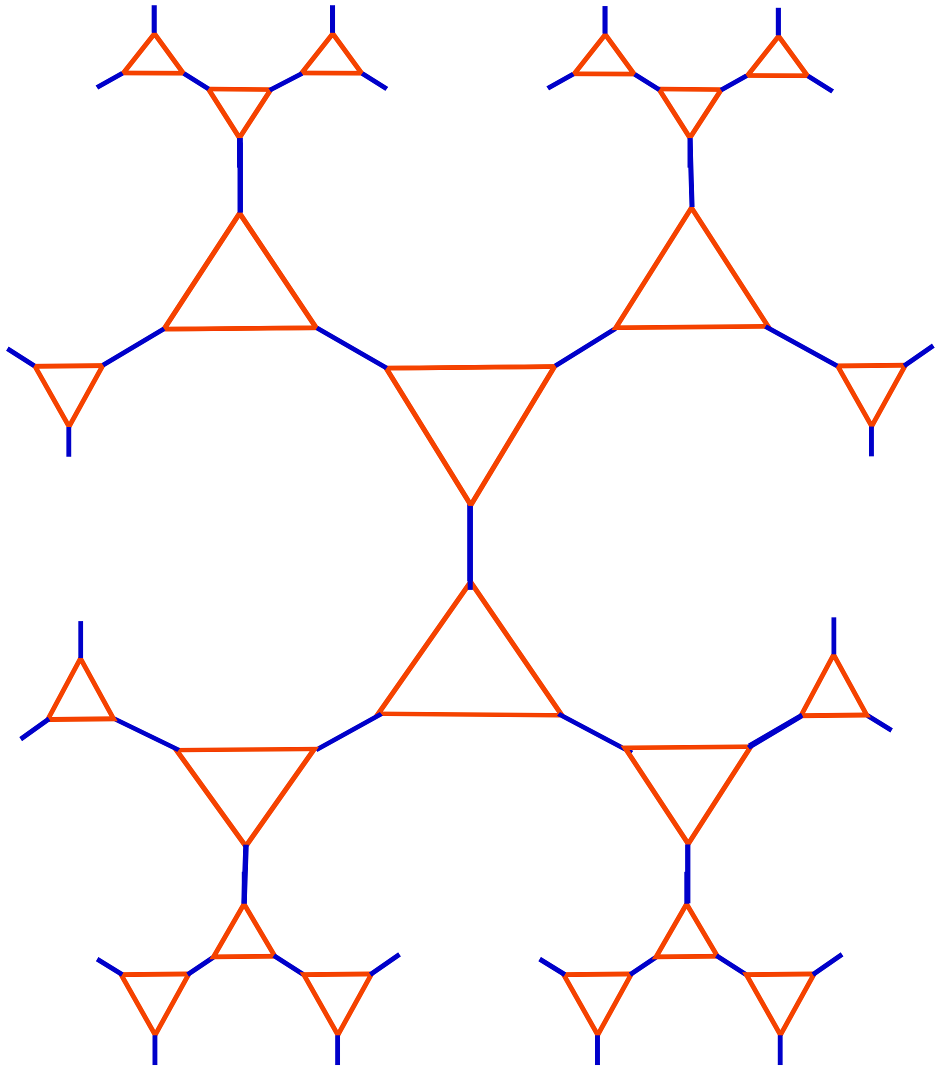
Topics In Groups – MT5824

Fall 2016



Dr. Collin Bleak

Contents

Introduction

In this course, we will undertake an exploration of a fair set of the methods of current group theoretical praxis. We will introduce topics in an independent fashion, as much as possible, but "in secret" we will slowly trace our way through one half of a proof of a celebrated result from the early 1980's. We will also explore in some depth the ingredients that imply simplicity for a group (recall taht simple groups are like the prime numbers of Group Theory). Along the way, we will acquaint the reader with many foundational infinite groups (many of which do not fit naturally into a first course in groups such as MT 4003), which will be examples that we will revisit throughout the course. Indeed, we will see these groups from several perspectives: geometric, combinatorial, and categorial. We hope the student will come away from the course with a foundation in some of these most important infinite groups, and with an understanding of many core methods used in understanding these groups from diverse perspectives.

closer to $3/4^{th}$s

I doubt everything! - Joe Student

The result we will be studying is the proof that a finitely generated group is virtually free if and only if it has a "context free word problem." This is a celebrated theorem from the 1980's attributed to Muller and Schupp (see the original papers [9, 10] which we mention also uses Dunwoody's work on accessibility in an essential manner) which connects basic objects of group theory (free groups) to a basic class of formal languages (context free languages) using proofs that rely heavily on the geometric and graph-theoretical methods developed by Bass, Dunwoody, Serre, Stallings and many others.

Our own approach to the theorem above will be slightly different than that published by Muller and Schupp; we will be primarily following the proof as given by Diekert and Weiß in their article [6]. That article is essentially complete on the details of most of the deeper proofs, but it is perhaps light in some foundational material. Another superb reference for most of this material is Ian Chiswell's text [5], which has more material on the foundations (eg., formal language theory, automata, decidability and recursive sets) and perhaps less on the deeper results such as in Dunwoody's accessibility theory. To make these module notes more palatable, we will focus on the core results and methods as in Diekert and Weiß, proving some in depth, while glossing over others when convenient, and borrowing also from Chiswell in the development of the more foundational material. No material of the Diekert and Weiß article should be beyond the reader when we are finished, however, so the interested student can revisit areas of lighter treatment independently for complete understanding if so desired.

Current mathematical praxis often involves multiple “fields” of mathematics simultaneously. Very few know all there is to know about any given area, but we all should know a bit of many areas, and as much as possible in our areas of primary interest. While mathematics is built from our core set theoretical foundations, very few people can trace any modern theorem to it’s set theoretical roots. There is even a field of research in Computer Science called “proof-checking” that tries to implement and check major proofs in this fashion, and they have had success with some big theorems such as the Four-Colour Theorem, but, this is a very active area of work, and points to our frontiers, not to what every practicing mathematician can do. The point is that the developing mathematician has to develop a level of comfort with borrowing major theorems and methods from other fields, even when they do not personally know those fields very well themselves. In this course, the author reckons we will be studying groups directly around 75% of the time. The rest of the time will be filled with learning about decidability, graph theory, automata theory, and formal language theory. This is representative of current research, where it is hard to do anything of lasting value in a field without having to learn a fair bit of ancillary material along the way.

A note on exercises: It is said that 80% of mathematics learning is the learning of a language, and the only way that the author knows of digesting a language properly is by using it. Indeed these notes, and any serious book on mathematics, can be quite intimidating if one just flips ten pages into the material and discovers you cannot read anything with understanding! Mathematical language builds quickly, and so one often finds that the language is dense; it can be hard to accurately decribe simple ideas! Thus, exercises play an essential rôle in digesting the material of this module in a proper manner; one becomes fluent with definitions and known theory by using them, and seeing the ideas interact “First-hand.” The author has worked hard to structure these notes so that the core ideas are in the reading, but with various proofs of lemmas and theorems broken down into palatable exercises for easier and deeper comprehension. Also in the exercises, there are routes into related theory and examples that are away from the “virtually free” groups which are the essential stars of these notes. The exercises span from basic testing of the use of definitions through problems with a bit more substance, to questions which are more difficult, and even some questions which are currently “open” (but, they are marked as such!). The author believes that proper mastery of the material can only be found through the sometimes-difficult job of taking the exercises and writing out complete answers whenever possible. Early success will breed more interest and drive for the harder depths... let your curiosity lead you forward!

A note on writing mathematics: While you should feel free to write your solutions any way you choose, this author strongly suggests the use of LaTex. After one hour of learning, it will give you free access to greek letters, subscripts and superscripts, an equation environment, matrices, integral symbols, and the insertion of figures. One can spend a lifetime learning more and more, but one hour will get you through 95% of what you will eventually use. When these notes were initially written, the author was using TexShop as his writing environment on the mac. TexShop is freely available and fully functional. For each system, (Linux, PC, mac, etc.) there are freely available environments for writing and compiling LaTex into beautiful pdf documents. In the author’s experience, every other math typesetting system (other than Latex) is eventually more work to use (often initially easier, but a real long-run headache!) and produces less beautiful typesetting. So, the cost of less than one hour’s set-up, and then one further hour’s learning efforts, will set you free as a mathematical writer.

But wait, there’s more!!! Following a wonderful example of Graham, Knuth, and Patashnik, student comments to the text will be inserted into the text in the margins of these notes for future generations (or the text will change dramatically in response to these comments). I hope the reader will take a pleasant 30 minute visit to the library, and look at the wonderful book “Concrete Mathematics” by Graham, Knuth, and Patashnik, to see how such comments have greatly added to an already beautiful book.

These comments are distracting!! Also, why are they so “Loud?”

1 L1: Free Groups I (Definitions and Constructions)

1.1 Constructing free groups I: $W(X)$ from X

Let X be a non-empty set. Denote by X^{-1} any set in bijective correspondence with X, where $X \cap X^{-1} = \varnothing$ (so, there is $\theta : X \to X^{-1}$ a bijection).

(*Note for the technically strong*: one can prove that given a set X, there is a set which we can call X^{-1}, which is disjoint from X, but in bijective correspondence with X, using the axioms of set theory. Also, given this X^{-1}, for $|X| > 1$ the map θ demonstrating the bijection is not unique. In any case, as such a set X^{-1} exists, it then becomes easy to show that many distinct such sets exist, so one can argue the set $W(X)$ we create below is not really well defined (we will have to have made a lot of arbitrary choices)! All of these issues require a lot of technical care to deal with correctly, but they can be addressed, and we will wilfully ignore these issues later.)

We build a map $invert : X \cup X^{-1} \to X \cup X^{-1}$, where we define and denote this map as follows:

$$invert(x) = \begin{cases} \theta(x) & \text{if } x \in X \\ \theta^{-1}(x) & \text{if } x \in X^{-1}. \end{cases}$$

Now, whenever we see a symbol $x \in X \cup X^{-1}$, we will take the corresponding symbol via the map *invert* when we write x^{-1}, that is define

$$x^{-1} := invert(x).$$

Also, we will sometimes denote any $x \in X \cup X^{-1}$ by x^{+1} or x^1; while this seems a waste of electrons or ink, in fact, it lays the foundation that leads to simplified writing of expressions later!

Before moving on, let us note that for all $x \in X \cup X^{-1}$, we have $(x^{-1})^{-1} = x$.

Following Unix notation, we denote by

$$(X \cup X^{-1})^* := \{w \mid \exists n \in \mathbb{N}, v_1, v_2, \ldots, v_n \in X \cup X^{-1}, w = v_1 v_2 \ldots v_n\}$$

the set of all finite words (finite sequences or strings of letters) using the set $X \cup X^{-1}$ as the alphabet. This includes the empty word, which we will typically denote by ε. For the rest of the course, given a set X, we carry out all of the process above and form the set

$$W(X) := (X \cup X^{-1})^*.$$

1.2 Constructing free groups II: equivalence.

Given words $w_1, w_2 \in W(X)$, we say that $w_2 = v_1 v_2 \ldots v_{k+2}$ is an *elementary expansion* of $w_1 = u_1 u_2 \ldots u_k$ if there is an index $j \leq k$ so that ...

1. for all $i \leq j$, we have $u_i = v_i$,
2. for all $i > j + 2$, we have $u_{i-2} = v_i$, and
3. there is $t \in X \cup X^{-1}$ so that $v_{j+1} = t$ and $v_{j+2} = t^{-1}$.

In this case, we write $w_1 \nearrow w_2$ or $w_2 \searrow w_1$. If $w_2 \searrow w_1$ we can also say w_1 is an *elementary collapse* of w_2. If w_1 is an elementary expansion or an elementary collapse of w_2, we say these two words are *related by an elementary move.*

We say two words $w_1, w_2 \in W(X)$ are *similar* if there is a finite sequence of words $(w_1 = x_1, x_2, \ldots, x_p = w_2)$ so that for all indices i, we have x_i is related to x_{i+1} by an elementary move. In this case we write $w_1 \sim w_2$. Similarity is an equivalence relation, and we denote the equivalence class of any word $w \in W(X)$ by $[w]_\sim$. We denote by $W(X)/\sim$ the set of equivalence classes in $W(X)$ under our equivalence relation $\sim$.

Exercises will be dispersed throughout the text, usually placed near the end of a subsection where the relevant material is discussed. For the exercises below, take X to be any set.

Exercise 1. *Suppose X is a set and consider $a \in W(X)/\sim$. Prove that there is exactly one word w which is irreducible (admits no elementary collapses) in a.*

We call this word w the *freely reduced representative of* a, or by similar language. This then gives us a well defined function from $W(X)/\sim \to W(X)$ which we call a *normal form for elements of* $W(X)/\sim$, that is, a unique choice of representative word in $W(X)$ for any element of $W(X)/\sim$.

Remark 1. *Given $a \in W(X)/\sim$, we call the unique w of minimal length with $w \in a$ the* normal form representative for a. *Note that for a set P and an equivalence relation $\sim$, any function from $\theta : P/\sim \to P$ such that for all $a \in P/\sim$ we have $\theta(a) \in a$ is technically also called a* normal form for $\sim$ on P. *Thus, normal forms for any given equivalence relation always exist by the axiom of choice. Unfortunately, typical normal forms are useless, you first would need an efficient way to compute the normal form, just to obtain a representative element from any equivalence class!*

Exercise 2. *Discuss the last remark. What is the point of a normal form? How could one be used?*

1.3 Constructing free groups III: multiplication.

The set $W(X)/\sim$ admits an associative binary operation.

If a, $b \in W(X)/\sim$, then there are words u and v so that $a = [u]_\sim$ and $b = [v]_\sim$. Let uv denote the concatenation of the words u and v. We boldly claim that the definition $a \cdot b = [uv]_\sim$ is well defined. (That is, the answer is independent of the choices of words u and v representing a and b, respectively. It is a homework exercise to verify the claim.) We further claim (more homework) that $W(X)/\sim$ becomes a group under this multiplication, which we denote F_X, and we describe F_X as *the free group on X* or as *the free group with basis X*.

$$F_X := W(X)/\sim \text{ (with the product given by concatenation of representative words).}$$

(For the technically careful, the phrase "the free group on X" is actually a mis-use of the definite article "the", since the set X^{-1} is not unique, but this is a distinction which we all like to ignore: the resulting groups are isomorphic for different choices of set X^{-1} and map θ from the previous subsection, of course!)

Remark 2. *The following statements pertaining to F_X might be too obvious to be homework, but should still be mentioned.*

1. *If $e = [\varepsilon]_\sim$ is the (similarity) class of the empty word, then e is the identity element of F_X, often written 1_{F_X}.*
2. *If $w \in W(X)$ is a word, written $w = u_1^{\epsilon_1} u_2^{\epsilon_2} \dots u_n^{\epsilon_n}$ where each $u_i \in X$ and $\epsilon_i \in \{-1, +1\}$, then the class of the word $\overline{w} = u_n^{-\epsilon_n} u_{n-1}^{-\epsilon_{n-1}} \dots u_1^{-\epsilon_1}$ represents the inverse of $[w]_\sim$ in F_X. In the future, we will use $w^{-1} := \overline{w}$.*

Exercise 3. *Verify that if a, $b \in W(X)/\sim$, and there are words u, $v \in W(X)$ so that $a = [u]_\sim$ and $b = [v]_\sim$, then $ab = [uv]_\sim$ is a well defined product. (That is, verify that the answer for ab would be the same element of $W(X)/\sim$ regardless of the choice of words $u \in a$ and $v \in b$.)*

Exercise 4. *Verify that $W(X)/\sim$ is a group under the product operation described in the last exercise.*

Exercise 5. *Suppose $w \in W(X)$ so that $w \not\sim \varepsilon$. Argue that for all positive integers n, the word*

$$w^n := ww \cdots w$$

(with n-copies of w concatenated) is not similar to ε. Thus, F_X has no non-trivial elements of finite order.

Note that for integer $n < 0$, we will by w^n mean

$$(w^{-1})^{|n|}.$$

In a group, we call any nontrivial element of finite order a *torsion* element. The last exercise demonstrates that F_X, the free group on X, is a *torsion-free* group.

1.4 Free group definition the second!

Definition 1. *A subset X of a group F is said to be a* free basis *for F if every function $\phi : X \to G$ from the set X to a group G can be extended uniquely to a group homomorphism $\tilde{\phi} : F \to G$ so that $\tilde{\phi}(x) = \phi(x)$ for every $x \in X$.*

$$\begin{tikzcd}
X \arrow[r, phantom, "\subseteq"] \arrow[drr, "\phi"'] & & F \arrow[d, dotted, "\exists!\tilde{\phi}"] \\
& & G
\end{tikzcd}$$

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

Surprising that you really do grow to love it and actually hope that it comes up in another module!! Categories for the win!!!

Note that if H is a group and $X \subseteq H$ so that every function $\phi : X \to G$ to a group G *which can be extended to a group homomorphism* can only be extended in a unique fashion, then we say X *is a basis for* H.

Uniqueness, above, is easy. More subtle is the point that **if H is a given group with a basis X, then there may be many functions $\phi : X \to G$ for G a group, which cannot be extended to a group homomorphism**. Coming to understand the 'ins and outs' of this statement is a major goal for this and then next section of our notes.

Exercise 6. *Suppose $X \subseteq H$ for some group H. The set X is a basis for H using the definition above if and only if every element of H can be written as a finite product of elements in the set $X \cup X^{-1}$ (where here, X^{-1} is the set of inverses in H of elements in X).*

We say a word w in generators and inverses is a *relation* if its product in the group evaluates to the identity.

Exercise 7. *If a group is free on basis X (i.e., with the universal property definition) then it is generated by the free basis and it has no non-trivial relations.*

1.5 Free groups exist

Theorem 1. *Given a non-empty set X, the group $F_X = W(X)/\sim$ is a free group with free basis X.*

Proof. Let G be any group and let $\phi : X \to G$ be a set function. For $a \in F_X$ let $w = x_1^{\epsilon_1} x_2^{\epsilon_2} \dots x_k^{\epsilon_k} \in a$ be any representative of a, and define $\tilde{\phi}(a) = \phi(x_1)^{\epsilon_1}\phi(x_2)^{\epsilon_2} \dots \phi(x_k)^{\epsilon_k}$, where $\epsilon_i = \pm 1$ for all indices i. By our homework which shows multiplication in F_X is well defined as word concatenation, we see that $\tilde{\phi}$ is a group homomorphism. As $\tilde{\phi}$ is determined on a set of generators for F_X by ϕ, we see that $\tilde{\phi}$ is the unique group homomorphism $F_X \to G$ extending ϕ. $\square$

In the remainder of these notes, we call such a map $\tilde{\phi}$ the *linear extension of the set map ϕ.*

1.6 A nice free group property

Corollary 2. *Every group is a quotient of a free group.*

Proof. Let G be a group and set $X = G$ as a set. Set $\phi : X \to G$ be given by the identity map. Now there is a unique group homomorphism $\tilde{\phi} : F_X \to G$ extending ϕ which is trivially surjective. $\square$

This construction is horrid; we should at least restrain ourselves to map from F_S where S is a minimal set of generators of G.

Definition 2. *The* rank *of a group G is the minimal cardinality over all sets of generators of G.*

$$d(G) := \min\{|S| \mid S \subseteq G \text{ and } \langle S\rangle = G\}.$$

Exercise 8. *Suppose G is a group, and $d(G) = k$ some natural number. Let X be a set so that $|X| = k$. Show that G is a quotient of F_X.*

We call such a group F_X as above a *free group of rank k*, and we might denote such as F_k given a natural number k. Of course, for given $k > 0$, F_k is not unique, but for two sets X and Y with $|X| = |Y| = k$, the groups F_X and F_Y are isomorphic (we will see this soon), so we treat F_k as a unique well-defined group.

Exercise 9. *Describe the groups F_0, F_1, and F_2.*

Exercise 10. *Prove that F_2 is not abelian.*

Exercise 11. *Suppose a group F is free on basis X and that there are a, $b \in F$ non-trivial elements so that $ab = ba$. Show there is $c \in F$ and integers j and k so that $a = c^j$ while $b = c^k$.*

2 L2: Free Groups II (Structure Theorems and Examples)

2.1 The equivalence of F_X with free group F with basis X

(In this section, and from now on, we release the name θ from the map used in the construction of $W(X)$, so we can use it for other purposes.)

Theorem 3. *Suppose X is free basis for a group F and consider the group $F_X = W(X)/\sim$. Then, $F_X \cong F$.*

Proof. By the universal property of a group with a free basis, we have the following diagram.

$$\begin{tikzcd}
X \arrow[r, phantom, "\subseteq"] \arrow[dr, "\phi(x)=[x]_\sim"'] & F \arrow[d, dotted, "\exists!\tilde{\phi}"] \\
 & F_X
\end{tikzcd}$$

where $\tilde{\phi}$ is surjective since ϕ is onto the generators of F_X.

There is also another map $\theta : F_X \to F$. Suppose $w = [x_1^{\epsilon_1} x_2^{\epsilon_2} \dots x_k^{\epsilon_k}]_\sim \in F_X$, where $x_i \in X$ and $\epsilon_i = \pm 1$ for all indices i. We define $\theta(w)$ to be $x_1^{\epsilon_1} x_2^{\epsilon_2} \dots x_k^{\epsilon_k} \in F$. One should check that θ is well defined (this last is a homework exercise).

Now if $w_1 = [u_1^{\delta_1} u_2^{\delta_2} \dots u_m^{\delta_m}]_\sim$ and $w_2 = [v_1^{\gamma_1} v_2^{\gamma_2} \dots v_n^{\gamma_n}]_\sim$ are two elements of F_X represented by words $u_1^{\delta_1} u_2^{\delta_2} \dots u_m^{\delta_m}$ and $v_1^{\gamma_1} v_2^{\gamma_2} \dots v_n^{\gamma_n}$ respectively (where u_I and $v_i \in X$ and δ_i and $\gamma_i = \pm 1$ for all valid indices i), then

$$\theta(w_1)\theta(w_2) = u_1^{\delta_1} u_2^{\delta_2} \dots u_m^{\delta_m} v_1^{\gamma_1} v_2^{\gamma_2} \dots v_n^{\gamma_n} = \theta(w_1 w_2),$$

so θ is a homomorphism of groups. Furthermore, θ is onto the generators of F so θ is surjective.

Direct computation now shows that for all $x \in X$, we have $\theta(\tilde{\phi}(x)) = x$ and $\tilde{\phi}(\theta([x]_\sim)) = [x]_\sim$ so θ and $\tilde{\phi}$ are left inverses of each other, and so both maps are injective, too. □

Thence, we have shown both that given a set X, a free group with basis X always exists, and that it is always isomorphic to our constructed group $F_X = W(X)/\sim$. We henceforth call any such constructed group $F_X = W(X)/\sim$ by the language 'The free group on basis X', noting that it is a group with free basis X.

Exercise 12. *Show the map θ in the proof above is well defined. That is, $\theta(w)$ is well defined for all $w \in F_X$; it does not depend on the choice of representative word for w.*

Exercise 13. *For any two groups G and H, if there are maps (group homomorphisms) $p : G \to H$ and $q : H \to G$ which are both right inverses (or both left inverses) of each other, then p and q are both monic (injective, or one-to-one).*

A note in passing: we also use the words *epic*, *surjective*, and *onto* interchangeably as well.

Exercise 14. *For any two groups G and H, if there are maps (group homomorphisms) $p : G \to H$ and $q : H \to G$ which are both right inverses (or both left inverses) of each other, then p and q are both epic (surjective, or onto).*

2.2 The rank of a free group

Theorem 4. *Let X and Y be sets. We have $F_X \cong F_Y$ if and only if $|X| = |Y|$.*

The author learned this proof from [4]

Proof. For any group F with free basis T, consider the 'Dirac projection' $d_F : F \to \bigoplus_T \mathbb{Z}_2$ which is the unique extension of the set function that takes each element $t \in T$ to the element $\delta_t \in \bigoplus_T \mathbb{Z}_2$ which is 1 in the coordinate t and 0 in all other coordinates. We observe d_F is surjective.

If T is finite, then $\bigoplus_T \mathbb{Z}_2$ has the cardinality of $\mathscr{P}(T)$, the power set of T, otherwise, $|\bigoplus_T \mathbb{Z}_2| = |T|$ (why is this not the cardinality of the power set of T in this case?).

Note that $ker(d_F) = \langle w^2 | w \in F\rangle$ (we will call this last subgroup the 'Square subgroup' for any group F, and it is independent of any generating basis of F). In particular, $|F_X/ker(d_{F_X})| = |F_Y/ker(d_{F_Y})|$ if and only if $|X| = |Y|$.

□

We now obtain this nice corollary.

Theorem 5. *All free bases for a free group F have the same cardinality.*

We mention a bit of notation. If G is a group, $a, b \in G$ and i is an integer, we set the notation $a^b := b^{-1} \cdot a \cdot b$ (the conjugate of a by b) and $a^i := a \cdot a \cdots a$ (with i occurences of a in the product – if $i = 0$ then $a^i := 1_G$).

We can now mention two further theorems, one with proof, and one without proof (until later in the course).

Theorem 6. *Let m, n be integers greater than 1. Then there are embeddings $F_m \to F_n$ and $F_n \to F_m$.*

Proof. If $m < n$ then it is easy to see the embedding of F_m into F_n. Therefore, we will argue the other direction.

First off, observe that $F_{\{a,b\}}$ embeds in both F_m and F_n. We show that F_n embeds in F_2.

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set with n elements. We embed it into $F_{\{a,b\}}$ as follows:

For each i, send x_i to a^{b^i}. This gives us a set function from the generators of F_n into $F_{\{a,b\}}$, and there is a unique extension of this to a homomorphism from F_n to $F_{\{a,b\}}$. Direct calculation should now show you that this homomorphism is actually injective. □

The last sentence above is a bit mysterious. Let us make a point of this!

Exercise 15. *Consider the group $H = \langle a, a^b, a^{b^2}\rangle \leq F_{\{a,b\}} \cong F_2$ has $H \cong F_3$, and define $\tau : F_{\{r,s,t\}} \to F_{\{a,b\}}$ to be the linear extension of the map ($r \mapsto a$, $s \mapsto a^b$, $t \mapsto a^{b^2}$). Show that τ is monic.*

That turned out to be more complicated than I thought...

Theorem 7 (Neilsen-Schreier Theorem)**.** *Every subgroup G of a free group F is a free group. Moreover, if G has finite index m in F, then $m(rank(F) - 1) = rank(G) - 1$.*

This last is a deep theorem. We will revisit it from different perspectives, and it will become more apparent as we go.

3 L3: Presentations I (Presentations of Groups and Monoids, and Cayley Graphs)

3.1 Normal Closure in Groups.

Definition 3. *Let G a group, and $S \subseteq G$. We define the* normal closure of S in G *to be the smallest normal subgroup of G containing S. We denote the normal closure of S in G by the expression*

$$\textit{Normal Closure}_G(S).$$

Before discussing the normal closure in more detail, let us give two further constructions.

If X, $Y \subseteq G$ a group, then we define the expression X^Y as follows.

$$X^Y := \{x^y | x \in X, y \in Y\}$$

Suppose G is a group and $S \subseteq G$. This situation arises, and we will often want to consider the group $\langle S^G \rangle$. As that notation is a bit clumsy (and ugly!), when the group G is understood, we will generally use a different notation, as follows:

$$\langle\langle X \rangle\rangle := \langle X^G \rangle.$$

Another construction that will often arise in the context of a subset S of a group G, is to take the intersection of all the normal subgroups of G which contain S, that is, to consider the set

$$\bigcap_{S \subseteq N \triangleleft G} N.$$

Why do we need all of these things? The first part of the answer is that they are all the same object.

Exercise 16. *Suppose G is a group, and S is a subset of G. Show the following are equivalent.*

1. *$\textit{Normal Closure}_G(S)$.*
2. *$\langle\langle S \rangle\rangle$.*
3. *$\langle S^G \rangle$.* ...Really? You ask us this???
4. *$\bigcap_{S \subseteq N \triangleleft G} N$.*

Which we formalise by the following lemma.

Lemma 8. *Suppose G is a group, and S is a subset of G. The following are equivalent.*

1. *$\textit{Normal Closure}_G(S)$.*
2. *$\langle\langle S \rangle\rangle$.*
3. *$\langle S^G \rangle$.*
4. *$\bigcap_{S \subseteq N \triangleleft G} N$.*

The second part of the answer is given in the next subsection.

3.2 Group Presentations – Definition

Definition 4. *Now suppose X is a set and $R \subseteq W(X)$. We define the symbol*

$$\langle X \mid R \rangle$$

to represent the group $F_X/\langle R^{F_X}\rangle = F_X/\langle\langle R\rangle\rangle$, and we call this group the group presented by $\langle X \mid R\rangle$.

We will often write something akin to $G = \langle X \mid R\rangle$. Of course, this means G IS the quotient of the free group on X by the normal subgroup $\langle\langle R\rangle\rangle$. In the definition above, we call every word in the set R a *generating relator*, so that R *is the set of generating relators.* Meanwhile, any word in $W(X)$ evaluating in F_X to an element in $\langle\langle R\rangle\rangle$ is a *relator for the group $F_X/\langle\langle R\rangle\rangle$*. If we have two words w_1 and w_2 so that $w_1 \cdot (w_2)^{-1}$ is a relator for the group $F_X/\langle\langle R\rangle\rangle$, then we say $w_1 = w_2$ is a *relation in the group $F_X/\langle\langle R\rangle\rangle$*.

We will write $H \cong \langle X \mid R\rangle$ when we mean a group H, which may be given to us in some other way (not via a presentation), is isomorphic to $F_X/\langle\langle R\rangle\rangle$.

Also, we might give X and R as lists of letters and words in $W(X)$ when naturally we mean the sets containing these lists, simply so that the notation is not to heavy with braces which can be understood without being written down. See the next subsection for examples of this.

3.3 Group presentations – Examples

We give some examples for discussion:

1. $\langle \{a\} \mid \varnothing\rangle = \langle a \mid \rangle \cong \mathbb{Z}$
2. $\langle a, b \mid \rangle = F_{\{a,b\}}$
3. $\langle a, b \mid [a, b]\rangle \cong \mathbb{Z} \times \mathbb{Z}$
4. $\langle a, b \mid a^2, b^3, abab\rangle \cong S_3$
5. $\langle a, b \mid a^2, b^2, (ab)^3\rangle \cong S_3$
6. $\langle a, b \mid a^2, b^n, abab\rangle \cong D_n$
7. $X = \{\sigma_1, \sigma_2, \ldots, \sigma_{n-1}\}$
   $R = \{\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}\,\forall i, [\sigma_i, \sigma_j]\,\forall |i - j| > 1\}$
   $\langle X \mid R\rangle \cong B_n$

Exercise 17. *Consider the group $Q := \langle x, i, j, k \mid i^2x^{-1}, j^2x^{-1}, k^2x^{-1}, x^2, ijkx^{-1}\rangle$.*

1. *Show that $\langle x\rangle$ is a subgroup of order two, and it is in the centre of Q.*
2. *Show that $|i|$ is 4.*
3. *Compute the normal closure of $\{i\}$.*
4. *The group $Q/\langle\langle x\rangle\rangle$ is a group you understand. What is that group?*
5. *The group Q is a group you know. What is that group?*

Given a presentation $G = \langle X \mid R\rangle$ of a group, we call any word $w \in W(X)$ with $w \in \langle\langle R\rangle\rangle$ a *consequence of the relations*. By definition, these are precisely the words which are equivalent in G to the identity element. We might write $w =_G 1_G$ or $w =_G \varepsilon$ when we want to emphasise that the equivalence is specifically in the group G.

For example, consider $X = \{a, b\}$ and $R = \{a^2, b^3, abab\}$ let $G = \langle X \mid R\rangle$, and $w = b^2$. Then we can write

$$aba =_G b^2.$$

But we note that

$$aba \neq_{F_X} b^2.$$

3.4 Normal closure in monoids.

Recall that (or "learn that...", if this is your first exposure) a *monoid* is a set with a binary operation, so that the binary operation admits an identity element, and where the binary operation is associative. In particular, monoids are like groups, except that one is not guaranteed the existence of an inverse to every particular element. Relatively speaking, monoids are much freer objects than groups: For large integer n, there are many more monoids of order n than there are groups of order n (up to isomorphism).

Now, taking quotients of monoids is a bit different to taking quotients of groups; we do not have the idea of a "normal submonoid," since conjugation makes little sense without inverses.

Suppose the monoid M is generated by a set X. Then, as in the groups case, the monoid is a (monoid) quotient from the free monoid on X. The free monoid on X is exactly the set X^* (since we cannot throw in inverse generators), where the monoid product is again concatenation.

To take a quotient, we want to put an equivalence relation on the free object X^* (just like in groups). So, we take a set of pairs of words $R = \{(u, v), (w, x), \ldots\}$ where u, v, w, x, ... are all elements of X^*, and we simply say that two words t_1, $t_2 \in X^*$ are equivalent if there are words a, b, c, $d \in X^*$ so that $t_1 = acb$ and $t_2 = adb$ where the pair (c, d) or the pair (d, c) is in R. We then extend this equivalence through transitivity. This forces any pair of words in R to be equivalent in the quotient, and any two words will represent the same class if there is a finite chain of substitutions from one word to the other, where each substitution is given by employing an equivalence listed in R. Just like in groups, it turns out that the product given by concatenation will still induce a well defined product on the equivalence classes.

For a monoid presentation, we will place a subscript "m" on the presentation symbol. Thus:

$$\langle a \mid \varnothing \rangle \cong \mathbb{Z}$$

while

$$\langle a \mid \varnothing \rangle_m \cong \mathbb{N}.$$

For a monoid quotient, we will stick with the normal closure notation $\langle\!\langle R \rangle\!\rangle$, but what this will now mean in the context of monoids is to modify the set R first by adding all inverse relations (so if $(u, v) \in R$ we add (v, u) as another relation in R), and then taking equivalence under the transitive closure of equivalences from (the new version of) R. Thus, we we still denote the monoid quotient of X^* by R as

$$\langle X \mid R \rangle_m \cong X/\langle\!\langle R \rangle\!\rangle.$$

Exercise 18. *Understand that if, for each generator x in X, there is a generator $y \in X$ so that $(xy, \varepsilon) \in R$ and $(yx, \varepsilon) \in R$, then $\langle X, R \rangle_m$ will actually be a group under this quotient operation.*

3.5 The Cayley Graph

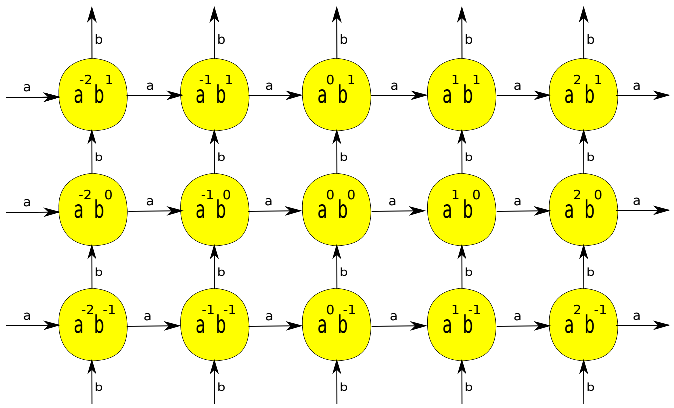
In this subsection, we will look at our first way to "visualise" a group. We will construct the Cayley Graph $\Gamma(G, X)$ for a pair (G, X), where G is a group, X is a set of generators of G, and $\Gamma(G, X)$ is a directed graph.

The vertices of $\Gamma(G, X)$ are precisely the elements of G, we denote these as $V(\Gamma(G, X))$. The directed edges of $\Gamma(G, X)$ are precisely the pairs (g, h) where, g, $h \in G$, and where $g^{-1}h = x$ for some $x \in X$. That is, we draw an edge from g to h (labelled by x, if you want) whenever $gx = h$.

Below is a presentation for $\mathbb{Z} \times \mathbb{Z}$, and immediately after that, a Cayley graph for this group.

$$\mathbb{Z}^2 = \langle a, b \mid a^{-1}b^{-1}ab \rangle$$

Below is a portion of the Cayley graph $\Gamma(\mathbb{Z}^2, \{a, b\})$. (Here, I have drawn very fat vertices, so as to ease the labelling. Normally, the labelling of the vertices is essentially forgotten, and the vertices are drawn as small dots (or not even emphasised at all, as on the cover of these notes), and different labelled edges might just have different colours as their "labels."



Exercise 19. *Consider this presentation below of D_4, the Dihedral group of order eight. Draw $\Gamma(D_4, \{a, r\})$ as neatly as you can:*

$$D_4 = \langle a, r \mid a^4, r^2, arar \rangle.$$

Exercise 20. *Consider the Cayley graph in the text given for $\mathbb{Z}^2$. What can you say about any closed path (a path ending where it began) in this complex? Do you think you can say a similar thing about the Cayley graph $\Gamma(G, X)$ for any group G generated by a set X?*

4 L4: Rewriting Systems

4.1 Graphs

Recall that a graph Γ is a tuple $\Gamma = (V, E, ends : E \to \mathscr{P}(V))$ where V is the set of vertices, E is the set of edges, and *ends* is a map that takes any edge to the set of vertices (one or two) which are at the ends of the edge (the vertices *incident to* the edge). If the graph is directed we replace *ends* by two maps s and t, where s takes an edge and produces the start or beginning vertex of the edge (the *source map*), and t takes an edge and produces the final or terminal vertex of the edge (the *target map*). Directed edges are usually represented by an ordered pair $e = (a, b)$ where e is the directed edge, a and b are vertices, and $s(e) = a$ while $t(e) = b$.

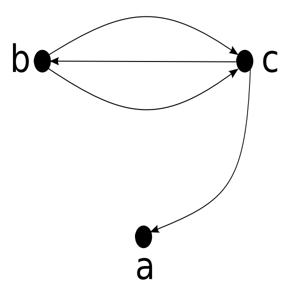
We single out special graphs, *lines*, which are directed graphs using as vertex set X a subset of $\mathbb{Z}$, where X is of the form $[a, b] \cap \mathbb{Z}$ for real $a \leq b$, or or the form $(-\infty, b] \cap \mathbb{Z}$ or of the form $[a, +\infty) \cap \mathbb{Z}$, or simply $X = \mathbb{Z}$. For two vertices $i-1$ and i in a line, there is precisely one edge (denoted e_i) in the line, and this edge has has $s(e_i) = i-1$, $t(e_i) = i$. The edges e_i so described are all of the edges of the line. Assuming a, b are integers, we then denote the resulting lines as $L_{[a,b]}$, $L_{(-\infty,b]}$, $L_{[a,\infty)}$ and $L_{\mathbb{Z}}$, respectively. We specify, for $n \in \mathbb{N}$ the line $L_n := L_{[0,n]}$ as our canonically favourite line of length n.

