Class Notes

Dylan Pentland

Contents

1 Introduction **2**

2 Bass-Serre theory **3**
2.1 Motivation and outline . . . 4
2.2 Trees and free groups. . . . 7
2.3 Graphs of groups . . . 12
2.4 Amalgams and HNN extensions . . . 18
2.5 Structure of groups acting on trees . . . 21
2.6 Amalgams and fixed points . . . 25

3 Applications to p-adic groups **34**
3.1 The Bruhat-Tits tree . . . 35
3.2 Structure theory . . . 42
3.3 Amalgams and SL_2 . . . 48
3.4 Ihara's theorem and Mumford uniformization . . . 51

4 Ramanujan graphs **58**
4.1 Spectra of graphs . . . 59
4.2 Motivation for Ramanujan graphs . . . 64
4.3 Connecting back to the tree . . . 66
4.4 Modular forms . . . 72
4.5 Hecke operators on modular forms . . . 80
4.6 Putting it all together . . . 87

1 Introduction

The star of this class will be on an extremely special tree, called the *Bruhat-Tits tree*. It looks like this:

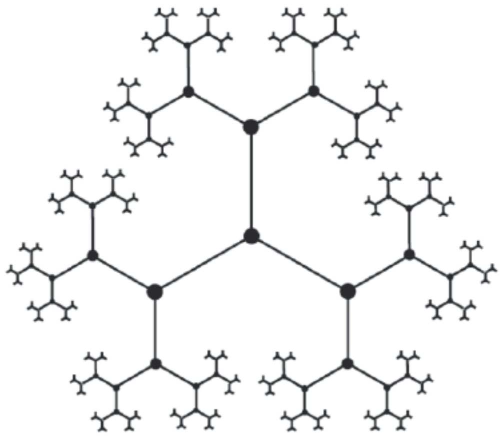


Figure 1: An example of a Bruhat-Tits tree.

Roughly, the class will be divided into three pieces. The first will focus on what is called *Bass-Serre theory*. Roughly, this aims to fully understand the structure of groups with nice actions on trees by carefully using a few tools from topology and some combinatorial arguments.

We will then see some amazing applications of these ideas to certain p-adic groups like $\mathrm{SL}_2(\mathbf{Q}_p)$ which act on trees: in particular, Bruhat-Tits trees.

Finally, we will see how the Bruhat-Tits tree connects even more deeply to the nature of the p-adic group by encoding some deep facts about its representation theory. We will apply this to construct Ramanujan graphs, which are extremely special graphs in combinatorics and computer science that have applications to error correcting codes. Roughly, these are d-regular graphs whose edges behave as if they are random and evenly distributed.

2 Bass-Serre theory

As mentioned in the introduction, Bass-Serre theory is about studying nice actions of a group G on a tree X and extracting whatever information we can from this about G. We will prove something like the following by the end of this section:

> THEOREM 2.1 (Very rough statement). Let G be a group which acts nicely on a tree X. There is a corresponding "homotopy quotient" space |X|//G such that
>
> $$\pi_1(|X|//G, v) \simeq G.$$
>
> Here, |X| denotes the topological space underlying a graph. We can compute this fundamental group from combinatorial data in a so-called "graph of groups" [X/G] describing the structure of this space. This recovers a presentation of G from a complete description of the action.

In the first subsection, we will sketch the very easiest case where the action on the tree is *free*. This means that no vertices are fixed by $g \neq 1$, and no edges are inverted. In this case, our group G will be identified with the fundamental group of a graph, which is in fact a free group. We'll then try to motivate the idea that we can extend to more general actions.

We'll then actually prove in detail the simplest case of Bass-Serre theory, when we have a free action.

Next, we construct the space |X|//G and explain how it behaves like a "graph of spaces", and upon taking fundamental groups of these spaces we get (inexplicitly) the combinatorial data of the graph of groups [X/G].

After this, we will need some group-theoretic preliminaries about *amalgams* and *HNN extensions*. We will use these later to explicitly write down $\pi_1(|X|//G, v)$.

Finally, we write down explicitly the graph of groups and use amalgams and HNN extensions to compute $\pi_1(|X|//G, v)$, thereby giving a (potentially new) presentation of G from the action.

At the end we'll show that the existence of fixed points in the action of G on X implies a certain level of complexity: namely, the group G is not an amalgam. We'll go through the example of $SL_3(\mathbf{Z})$.

2.1 Motivation and outline

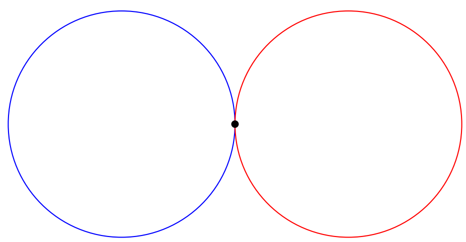
We can motivate Bass-Serre theory by just thinking about a very simple example where we can prove that a group is free. We can arrive at this idea by just thinking about how free groups arise topologically.

DEFINITION 2.2. A wedge of two spaces X and Y is a space $X \vee Y$ is defined as

$$X \coprod Y / \sim$$

where $\sim$ identifies chosen basepoints $x_0 \in X$ and $y_0 \in Y$ (imagine this as gluing these two points together).

For example, a wedge of two circles looks like the following:



If we were to calculate $\pi_1(S^1 \vee S^1)$, we'd see there are two generators for the blue and red loops. These have no relations.

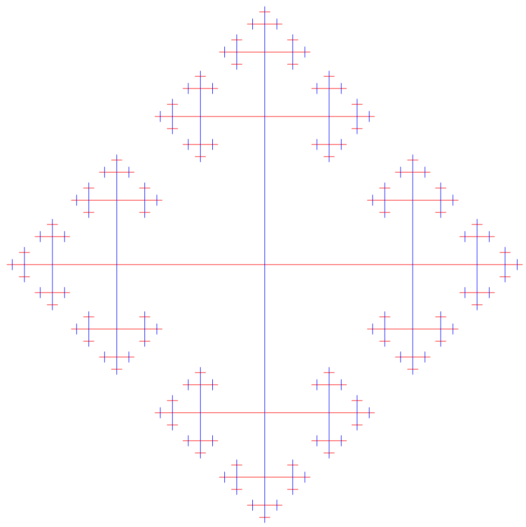
LEMMA 2.3. Let X be a wedge of n circles. Then $\pi_1(X)$ is a free group on n generators.

Proof. Use Van Kampen! □

Observe that X has a universal cover, which is an infinite $2n$-regular tree (due to also including inverses of the generators to draw the graph). The group $\pi_1(X)$ acts freely on this tree, via deck transformations.

By an action on a graph, we mean a homomorphism $G \to \mathrm{Aut}(X)$, the group of *graph automorphisms*. These are invertible graph homomorphisms, or functions between vertex sets of graphs that map adjacent vertices to adjacent vertices. By a free action we mean no vertices are fixed by a nontrivial element, and no edges are inverted (there is some orientation preserved by G). **We assume all group actions on trees are without inversions unless otherwise stated**.

In the case of $\mathbf{Z} * \mathbf{Z} = \mathbf{F}_{\{a,b\}}$, we'll explain an alternate construction coming quite directly from the group (which works for any free group). Take the vertex set to be $\mathbf{Z} * \mathbf{Z}/\langle a \rangle \coprod \mathbf{Z} * \mathbf{Z}/\langle b \rangle$. This is just all the words, broken up into those starting with b and a. The edge set is $\mathbf{Z} * \mathbf{Z}$, with a particular edge connecting the element modulo a to the element modulo b. The graph looks like this:



We'll see that we can use this construction of a tree in slightly more general situations. In particular, we can get a graph an amalgamated product $G_1 *_H G_2$ acts on by taking a quotient by H in the edge set.

So, we can realize free groups as fundamental groups. We wanted a technique to prove that a particular group is free, so what we would like to do is deduce this from some action on a tree that will play the role of the universal cover. This is indeed possible, and is the simplest case of Bass-Serre theory.

> PROPOSITION 2.4. Suppose a group G acts freely on a tree X. Then G is a free group.

Sketch. Write $|X|$ for the topological realization of a graph (essentially, the picture of a graph as a space). The general idea is that $|X| \to |X/G|$ is a covering map, which is deduced by the freeness of the action. Here, we define a quotient graph by taking vertex and edge sets modulo the equivalence relations $v \sim gv$ and $e \sim ge$ respectively.

Then, we can actually identify X as the universal cover. It turns out that this lets us identify $G \simeq \pi_1(|X/G|)$, and π_1 of a graph is always free since it is homotopy equivalent to a wedge of circles. $\square$

> Remark 2.5. We assume the action does not have inversions: otherwise, the swapping action on $* \to *$ by $\mathbf{Z}/2\mathbf{Z}$ would count as free since the nontrivial element fixes no vertices. We don't want to consider $\mathbf{Z}/2\mathbf{Z}$ a free group.

