^{2r}(2^{2r}-1)$ we may deduce that $2^{r-1}+1$ divides $2^{2r}-1 = (2^r+1)(2^r-1) = (2(2^{r-1}+1)-1)(2(2^{r-1}+1)-3)$. Thus $2^{r-1}+1$ divides 3, so that $r = 2$. □

In fact, both $\mathbb{P}^2(2)$ and $\mathbb{P}^2(4)$ are extendable. The extension of $\mathbb{P}^2(2)$ is the $3-(8,4,1)$ design whose blocks are the affine 2-spaces in $\mathbb{F}_2^3$. We see that the design obtained from this by omitting the zero vector and the planes not containing it gives $\mathbb{P}^2(2)$.

The $2-(21,5,1)$ design $\mathbb{P}^2(4)$ is extendable twice to give a $5-(24,8,1)$ design.

3.4 The Steiner system $S(5,8,24)$

A Steiner system with parameters $S(5,8,24)$ may be constructed as follows.

Theorem 3.4.1. *Let $(S,\mathcal{D})$ be a $t-(v,k,\lambda)$ design and G a group of automorphisms of $(S,\mathcal{D})$ that is t-transitive on S and transitive on $\mathcal{D}$. Suppose the action of G on S has a transitive extension to a permutation group $\overline{G}$ on the set $\overline{S} := S \cup \{\infty\}$ with $\overline{G}_\infty = G$. Then there exists a $(t+1)-(v+1,k+1,\bar{\lambda})$ design $(\overline{S},\overline{\mathcal{D}})$ preserved by $\overline{G}$, for some $\bar{\lambda}$.*

Proof. We choose a block $B \in \mathcal{D}$ and let $\overline{B} = B \cup \{\infty\}$, and $\overline{\mathcal{D}} = \{\bar{g}(B) \mid \bar{g} \in \overline{G}\}$. This family of subsets $\overline{\mathcal{D}}$ is independent of the B originally chosen, since $\overline{G}_\infty = G$ and G is transitive on $\mathcal{D}$. Since $\overline{G}$ is $(t+1)$-transitive on $\overline{S}$, it follows that $(\overline{S},\overline{\mathcal{D}})$ is a $(t+1)$-design. □

Corollary 3.4.2. *The Steiner system $S(2,5,21)$ extends twice to give 3, 4 and 5-designs.*

In fact, the numbers λ_t for these designs are all 1, but this depends on a closer examination of the action of the Mathieu groups: see Theorem 9.66 in Rotman's book. It is also the case that the $2-(9,3,1)$ design given by the affine plane $\mathbb{A}^2(3)$ extends twice to give Steiner systems $S(3,4,10), S(4,5,11)$ and $S(5,6,12)$.

Proof. We have seen that $PSL(3,4)$ acts 2-transitively on $\mathbb{P}^2(4)$, and because blocks in $\mathcal{D}$ are uniquely determined by any of their 2-element subsets, the action on blocks is transitive. Thus the conditions of Theorem 3.4.1 are satisfied, and since $PSL(3,4)$ does have a transitive extension to M_{22} we obtain a design $3-(22,6,\lambda)$ for some λ. We now repeat this process using M_{23} and M_{24}. □

Proposition 3.4.3 (The Leech triangle)**.** *In any Steiner system let $B = \{x_1, \ldots, x_k\}$ be a block, and put $B_i = \{x_1, \ldots, x_i\}$ for each $i \leq k$. Let a_{ij} be the number of blocks that intersect B_i in exactly B_j with $j \leq i$. Then $a_{ii} = \lambda_i$ and $a_{ij} = a_{i+1,j} + a_{i+1,j+1}$. These numbers form a triangle in which each term is the sum of the two below it.*

Proof. The fact that $a_{ii} = \lambda_i$ is immediate. To justify the equation $a_{ij} = a_{i+1,j} + a_{i+1,j+1}$ we observe that octads that intersect O_i in exactly O_j either do not contain x_{i+1} or do contain x_{i+1}. The number in the first case is $a_{i+1,j}$, and in the second case it is $a_{i+1,j+1}$, because this number is independent of the particular set of $i+1$ elements chosen in the block. For example, if we know (by induction) that 12 octads meet O_5 in 3 specific points and 4 octads meet O_6 in 4 specific points, regardless of choice, then $12 - 4 = 8$ octads meet O_6 in 3 specific points, regardless of choice. □

We now accept the existence of a Steiner system $S(5, 8, 24)$ and study its combinatorics. The blocks of the Steiner system $S(5, 8, 24)$ are called *octads*, and the triangle just constructed is called the *Leech triangle.*