A path of length n in a graph Γ is the image of a graph homomorphism p from L_n into Γ. Recall that a graph homomorphism is a function which takes vertices to vertices, and edges to edges, and preserves all senses of adjacency (a directed edge must go to a directed edge, if it is a homomorphism of directed graphs, and respect the orientation). One should think about what this should mean in both cases: Γ directed versus Γ undirected. **We will sometimes consider paths in a directed graph, but loosening the definition to allow ourselves to consider the path as a map into an undirected version of the graph.** If an edge e_i of a line L_* crosses a edge d in the directed Γ in the right orientation on the path $p : L_* \to \Gamma$, we will say $p(e_i) = d$. But if it crosses the edge d with the wrong orientation, then we will say $p(e_i) = \overline{d}$. We call this "overline" map on edges "edge involution."

A path in a directed graph which preserves all senses of orientation (so the homomorphism from L_* to Γ is a directed graph homomorphism) is called a *directed path* to emphasise that the path obeys all orientations in the range.

In any case, the start, beginning or initial vertex of a path p of length n in a graph Γ (so p is a graph homomorphism $p : L_n \to \Gamma$) is the image of the vertex 0, and the finish or ending or terminal vertex of the path is the image of the vertex n. We extend this language for infinite paths in the obvious ways.

We will also sometimes call a directed graph $\Gamma = (V, E, s : E \to V, t : E \to V)$ is called a (digraph). E.g., diagrammatically, a digraph.

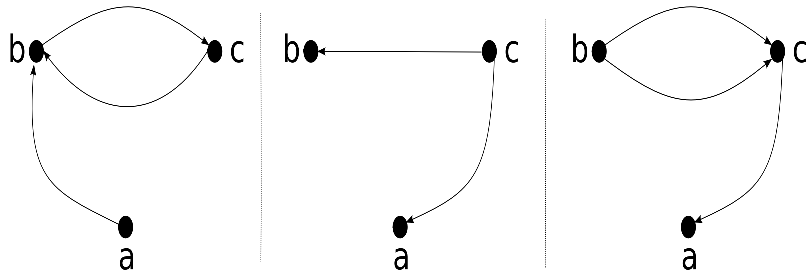


4.2 Noetherian digraphs and confluent digraphs

Let $\Gamma = (V, E, s, t)$ be a digraph.

We say Γ is *Noetherian* (terminating) if there are no infinite directed paths in Γ.

We say Γ is *confluent* if, for every vertex v, after having travelled away from v on any two finite directed paths, p_1 and p_2, there is a vertex z and two paths q_1 and q_2 starting at $t(p_1)$ and $t(p_2)$ respectively so that $t(p_1 \circ q_1) = t(p_2 \circ q_2) = z$.



4.3 Complete rewriting systems

Let A be a set, and $W \subseteq A^*$. We call a map $\to: W \to A^*$ a *string rewriting system.*

1. Build digraph $\Gamma(\to) = (A^*, E, s, t)$ where there is a directed edge from a string w (a *word in alphabet* A) to a string v whenever w and v admit subwords w' and v' so that $w' \to v'$. (A *substring substitution* takes w to v.)

2. We say two words w and v are equivalent if there is an undirected path of finite length m ($m = 0$ is allowed) in Γ taking w to v.

3. If Γ is Noetherian and confluent, then we say $\to: W \to A^*$ is a *complete string rewriting system.*

4.4 Newman's Lemma

We say a digraph is *locally confluent* if given any vertex v and two edges e_1 and e_2 with $s(e_1) = s(e_2) = v$, there is a vertex z and two directed paths p_1 and p_2 so that $t(p_1) = t(p_2) = z$ and $s(p_1) = t(e_1)$ while $s(p_2) = t(e_2)$.

Lemma 9 (Newman). *Let $\Gamma = (V, E, s, t)$ be a directed graph which is terminating and locally confluent. Then, Γ is confluent.*

M.H.A. Newman. On theories with a combinatorial definition of "equivalence". Annals of Mathematics, 43, no. 2, pp. 223–243, 1942.

Corollary 10. *Let $W \subseteq A^*$ for some set A, and let $\to: W \to A^*$ generate a Noetherian, locally confluent graph $\Gamma(\to)$. Then $\to$ generates a complete rewriting system.*

4.5 Normal forms

Given an equivalence relation on a set W, we call a choice of unique representative for each equivalence class a *transversal* of the partition generated by the equivalence relation, or a *normal form for the partition.*

From the point of view of a presented group $G = \langle X \mid R \rangle$, we are looking for a system of unique representatives for the cosets of $\langle\langle R \rangle\rangle$ in F_X. This gives us a unique representative word for each element of $G = F_X / \langle\langle R \rangle\rangle$; a *normal form* for elements of G.

Lemma 11. *Any class of equivalent words under a complete re-writing system has a unique irreducible element.*

Proof. Take any representative word, and travel any directed path in $\Gamma(\to)$ from your initial choice until you cannot travel any further. The Noetherian property guarantees you must stop (on an *irreducible*). The confluence property guarantees there are not two distinct irreducibles in the equivalence class. $\square$

The Point Is ...

Normal forms give us a unique way to express every element of a presented group. Thus, we can count elements, verify multiplication tables, and verify images of maps, etc.

Exercise 21. *Consider the group*

$$G := \langle a, b \mid a^2, b^3, abab \rangle$$

and the system of rewrite rules:

1. $a^{-1} \to a$, $b^{-1} \to b^2$.
2. $a^2 \to \epsilon$, $b^3 \to \epsilon$.
3. $ba \to ab^2$

Show this system is both terminating (Noetherian) and confluent for this presentation of G. What group is G? Can you prove it from the evidence gathered? [Hint: look up the wiki for local confluence, and learn what "critical pairs" are.]

The following exercise is a major piece of work, which ties together many of the ideas we have seen so far.

Exercise 22. *Let $G = \langle X \mid R \rangle$ be a presentation for a group.*

1. *Construct the set $C \subseteq W(X)$ which consists of words which can be built iteratively by applying the following elementary moves applied to the empty word ϵ.*
   (a) *At any location in the current word, and for any letter $x \in X$ insert a trivial cancelling pair xx^{-1} or $x^{-1}x$.*
   (b) *At any location in the current word, delete a trivial cancelling pair which already exists in the current word as a contiguous sub-word.*
   (c) *At any location in the current word, insert a word in R.*
   (d) *At any location in the current word, delete a word in R which already exists in the current word as a contiguous sub-word.*

   Argue that the words which can be so built are precisely the set of words in $W(X)$ which are in the equivalence class of the empty word in the definition of $G = F_X/\langle\langle R \rangle\rangle = (W(X)/\sim_1)/\sim_2$ where $\sim_1$ is equivalence under free reduction/insertion of trivial pairs, and $\sim_2$ is equivalence of those classes under deletion/insertion of a word in R.

2. *Let $\to$ represent a confluent terminating string rewriting system which includes the trivial reductions, and for each relator $r \in R$, at least one arrow of the form $w \to v$ where $s^g = wv^{-1}$ where $s = r$ or $s = r^{-1}$ and $g \in W(X)$. Now, let $s \in W(X)$.*
   (a) *Show that if $s \in R$, then s is in the same graph component as the empty word in $\Gamma(\to)$.*
   (b) *Show that if $s^{-1} \in R$, then s is in the same graph component as the empty word in $\Gamma(\to)$.*
   (c) *Show that if $s \in C$, the set of words constructed in part (1), then s is in the same graph component as the empty word in $\Gamma(\to)$.*
   (d) *Show that if $s \notin C$, then s is not in the same graph component as the empty word in $\Gamma(\to)$.*
   (e) *Conclude that as $\to$ is a complete string rewriting system, then $\to$ chooses a unique irreducible word representing each element of G.*

NB: In the previous exercise in part (2), you could remove the condition that the confluent system contains the trivial reductions, if you replace that condition with the condition that there are re-write rules which will allow all the consequences of these reductions. So, e.g., if x has order four, you could have the two rules: $x^{-1} \mapsto x^3$ and $x^4 \mapsto \varepsilon$.

5 L5: Presentations II (von Dyck, Tietze, and Markov)

Now that we have a strong handle on presentations, we might wonder what we can do with them, and what we can know about a presented group when we have a presentation in hand. This lecture will touch on three aspects of those questions.

5.1 von Dyck's Theorem

In this section, we explore the creation of homomorphisms between two presented groups (or at least, when we have a presentation for the domain group, and when we can check if a product of generators of the range group is trivial). The theorem in this section is due to the amazing Walther von Dyck.

Walther von Dyck is often credited as the first mathematician to give a formal definition of a group (see [1]). He was also instrumental in starting to properly understand non-orientable surfaces such as the Möbius band. The following core theorem essentially asserts that a function from a basis of a group G_1 to another group G_2 extends to a group homomorphism exactly when the relators of G_1 become trivial in the image group G_2, which is precisely the minimal condition for which one would hope!

Theorem 12 (von Dyck, 1882)**.** *Suppose $\langle X_1 \mid R_1\rangle = G_1$ and $\langle X_2 \mid R_2\rangle = G_2$ present two groups, and let $\theta : X_1 \to W(X_2)$ be a function. We have that θ extends to group homomorphism $\tilde{\theta} : G_1 \to G_2$ exactly when the image of each relator in R_1 is equivalent to the trivial word in the group $\langle X_2 \mid R_2\rangle$.*

Proof. First we note that there is a unique map $\tilde{\theta} : G_1 \to G_2$ extending θ which can pretend to be a group homomorphism. Elements of G_1 are equivalence classes of words written in the generators of G_1, where the group product is found by concatenating representative words, and then determining the equivalence class of the resulting word (that is, which coset of $\langle R_1^{F_{X_1}}\rangle$ contains the resulting word). Thus, if we know where each generator of G_1 goes under θ, we can pick a representative word in $W(X_2)$ for each such image. Now if we have an element $w \in G_1$, it is a (signed) product of generators, and for $\tilde{\theta}$ to be a group homomorphism, we need $\tilde{\theta}(w)$ to be the equivalence class of the appropriate concatenation of the image words of the generators in X_1.

It is necessary that the image of each relator in R_1 under $\tilde{\theta}$ is equivalent to the trivial word in G_2, else $\tilde{\theta}$ would send a word representing the trivial element of G_1 to a word not representing the trivial word in G_2, and thus $\tilde{\theta}$ would thus not be a group homomorphism.

It is also sufficient that the image of each relator in R_1 under $\tilde{\theta}$ is equivalent to the trivial word in G_2. Note that here is an induced homomorphism $\psi : F_{X_1} \to G_2$ from the free group on X_1 to G_2 (since we know where each element of X_1 goes), and this homomorphism has kernel $K_{(F_{X_1}\to G_2)}$ containing the kernel K_1 of the natural homomorphism $\eta_1 : F_{X_1} \to G_1$. Now the image of $K_{(F_{X_1}\to G_2)}$ in G_1 under η_1 is a normal subgroup of G_1 by the correspondence theorem, and this is exactly the kernel of the 'linear extension' $\tilde{\theta}$. That is, ψ factors as $\eta_1 \cdot \tilde{\theta}$. $\square$

The importance of von Dyck's theorem cannot be overstated.

Exercise 23. *Consider the presentations*

$$Q := \langle x, i, j, k \mid i^2x^{-1}, j^2x^{-1}, k^2x^{-1}, x^2, ijkx^{-1}\rangle$$

and

$$K := \langle a, b \mid a^2, b^2, abab\rangle.$$

Let $\theta : \{x, i, j, k\} \to K$ be given by the rules: $\theta(x) = 1_K$, $\theta(i) = a$, $\theta(j) = b$, and $\theta(k) = ba$. Show that θ extends to a group homomorphism $\tilde{\theta}$ from Q to K.

[1] von Dyck, Walther (1882), "Gruppentheoretische Studien (Group-theoretical Studies)", Mathematische Annalen 20 (1): 144, doi:10.1007/BF01443322, ISSN 0025-5831. (German)

5.2 Tietze Transforms

Let $G = \langle X \mid R \rangle$. In 1908, H. Tietze built the technology for changing the sets X and R in such a way as to create new presentations for the group G. That is, he gives four 'moves' (*Tietze transformations*) so that if one applies these moves the sets X and R change as $X \mapsto X'$ and $R \mapsto R'$, but

$$\langle X \mid R \rangle \cong \langle X' \mid R' \rangle.$$

We therefore say both $\langle X \mid R \rangle$ and $\langle X' \mid R' \rangle$ *present* G.

H. Tietze, *Über die topologischen Invarianten mehrdimensionalen Mannigfaltigkeiten*, Monatsh. f. Math. u. Physik, **19**, (1908), 1–118.

5.3 The Transformations

The moves are below.

1. (T1) If the words in $S \subseteq W(X)$ are derivable from those in R, then $\langle X \mid R \cup S \rangle$ presents G.
2. (T2) If the words in $D \subseteq R$ are derivable from the words in $R \backslash D$, then $\langle X \mid R \backslash D \rangle$ presents G.
3. (T3) Let K is a set disjoint from $X \cup X^{-1}$ and $N \subseteq W(X)$ so that $\phi : K \to N$ is a bijection, then $\langle X \cup K \mid \{k^{-1}\phi(k) \mid k \in K\} \cup R \rangle$ presents G.
4. (T4) If there is a set of generators $M \subseteq X$, and $\forall m \in M$, there is a relator of the form $m^{-1}W_m$ where the word W_m does not involve any generator (or inverse of a generator) in M, then the presentation

   $$\langle X \backslash M \mid Y \rangle$$

   presents G, where Y is the result, $\forall m \in M$, of removing all of the relators $m^{-1}W_m$ from R and transforming the remaining relators by $m \mapsto W_m$, $m^{-1} \mapsto W_m^{-1}$.

Note that we say a Tietze transformation is *elementary* if you only add or delete one new generator or relator with the move. (i.e., deleting one extraneous generator or relator with the move, or adding an extraneous generator or relator with the move.) Let us practice these moves a bit.

Recall we have two claimed presentations for S_3.

1. $\langle a, b \mid a^2, b^2, (ab)^3 \rangle$, and
2. $\langle a, b \mid a^2, b^3, abab \rangle$.

We show these both present the same group. (Recall from our practice with normal forms that the second presentation really does present a non-abelian group of order 6; i.e., S_3.) Here is a way to write the moves; Write the initial presentation, then an arrow, labelled by the type of move, and then the final presentation. E.g.,

$$\begin{gathered}
\langle a, b \mid a^2, b^2, (ab)^3 \rangle \mapsto^{T_3} \\
\langle a, b, c \mid a^2, b^2, (ab)^3, c^{-1}ab \rangle \mapsto^{T_1} \\
\langle a, b, c \mid a^2, b^2, (ab)^3, c^{-1}ab, b^{-1}a^{-1}c \rangle \mapsto^{T_2} \\
\langle a, b, c \mid a^2, b^2, (ab)^3, b^{-1}a^{-1}c \rangle \mapsto^{T_4} \\
\langle a, c \mid a^2, (a^{-1}c)^2, (aa^{-1}c)^3 \rangle \mapsto^{T_1} \\
\langle a, c \mid a^2, (a^{-1}c)^2, (aa^{-1}c)^3, c^3 \rangle \mapsto^{T_2} \\
\langle a, c \mid a^2, (a^{-1}c)^2, c^3 \rangle \mapsto^{T_1} \\
\langle a, c \mid a^2, (a^{-1}c)^2, c^3, (ac)^2 \rangle \mapsto^{T_2} \\
\langle a, c \mid a^2, c^3, (ac)^2 \rangle \mapsto T_3 \\
\langle a, b,c \mid a^2, c^3, (ac)^2, b^{-1}c \rangle \mapsto^{T_1} \\
\langle a, b, c \mid a^2, c^3, (ac)^2, b^{-1}c, c^{-1}b \rangle \mapsto^{T_4} \\
\langle a, b \mid a^2, b^3, (ab)^2, b^{-1}b \rangle \mapsto^{T_2} \\
\langle a, b \mid a^2, b^3, (ab)^2 \rangle.
\end{gathered}$$

The following problem is taken from an example of Baumslag, from his fantastic little book [3].

Exercise 24. *Prove that the following presented group is trivial.*

$$\langle a, b \mid a^{-1}bab^{-2}, b^{-1}aba^{-2} \rangle$$

5.4 Tietze's Theorem

With the moves, Tietze gives this very nice theorem.

Theorem 13 (Tietze,1908)**.** *Given two presentations for a group G,*

$$\mathscr{P}_1 = \langle X_1 \mid R_1 \rangle$$

and

$$\mathscr{P}_2 = \langle X_2 \mid R_2 \rangle$$

then $\mathscr{P}_2$ can be transformed into $\mathscr{P}_1$ by repeated applications of the transformations $(T_1) - (T_4)$.

Corollary 14. *In the above theorem, if both presentations are finite, then $\mathscr{P}_2$ can be changed into $\mathscr{P}_1$ in a finite number of elementary Tietze transformations.*

Proof of theorem (outline):

1. Let $\alpha_1 : \mathscr{P}_1 \to G$ and $\alpha_2 : \mathscr{P}_2 \to G$ be isomorphisms. Proceed as follows.
2. Set alphabets $A_1 = \{x\alpha_1 \mid x \in X_1\}$, $A_2 = \{x\alpha_2 \mid x \in X_2\}$. (Note, G is generated by A_1 and by A_2.)
3. Choose $w_{12} : A_1 \to W(A_2)$ choosing a word in A_2 representing a_1 in G for each $a_1 \in A_1$. Symmetrically define a choice function $w_{21} : A_2 \to W(A_1)$.
4. Let $\theta_1 : W(A_1) \to W(X_1)$ defined by applying α_1^{-1} repeatedly across the letters of any word in $W(A_1)$. Symmetrically define $\theta_2 : W(A_2) \to W(X_2)$.
5. Set $\tau_{12} : X_1 \to W(X_2)$ by $x_1 \mapsto x_1\alpha_1 w_{12}\theta_2$. Symmetrically define $\tau_{21} : X_2 \to W(X_1)$.
6. Apply T_3 to get $\langle X_1 \mid R_1 \rangle \to^{T_3} \langle X_1 \cup X_2 \mid R_1 \cup \{x_2^{-1} \cdot (x_2\tau_{21}) \mid x_2 \in X_2\}\rangle = \mathscr{P}_3$. Note G is presented by the new presentation using isomorphism μ mapping each $x_1 \in X_1$ by $x_1 \mapsto x_1\alpha_1$, and each $x_2 \in X_2$ by $x_2 \mapsto x_2\alpha_2$.
7. The relators R_2 are derivable from this last presentation, since they correspond to relations in G under α_2. So we can add them to the last presentation using T_1.

   $$\mathscr{P}_3 = \langle X_1 \cup X_2 \mid R_1 \cup \{x_2^{-1} \cdot (x_2\tau_{21}) \mid x_2 \in X_2\}\rangle \to^{T_1}$$
   $$\langle X_1 \cup X_2 \mid R_1 \cup \{x_2^{-1} \cdot (x_2\tau_{21}) \mid x_2 \in X_2\} \cup R_2 \rangle = \mathscr{P}_4$$

   which is still a presentation of G under μ.
8. By construction, for $\mathscr{P}_4$, we have that the words $x_1^{-1} \cdot (x_1\tau_{12})$ map under μ to the identity in G. Hence, these words must be derivable from the relations of $\mathscr{P}_4$. Hence we can apply T_1 again:

   $$\mathscr{P}_4 = \langle X_1 \cup X_2 \mid R_1 \cup \{x_2^{-1} \cdot (x_2\tau_{21}) \mid x_2 \in X_2\} \cup R_2 \rangle \to^{T_1}$$
   $$\langle X_1 \cup X_2 \mid R_1 \cup R_2 \cup \{x_2^{-1} \cdot (x_2\tau_{21}) \mid x_2 \in X_2\} \cup \{x_1^{-1} \cdot (x_1\tau_{12}) \mid x_1 \in X_1\}\rangle = \mathscr{P}_5$$
9. The last presentation is symmetric; we could have started from $\mathscr{P}_2$ and achieved this presentation by a process of three Tietze transforms. Thus, there is a chain of six Tietze transforms, $T_3 - T_1 - T_1 - T_2 - T_2 - T_4$ bringing $\mathscr{P}_1$ to $\mathscr{P}_2$!

5.5 Markov properties and the undecidable.

In the 1950's there was a spat of theorems about what can be proven about groups in general. There are two immediate flavours for these theorems. Flavour one: There is (or there is not) an algorithm which, when handed a (finite) presentation of a group, can answer decisively 'yes' or 'no' in finite time, whether or not the presented group has some property $\mathscr{P}$. Flavour two: Given a presentation of a group, can one produce an algorithm which, when given a finite list of words (w_i) in the generators of the presentation, will decide in finite time whether or not the given list has some property $\mathscr{P}$.

Informally (and that is all we can do in this course, but see Ian Chiswell's accessible text [5] for formal definitions), given an object O and a property $\mathcal{P}$, we say that we can decide if O has property $\mathcal{P}$ if there is an algorithm which, when given the object O as input, will terminate in finite time with a definite answer, 'yes' or 'no', that O has property $\mathcal{P}$.

Examples of things which we might want to know are:

1. Given a finite presentation $G = \langle X \mid R\rangle$ of a group G can we decide when given any word $w \in W(X)$, whether or not w equivalent to the empty word in G (the "Word Problem")?

2. Given two finite presentations $G_1 = \langle X_1 \mid R_1\rangle$ and $G_2 = \langle X_2 \mid R_2\rangle$ are the groups G_1 and G_2 isomorphic (the "Isomorphism Problem")?

3. Given a finite presentation of a group $G = \langle X \mid R\rangle$ and two words w_1 and w_2, is there an element $g \in G$ so that $[w_1]_\sim = ([w_2]_\sim)^g$? That is, are the elements in G which are represented by w_1 and w_2 *conjugate elements in* G (the "Conjugacy Problem")?

The three problems mentioned above were all proposed by Max Dehn, circa 1912, as problems which would be fundamental in group theory, and he was right!

Definition 5. *A property $\mathcal{P}$ is a* Markov Property for groups *if*

1. *the property $\mathcal{P}$ is preserved by isomorphism,*

2. *there is a finitely presented group with the property $\mathcal{P}$, and*

3. *there is a finitely presented group K which cannot embed as a subgroup of any finitely presented group G which has property $\mathcal{P}$.*

Well known examples of Markov properties are: Being trivial, being finite, being abelian, and being free. There are many other properties which are Markov properties.

The theorem about Markov properties is the following.

Theorem 15. *(Markov, 1951) If $\mathcal{P}$ is a Markov property, then there does not exist an algorithm which can take as input any finite presentation of a group and then decide if the given group has $\mathcal{P}$.*

There are many properties of a group which are Markov properties: Being trivial, being finite, having solvable word problem, having solvable conjugacy problem, etc..

So, for example, there is no general algorithm which we can use which will determine in finite time whether or not any particular handed in finite presentation of a group represents the trivial group!

NB: Generally, these properties are in fact *semi-decidable*: E.g., there actually is an algorithm which will check and answer positively in finite time that any given trivial finitely presented group actually is trivial. Unfortunately, if you hand into that algorithm a non-trivial group, it may never be able to decide that the handed in group is non-trivial!

Exercise 25. *Explain why your solution to Exercise 24 seemed "ad-hoc."*

The first paper of Markov's in the chain that lead to Markov's Theorem is this one:

A. Markov, "On the impossibility of certain algorithms in the theory of associative systems." C.R. (Doklady) Acad Sci. URSS (N.S.), vol 55, pp 583–586, 1947.

Exercise 26. *Use Tietze transformations to show that the following presentations represent the same group.*

- $\langle k, r | k^3, r^2, (kr)^3 \rangle$
- $\langle a, b, c | a^2, b^3, c^3, abc \rangle$

6 L6: On Graphs, Categories, and Actions

6.1 On Graphs

First, recall our earlier discussion of graphs from Lecture 4. This section has been edited quite a bit by Stuart Burrell from earlier versions!

We revisit our discussion of paths in a graph, connectedness and so forth.

We single out special graphs, *lines*, which are directed graphs using as vertex set X a subset of $\mathbb{Z}$, where X is of the form $[a, b] \cap \mathbb{Z}$ for real $a \leq b$, or or the form $(-\infty, b] \cap \mathbb{Z}$ or of the form $[a, +\infty) \cap \mathbb{Z}$, or simply $X = \mathbb{Z}$. For two vertices $i-1$ and i in a line, there is precisely one edge (denoted e_i) in the line, and this edge has has $s(e_i) = i-1$, $t(e_i) = i$. The edges e_i so described are all of the edges of the line. Assuming a, b are integers, we then denote the resulting lines as $L_{[a,b]}$, $L_{(-\infty,b]}$, $L_{[a,\infty)}$ and $L_{\mathbb{Z}}$, respectively. We specify, for $n \in \mathbb{N}$ the line $L_n := L_{[0,n]}$ as our canonically favourite line of length n.

A path of length n in a graph Γ is the image of a graph homomorphism p from L_n into Γ. Recall that a graph homomorphism is a function which takes vertices to vertices, and edges to edges, and preserves all senses of adjacency (a directed edge must go to a directed edge, if it is a homomorphism of directed graphs, and respect the orientation). One should think about what this should mean in both cases: Γ directed versus Γ undirected. **We will often consider paths in a directed graph, but here, we loosen the definition: we will allow ourselves to consider the path as a map into the undirected version of the graph.** If an edge e_i of a line L_* crosses a edge d in the directed Γ in the right orientation on the path $p : L_* \to \Gamma$, we will say $p(e_i) = d$. But if it crosses the edge d with the wrong orientation, then we will say $p(e_i) = \overline{d}$. We call this "overline" map on edges "edge involution."

In any case, the start, beginning or initial vertex of a path p of length n in a graph Γ (so p is a graph homomorphism $p : L_n \to \Gamma$) is the image of the vertex 0, and the finish or ending or terminal vertex of the path is the image of the vertex n. We extend this language for infinite paths in the obvious ways.

Recall that we can always forget the orientation of edges in a directed graph to produce an undirected graph. The map *ends* will take an edge e to the set $ends(e) = \{s(e), t(e)\}$ which might only have one member.

A graph is connected if given any two vertices in it, there is a finite (undirected) path which begins at one vertex and ends at the other. A graph is loop-less if there is no vertex v which has an edge e where v represents both ends of e. A path p in a graph is efficient if there is no index i so that the ith and $(i+1)$'st edges of p are the same edge of Γ. A loop-less connected graph is a tree if given any two vertices v_1 and v_2, there is only one efficient path from v_1 to v_2. A graph is a forest if each of its components is a tree. A path p from v_1 to v_2 in a graph Γ is a geodesic path in Γ if all other paths from v_1 to v_2 in Γ are at least as long as p. A graph is locally finite if for all vertices of the graph, the vertex has only finitely many incident edges.

Finally, a graph homomorphism $\theta : \Gamma \to \Delta$ is locally injective if for every vertex v of Γ, the set of edges of Γ incident on Γ are taken in an injective fashion to the set of edges incident on $\theta(v)$. A morphism of graphs which is surjective is called onto, a morphism of graphs which is injective is called one-one, and a morphism of graphs which is both injective and surjective is called an isomorphism of a graph. An isomorphism from a graph to itself is called an *Automorphism* of the graph.

Exercise 27. *Given a graph Γ, show the set of automorphisms of Γ form a group under composition.*

We denote by $Aut(\Gamma)$ the group of automorphisms of the graph Γ.

We will mostly need two results for graphs:

Lemma 16. *Every connected undirected graph has a spanning tree.*

and the well known König's Lemma.

Lemma 17. *(König) Let Γ be an infinite locally finite tree. Then there is a infinite geodesic path in Γ.*

6.2 Basic Categories, and Automorphisms

In this subsection, we will generalise the construction of the group $Aut(\Gamma)$ above so that one can build similar groups for objects Γ which are not graphs.

Given a mathematical object S, an *automorphism of S* is an isomorphism from S to itself. What is strange about this is that we are applying the word isomorphism to objects of very different types. What is an isomorphism of vector spaces? What is an isomorphism of a Platonic Solid? What is an isomorphism of a graph (OK, we know the answer to that one!)?

The general framework for this discussion lives in an area called Category Theory. An appendix at the back of these notes (written by erstwhile MT 5824 student Casey Blacker) contains a nice introduction to Category Theory. It is not required reading for this course, but it might give you a warm feeling to know that material well!

Loosely speaking, a category consists of a class of objects of a certain type, and for any two such objects in the same category, a set of maps between them (called morphisms). Composition of maps is well defined and is associative (if $f : X \to Y$ and $g : Y \to Z$ are maps in a category $\mathscr{C}$, then there will be a well defined map $fg : X \to Z$ which is given as the composition of f and g). The main criteria for a map in a category is that (informally speaking) it has to respect the structure that defines what are the objects of a category. This is best explained with some examples.

In the category of Vector Spaces, the morphisms are set functions which respect the additive structure of the vector spaces, and respect scalar multiplication by the relevant field. That is, the morphisms respect the linear structure of the vector spaces; these are precisely the linear transformations. In the category of Groups, the morphisms are the set functions which respect the binary operation of multiplying elements; the morphisms are group homomorphisms. In the category of topological spaces, the morphisms are the "continuous maps." In the category of sets, there is no further structure to respect, so the morphisms are simply set functions. In all of these examples, our categories have objects which are sets with extra structure to respect. There are categories that do not fit that general description, but we will not be running into them in this course.

Once a category is understood, we use the word map to mean a morphism in that category. When the category's objects are sets with extra structures, we can still consider set functions, but they will generally not be morphisms in the category. We will use the word functions for that sort of "forgetful" map.

With all of this established, we can get back to groups of automorphisms.

Invertible maps in a category are called isomorphisms. An invertible map from an object to itself is called an automorphism. Now, in a given category $\mathscr{C}$, if O is a object of $\mathscr{C}$, then

$$Aut(O) := \{\tau : O \to O \mid \tau \text{ is an isomorphism in } \mathscr{C}\}.$$

It is very easy to check that the set of maps $Aut(O)$ so defined will be a group under composition of maps (in the category).

If $\mathscr{C}$ is a category whose objects are sets with underlying structure, we will say a morphism from X to Y is injective if the underlying set map is one-one (we might also say the map is *monic* or is a *monomorphism*). We will say the morphism is surjective if the underlying set map in onto (we might also say the map is *epic*, or is an *epimorphism*), and we will say the morphism is bijective if the underlying set map is both one-one and onto. These are not the proper category theory definitions of these terms, but they are effective definitions for our situation. Again, see the appendix for all of this "done right."

Exercise 28. *Suppose X and Y are sets, and $f : X \to Y$ and $g : Y \to X$ are set functions.*

1. *Suppose $fg : X \to X$ is injective, then show f is injective. Give an example which shows that g need not be injective.*

2. *Suppose $fg : X \to X$ is surjective, then show g is surjective. Give and example which shows that f need not be surjective.*

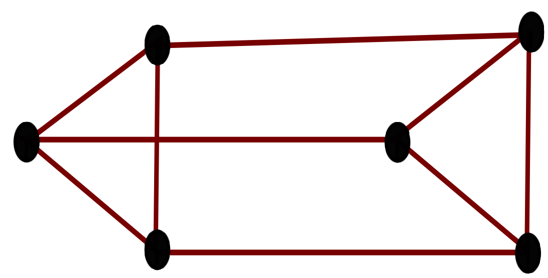
3. *if fg and gf are both bijective, then argue that f and g are bijective.*

The following is a very hard, but entirely basic, exercise. It is totally spoilt by reading the proof (which is available in many books and online). The exercise is more or less to make sure you remember the theorem, since I have seen graduate students work on the proof for months. The proof fits easily on one typed page and uses no deep mathematics (it can be done by third-year students if they are enthusiastic).

Exercise 29. *Prove the Cantor-Schröder-Bernstein Theorem.*

Theorem 18. *(Cantor-Schröder-Bernstein) Let X and Y be sets, and suppose there are functions $f : X \to Y$ and $g : Y \to X$ so that both f and g are monic. Show there is a bijection from X to Y.*

Exercise 30. *Give a finite presentation of the group $Aut(\Gamma)$ for the graph Γ below, with an explanation of what your generators do to the graph. (You do not have to prove that $Aut(\Gamma)$ is actually presented by your presentation. But, if you can, that is great!)*



6.3 Group Actions

6.3.1 Left, Right, and Sideways...

In what follows, we will often do many operations in a row. Usually, we use "right actions" as opposed to left actions. This means, we write things in left-to-right order for operations: For instance, if $f : X \to X$ and $g : X \to X$ are functions, then fg will mean "do f first, then do g." Of course, this is opposite to how you learn to compose functions in calculus, where left actions are used in order to support notation like "$f(x)$" which meant "do f to x."

Essentially, we will use right actions, but there are places where we need to "act from the other side," as we will see. Thus, we shall be explicit about each of the actions, whether it is a right action or a left action. Likely, we will write $f \circ g$ for left composition "do g first, and then do f." We might write parenthesies if we really wish to group our operations in a specific way, or to make a longer expression more clear. We do sometimes write $f(x)$ to mean apply f to x in a left action context (in the similar right action context, we would just write "xf" meaning "plug x into f"). We will see below that we need to pay careful attention to these issues, and that some notations mean different things in the different contexts. In general, there is not a good hard-and-fast rule we can always use, and things need to be made clear in context. The point is, be careful that you are always clear on the order of operations in the notation being used; both in your writing, and in your reading.

6.3.2 Definitions of Actions

Definition 6. *Let G be a group and let O be an object in a category $\mathscr{C}$. We say G* acts on *O if there is a group homomorphism*

$$\phi : G \to Aut(O).$$

In this case, for any $x \in O$ (imagine O is a set with extra structure, so x is an element of that set), and $g \in G$, we write $x \cdot g := x(g\phi) = xg\phi$ which is the image of x under the automorphim $g\phi$ of O. Observe that we are using 'right action notation,' so that we would say 'G acts on O from the right' (denoted $O \curvearrowleft G$). There is an equivalent notation for left actions, but as the course moves forward, we will prefer the right action notation, which allows one to read compositions in a left-to-right fashion (which is perhaps more natural).

Note that we will drop the '$\cdot$' in most expressions when the meaning is clear, (e.g., $x \cdot g$ might just be written as xg).

Some examples of groups and objects they act on are given below.

1. S_n acts on the set $\{1, 2, \ldots, n\}$,
2. D_4 acts on the set of corners of the square (so, as a subgroup of $S_{\{corner\, set\}}$).
3. $GL_n(\mathbb{R})$ acts on the vector space $\mathbb{R}^n$. (The group $GL_n(\mathbb{R})$ is the group of invertible $n \times n$ matrices of real numbers, where the group product is given by matrix multiplication. Obviously, elements of $GL_n(\mathbb{R})$ preserve the linear structure of the vector spaces $\mathbb{R}^n$, as we learn in linear algebra!)
4. The group $\mathcal{A}_1 := \{f\colon \mathbb{R} \to \mathbb{R} \mid xf = mx + b \text{ for } m \neq 0, b \in \mathbb{R}\}$ acts on the real line. What do you think the subscript '1' denotes? Also, what structure does this group preserve in the reals (under its action)?

Exercise 31. *1. Prove $\mathcal{A}_1$ is a group under composition.*

2. We claim that $\mathcal{A}$ acts on the topological space $\mathbb{R}$. What is the group $Aut(\mathbb{R})$? How can you verify that $\mathcal{A}_1$ acts on $\mathbb{R}$?

7 L7:Actions part Deux

Note that there is an alternative definition of a group action, when your objects are sets (if your category has objects which are sets with more structure, such as vector spaces, one temporarily 'forgets' the extra structure).

Definition 7 (Group G acting on a set B)**.** *A group G is said to act on a set B if there is a function $\cdot : B \times G \to B$ (we write '$b \cdot g$' for the image of the pair (b, g) under the function $\cdot$, instead of the formally correct notation $(b, g)\cdot$) so that the following rules are satisfied:*

1. *For all $b \in B$, we have $b \cdot 1_G = b$, and*
2. *for all g, $h \in G$ and $b \in B$, we have $(b \cdot g) \cdot h = b \cdot (gh)$.*

Exercise 32. *Show that in the category Set of sets (where the automorphisms of a set B are the bijections from B to itself (the permutations of B)) the two definitions of an action above are equivalent.*

Note that if you use the second definition of an action, then when your objects are in a category with more structure, one then needs to 'remember' the extra structure to show that your action (as a group of permutations) resolves to a subgroup of $Aut(\hat{B})$ (here, $\hat{B}$ is the original object which has more structure than the set B (recall that an action has to preserve the structure of your object!)).

A failure of the above sort is to consider the group of order two. We can act on the integers as a set by having each integer move to itself ± 1, depending on whether the integer is even or odd, for the non-identity element of C_2, and use the identity permutation for the identity element of C_2. This action of C_2 on the set $\mathbb{Z}$ IS an action in the category Set of sets, but it does not respect the group structure of $\mathbb{Z}$, so if we are thinking of $\mathbb{Z}$ as a group, then the permutation of $\mathbb{Z}$ given above is not an element of $Aut(\mathbb{Z})$, and so the described maps are not the image of a homomorphism from C_2 to the group $Aut(\mathbb{Z})$ in the category Group. (What is $Aut(\mathbb{Z})$ in the category Group of groups?)

7.0.1 Cayley's Theorem

Notation 8. *Given a set X, we use the symbol S_X to denote the full group of permutations on X.*

Theorem 19 (Cayley)**.** *Let H be a subgroup of a group G and let M be the set of all right cosets of H in G. Define the mapping $\phi : G \to S_M$ by the rule: for any $g \in G$ the permutation $g\phi$ sends a coset Hx to the coset Hxg.*

Then ϕ is a homomorphism with kernel

$$ker(\phi) = \cap_{x \in G} \left(x^{-1}Hx\right).$$

(Note: the group $\cap_{x \in G} \left(x^{-1}Hx\right)$ is called the core *of H* in *G. One related construction is to take the intersection of all proper finite index subgroups of a group; this a subgroup which is preserved by all automorphisms of G (a homework exercise).)*

Proof. Let H, G, M and $\phi : G \to S_M$ as in the statement of Cayley's Theorem. We first will show that ϕ is a group homomorphism, and then we will verify the claim about the kernel of ϕ.

Suppose r, s, and $t \in G$ so that $Hr \in M$. We wish to show that $(Hr)((st)\phi) = ((Hr)(s\phi))(t\phi)$. We compute

$$(Hr)((st)\phi) = (Hr)(st) = (H(rs))t = ((Hr)s)t = ((Hr)(s\phi))(t\phi)$$

by the associativity of the group product in G. Therefore, ϕ is a group homomorphism.

In the rest of this argument, we will drop the parentheses which are used to indicate the order of products, recognising that groups ARE associative, after all.

We now wish to show $ker(\phi) = \cap_{x \in G} \left(x^{-1}Hx\right)$.

Recall the general method that in order to show two sets are equal, one can show that each set contains the other.