To summarize, in the situation of a free action on a tree we have a covering map

$$X \to X/G$$

and the fundamental group of the graph X/G is G, and can also be computed as a free group. So, we can recover a description of G acting freely on a tree via its action.

What about groups which are not free? Ideally, we'd like to be able to combinatorially reconstruct G with action on a tree X without inversions by finding a space which has a fundamental group computable via the action. A key thing to notice is that π_1 is a homotopy invariant, so we might consider modifying the topological space $|X|$ up to homotopy equivalence.

> **Idea**: Construct a contractible space X' homotopy equivalent to $|X|$ with a free G-action such that
> $$X' \to X'/G$$
> is a covering map, so $\pi_1(X'/G) \simeq G$.

The proposal here is that we should produce some notion of a *homotopy* quotient, where we replace objects up to homotopy to make the action nicer. After understanding how this construction works, it will become clear that $\pi_1(X'/G)$ can be computed from the "graph of groups" $[X/G]$, which is roughly the naive quotient graph but where we remember the data of stabilizers of vertices and edges that the naive quotient forgets.

Another perspective you might have on this is starting from the graph of groups $[X/G]$, and thinking of that as the appropriate quotient graph since it now remembers stabilizers (which is precisely what we lose in the naive quotient). Then, you could try and find a "homotopy type" of this space, which turns out to be X'/G. [1]

[1] We won't be discussing it here, but the idea is that $[X/G]$ is a Deligne-Mumford topological stack. These have a homotopy type, which comes out to be the homotopy quotient.

2.2 Trees and free groups.

In this section, we see a detailed proof based on the previous outline that G acts freely (that is, nontrivial g fix no vertices and it acts without inversions) on a tree X if and only if G is a free group.

First, we make precise the notion of a topological realization of a graph. Serre defines a graph in a slightly nonstandard way; we're going to adapt the theory to work with the standard treatment of a graph. In this definition, we define a graph to be a pair (V, E) consisting of vertices and edges. The edges are equipped with boundary vertices $\partial_0 e$ and $\partial_1 e$. Implicitly, all graphs will be connected graphs here, and labeling the bounding vertices gives an implicit orientation. A tree is defined as a connected graph without any cycles (people often call a general graph without cycles a forest, since it consists of a union of trees).

To make the topological realization of a graph, we simply take each edge and replace it with $[0, 1]$. Then, we glue these intervals to vertices the edge connects to. This creates a topological space, which looks like the usual picture you draw of a graph. We denote this by $|\Gamma|$ for a graph Γ.

The homotopy type of a graph is actually quite simple.

> LEMMA 2.6. Let Γ be a connected graph. The topological realization $|\Gamma|$ is homotopy equivalent to a wedge of circles.

Proof. Let $T \leq \Gamma$ be a maximal subtree. We first claim that the map

$$\pi : |\Gamma| \to |\Gamma/T|$$

obtain by contacting T is a homotopy equivalence. This isn't too hard to see intuitively, since we can imagine picking a base for the tree and contracting the edges one by one.

However, if we are being rigorous there's a problem. We're essentially assuming that collapsing a contractible subspace to a point does not change the homotopy type. **This is false in general**. For example, take $X = S^1$ and $A = S^1 \setminus \{p\}$. Collapsing A by passing to the quotient actually gives a contractible space (seeing this takes a bit of work, however)

In our case, everything is fine. The closed subspace $|T| \subseteq |\Gamma|$ is what's called a cofibration. This means that if we have an extension $f' : |\Gamma| \to S$ of $f : |T| \to S$, any homotopy

$$H : |T| \times [0, 1] \to S$$

with $H(t, 0) = f(t)$ can be lifted to $H' : |\Gamma| \times [0, 1] \to S$ where $H'(g, 0) = f'(g)$.

In particular, take the homotopy H from $1_{|T|}$ where H_t contracts the tree to a point. Extending to the global identity $1_{|\Gamma|}$, we have a map $H'_1 : |\Gamma| \to |\Gamma|$. This factors though the quotient map $|\Gamma| \to |\Gamma/T|$, since on the tree H'_t was contracting it.

Further, there is an induced homotopy $\overline{H'_t} : |\Gamma/T| \to |\Gamma/T|$ on the quotient as it sends the tree T to itself.

We then have a commutative diagram

$$\begin{tikzcd} |\Gamma| \arrow[r, "H'_1"] \arrow[d, "p"] & |\Gamma| \arrow[d, "p"] \\ |\Gamma/T| \arrow[r, "\overline{H'_1}"] \arrow[ur, "f"] & |\Gamma/T| \end{tikzcd}$$

where f results from H'_1 factoring through the quotient. Commutativity of the bottom triangle follows from p being surjective. Via H_t, we deduce $f \circ p$ is homotopic to the identity. We can show $p \circ f$ is as well, using $\overline{H'_t}$. The initial claim follows that we can safely contract T without changing the homotopy type.

After this, the proof is simple: the graph Γ/T has a single vertex. Thus, we only have self-loops. $\square$

With this in hand, the proof strategy before makes more sense. If we just prove that we actually get a covering map $|X| \to |X|/G$ for a group G acting freely on a tree X, then the quotient is again a graph and we identify G with the fundamental group which we know is free. Let's do this rigorously now.

> **Proposition 2.7.** Suppose a group G acts freely on a tree X. Then G is a free group.

Proof. The map $|X| \to |X|/G$ satisfies a slightly stronger than non-identity elements of G not fixing vertices. Namely, every $x \in |X|$ (as a topological space) has a neighborhood U such that gU and U are disjoint except when g is the identity. This is called being an even action. The condition is obviously true for vertices. On edges, it is true because of the condition on inversions (if we had an inversion, think about the midpoint of the edge).

The quotient X/G is a graph again, and since $|X|$ is simply connected it is the universal cover of this graph via the covering map $|X| \to |X|/G$. It follows from covering space theory that

$$G \simeq \pi_1(|X|/G).$$

The required lemma is the following:

> Lemma 2.8. Let $p : X \to X/G$ be a quotient map where X is equipped with an even action. Then the quotient map is a covering map, and moreover $\pi_1(X/G) \simeq G$ if X is simply connected.

Being even makes this a covering map, and X being simply connected makes X the universal cover. This holds because $\pi_1(X/G)$ is the group of deck transformations of the universal cover (self-maps as a covering space). To see this is the case, we use the lifting theorem. Given a loop $[\gamma] \in \pi_1(X/G)$, lift this to X using the lifting theorem. We get a path $\tilde{\gamma}$, and so use the lifting lemma on

$$\begin{tikzcd}
 & (X, \tilde{x}) \arrow[d, "p"] \\
(X, \tilde{\gamma}(1)) \arrow[ur] \arrow[r, "p"] & (X/G, x)
\end{tikzcd}$$

where $\tilde{x}$ is the basepoint of the universal cover of $(X/G, x)$. This tells us there is a unique lift, which provides a deck transformation associated to the class $[\gamma]$. Deck transformations are the same if they agree at a single point, by uniqueness in the lifting lemma so this construction only depends on the homotopy class of γ. Thus, we get a map of sets $\pi_1(X/G, x) \to \mathrm{Deck}(X, X/G)$. This is seen to be a group homomorphism by looking at the endpoint of $\widetilde{\gamma \cdot \gamma'}(1)$, and noting that we get the same endpoint by following both lifted paths one after another. It is an isomorphism as well: it is injective, since if $[\gamma]$ and $[\gamma']$ in $\pi_1(X/G, x)$ have lifts with the same endpoint, they are homotopic since X is simply connected. Moreover, it is surjective since a deck transformation f is determined by where it sends $\tilde{x}$ in the fiber (uniqueness in the lifting lemma), and we can just draw a path from $\tilde{x}$ to $f(\tilde{x})$ and apply p to produce the corresponding loop $[\gamma]$.

Every $g \in G$ induces a deck transformation by acting on X. Letting f be a deck transformation, we see x and $f(x)$ are in the same orbit by definition. Then there exists some $g \in G$ sending $x \mapsto g \cdot x = f(x)$. A deck transformation is determined by where it sends a single point, since we have a unique lift of $p : X \to X/G$ for a given choice of basepoint (a lift of a given point). By the lifting theorem, there is a unique such lift!

Now apply the lemma to our $|X|$ which is simply connected as it contractible.

Since X/G is a graph, its topological realization is homotopy equivalent to a wedge of circles by contracting a maximal spanning tree. By Van Kampen's theorem, $\pi_1(G\backslash X)$ is a free group. $\square$

This has some useful consequences. For example, we get the following corollary now.