Corollary 3.4.4. *The Leech triangle for the Steiner system $S(5, 8, 24)$ is*

$$\begin{array}{ccccccccccccccccc}
&&&&&&&&759 \\
&&&&&&&506&&253 \\
&&&&&&330&&176&&77 \\
&&&&&210&&120&&56&&21 \\
&&&&130&&80&&40&&16&&5 \\
&&&78&&52&&28&&12&&4&&1 \\
&&46&&32&&20&&8&&4&&0&&1 \\
&30&&16&&16&&4&&4&&0&&0&&1 \\
30&&0&&16&&0&&4&&0&&0&&0&&1
\end{array}$$

Exercise 3.4.5. Calculate the Leech triangle for a Steiner system $S(3, 4, 16)$.

Corollary 3.4.6. *If O_1 and O_2 are octads then $|O_1 \cap O_2| = 0, 2, 4$ or 8.*

Proof. This is the interpretation of the fact that in the bottom row of the Leech triangle, the only non-zero entries are in positions 0, 2, 4 and 8. □

Definition 3.4.7. We define a vector space $\mathcal{C}$ over $\mathbb{F}_2$. For each finite set X we denote by $\mathcal{P}(X)$ the set of subsets of X. There is a bijection

$$\begin{aligned} \mathcal{P}(X) &\leftrightarrow \text{elements of } \mathbb{F}_2^{|X|} \\ A &\leftrightarrow \chi_A, \text{ the characteristic function.} \end{aligned}$$

Under this bijection $\chi_A + \chi_B = \chi_{A+B}$ where $A+B = (A \cup B) - (A \cap B)$ is the *symmetric difference* of A and B. We now define $\mathcal{C}$ to be the subspace of $\mathcal{P}(X)$ spanned by the octads, where now X is the set of 24 points of $S(5, 8, 24)$. This subspace is called the *extended binary Golay code.*

We now use the Leech triangle in studying the sets that lie in $\mathcal{C}$.

Proposition 3.4.8. *$X \in \mathcal{C}$.*

Proof. $\lambda_1 = 253$ is odd, so the sum of all octads is X since $\chi_{A_1+\cdots+A_t}(x) = 1$ if and only if x lies in an odd number of the A_i. □

Proposition 3.4.9. *if $Y \in \mathcal{C}$ and O is an octad then $|O \cap Y|$ is even.*

Proof. We use induction on the number of terms in an expression for Y as a sum of octads. When Y is an octad the result is true from the Leech triangle, so the induction starts. If Y_1 and Y_2 lie in $\mathcal{C}$ and have even intersectiions with O then $|O \cap (Y_1 + Y_2)| = |O \cap Y_1| + |O \cap Y_2| - 2|O \cap Y_1 \cap Y_2|$ and this is even. □

Proposition 3.4.10. *Every 8-element set in $\mathcal{C}$ is an octad.*

Proof. Let $Y \in \mathcal{C}$ have size 8. Any 5-element subset of Y is contained in a unique octad O, and if Y is not an octad then $|Y \cap O| = 6$. It follows that sets of the form $O \cap Y$ of size 6 where O is an octad form a Steiner system with parameters $S(5, 6, 8)$. The number of blocks in such a Steiner system is

$$\frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2} = \frac{28}{3}$$

which is not an integer, so no such Steiner system can exist. Hence Y is an octad. □

Corollary 3.4.11. *If O_1 and O_2 are octads with $|O_1 \cap O_2| = 4$ then $O_1 + O_2$ is an octad.*

Corollary 3.4.12. *If O_1 and O_2 are octads with $|O_1 \cap O_2| = \emptyset$ then $X - (O_1 \cup O_2)$ is an octad.*

Proposition 3.4.13. *Let S_1 be any 4-element subset of X. The five octads containing S_1 have the form $S_1 \cup S_2$, $S_1 \cup S_3, \ldots, S_1 \cup S_6$ where the S_i are 4-element subsets. These subsets have the properties that $S_1 \cup \cdots \cup S_6 = X$, and for each pair $i \neq j$, $S_i \cup S_j$ is an octad.*

Such a configuration of six 4-element subsets is called a *sextet*.

Proof. The fact that there are five octads containing a four element set follows from the Leech triangle. Their union is the whole of X since any point of X may be adjoined to the four to give a five element set that is contained in an octad. The union of any pair of the S_i is an octad since it lies in $\mathcal{C}$ and has size 8. □

Suppose two octads O_1 and O_2 have $|O_1 \cap O_2| = 2$. The 12-element set $O_1 + O_2$ is called a *dodecad* (or sometimes an *umbral dodecad*).