$\underline{ker(\phi) \subseteq \cap_{x\in G} \left(x^{-1}Hx\right)}$

Let $k \in ker(\phi)$. Then, for all $x \in G$, we have that $Hxk = Hx$. This is the same as claiming that $Hxkx^{-1} = H$, or that $xkx^{-1} \in H$ for all $x \in G$. Multiplying this last inclusion on the left and the right by x^{-1} and x, respectively, we see that $k \in x^{-1}Hx$ for all $x \in G$.

$\underline{\cap_{x\in G} \left(x^{-1}Hx\right) \subseteq ker(\phi)}$

Suppose that $p \in x^{-1}Hx$ for all $x \in G$, and suppose $y \in G$. Note in passing that there is $h \in H$ so that $p = y^{-1}hy$. Now,

$$Hyp = Hy(y^{-1}hy) = Hhy = Hy$$

Since y was arbitrary in G we see that p must not move any coset of H, thus $p \in ker(\phi)$. $\square$

If $H = \{1\}$, the homomorphism ϕ from Cayley's theorem is called the (*right*) *regular representation* of the group G.

Corollary 20. *Let G be a group.*

1. *There is an embedding of G into the group S_G. The image of any non-trivial element of G under this embedding is a permutation, which sends each element of G to a different element of G.*
2. *If G is finite with $|G| = m$ then it can be embedded into the group S_m.*
3. *If G is finite with $|G| = m$ then it can be embedded into the group $GL_m(R)$, where R is any ring (with identity element 1).*

Proof. The first two points are obvious. To see the third point, realise any finite permutation group (say, on m points), as a group of corresponding permutations of the canonical basis vectors $\{\vec{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)\}$. The matrices for these permutations have the property that each row and column consists of all zeroes excepting a single entry which is a 1, and the determinant of such a matrix will be ± 1.

$\square$

Finally, we have a corollary originally due to Poincaré.

Corollary 21 (Poincaré)**.** *Every subgroup H of finite index m in a group G contains a subgroup N which is normal in G and has finite index k in G such that $m|k$ and $k|(m!)$.*

Proof. The action of G on the cosets of H factors through the quotient group $Q = G/ker(\phi)$ which is isomorphic to a subgroup $P \leq S_m$ (there are m cosets of H in G). Now, the cosets of $ker(\phi)$ are the elements of Q, so in particular $k = |Q| = |P|$ and so $k|m!$. Likewise, the image of H in $G/ker(\phi)$ is a subgroup of Q with index m (since $ker(\phi) \triangleleft H$), so $m|k$. $\square$

7.0.2 The Conjugacy Action and the Class Equation

Let G be a group, and let it act on itself from the right via conjugation. I.e., if $g, h \in G$, we set $g \cdot h = g^h = h^{-1}gh$. (Note: $g^{1_G} = g$ and $g^{hk} = k^{-1}h^{-1}ghk = (g^h)^k$ so this is indeed an action of G on the set G.)

Exercise 33. *We have just seen that G acts on the underlying set of G as a group acting on a set. Verify that the conjugation action of G on itself is still an action if we consider the object acted on (G) to be a group, and not just a set. What do you need to verify?*

The following definitions are standard for any action, not just via conjugacy in a group.

Suppose $X \curvearrowleft G$. The *orbit of $x \in X$ under the action of G* is written $\mathscr{O}_G(x)$ or as xG and is defined as follows.

$$\mathscr{O}_G(x) := \{y \in X \mid \exists g \in G, x \cdot g = y\} = \{xg \mid g \in G\}$$

The orbits of an action on a set X partition X.

We also define the subgroup of G which fixes a set $Y \subseteq X$ set-wise as

$$\mathrm{Stab}_G(Y) = \{g \in G \mid Y \cdot g = Y\}$$

This group is often called the *stabiliser of Y under the action of G*. Note that for $Y = \{y\}$ a single point we also call this the *point stabiliser of y* and denote it as $\mathrm{Stab}_G(y)$ or equivalently as G_y. (This highlights also that the set-wise stabiliser of Y in X under the action of G on X can also be thought of as the point stabiliser of the point Y in the power set of X under the induced action of G on the subsets of X.)

Returning now to the action of G on itself via conjugation we have the following definitions.

Suppose $S \subseteq G$. We define

$$C_G(S) := \{g \in G \mid \forall s \in S, s \cdot g = s\}$$

the *centraliser of S in G*. It is easy to note that $C_G(S) \leq G$.

Similarly, we define

$$N_G(S) = \{g \in G \mid S \cdot g = S\}$$

the *normaliser of S in G*. Note that if S is a subgroup then $N_G(S)$ is the largest subgroup of G in which S is normal.

Exercise 34. *Let G be a group and suppose $R < G$. What is the difference between $\langle\langle R\rangle\rangle$ and $N_G(R)$?*

7.0.3 Normalizers and centralisers; properties

Note that $C_G(S)$ and $N_G(S)$ can easily be different, although $C_G(S) \triangleleft N_G(S)$.

Exercise 35. *Prove the following.*

1. $Z(G) = C_G(G)$,
2. *It is possible that* $S \not\subseteq C_G S$, *however,* $S \subseteq C_G(C_G(S))$,
3. *if* $S, T \subseteq G$, *then* $S \subseteq C_G(T) \iff T \subseteq C_G(S)$.

The remaining material of this lecture 7 is here for reference for later in the course. We do not need it now, but we will introduce these terms later in the course when they arise.

7.0.4 Cayley's Orbit Stabilizer Theorem

Theorem 22 (Orbit Stabilizer)**.** *Suppose that a group G acts on a set X, then*

$$|xG| = |G : G_x|.$$

If G is finite, then

$$|xG| = |G|/|G_x|.$$

In particular, the size of the orbit divides the order of the group.

Proof. Define the set function $\theta : xG \to G/G_x$, where the domain and range are sets (G_x is generally not normal in G), by the rule $y = x \cdot g \mapsto G_x g$. If g_1 and g_2 have $y = x \cdot g_1 = x \cdot g_2$ then by the rules of group actions we see that $x \cdot (g_1 g_2^{-1}) = x$, hence $G_x g_1 = G_x g_2$, and we see θ is well defined. If $G_x g_1 = G_x g_2$ then $G_x(g_1 g_2^{-1}) = G_x$ so $x \cdot g_1 = x \cdot g_2$, hence θ is monic (one-to-one). Finally, if $G_x g$ is a coset of G_x, then the point $x \cdot g \in xG$ has $\theta(x \cdot g) = G_x g$, so θ is epic (onto).

Now, as $[G : G_x]$ represents the index of G_x in G, and as this is the cardinality of the set of right cosets of G_x in G, we see that $|xG| = [G : G_x]$. The remaining statements of the theorem are direct consequences of this fact in those cases when G is finite. □

7.0.5 The Class Equation

Theorem 23 (The Class Equation)**.** *Suppose G is a finite group, and let Γ be a set with precisely one representative element from each conjugacy class in G which is not a conjugacy class of an element in the centre of G. Then*

$$|G| = |Z(G)| + \sum_{x\in\Gamma}[G : C_G(x)].$$

Proof. The orbits under conjugation partition G. Elements of the centre of G have orbits of size 1, while the other elements of G have larger orbits, with size corresponding to the index of the stabiliser of the element. The equation then follows, as we are simply summing the total sizes of the orbits under conjugation. □

Exercise 36. *Prove again (from MT 4003), using the Class Equation, that if p is a prime, $k > 2$ is an integer, and if G is a group of order p^k, then G is not a simple group.*

7.0.6 Language of Actions

In this subsection, we offer a few adjectives, commonly used to describe aspects of a group's action on a set or a topological space. Some of these terms will have definitions which will still not make sense to the initiate. **Do not let that worry you for now**; this is merely a list of terms that you will now at least recognise as adjectives you have seen. If you run into these terms again, you will have a reference point in these notes. We will start with some general adjectives that are commonly used when discussing a group action.

Definition 9. *Let G be a group and X a set so that $X \curvearrowleft G$. We say that the action is*

1. faithful *if the homomorphism from G to S_X has trivial kernel (i.e., every non-trivial group element actually moves SOME point in X),*
2. free *if every point of X is moved by every non-trivial element of G,*
3. transitive *if given any two points x, $y \in X$, there is $g \in G$ so that $x \cdot g = y$,*
4. k-transitive *if for some positive integer k, given any two k-tuples $(x_1, x_2, \ldots, x_k)$ and $(y_1, y_2, \ldots, y_k) \in X^k$ (where each tuple consists of pairwise distinct elements) there is some element $g \in G$ so that for all indices i we have $x_i \cdot g = y_i$,*
5. regular *if the action of G on X is both transitive and free, and*
6. primitive *if the action of G on X is transitive and it fails to preserve any non-trivial partition of X (given any nontrivial partition $\mathscr{P}$ of X there is some element $g \in G$ and a set $U \in \mathscr{P}$ so that Ug is not an element of $\mathscr{P}$).*

◇◇

The following bit of this subsection is really beyond the scope of this course, it is only here for those that are interested.

If we have, e.g., topological structures about, then some other adjectives might come into play.

Definition 10. *Let G be a group and Y a topological space so that $Y \curvearrowleft G$. We say that the action is*

1. properly discontinuous *if G itself is a discrete group and for any point $y \in Y$ there is an open neighbourhood U of y, such that the set of all $g \in G$ for which $Ug \cap U \neq \varnothing$ consists of the identity only, and*
2. cocompact *if Y is locally compact and there is a compact set $A \in Y$ so that $AG = Y$. (If G acts properly discontinuously, the we can just say that Y/G is compact in the quotient topology.)*

To discuss these last definitions in any meaningful way, we need to learn briefly about topological groups. A group G is a topological group if it is also given a topology, which turns it into a topological space, where the maps (multiply by any given element) and (inversion) are both continuous maps from the group to the group. Topological groups 'arise in nature' from physics, for example. Indeed, the group of all invertible 3×3 matrices with real entries is just the sort of group of which we speak. Give these matrices the subspace topology from $\mathbb{R}^9$, and look at matrix multiplication (for any specific matrix) and inversion: these are continuous maps! The group of rotations of the circle S^1 is another (abelian, this time) topological group.

The main problem with the groups above are that they are really, really big (uncountable) and one can smoothly transition amongst elements. So, we like to pass to countable subgroups that still contain a lot of structure (this is basically passing to a manageable approximation of the big group). Examples of this are the Thompson groups, as you might be able to see.

Still nicer is to pass to subgroups where you always have a certain distance from one element to any other (if your topology allows a nice metric). For instance, one can consider the integers as an additive subgroup of the topological group the real numbers under addition. As a subspace of $\mathbb{R}$, the integers $\mathbb{Z}$ pick up the 'discrete topology' (any subset of the integers is going to be an open set in the integers). So, we perceive the integers as a 'discrete topological group.' As a subgroup of $\mathbb{R}$ which inherits the discrete topology from the subspace topology, they are called a 'lattice' (this word gets generalised quite a bit, but for our purposes, you can carry on quite nicely with this intuitive description). To make the point more clear, the rationals $\mathbb{Q} \leq \mathbb{R}$ is a perfectly happy topological subgroup of $\mathbb{R}$, but it is not a lattice.

Well, any group can be given the discrete topology (just define all subsets to be open sets!), so the condition that a group be 'discrete' in the definition of 'properly discontinuous' is not such a big deal. Thus, we see that the main thrust of the definition of 'properly discontinuous' is that small open sets in our target space move 'far away' when we act by our discrete group (e.g., any ball of radius less than one in $\mathbb{R}$ is moved entirely off of itself by translation by any integer which is not 0; so, e.g., we are describing of how the lattice $\mathbb{Z}$ in $\mathbb{R}$ acts on $\mathbb{R}$–that is, properly discontinuously!).

In this latter topological definition we refer to the space Y/G. Note that this space is called the *quotient space of* Y *under the action of* G, and points in this space correspond to the orbit of a point in Y under the action of G (recall that two points y_1, $y_2 \in Y$ are identified as being in the same orbit if there is a $g \in G$ so that $y_1 g = y_2$; the orbits in Y under the action of G form a partition of Y), and a set U is open in Y/G if and only if the pre-image of U in Y (this is the union of the equivalence classes which make the points of U) is an open set in Y.

Note that as our course is not a course in topology, we will only expect the reader to have a general intuitive idea about these latter two definitions. The point is that you needn't run out of the room screaming if someone says 'Lattice!' or 'quotient space'. In fact, if you think about how the lattice $\mathbb{Z}$ acts on $\mathbb{R}$, you might come up with the quotient space $\mathbb{R}/\mathbb{Z} \cong S^1$.

8 L8: Formal Language Theory I (Regular Languages/Rational Sets/Kleene's Theorem)

Assume that Σ is a fixed finite set. Recall that a *language L over Σ* is any subset of Σ^*. That is, if $L \subseteq \Sigma^*$, then L is a language over Σ.

We will be discussing formal language theory here, and we start with some of the most basic language types (according to theoretical Computer Science).

8.1 Recognisable and rational sets

Let M and N be monoids, with N finite. We say that a subset L of M is *recognisable* if there is a monoid homomorphism $\tau : M \to N$ so that $L\tau\tau^{-1} = L$. Note that τ need not be injective, so the inverse operation we are applying could easily expand the result to simple contain L properly.

Exercise 37. *1. Show that the class of recognisable subsets of monoids is closed under inverse homomorphisms. (Hint: Suppose that L is a recognisable subset of a monoid M, and $\theta : K \to M$ is a monoid homomorphism, then show $L\theta^{-1}$ is recognisable in K.)*

2. Show that the recognisable sets in a monoid M form a Boolean Algebra (they are closed under finite unions and complementation).

For a group, G we would say a recognisable subset of G is any set L which has $L\tau\tau^{-1} = L$ for some group homomorphism $\tau : G \to Q$ for Q a finite group. This is not different than the definition using a monoid homomorphism to a monoid M; if $\tau : G \to M$ is a monoid homomorphism and G is actually a group inside the (finite) monoid M under the binary operation, then the image of τ is actually a subgroup in M under M's binary operation. The restriction of τ is then a group homomorphism onto a (finite) group.

Exercise 38. *1. Show that a recognisable subset of a group G is precisely any set which is a union of a set of cosets of any finite index subgroup H in G.*

2. Show that the set $\{1_G\}$ is a recognisable subset of a group G if and only if G is finite.

Thus, recognisable subsets of a group G are not too interesting; they are easily described! We will now broaden out to a different technical class of subsets.

Definition 11. *Let M be a monoid. The set $RAT(M)$ of* rational subsets *of M is defined inductively to be:*

1. *Finite subsets of M are rational.*
2. *If K, $L \in RAT(M)$ then $K \cup L$ is rational.*
3. *If K, $L \in RAT(M)$ then $KL := \{kl \mid k \in K, l \in L\} \in RAT(M)$ (we might write $K \cdot L$ to emphasise the product).*
4. *If $L \in RAT(M)$ then the generated sub-monoid $L^* \in RAT(M)$.*

Why are we working in monoids? Well, recall that if Σ is a finite alphabet, then Σ^* IS the free monoid on the alphabet Σ. Thus, we really are talking about when any particular formal language over a finite alphabet is recognisable or rational, etc..

Example 1. *Consider the alphabet $\Sigma := \{0, 1, 2\}$. Let w_1, w_2, and w_3 be words in the alphabet Σ. Observe that the expression:*

$$\{0\} \cup \{0112\} \cup w_1^*\{w_2\}\{0\} \cup \{w_2\}\{w_1\}\{w_3, 01\}^*\{w_1\}$$

determines a regional language (why?) over the alphabet Σ. Expressions of this sort are called rational expressions*, that is, expressions using the four defining features of the inductive definition of rational subsets* *of a monoid (in this case, Σ^*). Here, we have been careful to emphasise the products of rational sets with our "$\cdot$" symbols. Common writing of the above might (slightly incorrectly) be*

$$\{0\} \cup \{0112\} \cup w_1^* \cdot w_2 \cdot 0 \cup w_2 \cdot w_1 \cdot \{w_3, 01\}^* \cdot w_1$$

for the same language.

8.2 Automata

Definition 12. *We call a tuple $\mathcal{A} = (Q, \Sigma, \delta, I, F)$ a non-deterministic finite automaton whenever*

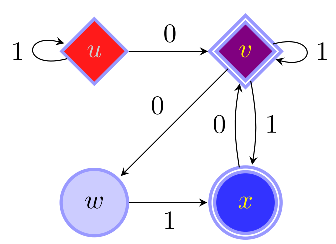
1. *Q is a finite set (called the* set of states of $\mathcal{A}$*),*
2. *Σ is a finite set (called the* alphabet of $\mathcal{A}$*),*
3. *$\delta \subseteq (Q \times \Sigma) \times Q$ (called the* transition relation of $\mathcal{A}$*),*
4. *$I \subseteq Q$ (called the* initial states of $\mathcal{A}$*),*
5. *$F \subseteq Q$ (called the* final (or 'accept') states of $\mathcal{A}$*).*

One can represent such an automaton as a directed, labelled graph with vertices the states Q, and where for each $((q_1, a), q_2) \in I$ we draw a directed edge from q_1 to q_2 with label a.

The idea is that such an automaton will determine a language $L \subseteq \Sigma^*$. To see this, determine L to be all words $w \in \Sigma^*$ which are *accepted by the automaton.* **A word is accepted if there is an initial state q_0 and a final state q_t and a directed path p in the automaton from q_0 to q_t so that the concatenation of the edge labels of the edges appearing in the path p produces precisely the word w.**

Example 2. *Set $Q = \{u, v, w, x\}$, $\Sigma = \{0, 1\}$, $I = \{u, v\}$, $F = \{v, x\}$, and set the relation δ as*

$$\delta := \left\{ \begin{matrix} ((u,0),v), ((u,1),u), \\ ((v,0),w), ((v,1),x), \\ ((x,0),v), ((w,1),x), \\ ((v,1),v) \end{matrix} \right\}.$$



The automaton $\mathcal{A}$.

In the picture of the automaton above, we use red diamonds for the initial state u (and it has a light grey lettering), light blue for the ordinary state w (neither initial nor accepting, and it's label is black), a dark blue circle with a double border for the accepting state x (yellow text), and finally, the state v is both initial and accepting thus it has a purple diamond shape and coloration with a double border and yellow text. We will use these shape and colour indications for all of our automata in these notes.

Exercise 39. *Characterise the language L accepted by the automaton $\mathcal{A}$ above.*

The automaton above is non-deterministic since the state v has two outgoing edges with label 1; thus, if one is in state v and reads a '1', then one would travel either to x or v, but the choice is not determined. More formally, an automaton is deterministic if the transition relation δ has that for each $q \in Q$ and $l \in \Sigma$, there is at most one transition of the form $((q, l), *)$. We can now state a wonderful theorem of Kleene.

Theorem 24. *(Kleene) Let Σ^* be a finitely generated free monoid and let $L \subseteq \Sigma^*$ be some language. Then the following statements are equivalent.*

1. *L is recognisable.*
2. *L is accepted by some deterministic finite state automaton.*
3. *L is accepted by some non-deterministic finite state automaton.*
4. *L is rational.*

A language L satisfying one of these conditions is also called *regular*. Regular languages represent the most basic type of formal languages that have received serious research attention, but they are quite restrictive, as will become clear.

Proof. First let us mention that unfortunately the proof as given in [6] is at best incomplete, and that might be confusing to less experienced readers such as the students for whom this set of notes is being written! However, if you refer to that text there is a construction where matrices are used in the (3)$\Longrightarrow$(4) proof. This usage of matrices is quite a nice and powerful method, which we do not employ below, but it might be of interest to our readers, so we can recommend looking at that text in order to see a slightly different approach. But as implied above, Diekert and Weiß's use of the sets $L^k_{i,j}$ (whose union is supposed to be the language L) might in fact produce a larger language, so in the proof below we add some admittedly small steps in order to focus down on the correct allowable sets.

(1)$\Longrightarrow$ (2):
Let Σ be a finite set and suppose $L \subseteq \Sigma^*$ is recognisable. Then there is $\tau : \Sigma^* \to N$ a monoid homomorphism to some finite monoid N so that $L\tau\tau^{-1} = L$ (so L is a recognisable language).

We will build a deterministic finite state automaton (DFSA) which also recognises L.

Set $\mathcal{A} := (N, \Sigma, \delta, \{1\}, L\tau)$ where $\delta = \{((n, a), n \cdot (a\tau)) \mid n \in N, a \in \Sigma\}$. We claim this automata accepts exactly the language L, and that this automaton is deterministic.

The automaton $\mathcal{A}$ is deterministic by definition, since given a state n and a letter $a \in \Sigma$, we precisely add in only the element $((n, a), n \cdot (a\tau))$ to the defining relation δ.

We now wish to show that $\mathcal{A}$ accepts precisely the language L.

Suppose $w_1 = a_1a_2 \ldots a_k \in L$. Then, starting at the state $1 \in N$ we can begin tracing a path through $\mathcal{A}$ according to the letters a_1 , a_2, and etc.. For each letter a_i we read, there is an edge to travel, as whatever state n we are on has afforded us to put in the edge $((n, a_i), n \cdot (a_i\tau))$. Thus upon reading the letter a_k we will come to some state p in the automaton $\mathcal{A}$. The state p has $p =_N (a_1\tau)(a_2\tau)\cdots(a_k\tau) = (a_1a_2 \ldots a_k)\tau$ since τ is a monoid homomorphism. Hence, $p = w_1\tau$ and so p is an accepting state. Thus, the language L is a subset of the language accepted by $\mathcal{A}$.

Suppose $\mathcal{A}$ accepts a word $w_2 = b_1b_2 \ldots b_j \in \Sigma^*$. Then, it must be the case that the path $(b_1\tau)(b_2\tau)\cdots(b_j\tau)$ (using the labels of the edges to describe which edges we are traversing) starts at 1 and ends in an accept state p of $\mathcal{A}$. Therefore, there is a word $w_3 \in L$ so that $p = w_3\tau$. Noting that $w_2\tau = (b_1\tau)(b_2\tau)\cdots(b_j\tau) = p$, we see that $w_2 \in p\tau^{-1}$, and thus $w_2 \in L\tau\tau^1 = L$. In particular, $w_2 \in L$.

(2)$\Longrightarrow$ (3):
We obtain this implication for free, since deterministic FSA are simply special types of non-deterministic FSA.

(3)$\Longrightarrow$ (4):
Let us assume now that L is accepted by a new automaton $\mathcal{A} = (Q, \Sigma, \delta, I, F)$ which might not be deterministic. We want to show that L is rational. We do this by characterising the language accepted by $\mathcal{A}$ in accordance with the inductive definition of rational languages.

We will express L as a finite union of rational languages. We start by assuming there are n states in $\mathcal{A}$. Order them, and refer to these states as 1, 2, …, n.

Define, for all indices $i, j, k \in \{1, 2, \ldots, n\}$, the language $J^k_{i,j}$ to be all words in Σ^* which when read starting from state i in $\mathcal{A}$ traverse a path through $\mathcal{A}$ to j which path never visits a vertex m with $m > k$

(excepting the start and finish of the path which are allowed to be at i and j respectively, even if $i > k$ or $j > k$).

Now we observe that $J^0_{i,j}$ is the set of letters a so that $((i,a),j) \in \delta$. If $i = j$ then $\varepsilon \in J^0_{i,j}$ as well.

We also observe that

$$J^k_{i,j} = J^{k-1}_{i,j} \cup J^{k-1}_{i,k} \cdot (\bigcup_{p \leq k} J^p_{k,k})^* \cdot J^{k-1}_{k,j}.$$

Now define

$$L^k_{i,j} := \begin{cases} J^k_{i,j} \text{ if } i \in I, j \in F \\ \varnothing \text{ otherwise} \end{cases}$$

As each $J^k_{i,j}$ is rational, each $L^k_{i,j}$ is rational.

Now we simply observe that

$$L = \bigcup_{i,j,k} L^k_{i,j},$$

and that this is a union of finitely many rational languages and hence is rational. □

Exercise 40. *Complete the proof above*

1. *by explaining why you know the expression*

$$J^{k-1}_{i,k} \cdot (\bigcup_{p \leq k} J^p_{k,k})^* \cdot J^{k-1}_{k,j}$$

   represents a rational language in the above proof, and

2. *by showing that (4)$\Longrightarrow$ (1).*

Exercise 41. *Determine all of the sets $L^k_{i,j}$ from the above proof of (3) $\Longrightarrow$ (4) for the language L accepted by the automaton $\mathcal{A}$ of Example 2.*

9 L9: Groups and Languages I (Anisimov's Theorem)

In this section, we will learn about a strange relationship between the language of words in a generating set of a group, and higher properties of the group. The properties we will be interested in are the properties which are preserved by group isomorphisms. The language properties we will be interested in will be independent of the choice of (finite) generating set for the group.

Definition 13. *Given a group G with set X of generators, we define the* word problem $WP(G, X)$ *of G to be the set*

$$WP(G, X) := \{w \in W(X) \mid w =_G 1_G\} .$$

That is, $WP(G, X)$ is the set of words in the generators X which are equivalent to the identity in G. Obviously, this is a language in the alphabet $(X \cup X^{-1})$ where here we mean by X^{-1} the maximal set of elements of G which are inverses of elements in X.

The amazing thing about the word problem of a group, as hinted above, is that its formal language properties can sometimes tell you powerful things about the group!

Theorem 25. *(Anisimov, 1971) Let G be a group generated by a finite set of generators X. G is finite if and only if $WP(G, X)$ is a regular language.*

Proof. Suppose G is a group generated by a finite set X of generators.

First we show that if G is finite, then it's word problem is regular.

Let $Y = X \cup X^{-1}$, and consider the right Cayley graph $\Gamma(G, Y)$ as a deterministic automaton, with initial state set simply $\{1_G\}$ and final state set $\{1_G\}$. Any element in $WP(G, X)$ will trace a path in this automaton which starts on 1_G and ends on 1_G, thus $WP(G, X)$ is a subset of the regular language accepted by this automaton. However, any word in Y which traces a path in the automaton from 1_G to 1_G is a word whose product IS 1_G, so the language accepted by the automaton is a subset of $WP(G, X)$, and thus these two languages are equal and we can conclude $WP(G, X)$ is a regular language in this case.

Now we show that if $WP(G, X)$ is regular, then G is a finite group. First, we know by Kleene's Theorem that $WP(G, X)$ is recognisable. Hence, there is a monoid homomorphism $\theta : (X \cup X^{-1})^* \to M$ for some monoid M and a finite subset $S \subseteq M$ so that $WP(G, X) = \theta^{-1}(S)$, and it is immediate for instance that $1_M \in S$, since $WP(G, X)$ contains expressions like xx^{-1} for each $x \in X$. However, we also know that $WP(G, X)$ is the kernel (preimage of the identity) of the monoid homomorphism from $q : (X \cup X^{-1}) \to G$. However, the kernel of the map θ is a subset of $WP(G, X)$, so in particular there is a monoid homomorphism $\psi : \text{Image}(\theta) \to G$ (from the image $\text{Image}(\theta)$ of θ to G) so that q factors as $\theta\psi$, and therefore G is finite since it is the image of a finite monoid ($\text{Image}(\theta)$ is finite since M is finite) under a monoid homomorphism (since q is surjective we see ψ must also be surjective).

We offer a second proof, which does not rely on the the monoid version of the Homomorphism Theorem. This proof is machine theoretic (the author learned this paper from the paper [9]). Suppose $WP(G, X)$ is regular but that G is not finite. Since G is not finite there must be words in $(X \cup X^{-1})$ of arbitrarily long length which do not contain sub-words which product to the identity in G (if not, then there is length n so that any word of length n has a sub-word which is equal to the identity, so all elements of the group can be written as a product of length less that n, which means G is finite). Let (w_i) be a sequence of words, so that w_i has length at least i and has no sub-word with product equal to the identity. Suppose $\mathcal{A}$ is an automaton on alphabet $(X \cup X^{-1})$ which accepts precisely the word problem of G. By the "powerset construction" (see [8]), we can assume that $\mathcal{A}$ has precisely one start state and is deterministic (see homework exercise below). Now consider w_m for some m larger than the number n of states of $\mathcal{A}$. There are distinct prefixes u and uv of w_m so that on reading either u or uv from the start state, we will be in the same state of $\mathcal{A}$. Now, we know then that continuing with the word u^{-1} must get us to an accept state for $\mathcal{A}$, since $uu^{-1} \in WP(G, X)$. But then, reading the word uvu^{-1} must also get us to this same accept state, while the word $uvu^{-1} \notin WP(G, X)$ since it is a conjugate of v which is not trivial in G. Thus $\mathcal{A}$ must accept words which are not in $WP(G, X)$. $\square$

Exercise 42. *Suppose $\mathcal{A}$ is an accepting automaton with more than one initial state. Build a new (non-deterministic) automaton $\mathcal{B}$ which has only one initial state but which accepts exactly the same language as $\mathcal{A}$.*

Exercise 43. *Read a resource (wikipedia or, e.g., [8]) on the "powerset construction" which takes a non-deterministic accepting automaton with one initial state and produces a deterministic accepting automaton on one initial state which accepts precisely the same language. Write a reasonably detailed outline of how this process is carried out.*

The final paragraph of the proof of Anisimov's Theroem indicates a method for showing that a language is not regular.

Lemma 26. *(Pumping Lemma for Regular Languages.) Suppose Σ is a fixed finite alphabet and L is a regular language over Σ. Then there is $N \in \mathbb{N}$ so that for any word $w \in L$ with $|w| > N$ we have that there are words x, y, $z \in \Sigma^*$ with $|y| > 0$ so that $w = xyz$ and for all natural numbers n we have $xy^n z \in L$.*

The following is a standard exercise in using the pumping lemma above.

Exercise 44. *Prove that the language $\{a^n b^n \mid n \in \mathbb{N}\}$ is not a regular language over the alphabet $\{a, b\}$.*

Exercise 45. *Prove the Pumping Lemma for regular languages.*

Exercise 46. *Explain how you know that, given a group G, if there are two finite generating sets X and Y for G, then $WP(G, X)$ is rational if and only if $WP(G, Y)$ is rational.*

10 L10: Formal Language Theory II (Chomsky Hierarchy and Context Free Languages)

10.1 Chomsky Hierarchy

Noam Chomsky in 1957 proposed a hierarchy of formal languages. This hierarchy has four classes of languages. These are:

1. Type 0

   These are all formal languages that can be recognised by a Turing machine. These are the *recursively enumerable* languages.

2. Type 1

   These are more restricted than Type 0. These languages can be recognised by a "linear bounded non-deterministic Turing machine." and are called called *Context Sensitive Languages*.

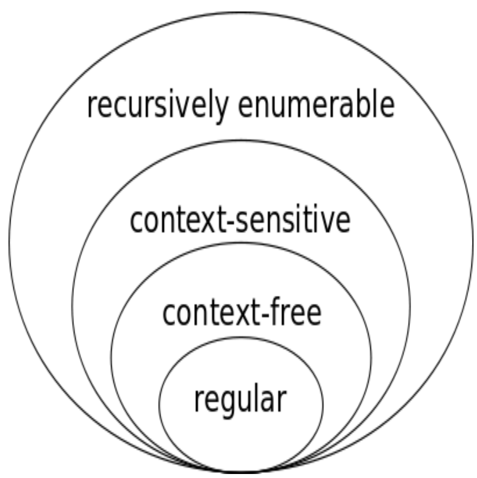
3. Type 2

   These are more restrictive still; these languages can be recognised by non-deterministic pushdown automata. They are called the *Context Free Languages*.

4. Type 3.

   The most restricted class of languages. These languages are recognised by finite automata, and are called the *Regular Languages*.

In this course, we have been looking at Type 3 languages, and we will look at Type 2 languages as well, particularly as the Muller-Schupp result concerns groups with context free word problem, The following diagram was stolen from Wikipedia's page on the Chomsky Hierarchy.



10.2 Context Free grammars and languages

A *context free grammar* is a four tuple

$$\mathcal{G} := (V, \Sigma, P, S)$$

where V is a finite set of *non-terminal symbols*, Σ is a finite set of *terminal symbols* (or letters), P is a finite relation $P \subseteq V \times (V \cup \Sigma)^*$ called *productions*, and $S \in V$ is the *start symbol*.

Touching on where we have been, one should think of productions simply as rewrite rules. Starting with the symbol S, we carry out productions until we arrive at a word containing only terminal letters. This word is then in the language generated by the grammar. Here is an example:

$$\mathcal{G} := (\{S\}, \{a, b\}, \{S \mapsto aSb, S \mapsto \varepsilon\}, S)$$

The language $L(\mathcal{G})$ generated by the grammar $\mathcal{G}$ is

$$L(\mathcal{G}) := \{w \in \Sigma^* \mid S \mapsto^* w\}$$

and in the case of our example, this language is $L(\mathcal{G}) = \{a^k b^k \mid k \in \mathbb{N}\}$.

The language is called "Context free" because the substitutions on the non-terminals do not have any care for the context of the non-terminal in any larger word. A context-sensitive grammar would, for instance, have a finite set of productions where the left hand sides of the rewrite rules would be words from $(V \cup \Sigma)^*$ so that specific productions for a non-terminal would care about the context: a production could only run if the word around the non-terminal fit the left hand side (of the production) recipe.

Exercise 47. *Build a context free grammar $\mathcal{G}$ with $\Sigma := \{(,), [,]\}$, whose language would be all correctly formatted double-parenthesised expressions. Eg., "([()()][])" would be in the language, but "([)]" would not be allowed (so no square-bracketed interval could contain "half" a round-bracketed block, and vice-versa).*

Exercise 48. *Can you think of general rules for a grammar, so that any grammar in your class would produce only regular languages and so that every regular language could be produced by one of your grammars?*

10.3 Pushdown automata

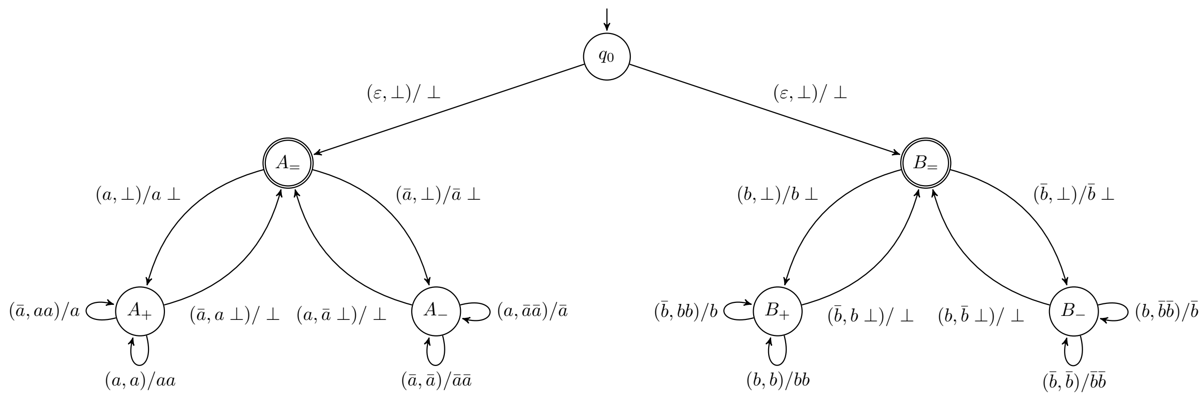
Just as regular languages are accepted by Finite State Automata, so context free languages are accepted by a particular type of automata. Context free languages are precisely the languages accepted by non-deterministic pushdown automata. Here is our definition:

A 7-tuple $P := (Q, \Sigma, \chi, \delta, q_0, \perp, F)$ is a non-deterministic pushdown automata if

1. Q is a finite set (of *states*),
2. Σ is a finite set (the *tape alphabet*),
3. χ is a finite set (the *stack alphabet*),
4. δ is a finite relation ($\delta \subseteq (Q \times (\Sigma \cup \{\varepsilon\}) \times \chi^*) \times (Q \times \chi^*)$ is the *transition relation*),
5. $q_0 \in Q$ (the *initial state*),
6. $\perp \in \chi$ (the *bottom of stack* symbol), and
7. $F \subseteq Q$ (the set of final, or *accept states*).

The idea is that a (non-deterministic) pushdown automaton is a more powerful type of automata (than e.g., a FSA), as it has a stack for memory. The PDA looks at its current state, the "plates" at the top of the stack, and its current input letter (on the input tape), and makes a transition to a new state while advancing the tape to the next letter, and possibly replacing a finite word on the top of the stack with another word. Once the PDA has read the whole input tape, if it is in an accept state, then the word is accepted. That is, the language accepted by the PDA is the set of all words so that there is a path through the automata, while reading the given word, which results in ending in an accept state.

The non-determinism comes in again by the fact that it is perfectly possible to have two transitions with the same conditions of transition (same state, same letter on tape, and same word on top of stack) but which correspond to two distinct transitions. The addition of ε to the tape alphabet in the transitions allows some transitions to be made without reference to the tape, which transitions DO NOT advance the tape (ε-moves).



Exercise 49. *Determine the language of the NPDA in the included figure (top of page). For this problem (but not in general) if no transition is drawn for an input letter for some state (other than q_0), assume a loop to the same state with no effect on the stack. (This avoids some clutter in the drawing).*

[Thanks to SH for this figure!]

Exercise 50. *Build a PDA which accepts exactly the language $\{a^k b^k \mid k \in \mathbb{N}\}$.*

Exercise 51. *Build a PDA which accepts exactly the language of Exercise 47.*

Exercise 52. *Draw a diagram of the following NPDA P. Then, describe precisely the language recognised by P.*

$$P := (\{\overline{A}, A, q_0, \overline{B}, B\}, \{a, \overline{a}, b, \overline{b}\}, \{a, \overline{a}, b, \overline{b}, \perp, \}, \delta, q_0, \perp, \{A, B\})$$

Where δ is given by the following set of 28 (!) transitions:

$$\left\{ \begin{array}{c}
((q_0,\varepsilon,\perp),(\overline{A},\perp)),((q_0,\varepsilon,\perp),(\overline{B},\perp)),\\[1em]
((A,a,a),(A,aa)),((A,\overline{a},\overline{a}),(A,\overline{aa})),((B,b,b),(B,bb)),((B,\overline{b},\overline{b}),(B,\overline{bb}))\\[1em]
((A,a,\overline{aa}),(A,\overline{a})),((A,a,\overline{a}\perp),(\overline{A},\perp)),((A,\overline{a},aa),(A,a)),((A,\overline{a},a\perp),(\overline{A},\perp))\\
((B,b,\overline{bb}),(B,\overline{b})),((B,b,\overline{b}\perp),(\overline{B},\perp)),((B,\overline{b},bb),(B,b)),((B,\overline{b},b\perp),(\overline{B},\perp))\\[1em]
((\overline{A},b,\perp),(\overline{A},\perp)),((A,b,a),(A,a)),((A,b,\overline{a}),(A,\overline{a})),\\
((\overline{A},\overline{b},\perp),(\overline{A},\perp)),((A,\overline{b},a),(A,a)),((A,\overline{b},\overline{a}),(A,\overline{a})),\\[1em]
((\overline{B},a,\perp),(\overline{B},\perp)),((B,a,b),(B,b)),((B,a,\overline{b}),(B,\overline{b})),\\
((\overline{B},\overline{a},\perp),(\overline{B},\perp)),((B,\overline{a},b),(B,b)),((B,\overline{a},\overline{b}),(B,\overline{b})),\\[1em]
((\overline{A},a,\perp),(A,a\perp)),((\overline{A},\overline{a},\perp),(A,\overline{a}\perp)),\\
((\overline{B},b,\perp),(B,b\perp)),((\overline{B},\overline{b},\perp),(B,\overline{b}\perp))
\end{array} \right\}$$

The point of the last exercise is that once you "understand" the automaton, then it is not too hard to understand the language it accepts.