> COROLLARY 2.9 (Nielsen-Schreier). Any subgroup of a free group is free.

Proof. Being a free group on some set S of generators, it acts freely on the universal cover of a wedge of circles (one for each $s \in S$). This universal cover is a tree. Any subgroup still acts freely, so by the previous result it is again a free group. $\square$

> COROLLARY 2.10 (Schreier index formula). Let G be a free group on a finite number of generators r_G, and $H \leq G$ a subgroup of finite index n. Then if H has r_H generators,
> $$n(r_G - 1) = r_H - 1.$$

Proof. The group G acts on a graph X, forming a quotient graph X/G. The quotient graph X/H is an n-fold cover of X/G.

We can read off the rank/number of generators of $\pi_1(|\Gamma|)$ for a connected graph Γ quite easily: it is $1 - \chi(\Gamma)$, where χ is the Euler characteristic as a topological space. Here, we can define χ as $\#V - \#E$ (there is a general definition in topology as well, so it is homotopy invariant). Thinking about a maximal spanning tree, the overall contribution to this number is 1. Then each additional edge forms a loop, so $1 - r_{\pi_1(|\Gamma|)} = \chi(\Gamma)$.

Now the proof, and appearance of $r_G - 1$, becomes clear. In an n-fold covering map of topological spaces, it is a fact that the Euler characteristic gets multiplied by n. The reason is just that we can count lifts of edges and vertices. In particular,

$$\chi(|X/H|) = n\chi(|X/G|).$$

From $r_G - 1 = -\chi(|X/G|)$ and similarly for H, the claim follows. $\square$

Both of these statements about free groups are really not obvious without the topological viewpoint. Particularly with the Schreier index formula, a non-geometric proof might obscure the "moral" reason why it is true, coming from χ.

One thing that might seem confusing at first with the Schreier index formula is that the roles of G and H seem reversed. You might initially expect that a subgroup should have less generators, as is the case with the obvious infinite index subgroups of a free group. With covering space theory, we can write down some explicit examples by producing finite covers of wedges of circles.

Let's return to our previous example of $S^1 \vee S^1$. In this case, we know the fundamental group is $\mathbf{F}_{\{a,b\}}$. Consider the following covering space from Hatcher's book:

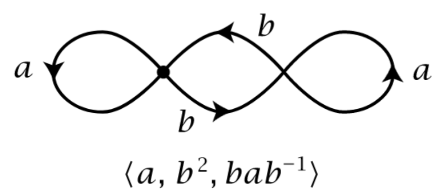


Figure 2: 2-fold cover of $S^1 \vee S^1$ from Hatcher's *Algebraic Topology*.

The covering map sends each vertex to the point of intersection point $*$ of the circles, and the edges labeled a and b are sent to the two generators of $\pi_1(S^1 \vee S^1, *)$. This is evidently a twofold cover: there are two copies of $*$. and two copies of each of a and b. We see that the Euler characteristic of the graph is $2\chi(S^1 \vee S^1)$ as a result. It follows the number of generators of the fundamental group of this graph is

$$1 - 2\chi(S^1 \vee S^1) = 3.$$

This is indeed the case, and generators are given by a, b^2, and bab^{-1}. We see now why the rank is nearly always higher: we can produce lots of words in a free group which have no relations!

2.3 Graphs of groups

Having seen the case where G acts freely on a tree X in detail, we will now begin to tackle the general case where G only acts without inversions.

Recall that for a general action without inversions, we were unsatisfied that $\pi_1(X/G)$ does not recover G. For example, take the extreme case of G acting trivially on $*$. The naive quotient still has trivial fundamental group.

However, if we are able to replace X with a homotopy equivalent (again contractible/homotopy equivalent to a point) space X', the hope is that now we can have G act freely again. Then we get a legitimate covering map

$$X' \to X'/G$$

and covering space theory, like in the case of a free action, allows us to deduce $\pi_1(X'/G) \simeq G$.

We call X'/G the homotopy quotient, and denote it $|X|//G$. We will first go about constructing it, and then connecting to the graph of groups $[X/G]$ alluded to earlier.

First, we ought to be able to construct the simplest case of a homotopy quotient $*//G$. Such a space should have a contractible universal cover, and we want to have $\pi_1(*//G) = G$. This actually already determines the homotopy type.

> Theorem 2.11. Up to homotopy equivalence, there is a unique space BG with a contractible universal cover.

Proof. See Theorem 1B.8 in Hatcher's *Algebraic Topology*. □

Up to homotopy equivalence, there is then also a unique space $|X|//G$ with the desired properties since $|X|$ is contractible. However, we want a better description than BG. The idea will be to build this out of $*//G_v$ and $*//G_e$ for stabilizer groups of vertices and edges.

All we need to do now in the case of $*//G$ is provide a good construction of the space. As a first attempt, maybe we try to manually make a space X with $\pi_1(X) = G$ and then consider the universal cover.

> Lemma 2.12. We can make a nice space with fundamental group G.

Proof. We know that we can write $G = \langle S : R \rangle$. Making the generators is easy: take a wedge of $|S|$ circles (this doesn't actually require S to be finite, we just let the set circles be in bijection with S). Denote the wedge of circles by X^1.

We observe that we can create arbitrary relations by taking a disk $D^2 \subseteq \mathbf{R}^2$ and gluing its boundary (a loop S^1) onto the graph in a particular way. Let's see an example.

Say we want to make X so $\pi_1(X) = \mathbf{Z}^2$. We expect to get a torus out of this. First, take two circles and wedge them together, and label the generators of the fundamental group a and b. Then, consider the nontrivial loop $\gamma = aba^{-1}b^{-1}$. If we trivialize just this, then we've made the generators commute. Being a loop, we can identify it with S^1 via a map $f : S^1 \to \gamma$. We glue a disk via this map, i.e. we attach $\theta \in S^1$ to $f(\theta) \in S^1 \vee S^1$.

This loop now becomes trivial: we can just slide it over the disk to make it trivial. To see this geometrically, imagine you have a nontrivial loop on $\partial D = S^1$ and slide it though the disk to a point.

Now we will make this argument precise. For each relation word in R, we take the loop for this word and attach D^2 along this loop as in the example. Suppose we have attached some number of disks to make X, and are now attaching another disk D^2 for a relation r to make X'. Cover X' with a small open neighborhood of X and a small open neighborhood of the disk D^2. By Van Kampen, we have

$$\pi_1(X') = \pi_1(D) *_{\pi_1(S^1)} \pi_1(X).$$

The map $\pi_1(S^1) = \mathbf{Z} \to \pi_1(X)$ sends $1 \mapsto r$. Thus, we take $1 *_{\mathbf{Z}} \pi_1(X)$, and by our construction of the amalgamated product this just adds a relation r. □

Unfortunately, in this construction the universal cover need not be contractible. It is also not a very 'canonical' construction, since we had to choose a presentation. Let's give this a second shot, trying to put more emphasis on the free action on the universal cover.

> Theorem 2.13. There is a contractible space EG endowed with a G-action which is free, such that
> $$EG \to BG := EG/G$$
> is a covering map. Then we can identify the homotopy quotient $*//G$ with BG.

Proof. We build up EG by gluing together n-dimensional cells. We begin with a space $(EG)^0$ by taking G as a discrete set: this has an obvious free action by G via multiplication, but is far from contractible. To remedy this, glue copies of $D^1 = [0, 1]$ for each pair $[g_1, g_2]$ onto g_1 and g_2 to form $(\mathrm{EG})^1$. This is just $\mathrm{G} \times \mathrm{G}$, so we can extend the G-action. Each $[g]$ now has a path to $[e]$, so we can move these along the paths to attempt to contract the first component $(\mathrm{EG})^0$. However, we now have a problem: we added new segments $[g_1, g_2]$, and we don't have a way to contract these.

The solution is to do the same thing again, so we can now contract $(\mathrm{EG})^1$. We continue with this for each $n \in \mathbf{N}$, attaching n cells $[g_1, \ldots, g_{n+1}]$, and gluing onto $n-1$-cells by forgetting a group element.

Thus, we obtain a space EG with a free G-action: multiply on the left. We can think of this as taking the most obvious free G-action, multiplication of G on itself, and then iteratively correcting it to be contractible. The reason EG is contractible at the end is that we can slide each $x \in [g_1, \ldots, g_{n+1}]$ along the segment in $[e, g_1, \ldots, g_{n+1}]$ to the identity element $[e]$ (which now always exists, since we repeated the process indefinitely).