Proposition 3.4.14. *A dodecad does not contain any octad.*

Proof. Suppose we have octads $O_3 \subset O_1 + O_2$ where $|O_1 \cap O_2| = 2$. Then O_3 is distinct from O_1 and O_2 so $|O_1 \cap O_3| \leq 4$ and $|O_2 \cap O_3| \leq 4$ (since if an intersection had size 5 or larger the octads would be the same). Thus $|O_1 \cap O_3| = |O_2 \cap O_3| = 4$ and so $O_1 + O_3$ is an octad, and it contains 2 points that do not lie in O_2. Now $|(O_1 + O_3) \cap O_2| = 6$, which is a contradiction. □

Corollary 3.4.15. *Let D be a dodecad. The subsets $O \cap D$ of size 6 with O an octad form a Steiner system $S(5, 6, 12)$.*

The special sets $O \cap D$ of size 6 are called *hexads.*

Proposition 3.4.16. *The Steiner System $S(5, 6, 12)$ has Leech triangle:*

$$\begin{array}{ccccccccccccc}
 & & & & & & 132 & & & & & & \\
 & & & & & 66 & & 66 & & & & & \\
 & & & & 30 & & 36 & & 30 & & & & \\
 & & & 12 & & 18 & & 18 & & 12 & & & \\
 & & 4 & & 8 & & 10 & & 8 & & 4 & & \\
 & 1 & & 3 & & 5 & & 5 & & 3 & & 1 & \\
1 & & 0 & & 3 & & 2 & & 3 & & 0 & & 1
\end{array}$$

Corollary 3.4.17. *The complement of a hexad in D is a hexad.*

Proof. This is indicated by the 1 at the bottom left corner. □

Lemma 3.4.18. *The complement $X - D$ of a dodecad D is a dodecad.*

Proof. Let $D = O_1 + O_2$ with $|O_1 \cap O_2| = 2$, and let O_3 be any octad disjoint from O_1 (how do we know such $O(because_3$ exists?). Let O_4 be the complement $X - (O_1 \cup O_3)$. Then $O_2 \cap O_3 \subseteq O_2 - O_1$ which has size 6, so $|O_2 \cap O_3| = 0, 2$ or 4. Similarly for $O_2 \cap O_4$ and without loss of generality $|O_2 \cap O_3| = 2$ and $|O_2 \cap O_4| = 4$. Now $O_2 + O_4$ is an octad and $X - D = (O_2 + O_4) + O_3$. □

Lemma 3.4.19. *Let D be a dodecad and O_1 an octad so that $|O_1 \cap D| = 6$. Then $O_2 = O_1 + D$ is an octad such that $D = O_1 + O_2$. The sets $O_1 \cap D$ and $O_2 \cap D$ are complementary hexads.*

Proof. The set $O_2 = O_1 + D$ has size 8 and lies in $\mathcal{C}$, so is an octad. Hence $O_2 + O_1 = O_1 + D + O_1 = D$. □

Lemma 3.4.20. *The number of dodecads is 2576.*

Proof. Suppose that D is a dodecad and suppose that $D = H_1 \cup H_2$ is a decomposition into complementary hexads, where $H_i = D \cap O_i$. The pair of points $O_1 \cap O_2$ completely determine this decomposition, because if also $D = O_1' + O_2'$ is a different decomposition with $O_1 \cap O_2 = O_1' \cap O_2'$, we know that $O_1' \cap O_1$ and $O_1' \cap O_2$ both have size at most 4, and also that the two points of $O_1 \cap O_2$ lie in O_1' and are outside D, hence $|O_1' \cap D| \leq 4$, a contradiction.

Now D contains 66 pairs of complementary hexads, because the total number of hexads is 132, and they occur in complementary pairs. Also, $X - D$ contains $\binom{12}{2} = 66$ pairs of points. Each pair of points in $X - D$ is associated with at most one pair of hexads. Therefore pairs of complementary hexads in D biject with pairs of points in $X - D$.

The number of unordered pairs of octads O_1, O_2 such that $|O_1 \cap O_2| = 2$ is

$$\frac{759 \times \binom{8}{2} \times 16}{2}$$

this being computed by choosing first an octad (759), then a pair of points in the octad ($\binom{8}{2}$), and then the 16 octads intersecting the first octad in that pair. The number 16 appears in the bottom row of the Leech triangle. The number of decompositions of $O_1 + O_2$ into such a pair is 66. Therefore the number of dodecads is

$$\frac{759 \times \binom{8}{2} \times 16}{2 \times 66} = 2576.$$

This is because the product of the number of dodecads and the number of unordered hexad decompositions equals the number of unordered pairs of octads with intersection of size 2. □

Theorem 3.4.21. *The sets in $\mathcal{C}$ are the empty set, octads, dodecads, complements of octads and X.*

Proof. We show that these sets are preserved under symmetric difference with octads. We have already seen that the symmetric difference of two octads is of the specified form.