11 L11: Formal Language Theory III (Context Free Languages and the Pumping Lemma)

This is a practical section about aspects of context free languages, to help the reader to have a better intuitive feel for the domain. We start with the standard Pumping Lemma, which is useful for showing a language is not context free.

Before stating and proving the Pumping Lemma for context free languages, we will build some notation and language.

Let $\mathscr{G} = (V, \Sigma, P, S)$ be a context free grammar. We call a sequence of productions

$$S = z_0 \xmapsto{\rho_1} z_1 \xmapsto{\rho_2} z_2 \xmapsto{\rho_3} z_3 \xmapsto{\rho_4} \cdots \xmapsto{\rho_n} z_n = z$$

which produces the word $z \in L(\mathscr{G})$ via the productions ρ_i a *derivation chain for* z. For any index $i \in \{1, 2, \ldots, n\}$ and non-terminal symbol M appearing in the word z_{i-1}, if the production ρ_i replaces this occurence of the symbol M by a new string s_i, so that $z_{i-1} = u_{i-1}Mv_{i-1}$ and $z_i = u_{i-1}s_iv_{i-1}$, then we call s_i the *shadow of* M *in* z_i. We now extend this language so that if w_{i-1} is a contiguous substring of z_{i-1} then either the non-terminal variable M in the production is appearing in this subword or not. If M is not appearing, then the *shadow of* w_{i-1} *in* z_i will just be the copy of w_{i-1} that arises in z_i at the same location as in z_i (the position will be the same in z_i from one of (either) the left or the right end of the string z_{i-1}). If the symbol M being replaced is actually appearing in the string w_{i-1}, then if we write $w_{i-1} = u_{i-1}Mv_{i-1}$ for some strings u_{i-1} and v_{i-1} then *the shadow of* w_{i-1} *in* z_i would be $u_{i-1}s_iv_{i-1}$ (again, we assume the production used is $M \xmapsto{\rho_i} s_i \in P$). Now we inductively extend this definition as follows: if $i < j$ are indices and now w_i is a contiguous substring in z_i then we can (and do!) define by induction the *shadow of* w_i *in* z_j.

Now, suppose again that $\mathscr{G}$ is a grammar and

$$S = z_0 \xmapsto{\rho_1} z_1 \xmapsto{\rho_2} z_2 \xmapsto{\rho_3} z_3 \xmapsto{\rho_4} \cdots \xmapsto{\rho_n} z_n = z$$

is a derivation chain for a word $z \in L(\mathscr{G})$. We say the derivation chain is *efficient* if whenever there are indices $i < j$ so that the production ρ_{i+1} applies to a non-terminal M in z_i and where the shadow of this M in z_j contains again the symbol M, then the shadow of the M from z_i in $z = z_n$ properly contains the shadow of the M from z_j in z (so, these shadows are not the same). The idea of this definition is that for derivation chains that are not efficient (so, inefficient derivation chains), one can replace such a derivation chain by a shorter derivation chain. Thus, every word in $L(\mathscr{G})$ admits an efficient derivation chain.

We are now ready to prove the following theorem, which by tradition is called a lemma!

Lemma 27. *(Pumping Lemma for CFL's) Suppose* $L \subseteq \Sigma*$ *is a context free language over finite alphabet* Σ. *Then there is* $p \in \mathbb{N}$ *so that for all* $z \in L$ *with* $|z| \geq p$ *there is a decomposition of* $z = uvwxy$ *as a product of the five substrings* u, v, w, x, y *where* $|vx| > 0$ *and* $|vwx| \leq p$ *so that for all* $n \in \mathbb{N}$ *we have*

$$uv^nwx^ny \in L.$$

Proof. Let $\mathscr{G} = \{V, \Sigma, P, S\}$ be a context free grammar, and let $L = L(\mathscr{G})$ be the language generated by $\mathscr{G}$.

In our first step, we define a set $X \subseteq L$. The set X contains exactly the words z in L which admit an efficient derivation chain such as

$$S = z_0 \xmapsto{\rho_1} z_1 \xmapsto{\rho_2} z_2 \xmapsto{\rho_3} z_3 \xmapsto{\rho_4} \cdots \xmapsto{\rho_n} z_n = z$$

so that whenever there are indices $i < j$ and a non-terminal M so that the production ρ_{i+1} acts on an M in z_i, then all of the non-terminal M symbols which appear in z_j are actually not in the shadow of the M that was acted upon by the production ρ_{i+1}.

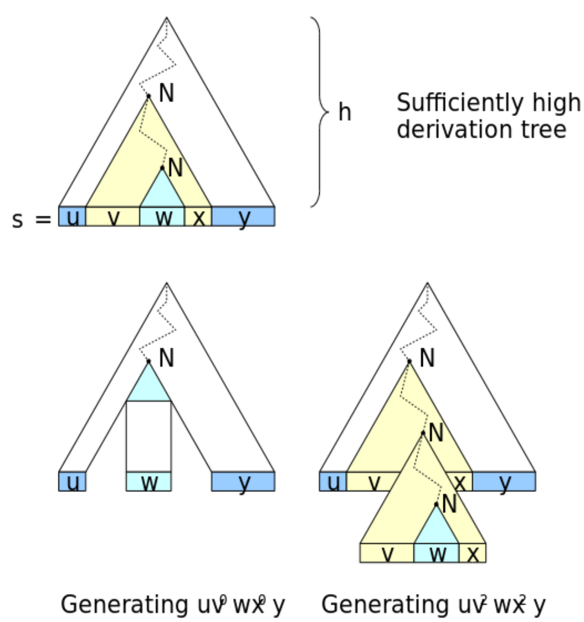
This may seem a very strange defining condition for the set X, but our reason for choosing it will become clear. In the meantime, notice that X is finite, since once you use a non-terminal in some production, then the non-terminal can never appear again in the shadow of itself, so all later productions acting in the shadow must be on different non-terminals, so in particular nested chains of productions can never be longer than the number of non-terminals that we have available in V, and thus there are only finitely many words we can produce in this way. Let r be the maximum length of a word in X, and if $L(\mathscr{G})$ has no words longer than r, set $p = r+1$, noting that this occurs if and only if $L(\mathscr{G})$ is actually a finite language (when it is clear that in this case and with this choice of p the Pumping Lemma holds true vacuously).

Thus, we will assume $L(\mathscr{G})$ is infinite. Let $z \in L$ with $|z| \geq p$, so that z is not in X, and assume

$$S = z_0 \xmapsto{\rho_1} z_1 \xmapsto{\rho_2} z_2 \xmapsto{\rho_3} z_3 \xmapsto{\rho_4} \cdots \xmapsto{\rho_n} z_n = z$$

is an efficient derivation chain for the word z. As z is not in X, there are some indices $i < j$ and a non-terminal M which occurs in the words z_i and z_j (the *progenitor* M and *descendant* M, respectively) so that the progenitor M is acted on by the production ρ_{i+1} while the descendant M is in the shadow word w_j of the progenitor M, which is a contiguous subword of z_j. Let us decompose the shadow word w_j as $v_j \cdot M \cdot x_j$, where the symbol M is the descendant M, and where at least one of v_j and x_j has a non-trivial shadow in z (because our derivation is efficient, and the progenitor M is acted upon by ρ_{i+1}, its shadow in z is strictly bigger than the smaller shadow w of the descendant M within z, and the difference is being produced as the shadows of v of v_j and x of x_j, in z).

But now we are done. Decompose z_i as $u_i \cdot M \cdot y_i$, and name the shadows of u_i and y_i in z by the names u and y respectively. Then, we have a full decomposition of z as $z = uvwxy$ where by replacing the derivations of the progenitor M by the derivations from the descendant M, we see that $uwy = uv^0wx^0y \in L$, and by replacing the derivation under (in the shadow of) the descendent M by the full derivation in the shadow of the progenitor M, we see that $uvvwxxy = uv^2wx^2y \in L$, and by repeating this many times we see by induction that for any natural number n we have $uv^nwx^ny \in L$. The picture below was taken from the wikipedia page on the pumping lemma for context free languages on Oct 11, 2014, and hopefully it clearly indicates the process.



∎

Example 3. *The following languages satisfy the Pumping Lemma for CFL's, and might be CFL's*

1. $L_1 = \{a^n b^n \mid n \in \mathbb{N}\}$

2. $L_2 = \{a^m b^n \mid m \leq n \in \mathbb{N}\}$

3. $L_3 = \{w \in \{a,b\}^* \mid |w|_a > |w|_b\}$

4. $L_4 = \{a^i b^j c^k \mid i \neq j \neq k \neq i\}$

The following languages do not satisfy the Pumping Lemma for CFL's, and cannot be CFL's

1. $L_5 = \{w \in \{a, a^{-1}, b, b^{-1}\} \mid Exp_a(w) = 0 \text{ } AND \text{ } Exp_b(w) = 0\}$ *where* $Exp_x(w)$ *is the sum of the exponents of the letter* x *in* w.

2. $L_6 = \{a^n b^n c^n \mid n \in \mathbb{N}\}$

3. $L_7 = \{w \in \{a,b,c\}^* \mid Exp_a(w) = Exp_b(w) = Exp_c(w)\}$

Interestingly enough, It is easy to find languages that satisfy the Pumping Lemma for CFL's, which languages are not context free; the pumping lemma for context free languages is very much valid only in one direction. Indeed, the language L_4 above is of this type as it is not context free.

Exercise 53. *Verify that languages* $L_1 - L_3$ *satisfy the pumping lemma for CFL's.*

Exercise 54. *Prove that Languages* L_6 *and* L_7 *above are not context free languages by using the Pumping Lemma for CFL's.*

Remark 3. *For those wanting a better Pumping Lemma, look up Ogden's Lemma on it's wikipedia page. One can easily use Ogden's Lemma to prove that* L_5 *is NOT a context free language. Still, even Ogden's Lemma fails to fully characterise the context free languages.*

Remark 4. *The class of context free languages are closed under inverse monoid homomorphisms; just like for regular languages.*

A consequence of the last remark is the following lemma, although this is also easy to prove using PDA's.

Lemma 28. *The union of finitely many context free languages is a context free language.*

The following exercise provides some context free languages for us to use to prove some interesting properties of context free languages.

Exercise 55. *In the following, you are asked either to produce a context free grammar or a NPDA which will accept or generate the following context free languages. Extra happiness for properly arguing that your constructions actually generate the languages you claim they do.*

1. $F := \{0, 1, 2\}^*$,

2. $R := \{0^i 1^j 2^k \mid i, j, k \in \mathbb{N}\}$,

3. $W := \{0^i 1^j 2^k \mid i \neq j, k \in \mathbb{N}\}$

4. $A := \{a^k b^m c^m \mid k, m \in \mathbb{N}\}$,

5. $B := \{a^m b^m c^k \mid m, k \in \mathbb{N}\}$,

6. $F \backslash R$.

However, and perhaps surprisingly given the last point of the exercise above, context free languages are very sensitive under intersection and complementation.

Lemma 29. *1. There exist context free languages* X *and* Y *so that* $X \cap Y$ *is not context free.*

2. There exist context free languages X *and* Y *so that* $X \backslash Y$ *is not context free.*

Proof. For this, we will use the languages A, B, F, R, and W of the previous exercise and also L_6 of Example 11.3, above. For the first point, consider the languages A and B. $A \cap B = L_6$, which we know is not context free.

A bit trickier is the following.

We first observe that $F \backslash A = (F \backslash R) \cup W$. Therefore $F \backslash A$ is a union of two context free languages, and hence is context free. Similarly, $F \backslash B$ is also a context free language.

Now set $C := (F \backslash A) \cup (F \backslash B)$. Clearly again, C is a context free language. But Lo! $F \backslash C = L_6$ again!! (Check this using basic set theory.) Thus, the class of context free languages is not closed under complement. $\square$

Research Question 1. *Find a non-trivial set of conditions on an infinite family* (L_i) *of nested context free languages which guarantee that the union*

$$\bigcup_{i \in \mathbb{N}} L_i$$

is a context free language.

In Lecture 13 we will describe a language $Co(Z^2 * Z)$ where no one can tell (so far) whether or not the language is context free! The research question above is intended to inform the search for answering the concrete question about $Co(Z^2 * Z)$.

12 L12: Groups and Languages II (PDA's and Virtually Free Groups)

We are finally ready to do the first half of the Muller-Schupp result!!!

12.1 Virtually something

First off, we define the adjective *virtually*.

Let P be a property of groups (so, P is preserved under isomorphism). A group G is *virtually P* if G has a finite index subgroup H so that H has property P.

Exercise 56. *Show that finite groups are virtually trivial!*

12.2 Quasi-isometry

This definition is motivated by the geometric properties of infinite groups. One considers the Cayley graph of a group under some finite generating set X. This Cayley graph turns the group into a metric space (take any edge to be length one, and then the distance between two vertices (group elements) to be the length of the shortest path from one to the other). Two metric spaces (S_1, d_1) and (S_2, d_2) are then *quasi-isometric* if there is a (not necessarily continuous) function θ from S_1 to S_2 so that the distances in S_1 are scaled roughly linearly $(d_2(\theta(x), \theta(y)) \leq A \cdot d_1(x, y) + B)$, and where there is a further constant distance C so that for every point a in S_2 there is some point b in the image of θ so that $d_2(a, b) < C$. Roughly, this means that the large scale geometries of the two spaces are very similar. This is often called "coarse geometry."

Now, if we would like to think about a property of a (finitely generated) group which property is preserved by isomorphism, and which is detectable geometrically (as in, you can see this property in a Cayley graph for the group), then this property really should be detectable in the Cayley graph built using any finite generating set. This sets up the next exercise.

Exercise 57. *Let G be a finitely generated group with two finite generating sets X_1 and X_2. Define metrics d_1 and d_2 on G from these generating sets by the rules*

$$d_i(x, y) = \min\left\{|w| \mid w \in W(X_i), w =_G x^{-1} \cdot y\right\}.$$

Show that the identity map induces a quasi-isometry from the space (G, d_1) to the space (G, d_2). Note that multiplying x on the right by $x^{-1}y$ produces y, so we can think of w as representing the labelling of a path from x to y in the Cayley graph.

For the next bit of the discussion, we need the following, really useful lemma.

Lemma 30 (Schreier's Lemma). *Let G be a finite index subgroup of a group H. Then, G is finitely generated if and only if H is finitely generated.*

Proof. Suppose H is a group and G is a finite index subgroup of H, let us say with index m. Let us choose a traversal $T := \{1_H, t_2, t_3, \ldots, t_m\}$ for the cosets of G in H, so that

$$H = \amalg_{t \in T} H \cdot t$$

since the cosets of H partition G. (We denote by t_1 the element $1_H \in T$.)

Suppose firstly that G is finitely generated with generating set X_G, we will show that H is finitely generated.

Now, every element $h \in H$ can be written as $g_h \cdot t_h$ for some $g_h \in G$ and $t_h \in T$. Therefore, given such an $h \in H$ and expanding g_h as a product $g_h = \hat{g}_1^{\epsilon_1} \hat{g}_2^{\epsilon_2} \ldots \hat{g}_p^{\epsilon_p}$ where for each index i, we have $\hat{g}_i \in X_G$ and $\epsilon_i \in \{-1, 1\}$, we see we can write h as a product of elements in the finite set $X_G \cup X_G^{-1} \cup T$. Thus, H is finitely generated.

Suppose instead that H is finitely generated with *inverse closed* generating set $X_H = \{h_1, h_2, \ldots, h_k\}$, we will show that G admits a finite generating set. (A generating set X_H is inverse closed if and only if for every $h \in X_H$ we have $h^{-1} \in X_H$ as well.)

Firstly, for all $h \in X_H$ determine $g_h \in G$, $t_h \in T$ so that $h = g_h \cdot t_h$ noting that these terms g_h and t_h are completely determined since H is a disjoint union of the cosets of G.

Now let $i \in \{1, 2, \ldots, k\}$ and $r,\ s \in \{1, 2, \ldots, m\}$, and determine $g_{r,i}, \tilde{g}_{r,s} \in G$, $t_{r,i}, \tilde{t}_{r,s} \in T$ so that $g_{r,i} \cdot t_{r,i} = t_r \cdot g_i$ and $\tilde{g}_{r,s} \cdot \tilde{t}_{r,s} = t_r \cdot t_s$. Noting that we are using the tilde decoration to differentiate between when sets of indices are coming from $\{1, 2, \ldots, k\} \times \{1, 2, \ldots, m\}$ or from $\{1, 2, \ldots, m\}^2$.

Suppose now that $h \in H$, and we write h as a product of the generators $\{h_1, h_2, \ldots, h_k\}$ as

$$h = h_{\theta(1)} \cdot h_{\theta(2)} \cdot \ \ldots \ \cdot h_{\theta(n)}$$

for some n where θ is a function which picks the correct generators in the correct order so the product expression is correct. We then have

$$\begin{aligned}
h &= h_{\theta(1)} \cdot h_{\theta(2)} \ldots h_{\theta(n)} \\
&= g_{\theta(1)} t_{\theta(1)} \cdot g_{\theta(2)} t_{\theta(2)} \cdot \ \ldots \ \cdot g_{\theta(n)} t_{\theta(n)} \\
&= \prod_{i=1}^{n} g_{\theta(i)} t_{\theta(i)}.
\end{aligned}$$

We now show that we can migrate the symbols in T to the right so we are left with a product of elements in the set X_G defined as

$$X_G := \{g_1, g_2, \ldots, g_k\} \cup \{g_{r,i} \mid r \in 1 \leq r \leq k, 1 \leq i \leq m\} \cup \{\tilde{g}_{r,s} \mid 1 \leq r, s \leq m\}$$

followed by a single element from T. We now do this work and re-express h:

$$\begin{aligned}
h &= g_{\theta(1)} t_{\theta(1)} \cdot g_{\theta(2)} t_{\theta(2)} \cdot \ \ldots \ \cdot g_{\theta(n)} t_{\theta(n)} \\
&= g_{\theta(1)} {\color{red} t_{\theta(1)} \cdot g_{\theta(2)}} t_{\theta(2)} \cdot \ \ldots \ \cdot g_{\theta(n)} t_{\theta(n)} \\
&= g_{\theta(1)} {\color{blue} g_{\theta(1),\theta(2)} \cdot t_{\theta_1,\theta(2)}} \cdot t_{\theta(2)} \cdot \ \ldots \ \cdot g_{\theta(n)} t_{\theta(n)} \\
&= g_{\theta(1)} g_{\theta(1),\theta(2)} \cdot {\color{red} t_{\theta_1,\theta(2)} \cdot t_{\theta(2)}} \cdot \ \ldots \ \cdot g_{\theta(n)} t_{\theta(n)} \\
&= g_{\theta(1)} g_{\theta(1),\theta(2)} \cdot {\color{blue} \tilde{g}_{\phi(1,2,2)} \tilde{t}_{\psi(1,2,2)}} \cdot g_{\theta(3)} t_{\theta(3)} \ \cdots \ \cdot g_{\theta(n)} t_{\theta(n)} \\
&= g_{\theta(1)} g_{\theta(1),\theta(2)} \cdot \tilde{g}_{\phi(1,2,2)} {\color{red} \tilde{t}_{\psi(1,2,2)} \cdot g_{\theta(3)}} t_{\theta(3)} \ \cdots \ \cdot g_{\theta(n)} t_{\theta(n)} \\
&= \ldots \\
&= g_{\theta(1)} \cdot \left(\prod_{i=2}^{n} \left(g_{\pi(i-1,i)} \cdot \tilde{g}_{\rho(i-1,i,i)} \right) \right) \cdot t',
\end{aligned}$$

for some indexing functions θ, ϕ, ψ, π, and ρ, and some $t' \in T$. By simply observing that $h \in G$ if and only if $t' = 1_H$, we see that every element of G can be written as a product using only elements from X_G. $\square$

Remark 5. *Observe that in the above proof, $g_{1,r} = g_r$ and $\tilde{g}_{1,r} = \tilde{g}_{r,1} = 1_H$, since $t_1 = 1_H$, so in particular, X_G is actually:*

$$X_G := \{g_{r,i} \mid r \in 1 \leq r \leq k, 1 \leq i \leq m\} \cup \{\tilde{g}_{r,s} \mid 2 \leq r, s \leq m\}.$$

With Schreier's Lemma in hand, the following exercise is standard.

Exercise 58. *Let G be a group and H be a finite index subgroup. Suppose G is finitely generated by the (inverse closed) set X_G, while H is finitely generated by the (inverse closed) set X_H. Then, the inclusion map from H to G induces a quasi-isometry of the Cayley graphs $\Gamma(H, X_H)$ and $\Gamma(G, X_G)$.*

Thus, the properties that are detectable for quasi-isometry type are properties that are passed down to the finite index subgroups of a group. Suddenly, one just likes to throw away the "messiness" of a big group, and if possible, pass to it's "clean" finite index subgroups to see what is really going on! And so, there is an industry of understanding a group only up to understanding one of its finite index subgroups.

Of course, there are plenty of group theorists who still think there are hard problems to be studied in finite groups, but there is certainly also a set that thinks of those problems as being "virtually trivial."

12.3 Virtually free groups are CF groups.

So from the above, we know that one of the themes of our course is to have a new handle on the virtually free groups; the groups G which have a finite index free subgroup.

We are now ready to show the wonderful result:

Theorem 31. *Let G be a virtually free group generated by a finite set Y, then $WP(G,Y)$, the word problem for G on generating set Y, is a deterministic context free language.*

Proof. Suppose G is generated by an (inverse closed) finite set $X_G = \{g_1, g_2 \ldots, g_m\}$ for some $m \in \mathbb{N}$ and that G is virtually free with a finite index free subgroup H. By Poincaré's Lemma, there is a subgroup $N \leq H$ which is normal in G and finite index in G. As N is a subgroup of a free group, it too is free. Let us now assume that N is a maximal free, finite index, normal subgroup of G, and suppose it's index is n. Now, let $T = \{t_1, t_2, \ldots, t_n\}$ be a traversal of the set of (right) cosets of N in G so that $t_1 = 1_G$. Clearly the set $B = N \cup T$ generates G (since the cosets of N cover G) and by following the proof of Schreier's Lemma from the previous section we can find some finite subset $Y_G := \{n_{i,j} \mid (t_i, g_j) \in T \times X_G\} \amalg (T \backslash \{1_G\})$ which generates G, and so that $B_N := \{n_{i,j} \in N \mid (t_i, g_j) \in T \times X_G\}$ actually generates N.

Now, as N is finitely generated and free, it admits a free basis $X_N = \{x_1, \ldots, x_p\}$ where p is the rank of N. For each $n_{i,j} \in B_N$, set $w_{i,j}$ to be the unique freely reduced element of $W(X_N)$ representing $n_{i,j}$, which word, we note, is written in the alphabet $X_N \cup X_N^{-1}$.

Now we can consider the group $Q := G/N$. Our PDA $\mathcal{A}$ is going to have as state set the elements of T (which will represent the corresponding elements of Q) and one other state, which we will denote as ξ. The initial state of $\mathcal{A}$ will be the state 1_G. The automaton $\mathcal{A}$ will have one accept state, which is also the state 1_G.

The tape alphabet Σ of the automaton $\mathcal{A}$ will be the set X_G. A word will be accepted if and only if upon reading the word and reacting, the automaton will have active state the accept state 1_G. The automaton will be deterministic.

The stack alphabet χ of the automaton $\mathcal{A}$ will be the union of the alphabet $X_N \cup X_N^{-1}$ with a one element set containing a new symbol, which symbol we will denote as $\perp$.

For each state t_i, for $i > 1$, and for the state ξ, we now define the edges leaving that state. For each generator g_j, consider the word $w_{i,j}$ (use $i = 1$ for the state ξ). We will now describe the edges from t_i (or ξ) to $t_{i,j}$ (or $t_{1,j}$). First, let us suppose $t_{i,j}$ is not 1_G. For every prefix p of $w_{i,j}$, with $p \neq w_{i,j}$, write $w_{i,j} = p \cdot w_{p,i,j}$, where $w_{p,i,j}$ is the suffix of $w_{i,j}$ after p. Denote by $a_{p,i,j}$ the first letter of $w_{p,i,j}$. Build the finite set $\tilde{S}_{p,w_{i,j}}$ of words (in the stack alphabet) defined

$$\tilde{S}_{p,w_{i,j}} := \left\{x \cdot p^{-1} \mid x \neq a_{p,i,j}^{-1}, x \in X_N \cup X_N^{-1} \cup \{\perp\}\right\}.$$

Now carry on to define $S_{w_{i,j}}$ as

$$S_{w_{i,j}} := (\cup_{p<w_{i,j}, p\neq w_{i,j}} S_{p,w_{i,j}}) \cup \{xw_{i,j}^{-1} \mid x \in \chi \text{if } xw_{i,j}^{-1} \text{ is already freely reduced }\}.$$

Further define a function $\theta_{i,j} : S_{w_{i,j}} \to (\{\perp\} \cup X_N \cup X_N{-}1)^*$ by the rule $y \mapsto$ (the free reduction of $y \cdot w_{i,j}$), noting that some of these reduced words might begin with the symbol $\perp$. Modify $\theta_{i,j}$ so that whenever $t_{i,j} = 1_G$ and the output would have been "$\perp$", $\theta_{i,j}$ will instead produce the empty word ε. The transitions from the state t_i corresponding to the input g_j are all the transitions of the form $(g_j, y)/\theta_{i,j}(y)$ which of course go to the state $t_{i,j}$, unless $t_{i,j} = 1_G$ but $\theta_{i,j}(y) \neq \varepsilon$, in which case the transition goes to the state ξ.

For the edges leaving 1_G, on reading g_j, we add the word $\perp w_{1,j}$ to the stack and go to state $t_{1,j}$. For the edges leaving ξ, on reading g_j, we add the word $\perp w_{1,j}$ to the stack and go to state $t_{1,j}$.

The automaton so constructed will read the entire word of input, and will result with active state 1_G if and only if the input word is equivalent to 1_G. It will have active ξ precisely when the word read so far is equivalent to a non-trivial element of N. $\square$

We have not yet highlighted an important fact.

Lemma 32. *The complement of a deterministic context free language is a deterministic context free language, but the complement of a non-deterministic context free language need not be a context free language.*

Proof. We have seen that the compliment of a context free language need not be context free in the previous lecture. Thus, we need only show that the compliment of a deterministic context free language is in fact deterministic context free.

But, this is easy. Let L be a deterministic context free language, and let $\mathscr{A} = (Q, \Sigma, \chi, \delta, S, \perp, F)$ be a deterministic PDA accepting the language, with accept states F. Change the accept states to $Q \backslash F$, and the resulting accepted language will be the exact compliment of L. $\square$

From now on in these notes, groups G which have a finite generating set X so that $WP(G, X)$ is a context free language will be called *CF groups*.

Exercise 59. *Show that the free group $F_{\{a,b\}}$ on basis $\{a, b\}$ is context free by building an explicit deterministic PDA accepting the word problem of $F_{\{a,b\}}$.*

13 L13: Groups and Languages III (Closure properties for Context Free Groups)

This section follows Diekert and Weiß, pages 20–21, very closely.

Here we point out some of the nice closure properties of the class of finitely generated CF groups. We will discuss these properties in lecture and perhaps even prove them, if the conversation goes that way.

Lemma 33. *Suppose G is a finitely generated group and $H \leq G$ is a finite index group with H a CF group. Then, G is a CF group.*

Colloquially, "The class of CF groups is closed under passage to finite index over-groups."

Lemma 34. *Suppose G is a finitely generated group with finite generating set X so that $WP(G, X)$ is a CF language. Then if Y is another finite generating set of G we have $WP(G, y)$ is a CF language.*

Lemma 35. *Suppose G is a CF group and $H \leq G$ is a finitely generated subgroup. Then H is a CF group.*

"The class of CF groups is closed under passage to finitely generated subgroups."

Proposition 36. $\mathbb{Z} \times \mathbb{Z} \cong \langle a, b \mid [a, b]\rangle$ *is not a CF group, but the language of co-words in $\mathbb{Z} \times \mathbb{Z}$ is a CF language.*

Exercise 60. *Prove every statement above in this section which was not proven during class discussion.*

The following is Theorem 2.14 of Diekert and Weiß. The proof is not hard, but we have to skip it in the interest of time. All CF grammars admit a "reduced" CF grammar which generates the same language.

Theorem 37. *(Hotz Isomorphism) Let $\mathscr{G} = (V, \Sigma, P, S)$ be a (reduced) context free grammar and let $L = L(\mathscr{G})$ be its generated language, with $\varepsilon \in L$. Then the inclusion of Σ^* into $F_{V\cup\Sigma}$ induces a canonical isomorphism:*

$$\phi : F_\Sigma/\langle\langle L\rangle\rangle \to F_{V\cup\Sigma}/P.$$

In particular, the minimal number of defining relations for $F_\Sigma/\langle\langle L\rangle\rangle$ is a lower bound on the number of productions P.

The next result is a major consequence of Hotz's theorem.

Corollary 38. *Context free groups and their finitely generated subgroups are finitely presented.*

Let N be a subgroup of a group G where G has a symmetric finite generating set Σ (symmetric means $\Sigma =_G \Sigma^{-1}$). We say that N has a *context free enumeration* if there is a context-free language $L \subseteq \Sigma^*$ so that for all $w \in L$ we have $w =_G v \in N$. Note that if we replace "context free" with "regular" in this definition then N would just be a rational subgroup.

Theorem 39. *[7] Let G be a finitely generated group and and N be a normal subgroup. The N has a context-free enumeration if and only if $N = \langle\langle R\rangle\rangle$ for some finite set $R \subseteq G$.*

Proof. First, suppose G is finitely generated by some finite symmetric generating set Σ and suppose $N = \langle\langle R\rangle\rangle$ for some finite set $R \subseteq G$. The following CF grammar yields a context-free enumeration for N:

$$\mathscr{G} := (V, \Sigma, P, S), \text{ where } P \text{ is}$$

$$S \mapsto aSa^{-1} \mid r \mid \varepsilon \forall a \in \Sigma,\, r \in R.$$

The other direction follows from Hotz's theorem.

□

Corollary 40. *Let G be a finitely presented group and N be a normal subgroup. Then N has a context-free enumeration if and only if G/N is finitely presented.*

In a slightly different direction, for a finitely generated group $G = \langle X \mid R \rangle$, we set $coWP(G, X) := (X \cup X^{-1})^* \backslash WP(G, X)$. That is, the *coWord Problem* for the pair (G, X) is the complement of the word problem of the pair (G, X); all the words which do not resolve to the trivial element in G.

$$coWP(G, X) := \{w \in (X \cup X^{-1})^* \mid w \neq_G 1_G\}.$$

As we will eventually see, the finitely generated groups with context free word problem are precisely the virtually free groups. But then, these groups have deterministic word problems, and so their co-word problems are also context free. In particular, the class of coCF groups (finitely generated groups with context free co-word problem) contains the CF groups. Indeed, the class of coCF groups also contains the group $\mathbb{Z} \times \mathbb{Z}$, and so is a strictly larger class of groups (compared with the class of CF groups).

Research Question 2. *Determine whether $\mathbb{Z}^2 * \mathbb{Z} \cong G = \langle a, b, c \mid [a, b] \rangle$ is a coCF group.*

This last research problem is topical in the research maths community at the time of this writing. The language passes all known Pumping Lemma/Ogden's Lemma type tests, but also, there is a graveyard filled with the broken automata of frustrated researchers who had hoped to show $coWP(G, \{a, b, c\})$ is in fact a CF language. If the group is not coCF, then the class of coCF groups is not closed under "free product," as per a conjecture of Holt, Rees, Röver, and Thomas.

14 L14: Bass-Serre Theory I (Graphs of Groups – Definitions and Examples)

14.1 A mixed formal-informal introduction

We begin with a somewhat complex definition, choosing a definition which parallel's Serre's treatment of graphs of groups.

Definition 14. *A graph of groups* $\mathscr{G} = (\mathbb{G}, \Gamma)$ *is a pair where* $\Gamma = (V, E, s, t)$ *is a directed graph with an involution function* $\overline{} : E \to E$ *on the directed edges so that for all* $e \in E$ *we have* $\overline{e} = f$ *where* $e \neq f$ *and where* $s(e) = t(f)$ *and* $t(e) = s(f)$*, and where* $\mathbb{G} = (\{G_x \mid x \in E \cup V\}, \{\alpha_e : G_e \to G_{s(e)} \mid e \in E\})$ *is a pair where the first set is a set of groups associated bijectively with the vertices and edges of* Γ *and where the second set is a set of group homomorphisms associating to each edge group an injective homomorphism into the vertex group at the start of the edge, with the further condition that for all* $e \in E$*, we have* $G_e = G_{\overline{e}}$*.*

Note that one can think of the vertices and edges as actually being the appropriate associated groups. Also, we usually call the maps α_e above "Boundary maps" as they map an edge group into its incident boundary group (at the start of the edge).

In conversation, we might loosely refer to Γ in the above situation as the graph of groups, when everything else is clear.

These objects might not seem too interesting at face value. However, there is an associated "canonical" group which can be presented from a graph of groups. (The "canonical" bit is something like a lie; the associated group depends on a choice of spanning tree in the graph; different spanning trees produce different groups, but the results are all the same up to isomorphism.) The canonical group has the property that its structure is strongly characterised by the graph of groups; both by the groups involved, and by the structure of the graph relating them. This then allows one to see a large group as being built from smaller groups in a meaningful way. This decomposition has played a major role in modern group theory and topology for the last 30 years (and even a bit longer in a less formal way).

The following definition is somewhat informal. The "$*$" represents a "base point." We give a formal construction a bit later where the base point will be given by maximal spanning tree in Γ (assuming Γ is connected).

Definition 15. *There is a canonical group associated with any finite connected graph of groups, called its* fundamental group*. The* fundamental group $\pi_1(\Gamma, *)$ of a graph of groups Γ *acts on a tree* T_Γ *in an orientation preserving fashion, with quotient graph* Γ*, where the vertex and edge stabilisers of the action of* $\pi_1(\Gamma, *)$ *on* T *correspond to the vertex and edge groups of the graph of groups in the quotient.*

The resulting theory of groups given via graphs of groups is called *Bass-Serre Theory* after its chief protagonists Hyman Bass and Jean-Pierre Serre [1, 2, 11].

Bass, Hyman. "Covering theory for graphs of groups." Journal of Pure and Applied Algebra 89, (1–2),pp 3–47, (1993).

Bass, Hyman and Serre, Jean-Pierre. "Arbres, amalgames, SL_2." Astèrisque, 46, 3^{rd} ed. (1983)

Serre, Jean-Pierre. "Trees." Springer Monographs in Mathematics, Berlin: Springer-Verlag (2003). (Translated by John Stillwell.)

14.2 Free Products of Groups

We now define the free product of a set of groups. This definition generalises a free group, which is a free product of a collection of groups, each of which is infinite cyclic.

Definition 16. *Let I be an index set, and for all $i \in I$, let G_i be a group with generating set X_i, and suppose that for indices $i \neq j$, that $G_i \backslash \{1_{G_i}\} \cap G_j \backslash \{1_{G_j}\} = \varnothing$. We say a group G is a free product of the groups $\{G_i\}_{i \in I}$ if for all i we have that $\langle X_i \rangle = G_i \leq G$ and whenever there is a group H and a set of functions $\theta_i : X_i \to H$ so that for all i the function θ_i extend to a group homomorphism $\hat{\theta}_i : G_i \to H$ then we have a unique group homomorphism $\phi : G \to H$ which agrees with the functions θ_i over the X_i and so that the following diagram commutes for all i.*

$$\begin{tikzcd}
X_i \arrow[r, phantom, "\subseteq"] \arrow[dr, "\theta_i"'] & G \arrow[d, dotted, "\exists!\phi"] \\
& H
\end{tikzcd}$$

We denote such a group as

$$\mathop{*}_{i \in I} G_i$$

.

As in the case of free groups, free products of groups actually exist. Below we give a "normal form" for the free product of groups.

Definition 17. *Let I be an index set and assume that for all $i \in I$, X_i is a generating set for a group G_i, so that for all $i \neq j$ indices in $/i$, we have $X_i \cap G_j = \varnothing$. We say a word $w \in W(\cup X_i)$ is in normal form for the set $\{X_i \mid i \in I\}$ if there is a decomposition $w = w_1 w_2 \ldots w_k$ where for each w_i we have $w_i \in W(X_i)$ and $w_i \neq 1_{G_i}$ and if for all indices i, we have that the generating sets X_i and X_{i+1} are distinct. In this case we also say that the expression $w_1 w_2 \ldots w_k$ is a normal form decomposition for w (for the set $\{X_i \mid i \in I\}$).*

We now place an equivalence relation on words in normal form.

Definition 18. *Let I be an index set and assume that for all $i \in I$, X_i is a generating set for a group G_i, so that for all $i \neq j$ indices in i, we have $X_i \cap G_j = \varnothing$. Suppose $u = u_1 u_2 \ldots u_j \in W(\cup X_i)$ and $v = v_1 v_2 \ldots v_k \in W(\cup X_i)$ are normal form decompositions for u and v. We then say $u \sim v$ if and only if $j = k$ and for all indices i we have $u_i =_{G_i} v_i$.*

As in the case of the definition of free groups, we now have that the equivalence classes of words so defined form a set admitting a group product using concatenation followed by reduction under equivalence under $\sim$.

Definition 19. *Let I be an index set and assume that for all $i \in I$, X_i is a generating set for a group G_i, so that for all $i \neq j$ indices in I, we have $X_i \cap G_j = \varnothing$. Suppose $u = u_1 u_2 \ldots u_j \in W(\cup X_i)$ and $v = v_1 v_2 \ldots v_k \in W(\cup X_i)$ are normal form decompositions for u and v. We define a product on the classes of such words under sim as follows*

$$[u]_\sim \cdot [v]_\sim = [u_1 u_2 \ldots u_j v_1 v_2 \ldots v_k]_\sim .$$

Here, when taking the product, if u_i and v_i are in the same group G_i, then if $u_i v_j = 1_{G_i}$ for some index i then we formally (inductively) define

$$[u_1 u_2 \ldots u_j v_1 v_2 \ldots v_k]_\sim = [u_1 u_2 \ldots u_{j-1} v_2 \ldots v_k]_\sim$$

.

This then becomes a group product for a group G as in the case of free groups, and we have the following normal form theorem.