One can check that $\mathrm{EG} \to \mathrm{EG}/\mathrm{G}$ is a covering map (we need slightly more than freeness), and hence it makes sense to define $\mathrm{BG} := \mathrm{EG}/\mathrm{G}$. □

Exercise. Check that a homomorphism $\mathrm{G} \to \mathrm{H}$ induces a continuous map $\mathrm{BG} \to \mathrm{BH}$. Moreover, if it is an isomorphism then BG and BH are homeomorphic.

Remark 2.14. We can construct BG directly as the geometric realization of the nerve of the 1-object category G, with morphisms for each group element composing as $g \circ g' = gg'$. This makes it a better choice as opposed to some other homotopy equivalent space.

This space is known as a classifying space because the space of homotopy classes of maps $[\mathrm{X}, \mathrm{BG}]$ classifies principal G-bundles on X. For example, when $\mathrm{G} = \mathrm{GL}_n$ this classifies vector bundles. We can compute the isomorphism classes using the orthogonal group $\mathrm{O}(n)$ instead, and then computing invariants of $\mathrm{BO}(n)$ such as cohomology allows us to understand vector bundles.

We will now want to choose a specific space $|\mathrm{X}|//\mathrm{G}$, rather than being satisfied with it being BG up to homotopy equivalence.

Corollary 2.15. Let G act on a tree X without inversions. Then $\mathrm{EG} \times |\mathrm{X}|$ is a contractible space with a free G action, and

$$\mathrm{EG} \times |\mathrm{X}| \to |\mathrm{X}|//\mathrm{G} := (\mathrm{EG} \times |\mathrm{X}|)/\mathrm{G}$$

is a covering map. It follows $\pi_1(X//G) = G$ as well.

Proof. Since X is a tree, $|X|$ is contractible. Further, since the action of EG is free and $|X|$ is a space with G-action induced by the tree action (graph automorphisms are topological automorphisms), the action on $EG \times |X|$ remains free. We can easily see the quotient by this action then yields a covering map (by checking the action is even), and due to freeness by the same argument as in the free action on a tree case we deduce $\pi_1(|X|//G) = G$. $\square$

Remark 2.16. For a general space X with an action by G, realize the quotient of X by G as a category by taking the objects to be $x \in X$, and morphisms $(g, x) : x \to g \cdot x$. This is a groupoid: all morphisms are invertible. It is usually called the *action groupoid*. The geometric realization of the nerve of this category is $X//G$. Applying to a tree, we get the space in the corollary so it is again the "best" choice.

The reason we chose this particular space is to fill in a key component of Bass-Serre theory: we need to be able to give a recipe to compute $\pi_1(|X|//G)$ from data about the action on X, otherwise we have gained no information about G. For example, with a free action we just contracted a maximal spanning tree on the resulting quotient graph and counted loops.

The idea is to read off $\pi_1(|X|//G)$ from a graph of groups, after better understanding what it looks like. Recall we informally defined a graph of groups as the naive quotient graph X/G, with edges and vertices adorned with their stabilizers G_v and G_e. We now make this notion precise.

Definition 2.17. A graph of groups is a connected graph X along with a group G_v for each vertex v and a group G_e for each edge e.

For each edge e, we have boundary vertices $\partial_0 e$ and $\partial_1 e$. In a graph of groups, we also ask for the data of injective homomorphisms

$$\psi_{e,i} : G_e \to G_{\partial_i e}$$

for $i = 0$ or 1.

The reason for the homomorphisms is that in a quotient graph, the stabilizer of any $v, e \in X/G$ is well-defined only up to conjugacy because we could take another orbit representative (though then the isomorphism class stays the same). Thus, we need to explain how they are related. We can get injective homomorphisms because for a particular lift in X, $G_e \leq G_v$. Thus, the graph of groups should **not** just be X/G with the isomorphism classes of stabilizers labeled.

How does $|X|//G$ connect to the graph of groups? To understand this, we just need to understand that it looks like a "graph of spaces" already. Indeed, given any graph of groups, we can make a graph of classifying spaces by taking the induced maps

$$BG_e \to BG_v$$

from the injective homomorphisms $G_e \to G_v$. Using these, we take $G_e \times [0, 1]$ and use these induced maps to attach these cylinders. One can even define the fundamental group of a general graph of groups to be the fundamental group of this space.

Definition 2.18. A graph of spaces is a graph X along with a space X_v for each vertex and a space X_e for each edge e.

For each edge e, we have boundary vertices $\partial_0 e$ and $\partial_1 e$. In a graph of spaces, we also ask for maps

$$\psi_{e,i} : X_e \to X_{\partial_i e}$$

for $i = 0$ or 1 inducing injective homomorphisms on fundamental groups.

The previous construction turns a graph of groups into a graph of spaces. By taking fundamental groups and looking at the induced homomorphisms, we can also turn a graph of spaces into a graph of groups.

The following lemmas tell us that $|X|//G$ is a graph of spaces, and hence has an underlying graph of groups [X/G].

Lemma 2.19. For our construction of $|X|//G$, there is a natural projection map

$$|X|//G \to |X/G|.$$

The fiber above a vertex $v \in |X|/G$ is homotopy equivalent to BG_v, and for an edge e it is homotopy equivalent to BG_e. Here, G_v and G_e are the stabilizers under the action of G.

Proof. We note that in $(|X| \times EG)/G$, projecting onto the first component gives the desired map. The fiber over a vertex $v \in X/G$ consists of all of the G-orbits $(\tilde{v}, x) \in |X| \times EG$ where $\tilde{v}$ lifts v. We can make a particular choice of lift $\tilde{v} \in X$, and then up to the G-action all of these can be written as $(\tilde{v}, x)$ for this particular lift $\tilde{v}$. Then we are free to further act by elements of the stabilizer $G_{\tilde{v}}$, so the fiber is naturally $EG/G_{\tilde{v}} \simeq BG_{\tilde{v}}$. The homotopy type of this only depends on the isomorphism class of the group, so for the homotopy type the lift does not matter. $\square$

What we have missed here is how these classifying spaces are attached to each other. We will delay the actual description of attaching maps until later, since this is where the actual work is done. For now, we take this for granted so that we are satisfied that we actually have a graph of groups $[X/G]$ induced by taking fundamental groups in $|X|//G$. The idea is roughly that if we pick a lift $\tilde{v}$ and $\tilde{e}$ consistently, it is induced by the inclusion map on that lift. However, subtleties arise because we need to extend this to the whole graph.

> Lemma 2.20. By the previous lemma, $|X|//G$ can be obtained by gluing together classifying spaces. The attaching maps are obtained by maps $BG_e \to BG_v$ induced from an injective homomorphism $G_e \to G_v$ on the fundamental groups.

Now, the main idea is the following.

> **Idea**: Given that $|X|//G$ is a graph of spaces, we want to take the corresponding graph of groups $[X/G]$ and then compute $G \simeq \pi_1(|X|//G)$ from the combinatorial data in this. This will express G in terms of G_v, G_e, and the homomorphisms in the graph of groups.

2.4 Amalgams and HNN extensions

Having constructed an appropriate topological realization for a graph of groups, we will need to combinatorially compute its fundamental group from the groups G_v and G_e. What we will need to do this are two operations on groups: amalgams of groups, and HNN extensions. Both of these constructions have topological interpretations in terms of gluing, which is all we want to do on a graph of spaces to compute the fundamental group.

First, recall the definition of an amalgam presented in the background notes.

> Definition 2.21. For groups G_1, G_2 and H, equipped with homomorphisms
>
> $$\varphi_1 : H \to G_1, \varphi_2 : H \to G_2$$
>
> define $G_1 *_H G_2$ to be $G_1 * G_2$ (the free product: take the union of generators and relations for both groups) with extra relations
>
> $$\varphi_1(h)\varphi_2(h)^{-1} = 1.$$
>
> You should think of this as forcing agreement along the two inclusions of H. We call this an *amalgamated product* $G_1 *_H G_2$.

The relevance to Bass-Serre theory is the following:

> Proposition 2.22. Suppose we have a graph of groups which looks like
>
> $$\underset{G_v}{\bullet} \xrightarrow{G_e} \underset{G'_v}{\bullet}.$$
>
> Then the fundamental group is the amalgam
>
> $$G_v *_{G_e} G_{v'},$$
>
> using the canonical injections $G_e \to G_v$ and $G_{v'}$.

Proof. We apply Van Kampen's theorem! Pick U and V to be open neighborhoods of BG_v and $BG_{v'}$ which extend to slightly over the halfway point in $BG_e \times [0, 1]$. The intersection is $BG_e \times [1/2 - \varepsilon, 1/2 + \varepsilon]$. □

To better understand amalgams, we will discuss some of results presented in Serre's book about them.