Consider now $O + D$ were O is an octad and D is a dodecad. In this case $|O \cap D|$ is even and less than 8, and similarly with $|O \cap (X - D)|$, so $|O \cap D| = 2, 4, 6$ and $|O \cap (X - D)| = 6, 4, 2$, respectively. The case of intersections of size 6 has just been considered.

Suppose that $|D \cap O| = 4$, so $|D + O| = 12$. Let $H_1 = D \cap O_1$ be a hexad containing $O \cap D$, and let $H_2 = O_2 \cap D$ be the complementary hexad. Now $D = O_1 + O_2$ and $D + O = O_1 + O_2 + O = (O + O_1) + O_2$. Here $O + O_1$ is an octad, and so $D + O$ is of the specified form.

Finally the complement of any octad may be written as a union of octads $O_1 + O_2$ in many ways: consider a sextet decomposition associated to the octad, and take the unions of pairs of hexads not in the octad. Now the symmetric difference with an octad O reduces to the previous cases on considering $(O + O_1) + O_2$. □

Corollary 3.4.22. $\dim_{\mathbb{F}_2} \mathcal{C} = 12$.

Proof. The number of vectors in $\mathcal{C}$ is $1 + 759 + 2576 + 759 + 1 = 4096 = 2^{12}$. □

The rows of the following matrix form a basis for a subspace of $\mathbb{F}_2^{24}$ that after relabeling the columns is $\mathcal{C}$:

$$\begin{bmatrix}
1&1&0&1&0&0&0&1&0&0&0&0&0&0&0&0&1&1&0&1&0&0&0&1\\
0&1&1&0&1&0&0&1&0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&1\\
0&0&1&1&0&1&0&1&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&1\\
0&0&0&1&1&0&1&1&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1\\
0&0&0&0&0&0&0&0&1&1&0&1&0&0&0&1&1&1&0&1&0&0&0&1\\
0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&1&0&1&1&0&1&0&0&1\\
0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&1&0&0&1&1&0&1&0&1\\
0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1&0&0&0&1&1&0&1&1\\
0&0&0&1&0&1&1&1&0&0&0&1&0&1&1&1&0&0&0&1&0&1&1&1\\
0&0&1&0&1&1&0&1&0&0&1&0&1&1&0&1&0&0&1&0&1&1&0&1\\
0&1&0&1&1&0&0&1&0&1&0&1&1&0&0&1&0&1&0&1&1&0&0&1\\
1&0&1&1&0&0&0&1&1&0&1&1&0&0&0&1&1&0&1&1&0&0&0&1
\end{bmatrix}$$

A different approach to constructing $\mathcal{C}$ is described in Wikipedia: order 24-bit binary strings in lexicographic order; any such string that differs from some earlier one in fewer than 8 positions is discarded.

Definition 3.4.23. The *binary Golay code* $\hat{\mathcal{C}}$ is the vector subspace of $\mathbb{F}_2^{23}$ obtained from $\mathcal{C}$ by projecting $\mathcal{C}$ onto the first 23 coordinates of $\mathbb{F}_2^{24}$.

In coding theory the *Hamming distance* between two vectors is the number of coordinates in which those vectors differ. A *linear code* is a subspace of a vector space. Its *minimum distance* is the minimum distance between pairs of vectors in the code, and this equals the minimum support size of a non-zero vector in the code.

Theorem 3.4.24. *1.* $\dim \hat{\mathcal{C}} = 12$.

2. The minimum distance of $\hat{\mathcal{C}}$ is 7.

3. The binary Golay code is a perfect 3-error correcting code.

Proof. The $\mathcal{C}$ has minimum distance 8 and the minimum distance can only go down by 1 on losing one coordinate. It is now standard that this implies that the code corrects 3 errors. The number of vectors distance at most 3 from a given vector is

$$1 + \binom{23}{1} + \binom{23}{2} + \binom{23}{3} = 1 + 23 + 23 \cdot 11 + 23 \cdot 11 \cdot 7 = 2048 = 2^{11}$$

and so the number of vectors in the balls of radius 3 around all the vectors in $\hat{\mathcal{C}}$ is $2^{11} \cdot 2^{12} = 2^{23}$, which is the total number of vectors in the ambient vector space. □

Exercise 3.4.25. Kirkman's Schoolgirl Problem (1850): Fifteen young ladies in a school walk out three abreast for seven days in succession: it is required to arrange them daily so that no two shall walk twice abreast.