Theorem 41. *Let I be an index set and assume that for all $i \in I$, X_i is a generating set for a group G_i, so that for all $i \neq j$ indices in I, we have $X_i \cap G_j = \varnothing$. Suppose $u = u_1 u_2 \dots u_j \in W(\cup X_i)$ is a normal form decomposition for u, and let G be the group with elements $[v]_\sim$ for $v \in W(\cup X_i)$ under the product defined above. Then $[u]_\sim = 1_G$ if and only if $j = 0$ (so u is the empty word).*

We then observe that the group G built as above actually satisfies the universal criterion of the diagram, and we call it the *free product of the groups G_i*.

In particular, the **normal form for a non-trivial element of $\mathop{*}\limits_{i \in I} G_i$** is essentially a (non-empty) product of the form

$$g_1 g_2 \cdots g_k$$

where for all indices i, we have

1. there is index $\alpha(i)$ so that $g_i \in G_{\alpha(i)}$,
2. $g_i \neq 1_{G_{\alpha(i)}}$, and
3. $\alpha(i) \neq \alpha(i+1)$.

And also, that any non-empty product satisfying this form and these three rules is in fact non-trivial in $\mathop{*}\limits_{i \in I} G_i$.

Finally, we observe that if I is an index set and assume that for all $i \in I$, $G_i = \langle X_i \mid R_i \rangle$, and so that for all $i \neq j$ indices in I, we have $X_i \cap G_j = \varnothing$, then we have

$$\mathop{*}_{i \in I} G_i = \langle \bigcup_{i \in I} X_i \mid \bigcup_{i \in I} R_i \rangle.$$

Exercise 61. *Consider the group $G := C_2 * C_2$.*

1. *Give a presentation for G.*
2. *Describe a normal form for the elements of G.*
3. *Find an index two subgroup H in G. Give the set of elements of H as a set of words in your generators in your normal form.*
4. *Describe the isomorphism type of H.*
5. *Prove G is virtually free.*

14.3 Formal algebraic definition of the fundamental group of a graph of groups.

Let $\mathscr{G} = (\mathbb{G}, \Gamma)$ be a graph of groups. We will inherit all notation from the definition (so, e.g., the sets V and E are defined, as are our vertex and edge groups, the edge involution, and the injective edge "boundary maps").

One can build a presentation for the fundamental group of Γ as follows.

First, find a spanning tree T for Γ. The fundamental group of Γ with respect to T (denoted $\pi_1(\Gamma, T)$) is defined as a quotient of the free product of groups

$$(\mathop{*}_{v \in V} G_v) * (\mathop{*}_{e \in E} e)$$

The rules for the quotient are:

1. For all $e \in E$ and $g \in G_e$ we have $\overline{e} \alpha_e(g) e = \alpha_{\overline{e}}(g)$ (the Serre relations),
2. For all $e \in E$, $e\overline{e} = \overline{e}e = 1$, and

3. For all $e \in E \cap T$, we have $e = 1$.

Exercise 62. *Consider the graph of groups given as* $\Gamma = (\{x, y, z\}, \{(x, y), (y, x), (y, z), (z, y), (z, x), (x, z)\}, s, t)$. *Where the groups are:*

$$G_x = \langle a \mid a^2 \rangle$$

$$G_y = \langle b \mid b^3 \rangle$$

$$G_z = S_3$$

and where the edge groups for the edges (x, y) *and* (z, x) *are both trivial, while the edge group for* (y, z) *is* $C_3 = \langle d \mid d^3 \rangle$. *The map* $\alpha_{(y,z)}$ *sends* d *to* b *and the map* $\alpha((z, y))$ *sends* d *to the permutation* $(1\,2\,3)$ *(note that* $\overline{(y, z)} = (z, y)$*). Give a presentation for the group which is the fundamental group of this graph of groups.*

15 L15: Bass-Serre Theory II (Normal forms)

15.1 Normal forms for fundamental groups of graphs of groups.

Let $\mathscr{G} = (\mathbb{G}, \Gamma)$ be a graph of groups with underlying graph $\Gamma = (V, E, s, t)$. For $v \in V$ set G_v to be the group associated to v, and for $e \in E$ set G_e to be the group associated with e and $\alpha_e : G_e \to G_{s(e)}$ the edge monomorphism associated with e. Finally, let the function $\overline{} : E \to E$ be the ordinary edge involution associated with a graph for a graph of groups. Consider the basic presentation of $\pi_1(\mathscr{G}, T)$ as P'. Carry out the basic Tietze transformations to remove the free edge generators of the set $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$ which are forced trivial by the relations of the presentation P' anyway, to make a new presentation P for a group isomorphic with $\pi_1(\mathscr{G}, T)$. We will call the new presentation the *core presentation of* $\pi_1(\mathscr{G}, T)$.

15.2 Tree equivalence

Notice that if an edge e goes from vertex a to vertex b (so $\overline{e}$ goes from b to a), then if $g \in G_b$ so that $g \in Image(\alpha_{\overline{e}})$ then $g((\alpha_{\overline{e}})^{-1}\alpha_e) \in G_a$ is the "equivalent" element of G_a (and these elements really are equivalent in $\pi_1(\mathscr{G}, T)$). This way in which one element of $\pi_1(\mathscr{G}, T)$ can show up represented in several of the vertex groups is the major source of difficulty for describing the normal forms for elements of $\pi_1(\mathscr{G}, T)$ under the core presentation. In the next paragraph, we just expand on the idea above for a directed path using edges in $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$. (Note that there is only one minimal length *undirected* path in T connecting any two vertices. Such a path has a unique directed version using the edges of $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$ if we instead traverse the edge $\overline{e}$ each time the undirected path traverses an edge e but in the wrong direction.)

Suppose $g \in G_v$ for some vertex v, and $h = G_u$ for some other vertex u. Now, in $\pi_1(\mathscr{G}, T)$ the elements g and h will represent the same element if there is a directed path of edges

$$p = e_1 \circ e_2 \circ \cdots \circ e_k$$

in $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$ so that $s(p) = u$, $t(p) = v$, and where $g((\alpha_{\overline{e_k}})^{-1}\alpha_{e_k})\cdot((\alpha_{\overline{e_{k-1}}})^{-1}\alpha_{e_{k-1}})\cdots((\alpha_{\overline{e_1}})^{-1}\alpha_{e_1}) = h$. When this happens we say that *$g \in G_v$ is represented as h in G_u*.

Exercise 63. *Show that if $g \in G_v$ is represented as h in G_u, then also $h \in G_u$ is represented as $g \in G_v$, and indeed the "is represented by" relation here is an equivalence relation.*

15.3 Grouped forms

So now suppose g is an element of $G := \pi_1(\Gamma, T)$. Then g has can be written as a freely reduced product of the generators of the core presentation of $\pi_1(\Gamma, T)$. Group maximal substrings of this product expression for g so that each such substring is either a solitary edge (but not from the edges in $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$), or is a freely reduced word in $W(X)$ for X the generating set of G_v for some vertex group G_v; denoting each such maximal substring as an expression g_i. We can then re-express g as

$$g = g_1 g_2 \cdots g_n$$

for some $n \in \mathbb{N}$. By our condition on maximality of words using generators from the same vertex group, we know that if for some index i, both g_i and g_{i+1} are from vertex groups, then these two vertex groups are not from the same vertex. We call any such expression for g a *grouped form for g (with n parts)*.

15.4 Basic reductions

Suppose as above $g \in \pi_1(\Gamma, T)$ and g is written in grouped form as $g = g_1 g_2 g_3 \ldots g_n$. As alluded to above, we might be able to express g as a grouped form with fewer parts. In particular, if g_i and g_{i+1} represent elements in two vertex groups G_{v_i} and $G_{v_{i+1}}$ respectively, and if g_{i+1} can be represented as h in G_{v_i}, then we have that the expression $g_i g_{i+1} = g_i h \in G_{v_i}$, where k can be expressed as a word k_i in the generators of G_{v_i} and their inverses. We then have

$$g = g_1 g_2 \dots g_{i-1} g_i g_{i+1} \dots g_n = g_1 g_2 \dots g_{i-1} k_i g_{i+2} \dots g_n$$

which is g written in a grouped form with only $n-1$ parts. We call this type of change of grouped forms for g a *vertex conglomeration reduction*.

Similarly, suppose we instead have the grouped form $g_1 g_2 \dots g_n$ for g, and there is an index i so that g_i is an edge e, while g_{i+1} represents an element y of $G_{t(e)}$ which happens to be in the image of $\alpha_{\overline{e}}$. Then there is an element $x \in G_{s(e)}$ so that $x = eye^{-1}$. This allows us to use the equivalence $ey \mapsto xe$ which moves a vertex group element from the right side of an edge to a different vertex group element on the left side of an edge. In this circumstance we have the following chain of equalities:

$$g = g_1 g_2 \dots g_{i-1} e y g_{i+2} \dots g_n = g_1 g_2 \dots g_{i-1} x e g_{i+2} \dots g_n.$$

Now we can find a word k_i representing x in the vertex group $G_{s(e)}$, and either e will now cancel with g_{i+2} if $g_{i+2} = e^{-1}$ or otherwise we just re-express g in grouped form

$$g = g_1 g_2 \dots g_{i-1} k_i \cdot (e \cdot g_{i+2}) \cdot g_{i+3} \dots g_n.$$

The new form has fewer parts (if we cancelled out $(e g_{i+2})$) or otherwise it is "improved" as we moved a vertex group word leftwards. (Clearly we can only do this sort of thing a finite number of times). We will call this change of grouped forms a *direct conjugation reduction*.

Now we can also combine the two ideas above. Suppose again we have the grouped form $g_1 g_2 \dots g_n$ for g, and there is an index i so that g_i is an edge e, while g_{i+1} represents an element y of G_v for some vertex $v \neq t(e)$. Now let $\hat{f}$ be the unique path in T which connects the vertex v to the vertex $t(e)$, and let f be either $\hat{f}$ or $\overline{\hat{f}}$ so that f is the directed path from v to $t(e)$. Finally suppose that $y \in Image(s(f))$ (the starting vertex of the path f). In this case y is represented by some element x in $G_{t(e)}$. If we now find a word k_{i+1} in the generators of $G_{t(e)}$ and their inverses so that $x = k_{i+1}$, we then have the chain of equalities:

$$g = g_1 g_2 \dots g_i g_i g_{i+1} \dots g_n = g = g_1 g_2 \dots g_i g_i k_{i+1} g_{i+2} \dots g_n.$$

We can now replace the original grouped form for g with the grouped form on the right. This is an improvement even though the number of parts did not decrease, because at the junction $g_i k_{i+1}$ it might be possible that we can apply a direct conjugation reduction, which might then reduce the number of parts. Obviously again, we can only do a move like this a few times to a given grouped word. We call this kind of change an *indirect conjugation reduction* since morally, we are replacing an expression fyf^{-1} with x (but where f is in $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$ and so would not need to have been listed in the initial product).

Thus we have three types of reductions: vertex conglomeration reductions, direct conjugation reductions, and indirect conjugation reductions. These three types of reductions will be called the *basic reductions.*

15.5 The Normal Form Theorem

We are now ready for the theorem for normal forms for graphs of groups.

Theorem 42. *Let $\mathscr{G} = (\mathbb{G}, \Gamma)$ be a connected graph of groups and let T be a maximal tree in Γ, and $g \in \pi_1(\mathscr{G}, T)$. If there is an integer $n > 1$ so that g can be represented as a grouped product $g = g_1 g_2 \dots g_n$ which admits no basic reductions, then g is not trivial in $\pi_1(\mathscr{G}, T)$.*

Note that it pays to recall consciously what the basic reductions are. For instance, if one of the g_i represents the trivial element in some vertex group G_{v_i} then the product actually a basic reduction.

The above theorem has this nice corollary.

Corollary 43. *The canonical homomorphisms $G_v \to \pi_1(\Gamma, T)$ are injective.*

Note, the last corollary means we can think of the vertex groups as subgroups of G. Also, as the edge groups inject into the vertex groups, we also have that the edge groups inject into $\pi_1(\mathscr{G}, T)$.

Remark 6. *Suppose K is a group, and one can pick out subgroups of K which appear to correspond to the groups in a graph of groups $\mathscr{G}$, and with the corresponding boundary maps, and etc. Then you can verify that K is isomorphic to the fundamental group of $\mathscr{G}$ simply by verifying that any product in the reduced grouped form (using your nominated subgroups) will be non-trivial in your group, and that everything in your group can be written as such a reduced grouped product. In this way you can prove your group splits over the graph of groups $\mathscr{G}$.*

Finally, we will call any product of generators (and their inverses) from $\pi_1(\mathscr{G}, T)$, which product is in grouped form and admits no basic reductions, to be *a product in normal form for $\pi_1(\mathscr{G}, T)$*.

15.6 Moral of the story

Morally, what one is doing is finding a complete re-writing system for words expressing elements of $\pi_1(\mathscr{G}, T)$. The basic idea is that given any element $g = g_1 g_2 \dots g_n$ (written as a grouped word for the core presentation), we can insert "connective tissue" directed paths p_i from the tree (well, actually from $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$ by adding "trivial" edges so that path p_i connects the terminal vertex for g_i (which is the vertex of the vertex group supporting g_i or it is the vertex $t(g_i)$ if g_i is an edge) to the initial vertex of g_{i+1} (which is the vertex of the group containing g_{i+1} if g_{i+1} is in a vertex group, or it is the vertex $s(g_{i+1})$ if g_{i+1} is an edge). Note that we also add the path p_n which connects the terminal vertex of g_n to the initial vertex of g_1. The resulting word is now

$$g_1 p_1 g_2 p_2 \cdots g_n p_n$$

which evaluates in the basic presentation to g still, as the edges of the inserted paths are forced trivial in the presentation. Now each of the reductions is actually replacing a string $e^{-1}xe$ by y where x and y are elements of vertex groups, that is, we are reducing length here by applying a Serre relations. The resulting reduced path is still a circuit, but now the length is two shorter. In the case that the path used to read $e^{-1}xf$ we effectively go through the process of substitutions:

$$e^{-1}xf \mapsto e^{-1}xee^{-1}f \mapsto ye^{-1}f$$

which first lengthens that path by two, then applies a Serre relation to return the length to what is was, but we have pushed the occurrence of a vertex group element leftward. This might allow a cancellation between e^{-1} and f, or between x and whatever is to the left of y at the end. The circuit thus gets shorter eventually, since y will conglomerate with a vertex group element to the left of y eventually, since these vertex group elements eventually must collide, and when they do, they are in the same vertex group. (Notice that we are not allowed to move g_1 backwards throughout the path p_n, so we eventually must have an edge cancellation or a collision of vertex group words.)

When this process finishes, no vertex group word g_i is to the right of an edge g_{i-1} which has $g_i \in Image(\alpha_{\overline{g}_{i-1}}$. Having found this stable circuit word representing g, we now just throw out all edge letters from the set $\{e \in E \mid e \in T \text{ or } \overline{e} \in T\}$, and we have the normal form element.

The following is more to point out that something is still missing. The following exercise has no chance of appearing on our final exam!

Exercise 64. *Suppose $\mathscr{G} = (\mathbb{G}, \Gamma)$ is a connected graph of groups. Further suppose that T and T' are maximal subtrees in Γ. Show that $\pi_1(\mathscr{G}, T) \cong \pi_1(\mathscr{G}, T')$.*

16 L16: Bass-Serre Theory III (Splittings of groups, free products with amalgamation and HNN extensions)

16.1 Splittings over graphs of groups

An isomorphism between a group G and the fundamental group of a graph of groups $\mathscr{G} = (\mathbb{G}, \Gamma)$ is call a *splitting of G ("over $\mathscr{G}$" or commonly, but less correctly, "over Γ").* If the edge groups are of a certain class of groups (Abelian, cyclic, finite, whatever), then we say the splitting is, e.g., "Over finite groups" or similarly. As indicated before, finding a splitting of a group is akin to finding a decomposition of a group as being built from less complicated groups put together in ways we can understand.

16.2 Free products with amalgamation and HNN extensions.

In the previous lecture we learned to compute presentations of fundamental groups of graphs of groups. The construction there is actually the result of iterating two processes: making free products with amalgamation, and making HNN extensions.

Given two groups $G_1 = \langle X_1 \mid R_1\rangle$ and $G_2 = \langle X_2 \mid R_2\rangle$ (with $X_1 \cap G_2 = \varnothing = X_2 \cap G_1$) and a third group $K = \langle X_K \mid R_K\rangle$ with two injections $\iota_1 : K \to G_1$ and $\iota_2 : K \to G_2$, we can form the *free product of G_1 and G_2 amalgamated over K* which is denoted by $G_1 \underset{K}{*} G_2$ and which is the group presented as:

$$G_1 \underset{K}{*} G_2 := \langle X_1 \cup X_2 \mid R_1 \cup R_2 \cup \{\iota_1(x)(\iota_2(x))^{-1} \mid x \in X_K\}\rangle$$

Ignoring this poor notation (why is it poor?), the actual construction is of tremendous use.

Exercise 65. *Let $G_1 = \langle X_1 \mid R_1\rangle$ and $G_2 = \langle X_2 \mid R_2\rangle$ (with $X_1 \cap G_2 = \varnothing = X_2 \cap G_1$) and $K = \langle X_K \mid R_K\rangle$ be groups. Suppose there are injective group homomorphisms $\kappa_1 : K \to G_1$ and $\kappa_2 : K \to G_2$, and we consider the graph of groups $\Gamma = (V = \{1, 2\}, E = \{(1, 2), (2, 1)\}, s, t)$ with the vertex group of vertex i being G_i, and with edge group K for both edges $(1, 2)$ and $(2, 1)$ and with edge monomorphisms $\alpha_{(1,2)} = \kappa_1$ and $\alpha_{(2,1)} = \kappa_2$.*

Show that $\pi_1(\Gamma, T) \cong G_1 \underset{K}{*} G_2$.

[Hint: use Tietze transformations. Does it matter which maximal tree you pick?]

Paraphrasing, the above exercise is essentially how you recognise a free product with amalgamation as a splitting of a group over a graph of groups Γ where Γ has two vertices, and two edges (going back and forth between the vertices) and where the edge group is a generic isomorphic copy of the subgroups being amalgamated.

In our next construction, we look at Higman, Neumann, and Neumann's famous extension (the HNN extension). Suppose you have a group $G = \langle X \mid R\rangle$ and two subgroups $K_1 = \langle Y_1\rangle \leq G$ and $K_2 = \langle Y_2\rangle \leq G$ with a group isomorphism $\phi : K_1 \to K_2$. We then denote by $G*_\phi$ the HNN extension of G with respect to ϕ, which is the group given below (assuming the symbol t is not in G):

$$G*_\phi = \langle X \cup \{t\} \mid R \cup \{t^{-1}yt(y\phi)^{-1} \mid y \in Y\}\rangle.$$

Note that t almost acts freely with respect to G, BUT, the new relators say that conjugating an element $k \in K_1$ by t will produce exactly the element $k\phi \in K_2$.

Example 4. *Consider $G = \langle a, b \mid [a, b]\rangle \cong \mathbb{Z}^2$, the two isomorphic subgroups $K_1 = \langle a\rangle$ and $K_2 = \langle b\rangle$, and the isomorphism $\phi : K_1 \to K_2$ which is induced by the map $a \mapsto b$. Give a presentation for $G*_\phi$.*

Exercise 66. *Suppose H is a group with subgroups K_1 and K_2 and where there is an isomorphism $\phi : K_1 \to K_2$. Give a splitting of $H*_\phi$ over a graph of groups Γ with only one vertex.*

Exercise 67. *Consider the presentations of $H*_\phi$ and $G_1 *_K G_2$ and their splittings over graphs of groups as described above. The resulting presentations of the fundamental groups of the graphs of groups are VERY similar. Explain the differences and the similarities as best as you can.*

HNN extensions are the key ingredient in proving Higman's amazing embedding theorem. (We will not prove it here, but just present it for your examination.)

Theorem 44 (Higman Embedding)**.** *Suppose G is a finitely generated group with generating set X and with a recursive defining set $R \subseteq W(X)$ of relators. Then G embeds in a finitely presented group.*

Exercise 68. *Prove that a splitting of a group G over trivial groups corresponds to a factoring of G as a free product of groups. [Hint: Be careful to describe what happens if the graph you split over is not just a tree!]*

17 L17: The modular group

1. Define the Hyperbolic plane $\mathbb{H} := \{z \in \mathbb{C} \mid \Im(z) > 0\}$
2. Define the groups $\mathcal{M} \leq \widetilde{\mathcal{M}}$ of Möbius transformations.
   (a) $\widetilde{\mathcal{M}} := \left\{ z \mapsto \frac{az+b}{cz+d}, ad - bc \neq 0, a, b, c, d \in \mathbb{C} \right\}$ is the full group of Möbius transformations.
   (b) $\mathcal{M}$ is the subgroup with $a, b, c, d \in \mathbb{Z}$ with $ad - bc = 1$: the *Modular Group* (usually given as $PSL_2(\mathbb{Z})$).
   (c) Exercise: show composition of maps in $\widetilde{\mathcal{M}}$ corresponds to matrix multiplication (matrices with complex entries).
3. Exercise: show that for $\gamma \in \mathcal{M}$ defined by $z \mapsto \frac{az+b}{cz+d}$ for $z \in \mathbb{H}$, has $\Im\gamma(z) = \frac{\Im(z)}{|cz+d|^2}$. (DONE IN LECTURE)
4. Define the matrix groups $SL_2(\mathbb{R})$, $SL_2(\mathbb{Z})$, $PSL_2(\mathbb{Z})$.
5. Exercise: Decompose the matrix
$$\begin{pmatrix} 11 & -3 \\ 15 & -4 \end{pmatrix}$$
as a product of the matrices
$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
and
$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$
Explain how you can conclude that $PSL_2(\mathbb{Z})$ is generated by (the equivalence classes of) S and T.
6. Define $\mathcal{D}$ to be the region in the hyperbolic plane containing the point $2i$ and bounded by the curves $C_1 = \{z \in \mathbb{C} \mid |z| > 1, \Re(z) = -1/2\}$, $C_2 = \{z \in \mathbb{C} \mid |z| > 1, \Re(z) = 1/2\}$, and $C_3 = \left\{z \in \mathbb{C} \mid z = e^{\theta i}, \theta \in [\pi/3, 2\pi/3]\right\}$. Define $\mathcal{M}' < \mathcal{M}$ as the subgroup $\mathcal{M}' = \langle S, T \rangle$ where $S(z) = -1/z$ and $T(z) = z+1$. State the following two key theorems for $\mathcal{M}'$ and $\mathcal{M}$.

   Theorem 45. *For all $z \in \mathbb{H}$, there is $\gamma \in \mathcal{M}'$ so that $\gamma(z) \in D$.*

   (So, the orbit of $\mathcal{D}$ under the action of $\mathcal{M}'$ is all of $\mathbb{H}$.)

   Theorem 46. *Suppose z is in the interior of $\mathcal{D}$. Show that for all $\gamma \in \mathcal{M}$, $\gamma(z) \in \mathcal{D}$ implies that γ is the identity map.*
7. Discuss that we call $\mathcal{D}$ a *fundamental domain for the action of $\mathcal{M}'$ on $\mathbb{H}$* (since both theorems hold for $\mathcal{M}'$). Discuss alternative definitions of the fundamental domain. (We don't really want the whole boundary. Also, isn't it better to just call $\mathcal{H}/\mathcal{M}'$ the fundametal domain?)

18 L18: More Modularity

1. Prove the two theorems of stated in the previous lecture (Theorem 45 and Theorem 46).

2. Draw the following corollaries from the two theorems:

   Corollary 47. *The group $\mathcal{M}$ is generated by S and T. (That is, $\mathcal{M}' = \mathcal{M}$.)*

   Noting that $X = ST$ is an element of order 3, and that $\langle S, T\rangle = \langle S, X\rangle$ we have the following further corollary.

   Corollary 48. *The modular group $\mathcal{M}$ is a quotient of the free product $\langle a, b \mid a^2, b^3\rangle$.*

19 L19: Time out for free products, again!

In this section we briefly discuss one of the great tools for detecting free products of groups. Fricke and Klein's famous Ping-Pong Lemma.

Lemma 49. *Let G be a group acting on a set X. Suppose G admits two subgroups G_1 and G_2, so that $|G_1| > 1$ and $|G_2| > 2$, and X admits subsets X_1 and X_2 with $X_1 \not\subseteq X_2$. Define $H := \langle G_1, G_2 \rangle \leq G$. Suppose further that*

1. *for all $x_1 \in X_1$ and $g_2 \in G_2 \backslash \{1\}$ we have $x_1 \cdot g_2 \in X_2$, and*
2. *for all $x_2 \in X_2$ and $g_1 \in G_1 \backslash \{1\}$ we have $x_2 \cdot g_1 \in X_1$,*

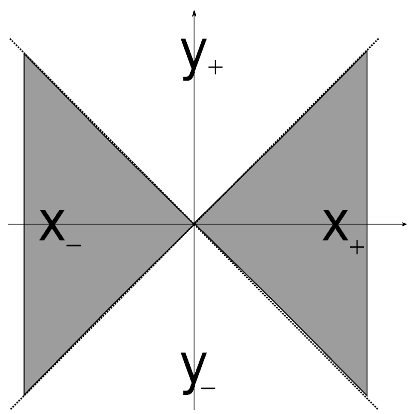
then $H = \langle G_1, G_2 \rangle \cong G_1 * G_2$.

In the above, we gave the statement of the Ping-Pong Lemma using right actions. Before going into the proof of this, work through this old exam question below, which uses left actions!

Exercise 69. *Consider the invertible matrices*

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

which together generate some group H (under matrix multiplication) which acts on the plane $\mathbb{R}^2$. Now let X_+, X_-, Y_+, and Y_- be the four open quadrants between the lines $y = x$ and $y = -x$ in the (xy)-plane $\mathbb{R}^2$ as depicted below.



Define $X := X_+ \cup X_-$ and $Y := Y_+ \cup Y_-$.

1. *Show that for any non-trivial element $\gamma \in \langle A \rangle$ and vector $\vec{v} \in Y$, we have $\gamma\vec{v} \in X$.*
2. *Find a rotation matrix M of order 4 acting on $\mathbb{R}^2$.*
3. *Show that $B = MAM^{-1}$ and conclude that for any non-trivial element $\tau \in \langle B \rangle$ and vector $\vec{w} \in X$, we have $\tau\vec{w} \in Y$.*
4. *Verify that $\langle A \rangle \cong \mathbb{Z} \cong \langle B \rangle$.*
5. *Deduce, using your work above, that $\langle A, B \rangle \cong F_2$, the free group on two generators.*

Good job!!

We will now prove the Ping Pong Lemma.

Proof. Assume G, G_1, G_2, H, X, X_1 and X_2 are as in the statement of the hypotheses of the Ping Pong Lemma. And let $g \in H$ be written as an alternating product:

$$g = g_1 g_2 g_3 \cdots g_n$$

for some natural number $n > 0$, where by an alternating product we mean that if $i < n$ is an index, then

1. if $g_i \in G_1$ then $g_{i+1} \in G_2$, and
2. if $g_i \in G_2$ then $g_{i+1} \in G_1$.

To conclude that H is the free product of G_1 and G_2, we need to argue that if g_i is not trivial for any index i, then g is not trivial.

Let us suppose therefore that g_i is not trivial for any index i. We will argue that g is not trivial. We first recall that if g is not trivial, then for any $h \in G$ we have g^h is also not trivial. Therefore our strategy will be to conjugate g by an element of G_2, and reduce to a new alternating product which has no trivial factors, and which begins and ends with an element from G_2. We can easily argue that such an element is non-trivial.

Consider g_1 and g_n. If they are both in G_2 then we conjugate by the identity element to recover exactly the alternating form $g_1 g_2 \ldots g_n$ which begins and ends with a non-trivial element of G_2.

If g_1 and g_n are both in G_1, then conjugating by any non-trivial element $k \in G_2$ will produce an alternating product $k^{-1} g_1 g_2 \ldots g_n k$ which has $n+2$ factors, non of which are trivial, and which begins and ends with an element from G_2.

If $g_1 \in G_1$ while $g_n \in G_2$, then there is a non-trivial element $k \in G_2$ which is not the inverse of g_n so that $k^{-1} g_1 g_2 \ldots g_{n-1}(gunk)$ is an alternating product of length $n+1$ of non-trivial terms ($(gunk)$ is a single factor in this product, and is nontrivial in G_2 by choice of k) which begins and ends with an element of G_2.

Similarly, if $g_1 \in G_2$ while $g_n \in G-1$, then there is nontrivial $k \in G_2$ so that $k \neq g_1$ so that the product $(k^{-1}g_1)$ is a non-trivial element of G_2 and so the product $(k^{-1}g_1)g_2 g_3 \ldots g_n k$ is an alternating product of $n+1$ nontrivial terms which begins and ends with an element of G_2.

Therefore we can assume now that $g_1 g_2 \ldots g_n$ is an alternating product of non-trivial terms from G_1 and G_2 which begins and ends with an element of G_2.

Now let $x \in X_1 \backslash X_2$. Such an element exists by hypothesis. We observe that n is odd, and further that all the odd index g_i move X_1 into X_2, while all the even index terms move X_2 into X_1. In particular, as x starts in $X_1 \backslash X_2$, after it is acted on by $g_1 g_2 \ldots g_n$ the result must be in X_2, and so cannot be x. Hence, $g_1 g_2 \ldots g_n$ cannot be the identity element. $\square$

Exercise 70. *The Modular group $\mathcal{M}$ is isomorphic to the free product $\langle a, b \mid a^2, b^3 \rangle$.*

20 L20: Combinatorial Two Complexes, part I

Here we give the definitions of a combinatorial two complex.

We define the presentation for the fundamental group of a connected combinatorial two complex. Read pages 1-6 of Brent Everitt's older notes for details, and sections 1-4 of Brent Everitt's new notes for a lot of details (also covering the next lecture).

For exam, I will expect you to know what a combinatorial two-complex is, and, how to go back and forth between a connected combinatorial two complex and a presentation of its fundamental group. (So, given a f.p. group, you can build a two complex with that group as its fundamental group, and given a complex, you can produce a presentation of its fundamental group.)

20.1 Definitions

A *combinatorial* 2*-complex* $\Gamma = (V, E, F, \partial)$ is a tuple of three sets V, E, and F and a map (*the boundary map*) ∂ which characterises relationships amongst the elements of the sets V, E, F (which we take as an explanation of how the "pieces" of the complex "fit together").

The set V is a set of vertices, which we call the 0*-skeleton of* Γ, and we denote the 0-skeleton as Γ^0.

The set E is a set of directed edges, which connect vertices of Γ^0 according to rules expressed by the boundary map ∂. If e is a directed edge connecting vertex v_0 to vertex v_1 (so, $s(e) = v_0$ and $t(e) = v_1$), then we write $\partial(e) = v_1 - v_0$. (So, in general, $\partial(f) = t(f) - s(f)$ for generic directed edge f.) The pair (V, E) together with induced maps $s : E \to V$ and $t : E \to V$ as determined by ∂ create a directed graph, which we refer to as the 1*-skeleton of* Γ and which we denote as Γ^1. We assume that for every edge $e \in E$, there is another edge $e^{-1} \neq e$ and use the rule that $(e^{-1})^{-1} = e$, in a fashion similar to how we set up our graphs in graphs of groups. We will think of e and e^{-1} as representing two different directions of the same edge, but formally they are different edges. For each such edge, we will choose one of the edges as representing the "normal direction" so that the other direction is the "inverse direction."

The set F is a set of *faces of* Γ. As in the case with the directed edges, faces also have "orientation," and we assume that for each face f, there is a different face f^{-1}, where $(f^{-1})^{-1} = f$. As a visual aid, one can think of a face of Γ as a two dimensional disc that we attach to the 1-skeleton Γ^1 by using a directed *attaching map* which goes from the boundary of the disc (an oriented circle) into the 1-skeleton Γ^1. This is determined again by the boundary map ∂. If f is a face, then $\partial(f)$ is a set of words representing a combinatorial cycle in the (directed)one skeleton Γ^1 (and, we call this set a *cycle*), and we are to envisage that we are gluing the face to the 1-skeleton by identifying the boundary circle of the face to the cycle in the graph Γ^1. Cycles in Γ^1 for this context can be one of two types. A cycle may be full set of cyclic rotations of the edge lables of some closed, directed cycle in Γ^1 (and here, we are recording the directions we went over those edges (by recalling that we have chosen some notion of preferred direction for each edge pair)). Alternatively, a cycle may have just a single element, which is a vertex from V.

In any case, if f is a face, and $\partial(f)$ is a set of finite words in the directed-edge alphabet, then $\partial(f^{-1}) = (\partial(f))^{-1} = \{w^{-1} \mid, w \in \partial(f)\}$ (for each word in $\partial(f)$, its inverse can be formed by inverting each letter and then writing the word in reverse order, just as in forming a group inverse). We are again to think of f and f^{-1} as one face in an intuitive way, but we use two faces since this enables some nice cooperation with Bass-Serre Theory in general.

Note that for any two-dimensional face f. Each element in $\partial(f)$ is called a *boundary word for* f.

Example 5. *Take* $V = \{v\}$*,* $E = \{a, a^{-1}, b, b^{-1}\}$*, and* $F = \{f, f^{-1}, g, g^{-1}\}$*, where we specify that* $\partial(a) = v - v$*,* $\partial(b) = v - v$ *(these are both technically written as zero, but that would be confusing just now), and where*

$$\begin{aligned}
\partial(f) =& \{a^{-1}b^{-1}ab, b^{-1}aba^{-1}, aba^{-1}b^{-1}, ba^{-1}b^{-1}a\}, \\
\partial(f^{-1}) =& \{b^{-1}a^{-1}ba, a^{-1}bab^{-1}, bab^{-1}a^{-1}, ab^{-1}a^{-1}b\}, \\
\partial(g) =& \{v\}, \text{ and} \\
\partial(g^{-1}) =& \{v\}.
\end{aligned}$$

Exercise 71. *Sketch carefully some two-dimensional picture which morally captures what this two-complex should "look like."*

20.2 Fundamental group of a two-complex

We will say two paths in Γ^1 are *homotopic rel endpoints*, and two cycles are *homotopic rel basepoint* if there is a finite chain of homotopy reductions (rel endpoints or basepoints) which change the first path into the second.

So, what are the valid homotopy reductions? There are two:

1. insert or remove a *spike* based at any vertex in the path, and
2. insert or remove a *boundary word* from a face with vertex on the path.

These are best seen diagramatically.

Now, for a connected combinatorial 2 complex, the (equivalence classes of) loops based at a chosen basepoint $*$ form a group, which we denote as

$$\pi_1(X, *)$$

is given as the set of homotopy classes of the cycles in Γ^1 based at the vertex $*$. The multiplication of equivalence classes is simply by concatenation of two (equivalence classes of) cycles as a single (equivalence class of a) longer cycle.

Given any connected 2-complex Γ and a choice of basepoint, there is an algorithm that gives a presentation for $\pi_1(\Gamma, *)$. The resulting isomorphism type of the group presented by the algorithm is independent of the choices in the algorithm, and indeed, even of the original choice of basepoint.

1. Choose a maximal tree T (agree that an edge in the tree means the inverese edge is "also" in the tree and is the "same" edge) in Γ^1.
2. Each edge e which is not in the tree determines a unique minimal length cycle of the form $w_e := c_1 \cdot c_2 \cdots c_k \cdot e \cdot c_{k+2} \cdot c_{k+3} \ldots c_n$ beginning and ending at $*$, where all of the c_i are in the tree T (here, it is fine to treat inverse edges as not being identified: Their corresponding cycle is the same as inverting the cycle for e in any case: e.g., $w_{e^{-1}} = c_n^{-1} c_{n-1}^{-1} \cdots c_{k+2}^{-1} \cdot e^{-1} \cdot c_k^{-1} \cdots c_1^{-1} = (w_e)^{-1}$ where this last expression is simply the inverse path of the cycle w_e in the graph Γ^{-1}. The generators of our group will be the set $X_\Gamma := \{w_e \mid e \notin T\}$. Note that this is an "inverse closed" generating set since we are allowning both cycles w_e and $w_{e^{-1}}$.
3. For each face f, and for each vertex v that is a vertex in the boundary cycle of f, there is a path in the tree T from the basepoint $*$ to a vertex v which is the start and end of a cycle in Γ^1 labelled by a boundary word w of f. We can then replace w by a word from the alphabet X_Γ, by deleting all edges in w that are in T and replacing each edge e in w that is not in T by the "letter" w_e. Denote by R_Γ the set of words in the alphabet X_Γ we can construct in this manner.
4. We can now present our group

$$\pi_1(\Gamma, *) = \langle X_\Gamma \mid R_\Gamma \cup \{w_e \cdot w_{e^{-1}} \mid e \in E, e \notin T\} \rangle .$$

Note that our presentation of "the fundamental group of Γ" is not very efficient. For each face f, we actually only need one boundary face relation (the other are conjugates), and we do not need any face boundary relations for the inverse face f^{-1}. We also only need one generator from each pair $(w_e, w_{e^{-1}})$, and then we can remove the extraneous $w_e \cdot w_{e^{-1}}$ relators. In any case, all of this can be done using T_2 and T_4 Tietze moves.

Morally, the group should really be written as $\pi_1(\Gamma, T, *)$, and not as $\pi_1(\Gamma, *)$, since our choice of T impacts the presentation (if not the isomorphism type) of the group.

Exercise 72. *Look up the shape of a (closed, oriented) surface of genus two (it looks like a donut with two holes instead of one).*

1. *Build a combinatorial* 2*-complex "representing" this shape,*
2. *compute its fundamental group using the method above, and*
3. *show your presentation is (Tietze) equivalent to the "one-relator" group*

$$\langle a, b, c, d \mid [a, b][c, d] \rangle.$$

21 L21: Tree decompositions, bags, and being neighbourly

In this section, we learn how to approximate a graph which is quasi-isometric to a tree by a tree. The key method is to use "Tree decompositions" which were used by Robertson and Seymour to show that any infinite collection of finite undirected graphs admits a pair of graphs one of which is the graph minor of the other (this result took Robertson and Seymour twenty papers to write up, and these papers appeared over the course of twenty years!). Recall that a graph G is a minor of a graph H if there is a finite sequence of edge contractions and edge deletions bringing H to G.

Throughout the section we will assume we are working with a graph $\Gamma := (V, E, s, t)$.