Suppose we are given a group A and injective homomorphisms

$$\mathrm{A} \to \mathrm{G}_i$$

for a collection of groups G_i where $i \in \mathrm{I}$. Define G to be the interated amalgamated product with respect to these (alternatively: the colimit of the diagram with these injective homomorphisms).

Pick coset representatives S_i for $i \in \mathrm{I}$ for G_i/A. A *reduced word* in this context is a family

$$(a; s_1, \ldots, s_n)$$

where $a \in \mathrm{A}$, and $s_j \neq 1 \in \mathrm{S}_{i_j}$. We require the sequence i_j to have $i_j \neq i_{j+1}$. Note the similarity to reduced words in free groups: we take $\mathrm{A} = 1$. We say this reduced word is of type $\mathbf{i} = \{i_1, \ldots, i_n\} \subseteq \mathrm{I}$.

Theorem 2.23 (Structure theorem). For $g \in \mathrm{G}$ there is a unique reduced word $m = (a; s_1, \ldots, s_n)$ of type $\mathbf{i}$, such that

$$g = f(a) f_{i_1}(s_1) \ldots f_{i_n}(s_n).$$

Here, f is the induced map $\mathrm{A} \to \mathrm{G}$ and f_{i_j} is the map $\mathrm{G}_{i_j} \to \mathrm{G}$.

A useful consequence of this structure result is the following.

Corollary 2.24. Every element of G of finite order is conjugate to an element in one of the G_i. In particular, if G_i are torsion free so is G.

The other construction we want to study, HNN extensions, naturally appears for a 1 edge graph of groups as well.

Proposition 2.25. Suppose we have a graph of groups which is a self-loop, with stabilizer G_v on the vertex and G_e on the edge and an injective homomorphisms $\alpha_i : \mathrm{G}_e \to \mathrm{G}_v$. Letting $\mathrm{G}_v = \langle \mathrm{S} : \mathrm{R} \rangle$, the fundamental group is the HNN extension

$$\langle \mathrm{S}, t : \mathrm{R}, t\alpha_1(g)t^{-1} = \alpha_2(g) \text{ for } g \in \mathrm{G}_e \rangle.$$

We will omit the proof of this. The intuition is that when we glue a subspace $\mathrm{Y} \subseteq \mathrm{X}$ via a homeomorphism $\mathrm{Y} \to f(\mathrm{Y}) \subseteq \mathrm{X}$, we get an HNN extension. The extra generator T comes from a path connecting the basepoint y to $f(y)$: conjugating allows us to switch between these basepoints. For a proof, see Proposition 1.2 in "Topological methods in group theory" by Scott and Wall.

With these two operations, we will be able to combinatorially compute $\pi_1(|X|//G)$.

2.5 Structure of groups acting on trees

In this subsection, we will put everything together and prove the following main theorem:

Theorem 2.26 (Main theorem). Let G be a group which acts on a tree X without inversions. We obtain

$$G \simeq \pi_1(|X|//G).$$

We compute $\pi_1(|X|//G)$ via amalgams and HNN extensions. Let V denote the vertex set of X/G, and E the edge set. We use $\psi_{e,i}$ denote the injective homomorphisms in the associated graph of groups.

First, we take

$$G_T = *_{v\in V} G_v / \sim$$

where $\sim$ means we adjoin the relations $\psi_{e,0}(g) = \psi_{e,1}(g)$ for $e \in T$ and $g \in G_e$ (alternatively, quotient by the normal closure N of the subgroup generated by $\psi_{e,0}(g)\psi_{e,1}(g)^{-1}$). Then G is obtained as

$$G = \langle G_T, E \setminus E(T) : e\psi_{e,1}(g)e^{-1} = \psi_{e,2}(g), e \in E \setminus E(T), g \in G_e\rangle.$$

This is meant to be combined with Proposition 2.29 which tells us explicitly how to write down the graph of groups.

We remark here that the condition about inversions is actually a very mild condition. Indeed, upon an initial barycentric subdivision of the tree (here this fancy term means put a vertex in the middle of an edge) we can force this condition to be true.

To prove this, we need to revisit the description of attaching maps in $|X|//G$. This will help us understand the homomorphisms in the graph of groups $[X/G]$ explicitly, and also allow for a proof of the previous theorem once completed.

Recall we have a map

$$|X|//G := (|X| \times EG)/G \to X/G.$$

To consider what happens above a vertex v with edge e connecting to it, again take orbit representatives in X. Say we pick $\tilde{v}$ and $\tilde{e}$ so the edge connects to the vertex. There is an inclusion map $G_{\tilde{e}} \to G_{\tilde{v}}$ from which it's not hard to see this induces the attaching map, by considering the previous argument made to compute the fiber over e and v.

However, we need to do this in a consistent way in order to actually get attaching maps for the whole space.

Let T be a maximal spanning tree of X/G. We have the following useful result from Serre's book:

> THEOREM 2.27. Let G be a group acting without inversions on a tree X. Then we can lift a spanning tree of X/G to a tree in X.

Proof. Take Ω to be the set of all subtrees of X which map down injectively to our spanning tree T in X/G. This is non-empty, since we can pick a lift of a vertex of T. Suppose $\tilde{T}_0$ is a maximal element of Ω (which exists by applying Zorn's lemma), and has image $T_0 \subseteq T$. Then if the image is not equal to T, there is an edge $e \in T$ not belonging to T_0. This can be chosen, by connectedness, to have a boundary vertex v_0 in T_0 and one not in T_0. The edge e is the image of $\tilde{e} \in X$, one of whose vertices is a lift $\tilde{v_0}$ of v_0. Both $\tilde{v_0}$ and the lift of v_0 in $\tilde{T}_0$ lie over v_0, hence $g\tilde{v_0}$ yields the lift in $\tilde{T}_0$ for some $g \in G$. But then adjoining the edge $g \cdot \tilde{e}$ to $\tilde{T}_0$ extends the tree and injectively maps into T, contradicting maximality. $\square$

Note that this process also allows us to build a tree containing a particular lift of a vertex $v \in T$. From here on, take T to be a spanning tree of X/G. Picking vertices and edges on the lift $\tilde{T}$ as the representatives, we now have a consistent way to make attachings on T by simply using inclusions. This works because of the assumption that the action does not have inversions, which makes them actual subgroups. For other edges, the attaching map is induced by conjugation.

Let's make this precise. First, because G acts on X without inversions, **this is equivalent to some orientation of X being preserved by the action**. This means that all lifts of $\partial_1 e$ can be chosen to be $\partial_1\tilde{e}$, and similarly for ∂_0. We will implicitly be using this below.

For each edge $e \in (X/G) \setminus T$, there's a unique lift $\tilde{e} \in X$ with $\partial_0\tilde{e} = \widetilde{(\partial_0 e)} \in \tilde{T}$. Moreover, there's a unique $g_e \in G$ such that

$$g_e^{-1} \cdot \partial_1\tilde{e} = \widetilde{(\partial_1 e)} \in \tilde{T}.$$

Now we can define the attaching map. This shows that $G_{\tilde{e}}$ can be regarded as a subgroup of $G_{\widetilde{(\partial_0 e)}}$, so the attaching map $\psi_{e,0}$ of classifying spaces is just induced by the inclusion of groups. On the other hand, $\psi_{e,1}$ is conjugation by g_e ($g_e^{-1}\iota g_e$, where ι is inclusion), since up to the action it is a subgroup of $G_{\widetilde{(\partial_1 e)}}$.

Note that of course we are making the auxiliary choices (e.g. T and its lift). These of don't affect the homotopy type of $|X|//G$, as we are just looking at slightly different ways of writing down the same space.

Remark 2.28. Recall in a previous remark that $|X|//G$ was obtained as the geometric realization of the nerve of the action groupoid. The above construction can be viewed as taking an equivalence of groupoids, and then doing nerve and geometric realization. Functoriality of these constructions then shows the choice doesn't matter.

Proposition 2.29. Let G act on a tree X without inversions. Choose a maximal tree T of X/G, and a lift $\tilde{T}$ in X.

We can make a choice of attaching maps to realize $|X|//G$ giving the following graph of groups $[X/G]$:

- The group for a vertex $v \in X/G$ is given by $G_{\tilde{v}}$ where $\tilde{v}$ is the lift in $\tilde{T}$.
- For edges $e \in T$, the group is given by $G_{\tilde{e}}$.
- For edges $e \notin T$, we use $G_{\tilde{e}}$ for the unique lift $\tilde{e} \in X$ with $\partial_0\tilde{e} = \widetilde{(\partial_0 e)} \in \tilde{T}$.

The injective homomorphisms are as follows:

- For edges $e \in T$, they are inclusions.
- For edges $e \notin T$, $\psi_{e,0}$ is the inclusion map.
- For edges $e \notin T$, $\psi_{e,1}$ is conjugation by the unique $g_e \in G$ such that

$$g_e^{-1} \cdot \partial_1\tilde{e} = \widetilde{(\partial_1 e)} \in \tilde{T}.$$

With a description of the attaching maps in hand, we are nearly done.

Proof of main theorem. Suppose that G acts on a tree X without inversions. We have a map

$$|X| \times EG \to |X|//G.$$

Using that $|X| \times EG$ is contractible and the G-action is even, this is a covering map and $|X| \times EG$ is the universal cover. Just as in the case with a free action, we obtain $\pi_1(|X|//G) \simeq G$.

Above, we described $|X|//G$ as a graph of spaces by describing the combinatorial structure of the corresponding graph of groups in Proposition 2.29. It is here that we crucially use that the action does not have inversions. Knowing the attaching maps, we just need to put together what we know about amalgams and HNN extensions.

Initially, take the maximal spanning tree and compute its fundamental group. We know the result is $*_{v \in X/G} G_v / \sim$, where we adjoin relations $\psi_{e,1}(h) = \psi_{e,2}(h)$, since after adjoining each edge we can use Van Kampen's theorem (essentially how we dealt with graphs of groups like $* \to *$) to see we just take an amalgamated product.

Next, we need to attach the remaining edges. Each of these gives an HNN extension, since we are gluing a copy of BG_e to itself along the edge. The homomorphisms are ψ_{e_1} and $\psi_{e,2}$: these inject into the fundamental groups associated to the adjacent vertices, which inject into the fundamental group of the space so far. $\square$

We have now given a way to combinatorially write down G from its action on a tree, assuming it is without inversions. This accomplishes our main task!

2.6 Amalgams and fixed points

Recall we saw that in the particularly simple case of the graph of groups

$$\underset{G_v}{\bullet} \xrightarrow{G_e} \underset{G'_v}{\bullet}.$$

the fundamental group was $G_v *_{G_e} G_{v'}$. For many interesting examples of groups acting on trees, we actually get such a graph of groups as the result.

Let us give an example that you might see in the theory of modular forms.

> DEFINITION 2.30. We use $\mathrm{SL}_2(\mathbf{Z})$ to denote the group of integer 2×2 matrices with determinant one.

This group has an action on the complex upper half place $\mathbb{H} := \{z \in \mathbf{C} : \mathrm{im}(z) > 0\}$ by *linear fractional transformations*. That is,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az+b}{cz+d}.$$

Verifying manually by multiplying such fractions, you can check this is indeed a group action.

The group $\mathrm{SL}_2(\mathbf{Z})$ is generated by the matrices S and T, where

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

The fundamental domain for $\mathbb{H}/\mathrm{SL}_2(\mathbf{Z})$ is the strip $[-1/2, 1/2]$ intersected with $|z| \geq 1$. We call this $\mathcal{D}$. In particular,

$$\mathbb{H} = \bigcup_{\gamma \in \mathrm{SL}_2(\mathbf{Z})} \gamma\mathcal{D}$$

which is almost a partition except on the boundaries.

At first glance, we might think there is no connection to Bass-Serre theory because we see no obvious tree. However, take the segment $[i, e^{i\pi/3}]$ bordering the fundamental domain $\mathcal{D}$. The action of $\mathrm{SL}_2(\mathbf{Z})$ extends this to a tree X where it acts without inversions.

Because of how we constructed the tree, the graph of groups consists of two vertices connected by a single segment. We just need to compute the stabilizers on the boundary of the edge we began with, and the edge itself. We obtain

$$\underset{\mathbf{Z}/4\mathbf{Z}}{\bullet} \xrightarrow{\mathbf{Z}/2\mathbf{Z}} \underset{\mathbf{Z}/6\mathbf{Z}}{\bullet}.$$

Thus,

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

This is a form of SL_2 which is not obvious without the application of Bass-Serre theory. Moreover, noting that $\pm\mathrm{I}$ act trivially the action descents to $\mathrm{PSL}_2(\mathbf{Z})$, upon which we get

$$\mathrm{PSL}_2(\mathbf{Z}) \simeq \mathbf{Z}/2\mathbf{Z} * \mathbf{Z}/3\mathbf{Z}.$$

Example 2.31. Here's another neat application. Consider the commutator subgroup of $\mathrm{SL}_2(\mathbf{Z})$. From the definition of $\mathrm{SL}_2(\mathbf{Z})$, it is not at all obvious this is an index 12 subgroup which is free on two generators.

First, we note that in the tree $\mathrm{SL}_2(\mathbf{Z})$ acted on, any stabilizer of a point is conjugate to $\mathbf{Z}/4\mathbf{Z}$ or $\mathbf{Z}/6\mathbf{Z}$. Now the previous result shows that the abelianization is $\mathbf{Z}/12\mathbf{Z}$, by looking at the amalgamated product. If a nontrivial element of the commutator subgroup landed in one of these stabilizers, since they all inject into the abelianization we get a contradiction.

Thus, the commutator subgroup is free. The index is 12 due to the abelianization. Now, what is the rank of this free group? We compute the Euler characteristic of the graph $\mathrm{X}/[\mathrm{SL}_2(\mathbf{Z}), \mathrm{SL}_2(\mathbf{Z})]$.

Due to having index 12 and acting freely, the commutator subgroup $\Gamma = [\mathrm{SL}_2(\mathbf{Z}), \mathrm{SL}_2(\mathbf{Z})]$ turns the $\mathbf{Z}/6\mathbf{Z}$ endpoint of the Bass-Serre tree representing a single orbit into $12/6 = 2$ Γ-orbits. Similarly, the other endpoint becomes 3 Γ-orbits. The edge becomes 6 Γ-orbits. The Euler characteristic is then $2 + 3 - 6 = -1$, and so the rank is $1 - \chi = 2$.

However, of course not all groups can be written as a single amalgam.

Theorem 2.32. The following are equivalent for G a countable group which is not finite:

- (FA) $\mathrm{X}^{\mathrm{G}} \neq \emptyset$ for any tree X on which G acts.
- The group G is not an amalgam, and further has no quotient isomorphic to $\mathbf{Z}$ and is finitely generated.

We call the first item "Property FA", following Serre.

Sketch. It's easy to check that Property FA implies the second item: if G is an amalgam, write it as a fundamental group of a graph of groups and produce the tree as a sort of Cayley graph (like we do for $G_1 * G_2$). If there's a quotient isomorphic to $\mathbf{Z}$, use this to make G act by translations on an infinite chain. Finally, if we have property FA, since G is countable and not finite, it is an increasing union of finitely generated subgroups $G - 1 \subset \ldots \subset G_i \subset G_{i+1} \subset \ldots$. Take the set $\coprod_i G/G_i$ and make it into a tree by connecting elements of G/G_i to their image under

$$G/G_i \to G/G_{i+1}.$$

This is a tree with a natural G-action on each collections of cosets, hence by FA has a fixed point under G. This lies in some G/G_n, hence $G = G_n$. Thus, FA implies the group G is not an amalgam, and further has no quotient isomorphic to $\mathbf{Z}$ and is finitely generated.

What we've seen of Bass-Serre theory so far might make it seem like it is too good to be true that the second item implies (FA). Well, the condition that no quotient is $\mathbf{Z}$ actually already imposes a *large* restriction. Recall there's a map

$$|X|//G \to |X/G|.$$

This induces a map of fundamental groups. But $|X/G|$ is actually a graph, being the naive quotient. The induced map on fundamental groups, being surjective, reveals that $\mathbf{Z}^{*n}$ is naturally a quotient of G. It follows $T = X/G$ can't be a very interesting graph: it must be a tree.

The main theorem of Bass-Serre theory then greatly simplifies as the HNN extensions go away. We get $G = G_T$, which means it is already an iterated amalgam (recall G_T is just an iterated amalgam of G_v over $v \in T$ over the groups G_e). Finite generation means that there's a finite subtree T' for which $G = G_{T'}$. If this finite subtree is not a point, then we obtain G as an amalgam by removing any endpoint and performing the last amalgamated product. Then T' consists of a single vertex, and must in fact be a fixed point. □

We will use the property FA to show that $SL_3(\mathbf{Z})$ is *not* an amalgam, in contrast to $SL_2(\mathbf{Z})$.

A nice consequence which Serre proves is the following:

> Corollary 2.33. Suppose that G is a countable group which is not finite, and

> assume it satisfies property FA. Then if
>
> $$\rho : \mathrm{G} \to \mathrm{GL}_2(\mathbf{C})$$
>
> is a homomorphism, the eigenvalues of $\rho(g)$ are algebraic integers.