Definition 20. *A tree decomposition of a non-empty connected graph* $\Gamma := (V(\Gamma), E(\Gamma), s, t)$ *is an undirected tree* $T = (V(T), E(T), ends)$ *together with a mapping* $\theta : V(T) \to \mathscr{P}(V(\Gamma))$ *defined by the rules* $p \mapsto X_p$ *where* $p \in V(T)$ *and where* X_p *is a finite subset of* $V(\Gamma)$ *such that the following conditions are satisfied.*

1. *For every node* $v \in V(\Gamma)$ *there is some* $p \in V(T)$ *such that* $v \in X_p$*,*
2. *For every edge* $e \in E(\Gamma)$ *there is some* $p \in V(T)$ *such that* $\{s(e), t(e)\} \subseteq X_p$*. (For simplicity, we say that* X_p *contains the edge* e*.)*
3. *If* $v \in X_p \cap X_q$ *for some* $v \in V(\Gamma)$*, then we have* $v \in X_r$ *for all vertices* r *of the tree which are on the geodesic from* p *to* q*. (So, the set* $\{t \in V(T) \mid v \in X_t\}$ *forms a subtree of* T*.)*

In the situation in the definition above, we will likely refer to the tree T AS a tree decomposition of Γ for ease of language. The sets X_p are called the *bags of the tree decomposition.* Given a bag, we call its cardinality its *bagsize*. Given the graph and following some unfortunate language in the literature, if we have a tree decomposition T of Γ we define

$$bs(T) = sup\{|X_t| \mid t \in V(T)\}$$

the *bagsize of* T.

We say Γ has *finite treewidth* if there is some $k \in \mathbb{N}$ for which there is a tree decomposition T for Γ with $bs(T) = k$. In this case (perhaps unfortunately), Robertson and Seymour defined the *treewidth of* Γ to be $p - 1$ where p is the infimum of the treewidths of all tree decompositions of Γ (a tree with at least two vertices will have treewidth 1, for instance).

Lemma 50. *Suppose* Γ *is a graph with finite treewidth* k *for some* k *and that* $\chi \leq \Gamma$ *is a subgraph of* Γ*, then the treewidth of* χ *is less than or equal to* k*.*

Proof. Any tree decomposition T for Γ induces a tree decomposition T' for χ simply by restricting the bags of T for Γ to subsets of vertices of χ. □

The following lemma gives connectivity properties of tree decompositions.

Lemma 51. *Suppose that* $\Gamma = (V(\Gamma), E(\Gamma), s, t)$ *has a tree decomposition* $T = (V(T), E(T), ends)$*.*

1. *Suppose* X_a*,* X_b*, and* X_c *are bags and that* b *is on the geodesic in* T *from* a *to* c*. Then all paths in* Γ *from any vertex in* X_a *to any vertex in* X_c *must visit a vertex in* X_b*.*
2. *If* X_a *and* X_b *are two bags so that* a *is connected to* b *by an edge in* T*, then* $X_a \cap X_b \neq \varnothing$*.*

Proof. (1): Let X_a, X_b and X_c as in the hypotheses of the lemma statement. Let x and z be vertices of X_a and X_c respectively, and let $p = e_1 e_2 \dots e_n$ be an undirected path in Γ from x to y. Let X' be a bag containing $ends(e_1)$. If X_b is on the geodesic from X_a to X', then x is in X_b. Otherwise b is on the geodesic from (the vertex in T corresponding to) X' to c, and we are done by induction on the length of p.

(2): Let a and b be adjacent vertices of T. Further, let $p = e_1 e_2 \dots e_n$ be a minimal length undirected path Γ so that $s(p) = x_0 \in X_a$ while $t(p) = x_n \in X_b$. If $n = 0$ (p is the constant path at x_0) then $x_0 \in X_a \cap X_b$. Thus we will assume below that $n > 0$.

Now for all e_i with $i > 1$ we have $s(e_i) \notin X_a$ and $t(e_i) \notin X_a$. Note also that in this last case $ends(e_1) = \{x_0, x_1\}$ where $x_1 \notin X_b$.

Since x_1 is not in X_a we see there is a vertex $q \in V(T)$ with $q \neq a$ so that $ends(e_1) = \{x_0, x_1\} \subseteq X_q$. As a and b are joined by an edge in T we see that either the geodesic form q to b in T contains the vertex a or the geodesic from q to a contains the vertex b.

If the geodesic from q to b contains the vertex a then all paths in Γ from a vertex in X_q to a vertex in X_b must intersect X_a. Then the path $e_2 e_3 \dots e_n$ visits X_a so our original path could not have been minimal length. Therefore the geodesic from q to b cannot contain the vertex a.

If the geodesic in T from q to a contains b then the constant path at x_0 (which is a path from the vertices of X_a to the vertices of X_q) must visit X_b, hence $x_0 \in X_b$ and we contradict $n > 0$.

Thus, in all cases assuming $n > 0$ lead to a contradiction and so $X_a \cap X_b \neq \varnothing$.

□

22 L22: Improving Tree decompositions, and, Observations on Cliques.

In fact, the defining properties of being a tree decomposition do not guarantee us that tree decompositions are easy to use.

Lemma 52. *Let $k \in \mathbb{N}$ and suppose $\Gamma = (V(\Gamma), E(\Gamma), s, t)$ is a locally finite graph having tree decomposition of bag size k. Then there is a tree decomposition $T = (V(T), E(T), ends)$ of bag-size k satisfying the following:*

1. *Each vertex $v \in V$ occurs in at most finitely many bags,*
2. *for all $t \in V(T)$, we have $X_t \neq \varnothing$,*
3. *the tree T is locally finite, and*
4. *whenever $X_p \subseteq X_q$ we have $p = q$.*

Proof. Let $k \in \mathbb{N}$ and suppose $\Gamma = (V(\Gamma), E(\Gamma), s, t)$ is a locally finite graph having tree decomposition T_0 of bag size k. We will steadily improve T_0 to T_1, T_2, … new tree decompositions for Γ each with bag-sae k until ultimately T_n will have all four of the desired properties.

(1): For each vertex $u \in V(\Gamma)$ set t_u to be a vertex in T_0 so that $u \in X_{t_u}$ for the tree decomposition T_0. Similarly, let $\mathcal{P} := \{endse \mid e \in E(\Gamma)\}$ be the set which has as elements the sets of ends of the edges in Γ (each such element is a set with either one or two vertices in it), and choose, for all $s \in \mathcal{P}$, a vertex $t_s \in V(T_0)$ with $s \subseteq X_{t_s}$. Now also define, for $u \in V(\Gamma)$ the set $S_u := \{e \in E(\Gamma) \mid u \in ends(e)\}$ and let $\mathcal{T}_u$ be the minimal tree in T_0 which contains the vertex u and each of the vertices b so that $b = t_s$ for $s \in endse$ for some $e \in S_u$. Since Γ is locally finite S_u is finite so $\mathcal{T}_u$ is a finite tree. Now, we define T_1 to be the tree decomposition which arises when, for each pair (u, X_a) for a vertex a in T_0 and $u \in X_a$, we remove the vertex u from X_a if a is not a vertex in the tree $\mathcal{T}_u$ which results (after carrying out this process for all $u \in X_a$) in a new bag we will denote $\overline{X}_a$. It is immediate that u will only now appear in finitely many bags $\overline{X}_b$.

We argue that T_1 is still a tree-decomposition for Γ with maximum bag-size at most k. Firstly, all bags either stay the same size or become smaller, so the new tree decomposition (if it is one) will still have maximum bag-size at most k. Suppose now that $e \in E(\Gamma)$, and consider the set $ends(e) = s \in \mathcal{P}$ and the vertex t_s. The bag X_{t_s} in T_0 contains $endse$. Since $u \in endse$ we see that $u \in \overline{X}_{t_s}$ since the vertex t_s will be in the finite subtree $\mathcal{T}_u$ of T_0.

Now we need to argue Let us suppose that p and q are vertices of T_1 and $u \in V(\Gamma) \in X_p \cap X_q$ and where $r \in V(T)$ is in the geodesic in T from p to q. We showed in Lemma 1 that any path from a vertex in X_p to a vertex in X_q must pass through a vertex in X_r. But the constant path at u is such a path, and it never leaves u, so $u \in X_r$.

For the second property, we recall that any two neighbouring vertices in the tree T_1 have bags which must share a vertex. Therefore the set of non-empty bags forms a subtree of T_1 so we can discard all vertices with empty bags and produce a connected subtree T_2 of T_1 which is still a tree decomposition for Γ.

The tree T_2 is locally finite since for any vertex a of T_2, a can only have finitely many outgoing edges (since each vertex in X_a appears in only finitely many bags, and $|X_a|$ is finite.

For the final property, we will need to improve the tree decomposition twice.

Firstly we pass to T_3 by improving the nesting properties of bags. Choose a root vertex r of T_3, and suppose there are bags $X \subseteq Y$ with $X \neq Y$. There is a nearest vertex s' to r in T so that $X_{s'} \subsetneq X_{t'}$ for some other vertex t'. Since all the vertices of $X_{s'}$ are contained in X_p for all vertices p on the geodesic from s' to t', there is a first such edge e in T so that the ends of e are s and t in T, with $X_s = X_{s'}$ and $X_s \subsetneq X_t$. We now replace X_s by X_t, and repeat this process inductively. As the vertices of $X_{s'}$ only appear in finitely many bags, we observe that there is no infinite chain of nested local improvements of this type. In the limit we have a tree T_4 which has that if p and q are vertices with $X_p \subseteq X_q$ then $X_p = X_q$.

We improve T_4 to T_5 by contracting subtrees of T with all bags identical to points. It is easy to see the result satisfies all of our original criteria. $\square$

Given a graph $\Gamma = (V(\Gamma), E(\Gamma), r, s)$ and a set $X \subseteq V(\Gamma)$ we set $\mathcal{N}(X)$*, the neighbourhood of* X *in* Γ, to be the set of all vertices p which are either in X or for which there is an undirected edge connecting p to a vertex in X. We inductively define $\mathcal{N}^0(X) := X$ and for all $k \in \mathbb{N}$ we set $\mathcal{N}^k(X) := \mathcal{N}(\mathcal{N}^{k-1}(X))$. The set $\mathcal{N}^k(X)$ is called the k^{th} neighbourhood of X.

The following two lemmas are easy enough to be taken as exercises for threader. Proofs can be found in Diekert and Weiß.

Lemma 53. *Let* $T = (V(T), E(T), ends)$ *be a tree decomposition of a graph* $\Gamma = (V(\Gamma), E(\Gamma), s, t)$*. If we replace each bag* X *by its neighbourhood* $\mathcal{N}(X)$*, we still have a tree decomposition of* Γ*.*

Recall that a *clique* in a graph Γ is an (undirected) subgraph H so that all pairs of distinct vertices in H admit a unique edge in H connecting them. A clique is maximal if it is not contained in any other clique.

Lemma 54. *Let* T *be a tree decomposition of a locally finite graph* $\Gamma = (V(\Gamma), E(\Gamma), s, t)$*. Then every clique of* Γ *is contained in some bag* X*.*

Exercise 73. *Prove Lemmas 53 and 54.*

Exercise 74. *Find a locally finite graph* $\Gamma = (V(\Gamma), E(\Gamma), s, t)$ *which admits a tree decomposition with all bags finite but where treewidth of* Γ *is infinite.*

Exercise 75. *Let* $a, b \in \mathbb{N}_1$ *and suppose* S_a *and* S_b *are the symmetric groups on* a *and* b *letters respectively. Take as generating sets* X_a *and* X_b *of* S_a *and* S_b *the sets of all nontrivial elements in these groups respectively. Consider the Cayley graph* $\Gamma = \Gamma(S_a * S_b, X_a \cup X_b)$*, and argue that* Γ *is a locally finite graph with a tree-decomposition* T *of* Γ *with treewidth* $max(a!, b!)$*.*

Exercise 76. *Use the previous exercise to show that if* G *and* H *are two finite groups, then there is a generating set* X *for* $G * H$ *so that* $\Gamma(G * H, X)$ *is a locally finite graph with finite tree-width.*

23 L23: Finite tree width and virtual freeness.

We showed earlier that Virtually free groups have CF word problem. In this section we indicate why finitely generated groups with CF word problem are virtually free.

The chain of theory involved is the following set of theorems. We will prove the first and last, and loosely indicate why the second is true.

Theorem 55. *Suppose $G = \langle X \mid R\rangle$ is a finitely generated group whose word problem $WP(G, X)$ is a CF language. Then G has a locally finite Cayley graph $\Gamma(G, X)$ with finite tree width.*

Theorem 56. *Suppose $G = \langle X \mid R\rangle$ is a finitely generated group whose Cayley graph $\Gamma(G, X)$ is a locally finite graph with finite tree width, then G splits over a graph of groups with finite vertex and edge groups.*

Theorem 57. *Suppose $\mathscr{G} = (\mathbb{G}, \Gamma)$ is a graph of groups with Γ connected and with all vertex groups finite, and with T is a maximal tree in Γ, then $\pi_1(\mathscr{G}, T)$ is a virtually free group.*

Theorem 55. Let G be finitely generated by generating set X and suppose that $WP(G, X)$ is a CF language. It is immediate that $\Gamma(G, X)$ is locally finite by the definition of $\Gamma(G, X)$. We will show that the Cayley graph $\Gamma(G, X)$ has finite treewidth.

If G is finite, then immediately the Cayley graph is finite so it certainly has finite treewidth. We thus assume G is infinite.

For the following proof we need a definition. For a graph Θ and a subset $\mathcal{C} \subseteq V(\Theta)$, we define the *vertex boundary of* $\mathcal{C}$ as $\beta\mathcal{C} := \{u, v \mid \exists e \in E(\Theta) s.t. \{u, v\} = ends(e), u \in \mathcal{C}, v \in \overline{\mathcal{C}}\}$.

Denote by B_n the ball of radius n about the origin vertex in $\Gamma(G, X)$. For $n \in \mathbb{N}$ define sets V_n so that $V_0 = V(\Gamma) - 1$ and $V_n = \{\mathcal{C} \subseteq V(\Gamma) \mid \mathcal{C} \text{ connected component of } \Gamma - B_n\}$ for $n \geq 1$.

The above definition defines a tree T with root B_1 as follows:

$$V(T) = \{\beta\mathcal{C} \mid \mathcal{C} \in V_n, n \in \mathbb{N}\} \cup \{B_1 = \beta V_0\}$$

$$E(T) = \{\{\beta\mathcal{C}, \beta\mathcal{D}\} \mid \mathcal{D} \subseteq \mathcal{C} \in V_n, \mathcal{D} \in V_{n+1}, n \in \mathbb{N}\} \cup \{\{\beta V_0, \beta\mathcal{D}\} \mid \mathcal{D} \in V_1\}.$$

Below, given any vertex $v \in V(T)$ we will refer to the subset of $V(\Gamma(G, X))$ corresponding to v as a bag.

Note we have had to fudge the structure slightly at the root, but βD for $D \in V_1$ will have vertices from B_1, so our result is connected.

Furthermore, every vertex of $\Gamma(G, X)$ is in the boundary set of some component of a V_n as each vertex admits a geodesic path to the origin (so there is always at least one edge incident on a vertex to a group element which is closer to the origin), thus we have the first criterion of being a tree decomposition.

Secondly, for any edge e in $\Gamma(G, X)$, e must connect to vertices which are either the same distance from the origin, or one of the vertices is distance 1 closer to the origin than the other. In the second case, it is immediate that the ends of such an edge are from two distinct balls (one in say B_n while the other is in $B_{n+1} \backslash B_n$, thus both vertices will be in the same bag. In the first case, if e connects g to h and $d_\Gamma(1, g) = d_\Gamma(1, h) = n$, then there are edges e_g and e_h which have ends $\{g', g\}$ and $\{h', h\}$ respectively where $g', h' \in B_{n-1}$ so we see g and h are in the same component $\mathcal{C}$ of $\Gamma(G, X)\backslash B_{n-1}$ (since they are connected by the edge e) but that both g and h are vertices in $\beta\mathcal{C}$ (since $\{g', g, h', h\} \subseteq \beta\mathcal{C}$). Thus in both cases, the ends of e are in $\beta\mathcal{C}$ for some $\mathcal{C}$.

Finally, coherence follows immediately from the definition of the graph, since by construction a vertex of $\Gamma(G, X)$ is in at most two bags, and in this case the two bags are connected by an edge in T.

We now argue that there is a fixed integer $k \geq 1$ so that $d_\Gamma(g, h) \leq 3k$ for any g and h in the same bag. Note that if $g, h \in B_1$ then $d_\Gamma(g, h) \leq 2 < 3k$ and we are done, so below we will assume g and h are in some bag $\beta\mathcal{C}$ for some component $\mathcal{C}$ of some V_n for $n > 0$.

In fact, this follows from the fact that G is a CF group. Let $\mathscr{G} = (V, P, \Sigma, S)$ be a CF grammar for $WP(G, X)$ (so $\Sigma = X \cup X^{-1}$) which we assume is in Chomsky normal form. (Recall that Chomsky normal form means that (1) the grammar consists of productions of the form $A \mapsto BC$ for $A, B, C \in V$, productions of the form $A \mapsto w$ for $A \in V$ and $w \in \Sigma*$, and, for any production of the form $B \Longrightarrow \varepsilon$ we have $B = S$, which production can only appear when ε is in the language generated by the grammar, (2) if $A \Longrightarrow BC$ is a production then neither B nor C can be the variable S, and (3), if $A \in V$, then there is a chain of productions from S so that eventually the variable A can appear.)

For each variable $A \in V$, set w_A to be a minimal length word in Σ^* so that $A \Longrightarrow^* w_A$. Set

$$k := max\{|w_A| \mid A \in V\}.$$

The value k is the length of the shortest terminal word generated by the variable $A \in V$ where A is the variable with the longest value for the length of its shortest word, amongst all variables in V.

Before proceeding with the proof, we make the small sub claim that if A is a variable in V and $A \Longrightarrow^* w_1$ and also $A \Longrightarrow^* w_2$ for two words w_1 and w_2 in $(X \cup X^{-1})*$, then $w_1 \equiv_G w_2$. This follows as there is a chain of productions $S \Longrightarrow^* r'At'$ for r' and t' words in $V \cup \Sigma$. Now we can resolve the variables in r' and t' to terminals r and t respective so that any terminal word in the language generated by the expression rAt is equivalent to 1 in G. But then $rw_1t = 1 = rw_2t$ so $w_1 = w_2$.

We are now ready to continue the main line of our proof.

Now suppose that g and h are in the same bag which is βC for some connected component C in $\Gamma(G, X)\backslash B_n$ for some n. Let μ and ω be geodesic paths from 1 to g and from h to 1 in $\Gamma(G, X)$ respectively, and let ν be a path in C from g to h. Suppose that the labels of the paths μ, ν and ω are the words u, v and w respectively.

Let $S \Longrightarrow^* V_1V_2 \ldots V_k$ be a sub-path in a chain of productions of the word uvw so that each V_i is a non-terminal variable, and so that for each V_i, the next evaluation is to produce a (non-empty) word v_i in Σ^*. If there is a maximal index i and minimal index j so that $v_i \ldots v_j$ corresponds to a sub-word of uvw which contains the word v. But, any chain of three or more variables can be replaced by a shorter chain of variables (by undoing a production of the form $A \Longrightarrow BC$) so that we can express uvw as the resulting word of a chain of productions from $V_1 \ldots V_{i-1}ABV_{j+1} \ldots V_k$, where we will eventually evaluate the variables A and B to produce words v' and v'' respectively with $v'v''$ a sub-word of uvw which contains the word v.

Now, this means we can factor uvw as $u'v'v''w'$ where $|u'| \le |u| \le |u'v'| < |uv| \le |u'v'v''|$.

Note that if BC is not the result of a single production $D \Longrightarrow BC$ then there are finite chains of merges; to B with variables to its left (e.g., currently V_{i-1}) and to C with variables to its right, so that the resulting new B and C admit a variable A so that $A \Longrightarrow BC$ and so that all of the analysis above holds after re-labelling the corresponding words u', v', v'', and w'.

By our sub-claim, we know that $uvw = u'v'v''w' =_G u'w_Bw_Cw' =_G u'w_Aw'$. In particular there are nodes x, y, and z in μ, ν and ω respectively so that $x =_G u'$, $y =_G u'v'$ and $z =_G u'v'v'' =_G w'^{-1}$, and we see that in $\Gamma(G, x)$ we have $d_\Gamma x, z \le |w_A| \le k$, $d_\Gamma(x, y) \le |w_B| \le k$, and $d_\Gamma(y, z) \le |w_C| \le k$.

But, as x is on a geodesic from 1 to g we have

$$d_\Gamma(g, x) = d_\Gamma(1, g) - d_\Gamma(1, x) \le d_\Gamma(1, y) - d_\Gamma(1, x) \le d_\Gamma(x, y)$$

where the first inequality holds since $d_\Gamma(1, g) \le n + 1 \le d_\Gamma(1, y)$. Similarly

$$d_\Gamma(h, z) = d_\Gamma(h, 1) - d_\Gamma(z, 1) \le d_\Gamma(1, y) - d_\Gamma(z, 1) \le d_\Gamma(x, y).$$

And so

$$d_\Gamma(g, h) \le d_\Gamma(g, x) + d_\Gamma(x, z) + d_\Gamma z, h$$

$$d_\Gamma(g, h) \le d_\Gamma(y, x) + d_\Gamma(x, z) + d_\Gamma(z, y) \le 3k.$$

$\square$

Sketch Theorem 56: This is the theorem we do not have time to do properly in this course (to obtain a full proof of the Muller-Schupp result).

If we have a locally finite Cayley graph with finite tree width we want to see that our group is the fundamental group of a finite graph of groups (with finite vertex groups). The core technical idea is to modify the group presentation and the bags chosen so that eventually you can partition the generators so that each group generates a finite group, and so that the bags are effectively covering cosets of these finite groups in the Cayley graph. These processes are not easy, but well within the capacity of anyone who has read these notes. In the literature, it was Dunwoody's accessibility theory that opened the doorway for this work. One can follow Diekert and Weiß for details.

Sketch Theorem 57: The proof of this theorem will be included here. It is pretty short, and can be found on ops 29–30 of Diekert and Weiß.

References

[1] Hyman Bass, *Covering theory for graphs of groups*, J. Pure Appl. Algebra **89** (1993), no. 1-2, 3–47. MR 1239551

[2] Hyman Bass and Jean-Pierre Serre, *Arbres, amalgames,* SL_2, Société Mathématique de France, Paris, 1977, Avec un sommaire anglais, Rédigé avec la collaboration de Hyman Bass, Astérisque, No. 46. MR 0476875

[3] Gilbert Baumslag, *Topics in combinatorial group theory*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1993. MR 1243634 (94j:20034)

[4] Oleg Bogopolski, *Introduction to group theory*, EMS Textbooks in Mathematics, European Mathematical Society (EMS), Zürich, 2008, Translated, revised and expanded from the 2002 Russian original. MR 2396717 (2009c:20001)

[5] Ian Chiswell, *A course in formal languages, automata and groups*, Universitext, Springer-Verlag London, Ltd., London, 2009. MR 2468045 (2009k:68001)

[6] V Diekert and A Weiß, *Context-free groups and Bass-Serre theory*, submitted (2013), 1–61.

[7] Christiane Frougny, Jacques Sakarovitch, and Paul Schupp, *Finiteness conditions on subgroups and formal language theory*, Proc. London Math. Soc. (3) **58** (1989), no. 1, 74–88. MR 969548 (90a:20068)

[8] John E. Hopcroft and Jeffery D. Ullman, *Introduction to automata theory, languages and computation*, ch. 5 *Push Down Automata*, pp. 114–115, Addison-Wesley Publishing Company, Reading, Massachusetts, 1979.

[9] David E. Muller and Paul E. Schupp, *Context-free languages, groups, the theory of ends, second-order logic, tiling problems, cellular automata, and vector addition systems*, Bull. Amer. Math. Soc. (N.S.) **4** (1981), no. 3, 331–334. MR 609044 (82m:03051)

[10] ———, *Groups, the theory of ends, and context-free languages*, J. Comput. System Sci. **26** (1983), no. 3, 295–310. MR 710250 (84k:20016)

[11] Jean-Pierre Serre, *Trees*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2003, Translated from the French original by John Stillwell, Corrected 2nd printing of the 1980 English translation. MR 1954121

A More on normal forms for graphs of groups.

We will now translate the normal forms of graphs of groups into normal forms for the two basic special cases: free products with amalgamation and HNN extensions.

A.1 Normal forms for free products with amalgamation

As we have alluded before, a free product of two groups with amalgamation of some specific isomorphic subgroups admits a splitting over a connected graph of groups with two vertices and two edges, where the two vertex groups are the main groups of the free product and where the amalgamation group is the edge group for the two edges (an edge and its involution). Thus, we can interpret the normal form lemma for fundamental groups of graphs of groups in the context of just such a splitting.

In this case the maximal tree contains an edge joining the two vertex groups, so this edge and its involution are both trivial in π_1. Thus our reduced grouped forms have no edges in them, and if i is an index so that the edge e_i goes from vertex v_i to v_{i+1} with $g_i \in G_{v_i}$ and $g_{i+1} \in G_{v_{i+1}}$ then if $g_{i+1} \in Image(\alpha_{\overline{e_1}})$ then the product $g_i g_{i+1}$ admits a vertex conglomeration reduction, contradicting our assumption that the grouped product was already reduced. Thus, in the reduced grouped product, not only are there no edges appearing in the grouped terms, but also, each g_i (excepting possibly g_1) is not in the image of the appropriate boundary monomorphism. Thus the normal form theorem for free products with amalgamation is as follows:

Theorem 58 (Normal form for free products with amalgamation)**.** *Let $G_1 = \langle X_1 \mid R_1 \rangle$ and $G_2 = \langle X_2 \mid R_2 \rangle$ be two groups. Suppose $K = \langle X_k \mid R_K \rangle$ be a third group so that there are injective group homomorphisms $\iota_1 : K \to G_1$ and $\iota_2 : K \to G_2$.*

Let $g \in G_1 \underset{K}{*} G_2$ be written in grouped form as $g = g_1 g_2 \dots g_n$ for some integer $n > 0$. Then if

1. *for all indices $i < n$ we have*
   - *if $g_i \in G_1$ then $g_{i+1} \in G_2$,*
   - *if $g_i \in G_2$ then $g_{i+1} \in G_1$,*

   and further we have
2. *for all indices $i > 1$ we have $g_i \notin Image(\iota_1) \cup Image(\iota_2)$,*

then g is not trivial in $G_1 \underset{K}{*} G_2$.

We call grouped forms satisfying the hypothesis above *reduced forms for the free product with amalgamation* or simply "reduced forms" when the context is clear.

In this exercise, you should assume that the normal form theorem for free products with amalgamation is true.

Exercise 77. *Suppose $G = G_1 \underset{K}{*} G_2$ is a free product with amalgamation where the groups G_1, G_2, and K, and the amalgamation maps ι_1 and ι_2 are given as above.*

1. *Show that any group element $g \in G$ can be written in such a reduced form.*
2. *Suppose that $g \in G$ is represented by two reduced products $g = g_1 g_2 \dots g_m$ and $g = h_1 h_2 \dots h_n$. Show that $m = n$ and for all indices i, we have that there is $k \in \{1, 2\}$ so that $g_i = h_i \in G_k$.*

Exercise 78. *Consider H, the free product with amalgamation of C_4 and C_6 amalgamated on their respective subgroups of order two.*

1. *Show there is only one way to do this.*
2. *Give a presentation H.*
3. *Describe a normal form for elements of H in some detail. What choices do you need to make to do this?*

A.2 Normal forms for HNN extensions

Similarly, an HNN extension admits a splitting over a graph of groups with one vertex and two edges, and again we can interpret the normal form lemma for fundamental groups of graphs of groups in this context.

Theorem 59 (Normal form for HNN extensions)**.** *Let $G = \langle X \mid R\rangle$ be a group and let $K = \langle Y \rangle$ be a subgroup of G. Let $\phi : K \to G$ be an injective group homomorphism, and present the free product with amalgamation as*

$$G*_\phi = \langle X \cup \{t\} \mid R \cup \{t^{-1}yt = y\phi \mid \forall y \in y\}\rangle.$$

Let $g \in G*_\phi$ be written in grouped form as $g = g_1g_2 \dots g_n$ for some integer $n > 0$. Then if for all indices $i < n$ we have

- *if $g_i \in G$ then $g_{i+1} \in \{t, t^{-1}\}$,*
- *if $g_i = t$ then $g_{i+1} = t$ or $g_{i+1} \in G\backslash K\phi$, and*
- *if $g_i = t^{-1}$ then $g_{i+1} = t^{-1}$ or $g_{i+1} \in G\backslash K$.*

then g is not trivial in $G*_\phi$.

We call grouped forms satisfying the hypotheses above *reduced forms for the HNN extension* or simply "reduced forms" when the context is clear. Note that it is fairly traditional to use the letter t to represent the new conjugation generator; one says "$G*_\phi$ is an HNN extension with stable letter t."

In this exercise, you should assume that the normal form theorem for free products with amalgamation is true.

Exercise 79. *Suppose $G = G*_\phi$ is an HNN extension with stable letter t and amalgamated group K (so $\phi : K \to G$ is an injective group homomorphism.*

1. *Show that any group element $g \in G$ can be written in such a reduced form.*
2. *Suppose that $g \in G$ is represented by two reduced products $g = g_1g_2 \dots g_m$ and $g = h_1h_2 \dots h_n$. Show that $m = n$.*

Exercise 80. *Consider H, the HNN extension of C_6 by stable letter t which conjugates the subgroup of order two to itself.*

1. *Show there is only one way to do this.*
2. *Give a presentation H.*
3. *Describe a normal form for elements of H in some detail. What choices do you need to make to do this?*

A.3 A final refinement for Normal Forms: Britton Reductions, and Splittings Revisited)

Here we give a final refinement for the normal form theorem for HNN extensions and we further explore splittings of groups over graphs of groups.

A.3.1 HNN extensions: "Play it again, Sam!"

Let X be a finite set and $R \subseteq W(X)$, and consider the group $G = \langle X \mid R \rangle$, with subgroups A and B and a group isomorphism $\phi : A \to B$. Recall then that we can form the HNN extension $G*_\phi$ (with stable letter t) as the group

$$G*_\phi := \langle X \cup \{t\} \mid R \cup \{t^{-1}at = \phi(a) \mid a \in A\} \rangle.$$

Now we are going to make some assumptions. In particular, we are going to assume that we can detect when an expression in $W(X)$ actually resolves to an element of A or of B (this is called the membership problem for the subgroups A and B), and also that G has an explicit normal form for its elements.

Now, for the subgroups A and B, choose traversals $T_A = \{1, u_1, u_2, \ldots\}$ and $T_B = \{1, w_1, w_2, \ldots\}$ of the cosets of A and of B respectively. Now, given any element g of G we can write $g = a_g \cdot u_g$ and $g = b_g \cdot w_g$, where $a_g \in A$ and $u_g \in T_a$ and$b_g \in G$ with $w_g \in T_B$ respectively, in a unique fashion (using the normal form of G to get unique expressions for a_g and b_g inside A and B respectively). We can now express a complete normal form for elements of G.

Theorem 60 (Normal form for HNN extensions and Britton's Lemma)**.** *Let X be a finite set and $R \subseteq W(X)$, and consider the group $G = \langle X \mid R \rangle$, with subgroups A and B and a group isomorphism $\phi : A \to B$, and consider the group*

$$G*_\phi := \langle X \cup \{t\} \mid R \cup \{t^{-1}at = \phi(a) \mid a \in A\} \rangle$$

which is the HNN extension of G with stable letter t and with amalgamated subgroups A and B.

Choose traversals $T_A = \{1_G, u_1, u_2, \ldots\}$ and $T_B = \{1_G, w_1, w_2, \ldots\}$ of the cosets of A and of B respectively, and assume we have a normal form for elements in G.

Now given $g \in G$, there is $n \in \mathbb{N}$ so we can write g in a unique fashion as

$$g = g_0 t^{\epsilon_1} g_1 t^{\epsilon_2} g_2 \ldots t^{\epsilon_n} g_n$$

where

1. *g_0 is an arbitrary element of G,*
2. *each ϵ_i is either 1 or -1, and for each index i, if $\epsilon_i = 1$ then $g_i \in T_B$ while if $\epsilon_i = -1$ then $g_i \in T_A$,*
3. *and where there is no index i so that $t^{\epsilon_i} g_i t^{\epsilon_{i+1}} = t^{-1} 1_G t^1$ or $t^{\epsilon_i} g_i t^{\epsilon_{i+1}} = t^1 1_G t^{-1}$.*

Furthermore (Britton's Lemma), any such expression for g satisfying the three points above represents the identity element if and only if $n = 0$ and $g_0 = 1_G$.

Some quick points. Firstly, the key idea is to use the Serre relations $t^{-1}at = \phi(a) \in B$ to replace expressions of the form $t^{-1}a$ with $\phi(a)t^{-1}$ for $a \in A$ and expressions of the form tb with $\phi^{-1}(b)t$ for $b \in B$. Secondly, the final expression is unique, since we assume the g_0 term is expressed in the unique normal form for G. Thirdly, we observe trivially that both G and $\langle t \rangle \cong \mathbb{Z}$ embed as subgroups in $G*_\phi$. Fourthly, any element of $G*_\phi$ CAN be written this way simply by grouping terms, and then resolving each of the G_i parts as an element of a coset of A or B respectively as required. Finally, uniqueness follows from Britton's Lemma (imagine you had two such expressions for g in hand, now multiply one versus the inverse of the other; this product can only resolve to the identity if the initial expressions are identical).

Example 6. *Consider $G = F_{\{a,b\}}$ the free group with basis $\{a, b\}$. Consider the subgroups A and B where $A = \langle a^2 \rangle \cong \mathbb{Z} \cong \langle b^3 \rangle = B$, and the isomorphism $\phi : A \to B$ defined by $\phi(a^2) = b^3$. Firstly we have*

$$G*_\phi = \langle a, b, t \mid t^{-1}a^2 t = b^3 \rangle.$$

A choice of transversal for A is the set of all reduced words which either begin with no a's or just one a, that is, $T_A = \{1, v, av \mid v \in W_{red}(\{a, b\}), v = b^{\pm 1}x\}$, while a choice of transversal for B consists of all reduced words that begin with no b's, one b or two b's, that is, $T_B = \{1, v, bv, b^2 v \mid v \in W_{red}(\{a, b\}), v = a^{\pm 1}x\}$.

To practice, use Britton reductions to resolve the expression $x = b^2 t^{-1} a^{-4} t b^5 a b t^{-1} a^4 b^3 a$ to its normal form $bab^7 t^{-1} b^3 a$. (This example is due to Bogopolski in his text "Introduction to group theory.")

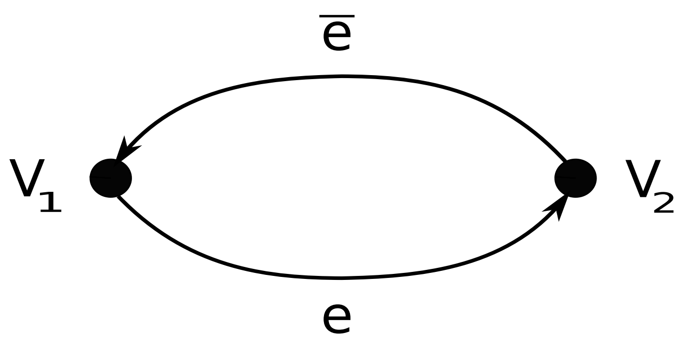
A.3.2 Splittings revisited for one-relator groups

Here we point out a little trick for spotting when a one-relator group might split as a free product with amalgamation. The trick generalises for more complex situations, but we will not worry about this here. We demonstrate the idea via an example.

Consider the group
$$H = \langle a, b, c, d \mid [a, b][c, d]\rangle.$$
That is, the one relator group on generators a, b, c, d with relator $a^{-1}b^{-1}abc^{-1}d^{-1}cd$. This relator has interesting properties. In particular, it can be split so as to show the equality of two words which are given in distinct generators. That is:
$$a^{-1}b^{-1}ab = d^{-1}c^{-1}dc$$

Now the group H is almost free: its ONLY relator is the splittable relator, and this relator involves all of the generators of the group. In particular, if $X \subseteq \{a, b, c, d\}$ with $|X| < 4$ then $\langle X\rangle$ must be free. That is, $\langle a, b\rangle \cong F_{a,b}$ and $\langle c, d\rangle \cong F_{c,d}$ as well.

Now the word $[a, b] = a^{-1}b^{-1}ab$ is non-trivial in $F_{\{a,b\}}$ and generates a copy of $\mathbb{Z}$ in this subgroup $F_{\{a,b\}}$ of H. Likewise, $[d, c] = d^{-1}c^{-1}dc$ also generates a copy of $\mathbb{Z}$ in $F_{\{c,d\}}$. Now, it is immediate that H splits as the free product of amalgamation of $F_{\{a,b\}}$ with $F_{\{c,d\}}$ amalgamated on these copies of $\mathbb{Z}$. That is, set $\iota_1 : \mathbb{Z} \to F_{\{a,b\}}$ by $+1 \mapsto [a, b]$ and $\iota_2 : \mathbb{Z} \to F_{\{c,d\}}$ by $+1 \mapsto [d, c]$, and consider the graph of groups in the diagram below:



With vertices v_1 and v_2 and edges e (connecting v_1 to v_2) and $\overline{e}$ (from v_2 to v_1), vertex groups $G_{v_1} = F_{\{a,b\}}$ and $G_{v_2} = F_{\{c,d\}}$ and edge groups $G_e = G_{\overline{e}} = \mathbb{Z}$ and with boundary maps $\alpha_e : G_e \to G_{v_1}$ given as $\alpha_e = \iota_1$ and $\alpha_{oveinee} : \mathbb{Z} \to G_{v_2}$ given as $\alpha_{\overline{e}} = \iota_2$. Now computing the presentation of the fundamental group of this graph of groups (say, with maximal tree $T = (\{v_1, v_2\}, \{\overline{e}\})$ gives a presentation that Tietze reduces to our original presentation of H.

The key point to observe is that it is the ability to split the relator into two pieces that clearly belong to different subgroups of H that allows this trick to see the original group as a free product with amalgamation.

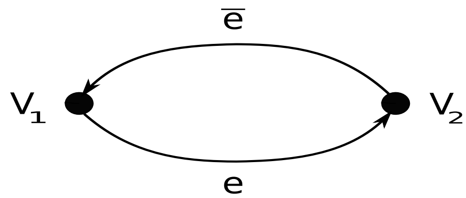
B Actions on trees

In these next subsections we explore aspects of how free products with amalgamation and HNN extensions, respectively, act on trees. From this we can make some conclusions about how graphs of groups act on trees. We are essentially following the presentation in Bogopolski's text.