Proof. Let K be the subfield generated by $\rho(g)$ for $g \in \mathrm{G}$. By the theorem, the group is finitely generated so we get a finitely generated field extension of $\mathbf{Q}$. Complete K at some absolute value to obtain a p-adic field K_v. We'll see later in the course that $\mathrm{GL}_2(K_v)$ acts on a tree X without inversions. In particular, $\mathrm{GL}_2(K)$ also acts on this tree. There is a homomorphism

$$v \circ \det : \mathrm{GL}_2(K_v) \to \mathbf{Z}.$$

Let $\mathrm{GL}_2(K)^0$ denote the kernel of this in $\mathrm{GL}_2(K)$. There is no quotient of G isomorphic to $\mathbf{Z}$, so it follows the image of ρ is contained in $\mathrm{GL}_2(K)^0$ rather than just $\mathrm{GL}_2(K)$.

Now again $\mathrm{GL}_2(K)^0$ acts on the tree X. There is a vertex invariant under the action of action of G (acting via $\rho(\mathrm{G})$). We will see upon analysis of the tree's stabilizers (conjugate to $\mathrm{GL}_2(\mathcal{O}_v)$ that this means $\rho(\mathrm{G})$ is contained in a conjugate of $\mathrm{GL}_2(\mathcal{O}_v)$. It follows for each $g \in \mathrm{G}$, the coefficients of the characteristic polynomial lie in $\mathcal{O}_v$.

However, $\bigcap_v \mathcal{O}_v = \overline{\mathbf{Z}} \cap K$, where v runs over non-Archimedean absolute values. This is a commutative algebra fact beyond the scope of the course. However, it implies that the coefficients are algebraic integers, hence so are eigenvalues (solve for them from the trace and determinant). $\square$

You're probably used to things like this, given that for finite groups we get roots of unity. However, for infinite groups it's not so obvious. Note that the result lets us deduce this for $\mathrm{SL}_3(\mathbf{Z})$.

We'll now prove that $\mathrm{SL}_3(\mathbf{Z})$ has property FA, following Serre's method. Our method of attack will be to deduce fixed points of G from fixed points of generators s_i and their products $s_i s_j$. We will prove that certain types of elements in nilpotent (close to abelian) subgroups always have fixed points, and we can find such elements generating $\mathrm{SL}_3(\mathbf{Z})$. We'll need to do a bit more work to deduce the products $s_i s_j$ of these generators have fixed points, showing that in fact the entire nilpotent subgroups within $\mathrm{SL}_3(\mathbf{Z})$ we consider have fixed points. We can then find one containing any product of generators $s_i s_j$.

Let's first cover some basic facts about geodesics in trees we'll need for the proof. For points p and q in a tree X, consider paths between them without backtracking which are called *geodesics*. If the path is not injective (repeats vertices), then between repeated vertices we have a loop which contradicts being a tree. If the path is not unique, then taking geodesics γ and γ' there is a look $\gamma\overline{\gamma'}$. Here, we mean to follow one geodesic, and backtrack on the second.

Thus, between vertices p and q in a tree X there is a unique geodesic. Call the length of this geodesic $\ell(p, q)$. Furthermore, if G acts on a tree X without inversions and fixes p and q in X, then the geodesic joining them is also fixed. It follows X^G is always a tree.

If T_1 and T_2 are two disjoint subtrees of a tree X, there is a minimal distance d between vertices of these trees. There is also a unique geodesic $\gamma : p \to q$ joining the trees of length d, so we can speak of the geodesic γ joining two subtrees.

> Lemma 2.34. Let G act on a tree without inversions, and suppose that G is generated by elements $\{a_i\}$ and $\{b_j\}$. Let A and B be the subgroups generated by the a_i and b_j respectively, and make the following assumptions:
>
> - X^A is nonempty, as well as X^B.
> - The elements $a_i b_j$ have fixed points.
>
> Then X^G is non-empty.

Proof. We assume $X^A \neq \emptyset$ and has $X^B \neq \emptyset$ as well. Also by assumption, the elements $a_i b_j$ all have fixed points.

Observe that $X^G = X^A \cap X^B$, and suppose for the sake of contradiction the two trees X^A and X^B are disjoint. Let γ be the geodesic between the trees, joining vertices $P \to Q$. Let P_1 be a vertex which is one edge away from $P \to Q$ along γ. We know there is a generator $a_i \in A$ which does not fix this particular point, since otherwise all of A would. We now consider the geodesic $a_i\gamma : P \to a_i Q$, which will also contain $a_i P_1$ since group actions without inversions carry geodesics to geodesics. The fact that $a_1 P_1 \neq P_1$ shows there is no backtracking, so putting together $a_i Q \to P \to Q$ is a geodesic. The situation we have looks like the following:

The path $a_i Q \to Q$ shown above is the unique geodesic joining $a_i Q$ to Q. Now we will use the key assumption that $a_i b_j$ has a fixed point. Note that $a_i b_j Q = a_i Q$.

We claim that the midpoint of our path $a_i b_j Q \to Q$ then has a midpoint fixed by $a_i b_j$. Indeed, consider an arbitrary element s with a fixed point. There are unique geodesics $Q \to X^s$ and $X^s \to sQ$ (using the fact X^s is a tree), the latter obtained by applying s and reversing. The union of these can be seen to have no backtracking by looking

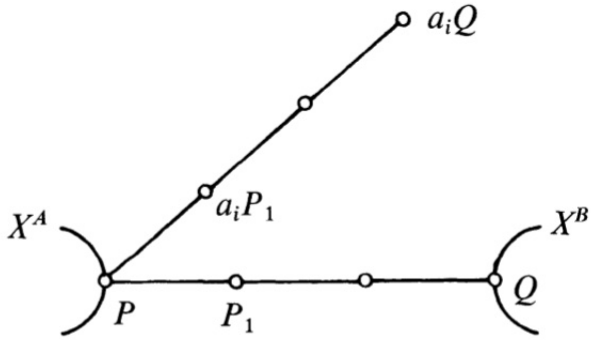


Figure 3: Diagram of the geodesics between the trees, from Serre's "Trees".

at the vertices neighboring X^s like we did with P_1: they cannot be the same vertex, otherwise they would belong to X^s.

Now use this observation on the present situation with $a_ib_jQ \to Q$, using the assumption that a_ib_j has a fixed point. The midpoint that is fixed must be P. Being fixed by the a_i already, we have $b_jP = P$ by acting by a_i^{-1}. It follows $P \in \mathrm{X}^{\mathrm{B}}$, which contradicts the trees being disjoint. Thus, X^{G} is nonempty. $\square$

A more practical form of this is the following, which is what we intend to use:

> Lemma 2.35. Let G act on a tree X without inversions. Then if s_i generate G and have fixed points, and s_is_j have fixed points as well, then X^{G} is nonempty.

Proof. Use induction on the number of generators, with $\mathrm{A} = \langle s_1, \ldots, s_{n-1}\rangle$ and with the new generator s_n generating B in the previous lemma. $\square$

Thus, we now have a possible way to extend fixed points to all of G.

We now focus on producing group elements that are ensured to have fixed points, by looking at nilpotent subgroups of G. A *nilpotent* group is one where the descending central series terminates to $\{1\}$ after finitely many steps.

Here, the lower central series is

$$\mathrm{G} \geq [\mathrm{G}, \mathrm{G}] \geq [[\mathrm{G}, \mathrm{G}], \mathrm{G}] \geq [[[\mathrm{G}, \mathrm{G}], \mathrm{G}], \mathrm{G}] \geq \ldots$$

where [G, G] means the subgroup generated by commutator elements $[a, b] := aba^{-1}b^{-1}$. A nilpotent group should be thought of as a group which is not too far from being abelian/commutative, its distance being measured by the length of the descending central series. For example, in an abelian group this series immediately terminates since [G, G] is generated only by the identity.

Nilpotent subgroups have the following property that allows us to far more easily construct fixed points.

> Theorem 2.36. Let G be a nilpotent group acting on a tree X without inversions. Then we have exactly one of the following:
>
> - $X^G \neq \emptyset$.
> - There is a straight path (infinite chain) T invariant under G on which G acts via translations, though a nontrivial homomorphism $G \to \mathbf{Z}$.

Sketch. These possibilities are mutually exclusive, since it turns out a group element s having no fixed points is equivalent to an infinite straight path T on which s induces a translation of nonzero amplitude. This is shown by Proposition 25 in section 6.4 of Serre's "Trees"; roughly, the idea is that if s has no fixed points then $P \neq sP$, so $m = \inf_P \ell(P, sP) > 0$. Taking T to be P such that $\ell(P, sP) = m$, this produces the desired infinite straight path. Conversely, if we already have the infinite straight path and the translation induced by s, we know from our previous argument that the geodesic $P \to sP$ has a midpoint in X^s if X^s is nonempty. But this simply cannot be the case for such a translation on an infinite path: the midpoint can be arbitrary within T if the translation is nontrivial, so in fact s fixes T and we get a contradiction.