NB: in this and the next subsections, all actions are left actions, and all cosets are left cosets. That is, if $H \leq G$ are groups, then we will denote the set of left cosets of H in G as G/H, defining this (even when H is not normal) as
$$G/H := \{gH \mid g \in G\}.$$

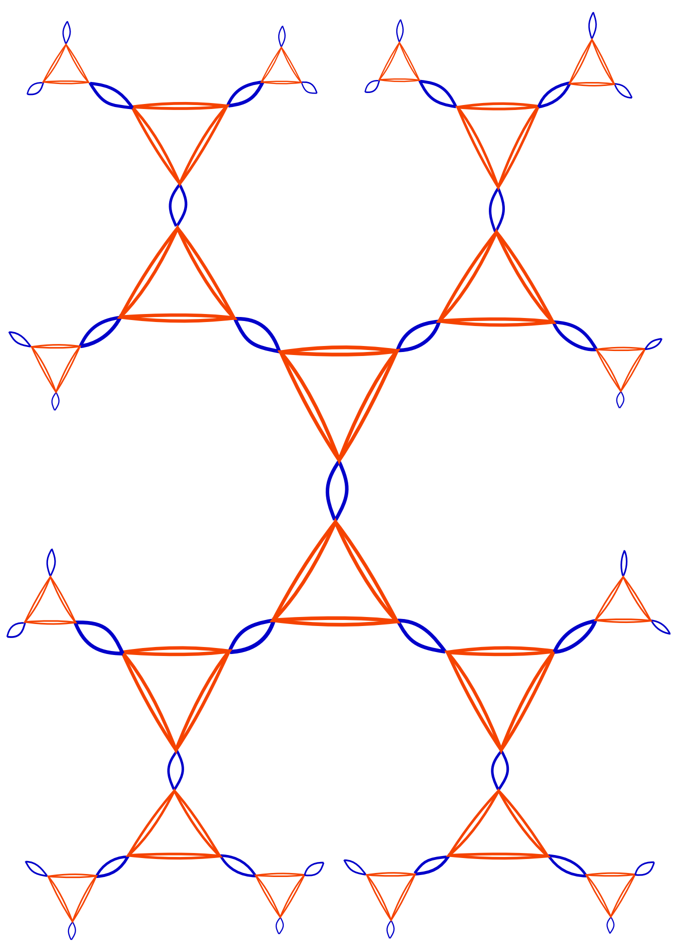
B.1 Segments and Graphs

Note that a connected graph, consisting of two vertices and two mutually inverse edges (under the Serre involution), is called a *segment*. As before, this graph is as follows.



For ease drawing, we henceforth will just draw a segment as the two vertices connected by a single edge (which might or might not be obviously directed, depending on the scale of the drawing), unless we explicitly say we are drawing all of the edges. One should imagine that for each subgraph which is a segment, we have chosen a favoured edge to draw, and ignored drawing the other, so the resulting drawn graph is a subgraph which still characterises the structure of the actual graph. Thus, the segment above would typically just be drawn as

Now the cover of our text is what we get when we take the graph below (which shows ALL edges e and $\overline{e}$) and we follow the convention above, where we imagine we are standing far enough away that we cannot see the orientation on the edges actually drawn.



B.2 The Bass-Serre tree for Free products with amalgamation

Theorem 61. *Let $G = G_1 *_A G_2$, a free product with amalgamation of G_1 with G_2 over subgroup A. Then there exists a tree T, on which G acts without inversion of edges such that the factor graph $G\backslash T$ is a segment. Moreover, this segment can be lifted to a segment X in T with the property that the stabilisers in G of the vertices and edges of X are equal to G_1, G_2 and A, respectively.*

First we recall one of the language above. Recall that we call a graph in Serre's sense a tree if, when we replace each segment by a single edge, the resulting graph is a tree in the traditional sense. Then, a group G acts on a graph Γ without edge inversion means that for all edges e of Γ, we have that $g \cdot e \neq \overline{e}$ for all $g \in G$. If G acts on a graph without edge inversion, then we can form a new graph called the factor graph or the quotient graph, where the edges of the new graph are the equivalence classes of the edges in the old graph, and the vertices of the factor graph are the equivalence classes of the vertices of the old graph.

Proof. Set the vertices of T to be

$$T^0 := G/G_1 \cup G/G_2,$$

the set of cosets of G_1 and of G_2 in G. Set the set T^1_+ of "positive edges" to be the set of cosets G/A (each coset is an edge). An edge $e_g = gA$ has $s_e = gG_1$ and $t(e) = gG_2$. The segment X is the subgraph with vertices $G_1 = 1G_1$, $G_2 = 1G_2$ and $A = 1A$. Note we assume the formal inverse edge $\bar(A)$ connecting the vertex G_2 to the vertex G_1 is also an edge, a "negative edge" in the set T^1_-.

It is immediate that G acts on T via multiplication on the left.

First we prove that T is connected. This is fairly easy to do. Firstly, the vertices G_1 and G_2 are connected by the edge e_1. Now consider any coset gG_2 for some element $g \in G$. Writing g in its normal form we have $g = g_1 g_2 \dots g_n$. Now, if g_n is in G_2, then we see $gG_2 = g_1 g_2 \dots g_{n-1} G_2$. We observe that the edge $g_1 g_2 \dots g_{n-1} A$ from $g_1 g_2 \dots g_{n-1} G_1 = g_1 \dots g_{n-2} G_1$ (since $g_{n-1} \in G_1$) connects to our vertex $g_1 g_2 \dots g_{n-1} G_2 = g_1 g_2 \dots g_n G_2$. Similarly, if $g_n \notin G_2$ then the edge $g_1 g_2 \dots g_n A$ connects $g_1 g_2 \dots g_n G_1 = g_1 g_2 \dots g_{n-1} G_1$ to $g_1 g_2 \dots g_n G_2$. So in both cases, we can find an edge to gG_2 to a vertex $g' G_1$ where g' has a shorter grouped form than g. Similarly, any vertex gG_1 admits an outgoing edge to a vertex $g' G_2$ where g' has a shorter grouped form than g. In particular, for any vertex gG_i we can find a path from $1G_1$ to the vertex gG_i. Hence, T is connected.

Now we show that T is a tree (in the sense of Serre). Since we know that T is connected, we just need to show that if $p = e_1 e_2 \dots e_n$ is a path in T with $s(p) = t(p)$ then there must be some index i so that e_i and e_{i+1} have $\overline{e_i} = e_{i+1}$. From the argument in the previous paragraph, we know that each time we travel over an edge, we are actually either shortening or lengthening the grouped form for the coset representative of the vertex by length exactly one (note, this is independent of choice of coset representative!). since the length of the coset representative at the start and finish of our path are the same, we must have that at some point we increase the length by one, and then in the next step, decrease the length by one, OR, at some step we decrease the length by one and then immediately increase the length by one. In either of these cases, we must in fact traverse an edge gA and its formal inverse edge in some order for some specific edge gA.

We now verify the conditions on the stabilisers of the parts of X. Note that this actually tells us the vertex and edge stabilisers for all vertices and exes of T using the conjugation trick for stabilisers (if p is in a set Y acted on by group H, and H_p is the stabiliser of p, then H_p^g is the stabiliser of H_{gp} (phrased as left actions)). Firstly, it is immediate that the stabiliser of the vertex G_1 is precisely the group G_1 (since G_1 is closed under left multiplication by elements of G_1, and the only elements of G_2 which stabilise G_1 are precisely those elements in A, which are already in G_1), while the stabiliser of G_2 is the group G_2, and the stabiliser of the edge A must stabilise both G_1 and G_2 and in particular is $G_1 \cap G_2 = A$. (So, for the segment $gG_1 - gA - gG_2$, we have that the conjugates G_1^g, A^g and G_2^g stabilise the respective parts.) $\square$

Note the valency of the vertex gG_1 is precisely the index of A in G_1, since the edge gg_1A connects gG_1 to gg_1G_2.

Exercise 81. *Verify that if gA and hA are distinct cosets of A in G_1, then the vertices gG_2 and hG_2 are distinct.*

Exercise 82. *Determine the valency of the vertices gG_2 in the tree.*

Theorem 62. *Let the group G act without inversions of edges on a tree $T = (T^0, T^1)$ and suppose that the factor graph $G\backslash T$ is a segment X. Let $\tilde{X}$ be an arbitrary lift of X in T. Denote the vertices of $\tilde{X}$ by P and Q and the edge by e, and let G_P and G_Q be the stabilisers of P and Q respectively, and with $A = G_P \cap G_Q$ being the stabiliser of $\tilde{X}$. Then the homomorphism $G_P *_A G_Q \to G$ which is the identity on G_p and G_Q respectively is an isomorphism.*

Proof. Firstly, let us prove that $G = \langle G_P, G_Q\rangle$. Write $H = \langle G_P, G_Q\rangle$ and suppose $H \neq G$. Then the graphs $H \cdot \tilde{X}$ and $(G - H) \cdot \tilde{H}$ are disjoint. This follows as the identity $hP = gQ$ is impossible for any $h \in H$, $g \in (G - H)$, simply because G does not act to send P to Q. Likewise, $hQ = gP$ is impossible for $h \in H$, $g \in (G - H)$. Finally, suppose $R \in \{P, Q\}$, the identity $gR = hR$ is also impossible for $h \in H$, $g \in (G - H)$ since then $g^{-1}h \in G_R$, which implies that $g^{-1}h \in H$, which can only happen if $g^{-1} \in H$. But clearly, T is connected, and $T = H\tilde{X} \cup (G - H)\tilde{X} = G\tilde{X}$. Thus it must be the case that $(G - H)$ is empty.

We now argue that the homomorphism above is injective. To do this, we simply need to argue that any expression $g_n g_{n-1} \dots g_1$ which is a grouped product in normal form for $G_P *_A G_Q$ we has this product is cannot be trivial in G (note that each g_i is in $G_P - A$ or $G_Q - A$ and that the terms alternate as to which set they are in if $n > 1$).

We argue two cases: $n > 1$ and $n = 1$. Firstly let us argue $n > 1$.

Suppose $n > 1$ and $g_1 \in G_P - A$, then using the distance in the tree (each edge is length 1) we have $d(P, g_1Q) = d(g_1P, g_1Q) = d(P, Q) = 1$, while since $g_1Q \neq Q$ we see that $d(g_1Q, Q) = 2$. That is, elements in $G_P - A$ act as a rotation of the tree about the vertex P. Similarly, elements of $G_Q - A$ act as rotations of the tree about the vertex Q. Now by induction it is easy to see that $d(Q, g_n g_{n-1} \dots g_1 Q)$ is n for even n and $n + 1$ for odd n. Likewise, if $g_1 \in G_Q - A$, then $d(P, g_n g_{n-1} \dots g_1 P)$ is n if n is even, and $n + 1$ if n is odd. In particular, we have that all such grouped products of length 2 or greater move some point (P or Q) a distance at least two.

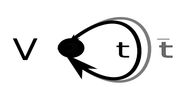
Now suppose $n = 1$. In this case, our homomorphism is already restricted to be injective over G_1 and G_2, so we must have $g_1 = 1_G$.

Thus in both cases, triviality of the product expression in G enforced triviality in the free product with amalgamation, so our homomorphism is injective. □

Corollary 63. *The free group on two generators acts on a tree without edge inversion so that the factor graph is a segment. This factor segment then provides a framework for a graph of groups, using the trivial group as the edge group, and $\mathbb{Z}$ for both vertex groups, which in turn expresses π_1 of this graph of groups as the free product with amalgamation of two copies of $\mathbb{Z}$ amalgamated over the trivial group.*

B.3 The Bass-Serre tree for HNN Extensions

As we have discussed before, the graph of groups representing an HNN extension is a single vertex group G and two edges $\{t, \bar{t}$ where the generator t conjugates a subgroup A to another copy B sitting in the vertex group; the edges both connecting the single vertex to itself. We call such a graph (consisting of a single vertex and two mutually inverse edges) a *loop*.



This loop will play, below, the rôle of the segment of the previous subsection.

Theorem 64. *Let $G = \langle H, t \mid t^{-1}at = \phi(a), a \in A\rangle$ be an HNN extension $H*_\phi$ of the group H with associated subgroups A and $B = \phi(A)$. Then there exists a tree T on which G acts without inversion of edges such that the factor graph $G\backslash T$ is a loop. Moreover, there is a segment $\tilde{Y}$ in T such that the stabilisers of its vertices and the edges in the group G are equal to H, tHt^{-1}, and A respectively.*

Proof. Let H, t, A, ϕ and G be as in the statement of the hypotheses.

Set $T^0 = G/H$ and $T^1_+ = G/A$ (the positive edges) similarly to Theorem 61, noting that all of these cosets are left cosets. Set $s(gA) = gH$ and $t(gA) = gtH$. Let $\tilde{Y}$ be the segment in T with vertices H and tH an connecting positive edge $1A = A$. Define the left action of G on the graph T as left multiplication on the cosets, observing that this IS an action on the graph. The rest of the proof is similar to that of Theorem 61, and we leave it to the reader (note that here, $G_1 = G_2 = G$, and the action on the vertices is actually transitive, which is why the factor graph is not a segment, but instead a loop). ☐

Theorem 65. *Let G be a group acting without edge inversion on a tree T and let the factor graph be $Y = G\backslash T$ be a loop. Let $\tilde{Y}$ be an arbitrary segment in T, and let P, Q, e and $\overline{e}$ be the vertices and edges of this segment. Finally, let G_P, G_Q, and $G_e = G_{\overline{e}}$ be the stabilisers of these edges and vertices in the group G. Let $x \in G$ be an arbitrary element such that $Q = xP$, and put $G'_e = x^{-1}G_e x$ and set $\phi : G_e \to G'_e$ to be the isomorphism induced by conjugation by x on the subgroup G_e of G. Then*

1. *$G'_e \leq G_P$ and*
2. *the homormorphism $\langle G_P, t \mid t^{-1}at = \phi(a), a \in G_e\rangle \to G$ which is the identity on G_P and which sends t to x is an isomorphism.*

Proof. This proof is like the proof of Theorem 62. ☐

B.4 HNN extensions, Free Products with Amalgamation, and classical products

So how are HNN extensions, and free products with amalgamation related?

Let $G*_\phi$ be an HNN extension, presented as

$$G*_\phi = \langle G, t \mid tat^{-1} = \phi(a) \forall a \in A\rangle$$

We can consider the subgroup $N = \langle t^i G t^{-1} \mid i \in \mathbb{Z}\rangle$. This subgroup is an infinitely iterated free product with amalgamation, and it is the kernel of a map from $G*_\phi$ to $\langle t\rangle \cong \mathbb{Z}$. The subgroup $\langle t\rangle$ acts on the kernel via conjugation, and we obtain

$$G*_\phi = N \rtimes \langle t\rangle.$$

A semi-direct product.

B.5 Actions and splittings and all that noise.

So we understand some things. Let $\mathscr{G} = (\mathbb{G}, \Gamma)$ be a graph of groups with underlying graph $\Gamma = (V, E, s, t)$.

For simplicity's sake, let us suppose the graph of groups is connected and has maximal tree T.

If we pass to the graph of groups one obtains by restricting attention to the vertices in the vertices in T, and add in the Serre-involute edges, we obtain a (likely simpler) graph of groups $\mathscr{G}_T$. Starting with some vertex group from T, we can inductively build up the fundamental group of $\mathscr{G}_T$ by performing iterated free products with amalgamation. Each new gluing can be seen as amalgamating to the current result on one side (which we can think of as a single (large) vertex group) the new vertex group to be added (gluing along the common subgroup).

What remains from the original graph Γ are the edges (and their involutes) which are not in the tree. Some of these will be loops from vertices of Γ to themselves, but some of them will connect differing vertices of Γ. In any case, if we tree the resulting group H_T from the previous paragraph as a vertex group for a new graph of groups (so, perform a contraction of the tree T to a single vertex), then all of the remaining edges of Γ simple will go from the vertex H_T to itself: they now appear as a collection of loops.

Inductively we can grow this vertex group by performing each HNN extension for these loops, one at a time. Each time we do this, the vertex group will grow, and the number of loops will decrease. We stop when our graph of groups is simply a vertex.

This also provides a way to build an action of the fundamental group of our graph of groups on a tree. At each stage, we build out the tree following the recipes of the induced actions on trees from the individual free products with amalgamation (we build a bi-coloured tree with vertices representing the two vertex groups, and with edge counts corresponding to the cardinality of the cosets of the included subgroups) and from the HNN extensions (we build the 'infinite line' tree).

When one extends from the current action on a tree, the new tree is built according to whether we are considering a free product with amalgamation or an HNN extension. For free products with amalagamation, each vertex stabiliser of a vertex of the induced tree corresponds to one of the vertex groups of the free product with amalgamation. We can therefore replace each such vertex with the tree that that group acts on, and get a much bigger resulting tree that the whole group acts on. And similarly for the HNN extensions.

C 2-Complexes, covers and subgroups

We won't have time for this in detail, but I wanted to leave it for you, if you are curious.

C.1 2-complex maps and covering maps.

Definition 21. *Let X and Y be 2-complexes. A map $p : X \to Y$ is a map of 2-complexes if it*

1. *assigns to each vertex of X a vertex of Y,*
2. *assigns to each edge of X an edge or vertex of Y,*
3. *assigns to each face of X a face, edge, or vertex of Y, and*
4. *and where all of this is carried out in an incidence preserving manner. (If a higher dimensional cell f maps into a lower dimensional skeleton, then $\partial(f)$ as a set of cells is sent into the image $p(f)$ as a set of cells.)*

Definition 22. *A map $p : X \to Y$ of 2-complexes X and Y is a covering map if*

1. *p preserves dimension,*
2. *if $\tilde{v}p = v$ for some $v \in V(Y)$, then p is a bijection from the set of edges $\tilde{e}$ with $s(\tilde{e}) = \tilde{V}$ to the set of edges e with $s(e) = v$, and*
3. *if f is a face and v is a vertex of Y, let $m(f, v)$ be the number of times that v appears in the boundary of f. Then for any $\tilde{v}$ with $\tilde{v}p = v$,*
$$\Sigma_{\tilde{f}\mapsto f} m(\tilde{f}, \tilde{v}) = m(f, v)$$
the sum being over all faces $\tilde{f}$ covering f.

Definition 23. *A covering map $p : X \to Y$ of two complexes X and Y is regular if whenever w is a loop at a base point v of K, then the lifts of w at each vertex in the fibre of v are either* all loops *or* all non-loops.

Definition 24. *A* deck transformation *for a covering map $p : X \to Y$ is 2-complex automorphism $d : X \to X$ so that $dp = p$ as maps. (That is, if x is a face, edge, or vertex of X, then $xdp = xp = y \in Y$.)*

C.2 results that come quickly...

Theorem 66. *Suppose $p : E \to B$ is a covering map of connected 2-complexes. Then, $p_* : \pi_1(E) \to \pi_1(B)$ is an injection.*

Theorem 67. *Suppose $p : E \to B$ is a covering map of connected 2-complexes. Then $Im(p_*)$ is a normal subgroup if and only if p is a regular cover.*

Theorem 68. *Suppose $p : E \to B$ is a covering map 2-complexes and B is connected. Then p is a k-to-1 map where k is the index of $Im(p_*)$ in $\pi_1(B)$.*

Theorem 69. *Let Γ be a connected graph seen as a 2-complex. Then $\pi_1(\Gamma)$ is a free group.*

Theorem 70. *Let K_1 and K_2 be 2-complexes, each with a single vertex v_i, and $\phi : \pi_1(K_1, v_1) \to \pi_1(K_2, v_2)$ a homomorphism of groups. Then there is a subdivision $\overline{K}_1$ of K_1 and a 2-complex map $p : \overline{K}_1 \to K_2$ such that*

$$
\begin{tikzcd}
\pi_1(K_1, v_1) \arrow[rr, "\phi"] \arrow[dr, "\cong"'] & & \pi_1(K_2, v_2) \\
& \pi_1(\overline{K}_1, v_1) \arrow[ur, "p_*"'] &
\end{tikzcd}
$$

commutes.

Theorem 71. *Any subgroup of a free group is a free group.*

Theorem 72. *If G is a free group of rank n and H is a subgroup of index k, then $rank(H) = kn-(k-1) = k(n-1)+1$.*

Proof. Realise the presentation complex of G as an n-leafed rose R. Let C be the connected cover of R corresponding to the subgroup H with covering map $p : C \to R$ (so that the image of $\pi_1(C)$ in $\pi_1(R)$ is index k).

C now must have k vertices,, and kn edges, and any maximal tree T in C will have $k-1$ edges in it. Hence, the generators of $\pi_1(C)$ come from the $kn-(k-1)$ edges not in the tree, and there are no faces to make relators. Thus we have

$$rank(H) = kn - (k-1) = k(n-1) + 1.$$

□

Above, we did the following exercise for free groups. Do it for some group with relators.

Exercise 83. *Work a detailed example using covering spaces theory to present a finite index subgroup of a finitely presented group. Examples might include $C_2 \leq S_3$, $A_4 \leq S_4$, or the group*

D Review of certain products.

Review of external direct products
Recall the following. Let H, K be groups. Construct the set

$$H \times K = \{(h, k) \mid h \in H, k \in K\}$$

and define a binary product $\cdot : H \times K \to H \times K$ by the rules

$$(h_1, k_1) \cdot (h_2, k_2) = (h_1h_2, k_1k_2), \forall h_1, h_2 \in H, k_1, k_2 \in K.$$

Then $(H \times K, \cdot)$ forms a group, which we call the *external direct product of H and K.* (Informally, we usually just write $H \times K$ for the external direct product of groups H and K.) **Review of internal direct products**
Let G be a group with subgroups H and K. Suppose the following:

1. $H \cap K = \{1_G\}$,
2. $G = HK$, and
3. $kh = hk$, for all $h \in H$, $k \in K$.

then we say G *is the internal direct product of its subgroups* H *and* K, and we recall that

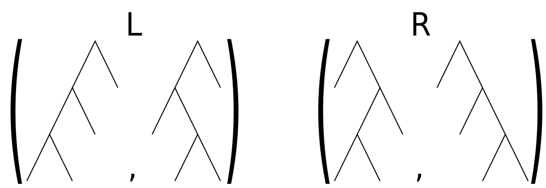
$$G \cong H \times K.$$

We note in passing that in this situation, both H and K are normal subgroups in G. **Internal semi-direct products**
Suppose now that G is a group and N and K are subgroups of G so that

1. $N \cap K = \{1_G\}$,
2. $G = NK$, and
3. for all $n_1 \in N$ and $k \in K$, there is $n_2 \in N$ so that $kn_1 = n_2k$.

In this case we say G *is the internal semi-direct product of* N *and* K, and we note that N is normal in G, but K might not be. **Examples of semi-direct products of subgroups**
Here are some examples:

1. $G = S_n$, $N = A_n$, $K = \langle (1\,2) \rangle$
2. $G = D_{10}$, $N = C_{10}$, $K = \langle r \rangle$, where r is a reflection.
3. $G = F$, the R. Thompson group. N the subgroup of elements which act as identity near 0 and 1, $K = \langle L, R \rangle$, with L and R given below:



Left action notation
We usually prefer right actions. Sadly, almost always, semi-direct products are treated as above, with shape NK. Multiplying elements looks like this:

$$g_1 \cdot g_2 = (n_1k_1) \cdot (n_2k_2) = n_1k_1n_2k_2 = n_1\underline{k_1n_2k_1{}^{-1}}k_1k_2 = n_1n_2^{k_1^{-1}}k_1k_2 = g_3$$

so, we effectively acted on n_2 by conjugation via k_1^{-1}. This is a standard trick, though. To turn a right action into a left action...
Big trick: left actions, from the right.
Let us assume we have a right action, can we make a left action?

$$X \curvearrowleft G$$

$$x \circ (gh) = ((x \circ g) \circ h)$$

induces...

$$((st)) \cdot x = (s \cdot (t \cdot x)) = (s \cdot (x \circ t^{-1})) = ((x \circ t^{-1}) \circ s^{-1})$$

$$= x \circ (t^{-1}s^{-1}) = x \circ ((st)^{-1})$$

So to change the sides of an action, act from the other side, with inverses!

External semi-direct products

Let N, K be groups. Construct the set

$$N \times K = \{(n,k) \mid n \in N, k \in K\}$$

and suppose you have an action $\phi : K \to Aut(N)$. Define a binary product $\cdot : N \times K \to N \times K$ by the rules

$$(n_1,k_1)\cdot(n_2,k_2) = (n_1(\phi(k_1))(n_2), k_1k_2), \forall n_1,n_2 \in H, k_1,k_2 \in K.$$

Then $(H \times K, \cdot)$ forms a group, which we call the *external semi-direct product of H and K under ϕ.* We denote this group by $H \rtimes_\phi K$. (Informally, we usually just write $H \rtimes K$ once the action ϕ is understood).

Checking external semi-direct products are well defined...

$$(n_1,k_1)\cdot(n_2,k_2) = (n_1(\phi(k_1))(n_2), k_1k_2) \text{ (def.)}$$
$$(h,k)\cdot(1,1) = (h\cdot(\phi(k))(1), k\cdot 1)) = (h\cdot 1, k\cdot 1) = (h,k) \text{ right id.}$$

$$(h,k)\cdot((\phi(k^{-1}))(h^{-1}),k^{-1}) = (h\cdot((\phi(k))((\phi(k^{-1}))(h^{-1}))), k\cdot k^{-1}) = (h\cdot((\phi(k)\circ\phi(k^{-1}))(h^{-1})),1) =$$
$$(h\cdot((\phi(kk^{-1}))(h^{-1})),1) = (h\cdot((\phi(1))(h^{-1})),1) = (h\cdot h^{-1},1) = (1,1) \text{ right inverse}$$

Associativity of external semi-direct products

$(n_1,k_1)\cdot((n_2,k_2)\cdot(n_3,k_3)) = (n_1,k_1)\cdot(n_2\cdot(\phi(k_2))(n_3), k_2k_3) =$
$(n_1\cdot((\phi(k_1))(n_2\cdot(\phi(k_2))(n_3))), k_1\cdot(k_2\cdot k_3)) =$
$(n_1\cdot((\phi(k_1)(n_2))\cdot((\phi(k_1))((\phi(k_2))(n_3)))), k_1(k_2k_3)) =$
$((n_1\cdot(\phi(k_1)(n_2)))\cdot((\phi(k_1))((\phi(k_2))(n_3))), (k_1k_2)k_3) =$
$((n_1\cdot(\phi(k_1)(n_2)))\cdot((\phi(k_1k_2))(n_3)), (k_1k_2)k_3)$
$((n_1,k_1)\cdot(n_2,k_2))\cdot(n_3,k_3) = ((n_1\cdot((\phi(k_1))(n_2))), k_1k_2)\cdot(n_3,k_3) =$
$((n_1\cdot((\phi(k_1))(n_2)))\cdot((\phi(k_1k_2)(n_3))), (k_1\cdot k_2)\cdot(k_3)) =$
$((n_1\cdot(\phi(k_1)(n_2)))\cdot((\phi(k_1k_2))(n_3)), (k_1k_2)k_3)$

The isomorphism

Theorem 73. *Let G be a group, and let N, K be subgroups of G so that G is the internal semi-direct product of N with K. Then*

$$G \cong N \rtimes_\phi K,$$

where $\phi : K \to Aut(N)$ is given by the rule $k \mapsto \theta_k \in Aut(N)$ where, for all $n \in N$, $\theta_k(n) = knk^{-1}$.

Examples

$$S_3 \cong C_3 \rtimes_\phi C_2$$

Here, the generator of C_2 acts by inverting elements of C_3, and (more theoretically), Let $N = \langle X_N \mid R_N\rangle$, and $K = \langle X_K \mid R_K\rangle$, with $W(X_N) \cap W(X_K) = \varnothing$, and let $\phi : K \to Aut(N)$ be a group homomorphism. Then

$$N \rtimes_\phi K = \langle X_N \cup X_K \mid R_N \cup R_K, knk^{-1} = (\phi(k))(n) \forall k \in X_K, n \in X_n\rangle$$

Exercise 84. *Let $G_1 = \langle X_1 \mid R_1\rangle$, $G_2 = \langle X_2 \mid R_2\rangle$ and $K = \langle X_K \mid R_K\rangle$, and imagine that you have injective group homomorphisms $\iota_1 : K \to G_1$, $\iota_2 : K \to G_2$ and $\tau : K \to G_1$. Set $A = image(\iota_1)$ and $B = image(\tau)$. Finally, suppose there is a group homomorphism $\phi : G_2 \to Aut(G_1)$. Describe, for each presentation below, the group you obtain.*

1. $\langle X_1 \cup X_2 \mid R_1 \cup R_2\rangle$,

2. $\langle X_1 \cup X_2 \mid R_1 \cup R_2 \cup \{[x_1, x_2] \mid x_1 \in X_1, x_2 \in X_2\}\rangle$,
3. $\langle X_1 \cup X_2 \mid R_1 \cup R_2 \cup \{x_2^{-1}x_1x_2 \cdot ((\phi(x_2))(x_1))^{-1} \mid x_1 \in X_1, x_2 \in X_2\}\rangle$,
4. $\langle X_1 \cup X_2 \mid R_1 \cup R_2 \cup \{\iota_1(k) \cdot (\iota_2(k))^{-1} \mid k \in X_k\}\rangle$, *and*
5. $\langle X_1 \cup \{t\} \mid R_1 \cup \{t^{-1}(\iota_1(k))t \cdot (\tau(k))^{-1} \mid k \in X_k\}\rangle$.

E Introduction to Category Theory

Casey Blacker

E.1 Introduction

This appendix follows and elaborates on Chapter 1, *Category Language*, of John R. Stallings' notes on geometric group theory as well as incorporating some material from *Categories for the Working Mathematician*, by Saunders Mac Lane. Both are excellent resources and the interested reader looking for a more comprehensive overview will find much additional material in either.

Much like the study of sets and relations, category theory can be used to form the groundwork of virtually any field of mathematics. In contrast to set theory, which begins with a formalisation sets and then proceeds to establish what is meant by the relations between them, category theory starts immediately with the notion of a relation (or *morphism*) and then conceives of the objects between which such relations hold as suitably-defined identity relations. This alternative perspective enables us to "see past the surface" and observe the deeper structure within, thus facilitating work between seemingly-disparate fields. Indeed, one of the motivations behind the early development of the subject was to forge a bridge between certain branches of topology and subfields of algebra.

In this paper, we begin with an overview of the basic terminology, proceed with a number of examples, explore the notion of duality and finish with an investigation of limits and their special cases.

E.2 Preliminary Material

We begin with a definition.

Definition 25 (Category)**.** *A* category $\mathcal{C}$ *is a* 5*-tuple* (O, M, s, t, c)*, where* O *and* M *are classes with* $O \subseteq M$*, and* $s : M \to O$ *and* $t : M \to O$ *are functions. By means of this, define a set* $D \subseteq M \times M$*:* $D = \{(\alpha, \beta) : t(\alpha) = s(\beta)\}$*. Then,* $c : D \to M$ *is a function. These are to satisfy the rules:*

$\forall \alpha, \beta, \gamma \in M,$

(1) $s(c(\alpha, \beta)) = s(\alpha)$

(2) $t(c(\alpha, \beta)) = t(\beta)$

(3) If $t(\alpha) = s(\beta)$ *and* $t(\beta) = s(\gamma)$*, then* $c(c(\alpha, \beta), \gamma) = c(\alpha, c(\beta, \gamma))$

(4) $\forall \alpha \in O, s(\alpha) = t(\alpha) = \alpha$

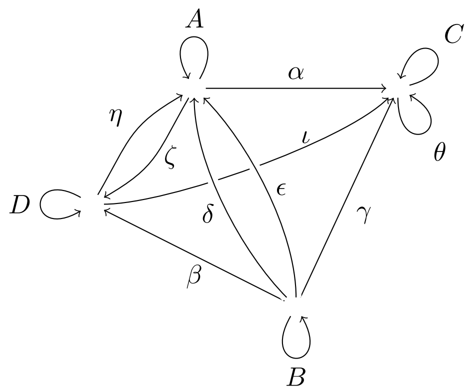
(5) $\forall \alpha \in M, c(s(\alpha), \alpha) = c(\alpha, t(\alpha)) = \alpha$

The elements of M *are called* $\mathcal{C}$-maps *(or* arrows *or* morphisms*). The elements of* O *are* $\mathcal{C}$-objects *or* identity maps*. The elements* $s(\alpha)$ *and* $t(\alpha)$ *are the* source *and* target*, respectively, of* α*. The set* D *describes the set of pairs of maps whose* composition $c(\alpha, \beta)$ *is defined. If* $A = s(\alpha)$ *and* $B = t(\alpha)$ *then we write*

$$\alpha : A \to B$$

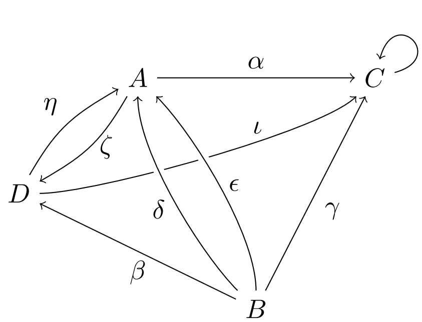
Definition 26 (Small Category)**.** *If in the category* $\mathcal{C} = (O, M, s, t, c)$ *the classes* O *and* M *are sets, then we call* $\mathcal{C}$ *a* small category*.*

Intuitively, we can think of a category $\mathcal{C}$ as a collection of "objects" $A, B, C, \ldots$ together with a collection of "arrows" $\alpha, \beta, \gamma, \ldots$ between them. Formally, we consider even the objects $A, B, C, \ldots$ themselves to be $\mathcal{C}$-maps, playing the role of the identity. To illustrate, we might consider:



the category $\mathcal{C}$

Though conceptually sound, this is usually depicted as



Here, the collection of composable pairs $D \subseteq M \times M$ contains precisely those pairs of arrows (x, y) for which the head of x matches up with the tail of y. Thus (β, η) and (A, α) are composable pairs, but (θ, η) and (δ, ϵ) are not. We compose two arrows x and x by joining them together. Hence $c(A, \alpha)$ is equal to α and $c(\beta, \eta)$ is one of either δ or ϵ.

With this framework in place, it easy to see the foundation for the axioms above. The reasoning behind (1) and (2) is apparent in from the following diagram:

$$\begin{tikzcd}
& B \arrow[dr, "\beta"] & \\
s\big(c(\alpha,\beta)\big) = s(\alpha) = A \arrow[ur, "\alpha"] \arrow[rr, "c(\alpha,\beta)"'] & & C = t(\beta) = t\big(c(\alpha,\beta)\big)
\end{tikzcd}$$

Axiom (3) is an associative law, stating that the matter of composition of arrows is irrelevant.

$$\begin{tikzcd}
& B \arrow[r, "\beta"] \arrow[drr, dashed, "c(\beta,\gamma)"'] & C \arrow[dr, dashed, "\gamma"] & \\
A \arrow[ur, dashed, "\alpha"] \arrow[urr, dashed, "c(\alpha,\beta)"'] \arrow[rrr, "c\big(c(\alpha,\beta),\gamma\big) = c\big(\alpha, c(\beta,\gamma)\big)"'] & & & D
\end{tikzcd}$$

Axioms (4) and (5) establish the notion of objects as distinguished arrows. Axiom (4) requires that every object is its own source and target, which implies that in a certain sense each object is self-contained. This ensures that we admit forms such as

$$\begin{tikzcd}[column sep=large] {} \arrow[loop above, "A"] & {} \arrow[loop above, "B"] & {} \arrow[loop above, "C"] \end{tikzcd}$$

while forbidding meaningless expressions along the lines of

$$\begin{tikzcd} {} \arrow[r, "A"] & {} \arrow[r, bend left, "B"] & {} \arrow[l, bend left, "C"] \end{tikzcd}$$

Finally, axiom (5) fixes each object as an identity map. Pictorially, we have

$$\underset{c\big(s(\alpha),\alpha\big)}{\begin{tikzcd} {} \arrow[loop left, "A"] \arrow[r, "\alpha"] & {} \arrow[loop right, "B"] \end{tikzcd}} \quad = \quad \underset{c\big(\alpha,t(\alpha)\big)}{\begin{tikzcd} {} \arrow[loop left, "A"] \arrow[r, "\alpha"] & {} \arrow[loop right, "B"] \end{tikzcd}} \quad = \quad \underset{\alpha}{\begin{tikzcd} {} \arrow[loop left, "A"] \arrow[r, "\alpha"] & {} \arrow[loop right, "B"] \end{tikzcd}}$$

It is important to stress that though we term $\alpha, \beta, \gamma, \ldots$ "maps" they might in fact differ considerably from the notion of a function. This will be made apparent in the next section.

E.3 Examples of Categories

In order to elucidate the definition, we provide a number of examples. Each example is thoroughly worked to a considerable level of detail.

Example 7 (Discrete Categories)**.** *A* discrete category *has as its morphisms $A, B, C, \ldots \in X$, for some arbitrary set X. The defining property in a discrete category is that every morphism is an object. We formalise this as $\mathcal{C} = \{O, M, s, t, c\}$ where*

$$\begin{aligned} O &= \{X, Y, Z, \ldots\} \\ M &= \{X, Y, Z, \ldots\} \\ s(A) &= A \\ t(A) &= A \\ c(A, A) &= A \end{aligned}$$

for all $A \in O$. The composable pairs $D \subseteq M \times M$ are

$$D = \{(A, A)\}_{A \in O}$$

The axioms follow immediately.

(i) $s(c(A, A)) = s(A)$*;*
(ii) $t(c(A, A)) = t(A)$*;*
(iii) $c(c(A, A), A) = c(A, A) = c(A, c(A, A))$*;*
(iv) $s(A) = A = t(A)$*;*
(v) $c(s(A), A) = c(A, A) = c(A, t(A))$*.*

Example 8 (Preorders)**.** *Let $X = \{a, b, c, \ldots\}$ be a set equipped with a preorder $\preceq$. Let $\mathcal{C} = (O, M, s, t, c)$ be given by*

$$\begin{aligned} O &= \{(x, x)\}_{x \in X} \\ M &= \{(x, y) \,:\, x \preceq y\}_{x,y \in X} \\ s[(x, y)] &= (x, x) \\ t[(x, y)] &= (y, y) \\ c[(x, y), (y, z)] &= (x, z) \end{aligned}$$

for all $x, y, z \in X$. Here, we take the composable pairs to be

$$D = \{\big((x, y), (y, z)\big)\}_{x,y,z \in X}$$

Observe that

(1) $s\big[c[(x, y), (y, z)]\big] = s[(x, z)] = (x, x) = s[(x, y)]$*;*

(2) $t\big[c[(x, y), (y, z)]\big] = t[(x, z)] = (z, z) = t[(y, z)]$*;*

(3) we have

$$\begin{aligned} c\big[c[(w, x), (x, y)], (y, z)\big] &= c[(w, y), (y, z)] \\ &= (w, z) \\ &= c[(w, x), (x, z)] \\ &= c\big[(w, x), c[(x, y), (y, z)]\big]; \end{aligned}$$

(4) $s[(x, x)] = (x, x) = t[(x, x)]$*;*

(5) it follows that

$$\begin{aligned} c\big[s[(x, y)], (x, y)\big] &= c[(x, x), (x, y)] \\ &= (x, y) \\ &= c[(x, y), (y, y)] \\ &= c\big[(x, y), t[(x, y)]\big] \end{aligned}$$

Example 9 (Relations)**.** *Let* Set *be the class of all sets and let $\rho_{A,B}$ be a binary relation between $A, B \in \mathrm{Set}$. Thus define $\mathcal{C} = (O, M, s, t, c)$ by*

$$\begin{aligned} O &= \{\Delta_X\}_{X \in \mathrm{Set}} \\ M &= \{\rho_{X,Y}\}_{X,Y \in \mathrm{Set}} \\ s(\rho_{A,B}) &= \Delta_A \\ t(\rho_{A,B}) &= \Delta_B \\ c(\rho_{A,B}, \sigma_{B,C}) &= (\rho \circ \sigma)_{A,C} \end{aligned}$$

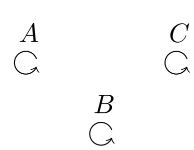


Figure E1: The discrete category with objects A, B and C.

$$
\begin{tikzcd}
 & (b,b) \arrow[dl, "{(b,c)}"'] & & \\
(c,c) & & (a,a) \arrow[ul, "{(a,b)}"'] \arrow[ll, "{(a,c)}"] \arrow[r, "{(a,d)}"'] & (d,d)
\end{tikzcd}
$$

Figure E2: The category corresponding to the preorder where $a \preceq b \preceq c$ and $a \preceq d$.

for all $A, B, C \in \mathrm{Set}$, where $\Delta_A = \{(x,x)\}_{x\in A}$ is the diagonal on A and $(\rho \circ \sigma)_{A,C} = \{(x,z) \in A \times C \mid \exists y \in B : (x,y) \in X$ and $(y,z) \in Y\}$ for relations $\rho_{A,B}$ and $\sigma_{B,C}$. The composable pairs are

$$D = \{(\rho_{A,B}, \sigma_{B,C})\}_{A,B,C\in \mathrm{Set}}$$

We have

(1) $s\big(c(\rho_{A,B}, \sigma_{B,C})\big) = s((\rho \circ \sigma)_{A,C}) = \Delta_A = s(\rho_{A,B})$;

(2) $t\big(c(\rho_{A,B}, \sigma_{B,C})\big) = t((\rho \circ \sigma)_{A,C}) = \Delta_C = t(\sigma_{B,C})$;

(3)

$$
\begin{aligned}
c\big(c(\rho_{A,B}, \sigma_{B,C}), \tau_{C,D}\big) &= c((\rho \circ \sigma)_{A,C}, \tau_{C,D}) \\
&= (\rho \circ \sigma \circ \tau)_{A,D} \\
&= c(\rho_{A,B}, (\sigma \circ \tau)_{B,D}) \\
&= c\big(\rho_{A,B}, c(\sigma_{B,C}, \tau_{C,D})\big);
\end{aligned}
$$

(4) $s(\Delta_A) = A = t(\Delta_A)$;

(5)

$$
\begin{aligned}
c\big(s(\rho_{A,B}), \rho_{A,B}\big) &= c(\Delta_A, \rho_{A,B}) \\
&= (\Delta \circ \rho)_{A,B} \\
&= (\rho \circ \Delta)_{A,B} \\
&= c(\rho_{A,B}, \Delta_B) \\
&= c\big(\rho_{A,B}, t(\rho_{A,B})\big)
\end{aligned}
$$

We elaborate on step (3). Given $X \subseteq A \times B$, $Y \subseteq B \times C$ and $Z \subseteq C \times D$, we have

$$
\begin{aligned}
(w,z) \in (\rho \circ \sigma)_{A,C} \circ \tau_{C,D} &\iff \exists y \in C : (w,y) \in (\rho \circ \sigma)_{A,C} \text{ and } (y,z) \in \tau_{C,D} \\
&\iff \exists y \in C : \Big(\exists x \in B : (w,x) \in \rho_{A,B} \text{ and } (x,y) \in \sigma_{B,C}\Big) \text{ and } (y,z) \in \tau_{C,D} \\
&\iff \exists x \in B : \exists y \in C : (w,x) \in \rho_{A,B} \text{ and } (x,y) \in \sigma_{B,C} \text{ and } (y,z) \in \tau_{C,D} \\
&\iff \exists x \in B : (w,x) \in \rho_{A,B} \text{ and } \Big(\exists y \in C : (x,y) \in \sigma_{B,C} \text{ and } (y,z) \in \tau_{C,D}\Big) \\
&\iff \exists x \in B : (w,x) \in \rho_{C,D} \text{ and } (x,z) \in (\sigma \circ \tau)_{A,C} \\
&\iff (w,z) \in \rho_{A,B} \circ (\sigma \circ \tau)_{B,D}
\end{aligned}
$$

Thus $(\rho_{A,B} \circ \sigma_{B,C}) \circ \tau_{C,D} = \rho_{A,B} \circ (\sigma_{B,C} \circ \tau_{C,D})$ and so expressions of the form $(\rho \circ \sigma \circ \tau)_{A,D}$ are well-defined.

Example 10 (Sets and Functions). *Here we consider sets and the functions between them as our objects and morphisms, respectively. Let $A, B \in \text{Set}$ and let $f_{A,B} : A \to B$ be a function. If we apply our functions on the right (i.e. $x \mapsto xf$) then we define $\mathcal{C} = (O, M, s, t, c)$ where*

$$\begin{aligned} O &= \{\mathrm{id}_A\}_{A \in \text{Set}} \\ M &= \{f : A \to B\}_{A,B \in \text{Set}} \\ s(f_{A,B}) &= \mathrm{id}_A \\ t(f_{A,B}) &= \mathrm{id}_B \\ c(f_{A,B}, g_{B,C}) &= (f \circ g)_{A,C} \end{aligned}$$

The composable pairs are

$$D = \{(f_{A \to B}, g_{B \to C})\}_{A,B,C \in \text{Set}}$$

where id_A denotes the identity function on A. The proof that this forms a category is exactly analogous to Example 9. In fact, the category of sets and functions is a subcategory of the category of sets and relations. We usually denote this category by **Set**.

Example 11 (Monoids). *Let S be a monoid with elements $\alpha, \beta, \gamma, \ldots$ and identity ϵ. Recall that the right regular representation $\rho : S \to T(S)$ is given by*

$$\big[\rho(\alpha)\big](\beta) = \beta\alpha$$

We will denote $\rho(\alpha)$ by $\bar{\alpha}$ for each $\alpha \in S$. Define $\mathcal{C} = (O, M, s, t, c)$ where

$$\begin{aligned} O &= \{\bar{\epsilon}\} \\ M &= \{\bar{\alpha}\}_{\alpha \in S} \\ s(\bar{x}) &= \bar{\epsilon} \\ t(\bar{x}) &= \bar{\epsilon} \\ c(\bar{x}, \bar{y}) &= \overline{xy} \end{aligned}$$

for all $x, y \in S$. For composable pairs, we have

$$D = \{(\bar{\alpha}, \bar{\beta})\}_{\alpha,\beta \in S}$$

We have

(1) $s(c(\bar{x}, \bar{y})) = \bar{\epsilon} = s(\bar{x})$*;*

(2) $t(c(\bar{x}, \bar{y})) = \bar{\epsilon} = t(\bar{x})$*;*

(3) it trivially holds that $t(\bar{x}) = s(\bar{y})$ and $t(\bar{y}) = s(\bar{z})$, moreover $c(c(\bar{x}, \bar{y}), \bar{z}) = c(\overline{xy}, \bar{z}) = \overline{xyz} = c(\bar{x}, \overline{yz}) = c(\bar{x}, c(\bar{y}, \bar{z}))$;

(4) $s(\bar{\epsilon}) = t(\bar{\epsilon}) = \bar{\epsilon}$*;*

(5) $c(s(\bar{x}), \bar{x}) = c(\bar{\epsilon}, \bar{x}) = \overline{\epsilon x} = \bar{x} = \overline{x\epsilon} = c(\bar{x}, \bar{\epsilon}) = c(\bar{x}, t(\bar{x}))$.

as required.

Example 12 (Directed Paths). *Recall that a* directed graph *is a 4-tuple $\Gamma = (V, E, s', t')$ where $V \subseteq E$ and $s, t : E \to V$ satisfy*

$$\forall v \in V \;:\; s'(v) = t'(v) = v$$

A directed path *in Γ is a finite-tuple $p = (e_1, \ldots, e_n)$ of edges in E such that $t'(e_i) = s(e_{i+1})$ for $1 \leq i \leq n-1$. Any path can be* reduced *by eliminating redundant entries, i.e. any elements of V for a path of length greater*

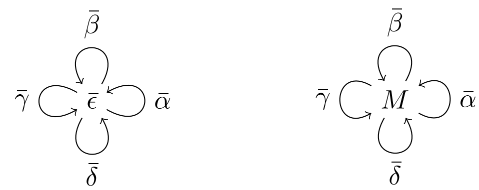


Figure E3: Two possible representations of the category corresponding to the monoid M, with elements $\alpha, \beta, \gamma, \delta$ and identity ϵ. The first reinforces the fact that every object is a morphism; the second is more standard.

than 1. *We can* concatenate *two paths p and q in the obvious way, and when we do so we will take it as understood that the result is fully reduced. Let the notation $p_{e,f}$ denote a reduced path p from $e \in V$ to $f \in V$. Note that this implies that $p_{(e,e)}$ is uniquely determined to mean (e,e), which we identify with e. Hence define $\mathcal{C} = (O, M, s, t, c)$ by*

$$\begin{aligned} O &= \{(x,x)\}_{x \in V} \\ M &= \{p_{x,y}\}_{x,y \in V} \\ s(p_{e,f}) &= (e,e) \\ t(p_{e,f}) &= (f,f) \\ c(p_{e,f}, q_{f,g}) &= pq_{e,g} \end{aligned}$$

for all $e, f, g \in V$. In this case, the composable pairs are

$$D = \{(p_{e,f}, p_{f,g})\}_{e,f,g \in V}$$

We have

(1) $s\big(c(p_{e,f}, q_{f,g})\big) = s(pq_{e,g}) = (e,e) = s(p_{e,f})$;

(2) $t\big(c(p_{e,f}, q_{f,g})\big) = t(pq_{e,g}) = (g,g) = t(q_{f,g})$;

(3) $c\big(c(p_{e,f}, q_{f,g}), r_{g,h}\big) = c(pq_{e,g}, r_{g,h}) = pqr_{e,h} = c(p_{e,f}, qr_{f,h}) = c\big(p_{e,f}, c(qr_{f,h})\big)$;

(4) $s\big((e,e)\big) = (e,e) = t\big((e,e)\big)$;

(5) $c\big(s(p_{e,f}), p_{e,f}\big) = c\big((e,e), p_{e,f}\big) = p_{e,f} = c\big(p_{e,f}, (f,f)\big) = c\big(p_{e,f}, t(p_{e,f})\big)$

Though this example admittedly lacks the formal rigour of the others, it is instructive to see a category in which the morphisms do not resemble functions.

Example 13 (Categories and Functors)**.** *Let $\mathcal{C} = (O, M, s, t, c)$ and $\mathcal{C}' = (O', M', s', t', c')$ be categories. A* functor *(or* covariant functor*) Φ is a function from M to M' – and hence, by restriction, from O to O' – such that*

$$\begin{aligned} \Phi\big(s(\alpha)\big) &= s\big(\Phi(\alpha)\big) \\ \Phi\big(t(\alpha)\big) &= t\big(\Phi(\alpha)\big) \\ \Phi\big(c(\alpha,\beta)\big) &= c\big(\Phi(\alpha),\, \Phi(\beta)\big) \end{aligned}$$

If we let Cat *be the collection of all categories, then let $\mathbf{I}_{\mathcal{D}} : M' \to M'$ be the identity on the morphisms of $\mathcal{D} = (O', M', s', t', c')$ for each $\mathcal{D} \in$* Cat*. Hence $\mathbf{I}_{\mathcal{D}}$ is a functor. We can form a category $\mathcal{C} = (O, M, s, t, c)$*

$$\begin{tikzcd} & t(\alpha)=B=s(\beta) \arrow[dr, "\beta"] & & \mathcal{C} \\ s(\alpha)=A \arrow[ur, "\alpha"] \arrow[rr, "c(\alpha,\beta)"'] & & C=t(\beta) & \end{tikzcd}
\qquad
\begin{tikzcd} & s^*(\alpha)=B=t^*(\beta) \arrow[dl, "\alpha"'] & & \mathcal{C}^* \\ t^*(\alpha)=A & & C=s^*(\beta) \arrow[ul, "\beta"'] \arrow[ll, "c^*(\alpha,\beta)"] & \end{tikzcd}$$

Figure E4: A category $\mathcal{C} = (O, M, s, t, c)$ and its dual $\mathcal{C}^* = (O, M, s^*, t^*, c^*)$.

where

$$\begin{aligned} O &= \{\mathbf{I}_{\mathcal{D},\mathcal{D}}\}_{\mathcal{D}\in\mathrm{Cat}} \\ M &= \{\Phi_{\mathcal{D},\mathcal{E}}\}_{\mathcal{D},\mathcal{E}\in\mathrm{Cat}} \\ s(\Phi_{\mathcal{D},\mathcal{E}}) &= \mathbf{I}_{\mathcal{D}} \\ t(\Phi_{\mathcal{D},\mathcal{E}}) &= \mathbf{I}_{\mathcal{E}} \\ c(\Phi_{\mathcal{D},\mathcal{E}}, \Phi_{\mathcal{E},\mathcal{F}}) &= \Phi_{\mathcal{D},\mathcal{E}} \circ \Phi_{\mathcal{E},\mathcal{F}} \end{aligned}$$

where $\Phi_{\mathcal{D},\mathcal{E}} \circ \Phi_{\mathcal{E},\mathcal{F}}$ *is standard composition of functions. For composable pairs, we obtain*

$$D = \{(\Phi_{\mathcal{C},\mathcal{D}}, \Psi_{\mathcal{D},\mathcal{E}})\}_{\mathcal{C},\mathcal{D},\mathcal{E}\in\mathrm{Cat}}$$

This category is termed **Cat**.

We can similarly show that other examples of categories are groups, rings, G-sets, Λ-modules and topological spaces – to name just a few.

E.4 Duality

Definition 27 (Dual of a Category)**.** *Informally, the dual* $\mathcal{C}^*$ *simply "reverses the arrows" of* $\mathcal{C}$*. Formally, given a category* $\mathcal{C} = (O, M, s, t, c)$*, the* dual *of* $\mathcal{C}$ *is the category* $\mathcal{C}^* = (O, M, s^*, t^*, c^*)$ *where*

$$\begin{aligned} s^* &= t \\ t^* &= s \\ c^*(\alpha, \beta) &= c(\beta, \alpha) \end{aligned}$$

The class of composable maps then becomes $D^* = \{(\beta, \alpha) \in M \times M \,:\, (\alpha, \beta) \in D\}$*, where* D *is the collection of composable maps in* $\mathcal{C}$*. Thus*

$$\alpha : s(\alpha) \to t(\alpha)$$

in $\mathcal{C}$*, becomes*

$$\alpha : t(\alpha) \to s(\alpha)$$

in $\mathcal{C}^*$*. Note that if* α *is a morphism in* $\mathcal{C}$ *then it is also a morphism in* $\mathcal{C}^*$*. However, in order to avoid confusion, we typically denote* α *by* α^* *when we wish to indicate that we are working in the context of* $\mathcal{C}^*$ *and not* $\mathcal{C}$*. Hence* $\alpha : A \to B$ *in* $\mathcal{C}$ *becomes* $\alpha^* : B^* \to A^*$ *in* $\mathcal{C}^*$*.*

This gives rise to the notion of dual concepts.

Definition 28. *Given two constructions* P *and* Q*, we say that* P *is the* dual *of* Q *if an instance of* P *in* $\mathcal{C}$ *becomes an instance of* Q *in* $\mathcal{C}^*$*, and vice versa.*

The following table lists some categorical notions and their corresponding duals.

Construction	Dual
retraction	coretraction
equivalence	equivalence
monic	epic
initial object	terminal object
cone above	cone below
limit	colimit
product	coproduct
pullback	pushout
equaliser	coequaliser

We will explore many of these concepts throughout the remainder of this paper.

Example 14 (Retraction, Coretraction and Equivalence)**.** *Let $\mathcal{C}$ be a category and $\alpha : A \to B$ be a $\mathcal{C}$-morphisms. If α has a right inverse, say $\beta : B \to A$, then α is called a* retraction*. If α has a left inverse, say $\gamma : B \to A$, then α is called a* coretraction*. Finally, if α has a both a left and right inverse, then α is called a $\mathcal{C}$-*equivalence*. We will show that the dual notion of retraction is coretraction, and that the notion of equivalence is self-dual. Thus suppose that α is a retraction and that $c(\alpha, \beta) = A$. Then $c^*(\beta^*, \alpha^*) = c(\alpha, \beta) = A$. Conversely, if α is a coretraction with $c(\gamma, \alpha) = B$, then $c^*(\alpha^*, \gamma^*) = B$. Hence a retraction in $\mathcal{C}$ becomes a coretraction in $\mathcal{C}^*$ and vice versa. We can thus conclude that an equivalence in $\mathcal{C}$ remains an equivalence in $\mathcal{C}^*$.*

Note that by definition every object A in $\mathcal{C}$ is both a monic and an epic and hence an equivalence.

Example 15 (Monic and Epic)**.** *Let $\alpha : A \to B$ be a morphism of $\mathcal{C}$. If $c(\alpha, x) = c(\alpha, y)$ implies that $x = y$, then we say that α is* monic*. Conversely, if $c(x, \alpha) = c(y, \alpha)$ always yields $x = y$, then we say that α is* epic*. Suppose that α is monic and observe that $c^*(x^*, \alpha^*) = c^*(y^*, \alpha^*)$ is equivalent to $c(\alpha, x) = c(\alpha, y)$ whence $x = y$ which the same as $x^* = y^*$. Now suppose that α is epic and note that $c^*(\alpha^*, x^*) = c^*(\alpha^*, y^*)$ entails $c(x, \alpha) = c(y, \alpha)$ from which $x = y$, that is $x^* = y^*$. Therefore monic and epic form a dual pair.*

We now provide some concrete examples.

Category	*Monic*	*Epic*
Sets and Functions	*injection*	*surjection*
Groups	*monomorphism*	*epimorphim*

Example 16 (Initial and Terminal Objects)**.** *If A is an object in $\mathcal{C}$ and for every object B in $\mathcal{C} = (O, M, s, t, c)$ there is a unique morphism $\alpha : A \to B$, then we call A an* initial object*. Conversely, if for every object B in $\mathcal{C}$ there is a unique morphism $\beta : B \to A$, then we call A a* terminal object*. Observe that*

$$\begin{aligned} A \text{ is initial in } \mathcal{C} &\iff \forall B \in O : \exists!\, \alpha \in M : s(\alpha) = A \text{ and } t(\alpha) = B \\ &\iff \forall B^* \in O : \exists!\, \alpha^* \in M : s^*(\alpha^*) = B^* \text{ and } t^*(\alpha^*) = A^* \\ &\iff A^* \text{ is terminal in } \mathcal{C}^* \end{aligned}$$

since $s^*(\alpha^*) = t(\alpha)$ and $t^*(\alpha^*) = s(\alpha)$. Thus the notions of intial and terminal objects are indeed dual.

Again, we illustrate these concepts with some concrete examples.

Category	*Initial Object*	*Terminal Object*
Groups	*trivial group*	*trivial group*
Topological spaces	*empty space* $\{\}$	*single-point space* $\{p\}$
Preorder	*minimal element*	*maximal element*
Nontrivial Monoid as Category	*none*	*none*
Trivial Monoid $\{\epsilon\}$ *as Category*	$\bar{\epsilon}$	$\bar{\epsilon}$

$$
\begin{tikzcd}
 & B & \mathcal{C} \\
A \arrow[ur] & & C \arrow[ul] \arrow[ll]
\end{tikzcd}
\qquad
\begin{tikzcd}
 & B^* \arrow[dl] \arrow[dr] & \mathcal{C}^* \\
A^* \arrow[rr] & & C^*
\end{tikzcd}
$$

Figure E5: The terminal object B in $\mathcal{C}$ becomes initial in $\mathcal{C}^*$.

$$
\begin{tikzcd}
\mathcal{C} \arrow[r, "\Phi"] & \mathcal{D} \\
A \arrow[d, "\alpha"'] & \Phi(A) \arrow[d, "\Phi(\alpha)"] \\
B & \Phi(B)
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathcal{C} \arrow[r, "\Psi"] & \mathcal{E} \\
A \arrow[d, "\alpha"'] & \Psi(A) \\
B & \Psi(B) \arrow[u, "\Psi(\alpha)"']
\end{tikzcd}
$$

Figure E6: A covariant functor $\Phi : \mathcal{C} \to \mathcal{D}$ and a contravariant functor $\Psi : \mathcal{C} \to \mathcal{E}$.

Definition 29 (Contravariant Functor)**.** *For two categories $\mathcal{C} = (O, M, s, t, c)$ and $\mathcal{C}' = (O', M', s', t', c')$, a* contravariant functor *$\Psi : M \to M'$ is a functions which satisfies*

$$
\begin{aligned}
\Psi\big(s(\alpha)\big) &= t\big(\Psi(\alpha)\big) \\
\Psi\big(t(\alpha)\big) &= s\big(\Psi(\alpha)\big) \\
\Psi\big(c(\alpha, \beta)\big) &= c\big(\Psi(\beta),\, \Psi(\alpha)\big)
\end{aligned}
$$

Example 17 (The Operator * as a Contravariant Functor)**.** *For a category $\mathcal{C} = (O, M, s, t, c)$, we can define a function $^* : M \to M^*$ which takes each morphism in $\mathcal{C}$ to its corresponding morphism in $\mathcal{C}^*$, i.e.*

$$^* : \alpha \mapsto \alpha^*$$

Then * is a contravariant functor. This is perhaps the most natural example of a such a functor.

E.5 Limits and Colimits

In order to discuss limits and colimits it is necessary to introduce some definitions. Up to now, we have been modelling our categories by means of visual diagrams. Our first task will be to make rigorous our intuitive notion of this construction.

Definition 30 (Diagram)**.** *Let $\mathcal{C} = (O, M, s, t, c)$ be a category and Γ a directed graph. A* diagram *in $\mathcal{C}$ modeled on $\Gamma = (V, E, s', t')$ is simply a functor $\mathbf{D} : \Gamma \to M$ such that*

$$\forall v \in V \; : \; \mathbf{D}(v) \in O$$

In other words, a diagram is a directed graph labeled with maps in $\mathcal{C}$.

Definition 31 (Commutativity)**.** *We will give an informal definition. We say that a diagram $\mathbf{D}$ is* commutative *if composition of arrows is "path independent". For example, the diagram*

$$
\begin{tikzcd}
A \arrow[r, "\alpha"] \arrow[d, "\gamma"'] & B \arrow[d, "\beta"] \\
C \arrow[r, "\delta"'] & D
\end{tikzcd}
$$

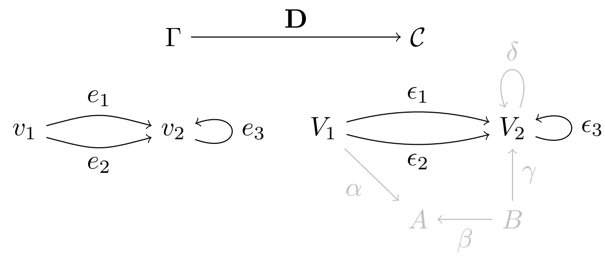


Figure E7: A diagram $\mathbf{D} : \Gamma \to \mathcal{C}$, where $\mathbf{D}(v_i) = V_i$ and $\mathbf{D}(e_i) = \epsilon_i$.

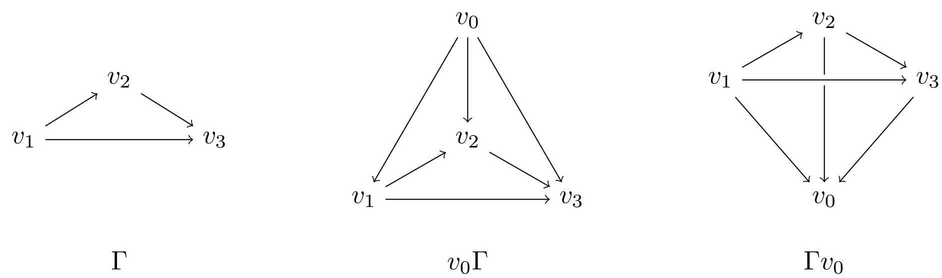
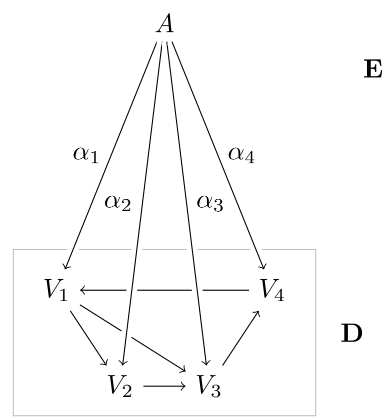


Figure E8: A cone above and and a cone below the graph Γ.

is commutative just in case $c(\alpha, \beta) = c(\gamma, \delta)$*.*

Definition 32 (Cone)**.** *Let* Γ *be a directed graph. Adjoin to* Γ *a new vertex* v_0 *and, for all vertices* w *of* Γ*, new edges* $e_w : v_0 \to w$*. This produces a new graph* $v_0\Gamma$*, the* cone above Γ*. The dual construction, with new vertex* v_1 *and new edges* $e'_w : w \to v_1$*, is written* Γv_1 *and called the* cone below Γ*.*

Definition 33 (Category Above **D** and Category Below **D**)**.** *Let* $\mathbf{D} : \Gamma \to \mathcal{C}$ *and* $\mathbf{E} : v_0\Gamma \to \mathcal{C}$ *be commutative diagrams. If* $\mathbf{E}$ *extends* $\mathbf{D}$*, then we call* $\mathbf{E}$ *and* object above $\mathbf{D}$*. Pictorially, this is*



We occasionally depict this as

$$\begin{tikzcd} A \arrow[d, "\boldsymbol{\alpha}"'] & \mathbf{E} \\ \mathbf{D} & \end{tikzcd}$$

A map above **D** *is a commutative diagram of the form:*

$$\begin{tikzcd} A \arrow[rr, "\phi"] \arrow[dr, "\boldsymbol{\alpha}"'] & & B \arrow[dl, "\boldsymbol{\beta}"] \\ & \mathbf{D} & \end{tikzcd}$$

where $\phi : A \to B$ is a single $\mathcal{C}$-map. The category above **D** *consists of all the objects and maps above* **D**. *The dual notion to this construction is the* category below **D**.

Remark 7. *In the previous definition we required that $v_0 \notin V$. Note that this* does not *imply that the associated object $A \in O$ is not in the image of* **D**. *Indeed, we might have*

$$\Gamma \xrightarrow{\mathbf{D}} \mathcal{C}$$

$$v_1 \xrightarrow[e_1]{} v_2 \qquad A \xrightarrow[\alpha]{} B$$

and

$$v_0\Gamma \xrightarrow{\mathbf{E}} \mathcal{C}$$

$$\begin{tikzcd} & v_0 \arrow[dl, "e_0"'] \arrow[dr, "e_0'"] & \\ v_1 \arrow[rr, "e_1"'] & & v_2 \end{tikzcd} \qquad A \xrightarrow[\alpha]{} B$$

where $\mathbf{E} : v_0\Gamma \to \mathcal{C}$ is the extension of $\mathbf{D} : \Gamma \to \mathcal{C}$ given by

$$\begin{aligned} v_0 &\mapsto A & e_0 &\mapsto A \\ v_1 &\mapsto A & e_0' &\mapsto \alpha \\ v_2 &\mapsto B & e_1 &\mapsto \alpha \end{aligned}$$

Definition 34 (Limit and Colimit)**.** *Let $\mathbf{D} : \Gamma \to \mathcal{C}$ be a diagram. If the diagram above* **D** *has a terminal object, then this object is called the* limit *of* **D** *and is unique up to equivalence in the category above* **D**. *Conversely, if the diagram below* **D** *has an initial object, then this is called the* colimit *of* **D** *and is also unique up to equivalence.*

We now give some special cases of limits and colimits.

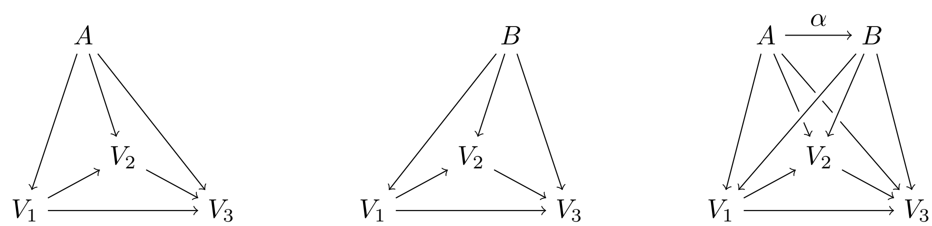
Definition 35 (Product and Coproduct)**.** *Let $\mathbf{D} : \Gamma \to \mathcal{C}$ be a diagram where Γ contains only vertices and no edges. The image of* **D** *is a discrete category with objects $\{A_v\}_{v\in\Gamma}$. Then the limit of* **D** *is called the* product *of $\{A_v\}_{v\in\Gamma}$. We usually denote this by*

$$\prod_{v\in\Gamma} A_v$$

or just $A \times B$ if the image of the diagram contains only two objects. Dually, the colimit of **D** *is called the* coproduct *of $\{A_v\}_{v\in\Gamma}$ and is typically denoted by*

$$\coprod_{v\in\Gamma} A_v$$

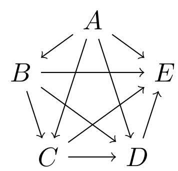
or $A \amalg B$ in the case that we are only considering two object.



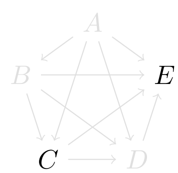
$\tilde{A}$ $\tilde{B}$ $\tilde{\alpha} : \tilde{A} \to \tilde{B}$

Figure E9: Two objects $\tilde{A}, \tilde{B}$ and a map $\tilde{\alpha} : \tilde{A} \to \tilde{B}$ above $\mathbf{D}$. By definition, each of $\tilde{A}$, $\tilde{B}$ and $\tilde{\alpha}$ is a commutative diagram. Holding $\tilde{B}$ fixed, if $\tilde{\alpha} : \tilde{A} \to \tilde{B}$ exists and is uniquely determined for every $\tilde{A}$ above $\mathbf{D}$, then $\tilde{B}$ is the limit of $\mathbf{D}$.

Example 18 (Product in a Preorder)**.** *Consider the preorder where $a \preceq b \preceq c \preceq d \preceq e$. We depict the category $\mathcal{C}$ corresponding to this preorder as follows:*

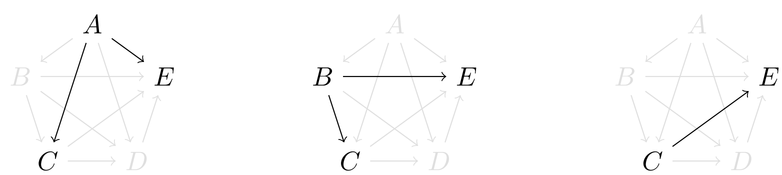


where we denote by X the pair (x, x) for each element x in our preorder. Hence the product $C \times E$ is the limit of the diagram $\mathbf{D} : \Gamma \to \mathcal{C}$, given by



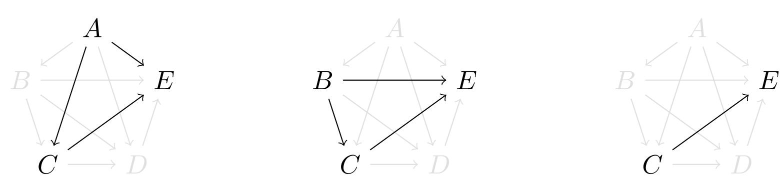
$\mathbf{D}$

The objects above $\mathbf{D}$ are the commutative diagrams



$\tilde{A}$ $\tilde{B}$ $\tilde{C}$

Observe that each of these diagram admits a unique morphism to $\tilde{C}$ in the category above $\mathbf{D}$:



$\tilde{A} \to \tilde{C}$ $\tilde{B} \to \tilde{C}$ $\tilde{C} \to \tilde{C}$

Since this is only true of $\tilde{C}$, we conclude that the limit of $\mathbf{D}$, and hence the product of C and E, is $\tilde{C}$.

Example 19 (Products in **Grp**)**.** *Let* **Grp** *be the category with groups as objects and group homomorphisms as maps. Fix $G, H \in \mathbf{Grp}$ and let $G \times H$ be the group-theoretic direct product of G and H. This yields the following commutative diagram:*

$$\begin{tikzcd} & G \times H \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ G & & H \end{tikzcd}$$

Now let $K \in$ **Grp** *be any group with group homomorphisms $h_1 : K \to G$ and $h_2 : K \to H$. We will show that there is precisely one homomorphism $\phi : K \to G \times H$.*

$$\begin{tikzcd} K \arrow[r, dashed, "\phi"] \arrow[d, "h_1"'] \arrow[dr, "h_2"' near start] & G \times H \arrow[dl, "\pi_1" near start] \arrow[d, "\pi_2"] \\ G & H \end{tikzcd}$$

By commutativity, we have

$$h_1 = \pi_1 \circ \phi$$
$$h_2 = \pi_2 \circ \phi$$

and hence ϕ is uniquely determined by

$$\phi(x) = \big(h_1(x),\, h_2(x)\big)$$

for each $x \in K$. Therefore the categorical product of G and H is the original diagram

$$\begin{tikzcd} & G \times H \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ G & & H \end{tikzcd}$$

This is easily extended to a collection of groups $\{G_i\}_{i \in I}$, in which case the categorical product is the diagram

$$\begin{tikzcd} & & \prod_{i \in I} G_i \arrow[dll, "\pi_1"'] \arrow[dl, "\pi_2"] \arrow[d, "\pi_3"] \arrow[drr, "\pi_j"] & & & \\ G_1 & G_2 & G_3 & \cdots & G_j & \cdots \end{tikzcd}$$

where the product at the top is the group-theoretic direct product.

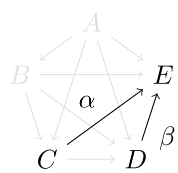
In a similar way, we can show that in **Set** the categorical product of a collection of sets $\{A_i\}_{i\in I}$ is the cartesian product $B = \prod_{i\in I} A_i$ together with the projection maps $\pi_i : B \to A_i$. In the category **Top** of topological spaces and continuous maps, the categorical product of spaces $\{(X_i, \mathcal{T}_i)\}_{i\in I}$ is the cartesian product $Y = \prod_{i\in I} X_i$ equipped with the product topology, together with the projection maps $\pi_i : Y \to X_i$.

Definition 36 (Pullback and Pushout)**.** *The limit of a diagram* $\mathbf{D} : \Gamma \to \mathcal{C}$ *with image*

$$A \xrightarrow{\alpha} B \xleftarrow{\beta} C$$

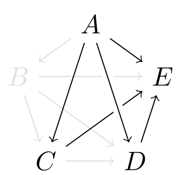
is called the pullback *of* α *and* β*. The dual concept is called the* pushout *of* α *and* β*.*

Example 20 (Pullback in a Preorder)**.** *Let* $\mathcal{C}$ *be the category from Example 18. The maps* $\alpha : C \to E$ *and* $\beta : D \to E$ *are uniquely determined to be* $\alpha = (a, d)$ *and* $\beta = (c, d)$*. The pullback of* α *and* β *is the limit of the diagram* $\mathbf{D} : \Gamma \to \mathcal{C}$*, where* $\mathbf{D}$ *is*

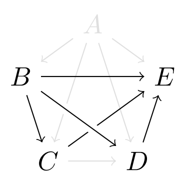


$\mathbf{D}$

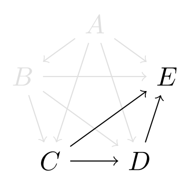
The objects above $\mathbf{D}$ *are*



$\tilde{A}$

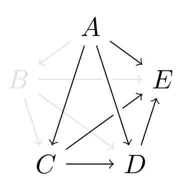


$\tilde{B}$

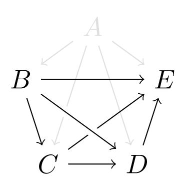


$\tilde{C}$

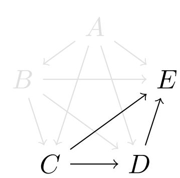
Note that there is a unique morphism from each object in the category above $\mathbf{D}$ *to* $\tilde{C}$*:*



$\tilde{A} \to \tilde{C}$



$\tilde{B} \to \tilde{C}$



$\tilde{C} \to \tilde{C}$

Therefore the limit of $\mathbf{D}$ *is the diagram* $\tilde{C}$*, and thus we conclude that* $\tilde{C}$ *is the pullback of* α *and* β*.*

Example 21 (Pullbacks in **Grp**)**.** *Fix* $G_1, G_2, H \in \mathbf{Grp}$ *and let* $f_1 : G_1 \to H$ *and* $f_2 : G_2 \to H$ *be homomorphisms. The corresponding diagram is automatically commutative:*

$$G_1 \xrightarrow{f_1} H \xleftarrow{f_2} G_2$$

Let $(G_1 \times G_2)_{f_1 = f_2} = \{(g_1, g_2) \in G_1 \times G_2 \,:\, f_1(g_1) = f_2(g_2)\}$*, and let* $K \in \mathbf{Grp}$ *be arbitrary with homomorphisms* h, h_1, h_2 *as illustrated*

$$
\begin{tikzcd}
K \arrow[rr, dashed, "\phi"] \arrow[d, "h_1"'] \arrow[dr, "h"] \arrow[drr, "h_2", near start] & & (G_1 \times G_2)_{f_1 = f_2} \arrow[dll, "\pi_1"', near start] \arrow[dl, "f_2 \circ \pi_2"] \arrow[d, "\pi_2"] \\
G_1 \arrow[r, "f_1"] & H & G_2 \arrow[l, "f_2"']
\end{tikzcd}
$$

Since

$$
\begin{aligned}
h_1 &= \pi_1 \circ \phi \\
h_2 &= \pi_2 \circ \phi
\end{aligned}
$$

we conclude that, if ϕ *exists, it must be defined by*

$$
\phi(x) = \big(h_1(x),\, h_2(x)\big)
$$

It remains to show that this definition of ϕ *yeilds a commutative diagram. Indeed we have*

$$
\begin{aligned}
h(x) &= f_2 \circ h_2(x) \\
&= f_2 \circ \pi_2\big(h_1(x), h_2(x)\big) \\
&= (f_2 \circ \pi_2) \circ \phi(x)
\end{aligned}
$$

Thus ϕ *exists and is uniquely determined. We conclude that the pullback of* f_1 *and* f_2 *is*

$$
\begin{tikzcd}
& (G_1 \times G_2)_{f_1 = f_2} \arrow[dl, "\pi_1"'] \arrow[d, "f_1 \circ \pi_1"', "f_2 \circ \pi_2"] \arrow[dr, "\pi_2"] & \\
G_1 \arrow[r, "f_1"] & H & G_2 \arrow[l, "f_2"']
\end{tikzcd}
$$