Then we just need to show one of them must occur. This is where being nilpotent is actually used. Take a composition sequence

$$1 \leq G_1 \leq \ldots \leq G = G_n$$

where successive quotients are cyclic. If $n = 0$, we're done as G is trivial. Otherwise, apply the induction hypothesis on G_{n-1}. If G_{n-1} has a fixed point, use Lemma 2.34 to deduce X^G is nonempty. Namely, we look at the action of the cyclic group G/G_{n-1} on $X^{G_{n-1}}$.

If G_{n-1} does not have a fixed point, it has an infinite straight path T. As it is normal in G, the path is stable under G and acts by a homomorphism $G \to \mathrm{Aut}(T)$ containing a non-trivial group of translations in its image.

If G did not fix T, then we would have a line gT. This would have to be disjoint from T, since by normality $gG_{n-1}g^{-1} = G_{n-1}$, hence G_{n-1} still acts by the same nontrivial translation on the tree. The point of intersection would give a contradiction, as the point would be translated to two different points by the same group element. Now if we have two disjoint lines where G_{n-1} acts by a shift, the geodesic connecting two points on these lines is also shifted. But this creates a cycle, and so we can only have a single copy of T. It follows G fixes T, and we have a map $G \to \mathrm{Aut}(T)$ describing the action.

As T is a straight line, this rules out everything except $\mathbf{Z}$ and the infinite dihedral group (the full isometry group of a straight line), the latter of which is not nilpotent and so is ruled out. The claim then inductively follows. □

Now we come to the key result allowing us to produce fixed points.

> Lemma 2.37. Let G be a finitely generated nilpotent group acting without inversions on a tree X.
>
> - If $g \in G$ has a power $g^n \in [G, G]$, then it has a fixed point.
> - If G is generated by elements which have fixed points, it has a fixed point.

Proof. For the first item, if G has a fixed point we're done. Otherwise, we look to the infinite straight path T. By some power landing in $[G, G]$, the image of g is trivial under the homomorphism $G \to \mathbf{Z}$ used to act on T. Thus, T is fixed and g has a fixed point.

For the second item, we are again done if G has a fixed point so we look to the infinite straight path. If G is generated by some s_i, at least one has a nontrivial image under $G \to \mathbf{Z}$ acting on the tree. But then we know this is equivalent to s_i not having a fixed point, violating the hypothesis. We are therefore only in the first case of Theorem 2.36. □

> Theorem 2.38. $SL_3(\mathbf{Z})$ has property FA.

Proof. We will use the previous lemma to produce elements with fixed points, and then extend them to nilpotent subgroups with fixed points. This will allow us to apply the original criterion to deduce $X^{SL_3(\mathbf{Z})} \neq 0$ by finding generators s_i with fixed points, such that $s_i s_j$ also have fixed points (via showing subgroups containing s_i, s_j have fixed points).

Let $\mathrm{SL}_3(\mathbf{Z})$ act on a tree X without inversions. Let $i, j \in \{1, 2, 3\}$, so and let e_{ij} denote the elementary matrix which is zero except for the ij entry, which is one.

We have $\mathrm{SL}_3(\mathbf{Z}) = \langle 1 + e_{ij} \rangle_{i \neq j}$. These six generators are ordered as

$$z_1 = 1 + e_{13}, z_2 = 1 + e_{23}, z_3 = 1 + e_{21}$$

$$z_4 = 1 + e_{31}, z_5 = 1 + e_{32}, z_0 = z_6 = 1 + e_{12}.$$

Indexing over $\mathbf{Z}/6\mathbf{Z}$, z_i commutes with $z_{i\pm1}$. The commutator $[z_{i-1}, z_{i+1}]$ is $z_i^{\pm1}$, so in

$$\mathrm{B}_i := \langle z_{i-1}, z_{i+1} \rangle$$

the element z_i is in $[\mathrm{B}_i, \mathrm{B}_i]$. It follows that each z_i has a fixed point on X by the previous lemma. In particular, B_i is a nilpotent group generated by elements with fixed points so again by the previous lemma $\mathrm{X}^{\mathrm{B}_i} \neq \emptyset$.

It follows that z_i have fixed points, and $z_{i-1}z_{i+1}$ have fixed points as well (being elements of B_i).

Because of the commutator relations among the z_i, it follows $\mathrm{SL}_3(\mathbf{Z}) = \langle z_1, z_3, z_5 \rangle$. Moreover, these elements z_i all have fixed points; using the fact that all products of pairs of generators are of the form $z_{i-1}z_{i+1}$ for some i, we know all of these have fixed points. Now use Lemma 2.34 to conclude $\mathrm{X}^{\mathrm{SL}_3(\mathbf{Z})}$ is nonempty for any tree X. $\square$

Thus, $\mathrm{SL}_3(\mathbf{Z})$ is not an amalgam.

3 Applications to p-adic groups

Having sufficiently developed Bass-Serre theory, we will now apply it to understand the structure of p-adic groups.

I will again be loosely following Serre's book. One large difference will be that Serre works with skew fields (where multiplication need not commute, for example quaternions) rather than only fields. It's fine to do this since most properties of local skew fields we need are the same and the theory of linear algebra largely carries over, but for the sake of getting the point across with less confusion I'll avoid this.

We'll start with the construction of the Bruhat-Tits tree X, and proving some basic properties about it (e.g. an explicit description of the tree). This first part follows some of Bill Casselman's notes, which I recommend looking at: https://ncatlab.org/nlab/files/CasselmanOnBruhatTitsTree2014.pdf.

We will then study the action of $\mathrm{GL}_2(\mathbf{Q}_p)$ on the tree, seeing how a subgroup $\mathrm{GL}_2(\mathbf{Q}_p)^+$ acts without inversions. We'll see how many important subgroups of $\mathrm{GL}_2(\mathbf{Q}_p)$ reveal themselves as stabilizers under the $\mathrm{GL}_2(\mathbf{Q}_p)$ action on X. Then, using Bass-Serre theory, we'll write down certain subgroups $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^+$ as amalgams.

Finally, we'll prove Ihara's theorem. We'll interpret this in terms of p-adic Schottky groups, and discuss how you can construct some curves from these.

3.1 The Bruhat-Tits tree

We use $\mathbf{Q}_p$ to denote the p-adic numbers, and $\mathbf{Z}_p$ for the p-adic integers.

We will largely be concerned with the groups

$$\mathrm{GL}_2(\mathbf{Q}_p) : \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc \neq 0, a, b, c, d \in \mathbf{Q}_p \right\}$$

and $\mathrm{SL}_2(\mathbf{Q}_p)$ where $ad - bc = 1$. We will also study $\mathrm{PGL}_2(\mathbf{Q}_p)$, the quotient of $\mathrm{GL}_2(\mathbf{Q}_p)$ by the scalar matrices $\lambda \mathrm{I}_2$.

Both $\mathrm{GL}_2(\mathbf{Q}_p)$ and $\mathrm{SL}_2(\mathbf{Q}_p)$ have open compact subgroups $\mathrm{GL}_2(\mathbf{Z}_p)$ and $\mathrm{SL}_2(\mathbf{Z}_p)$. These groups act on lattices, which we will now study.

Definition 3.1. A lattice in $\mathbf{Q}_p^2$ is a finitely generated $\mathbf{Z}_p$-submodule of $\mathbf{Q}_p^2$ which spans it as a vector space.

For example, $\mathbf{Z}_p^2$ is a lattice.

Lemma 3.2. Every lattice in $\mathbf{Q}_p^2$ is equivalent to $\mathbf{Z}_p^2$ via a change of basis. Specifically, we can write them all as

$$\mathbf{Z}_p \alpha + \mathbf{Z}_p \beta$$

for $\alpha, \beta \in \mathbf{Q}_p$.

The following definition will make vertices of the Bruhat-Tits tree.

Definition 3.3. We say lattices L_1 and L_2 are equivalent if

$$L_1 = \lambda L_2$$

for $\lambda \in \mathbf{Q}_p$.

The group $\mathrm{GL}_2(\mathbf{Q}_p)$ acts on a lattice by applying the linear operator to each element of the lattice. The action is transitive, as we can send basis elements anywhere we want. We see that $\mathrm{PGL}_2(\mathbf{Q}_p)$ acts on lattices up to equivalence since the scalar matrices in $\mathrm{GL}_2(\mathbf{Q}_p)$ don't change an equivalence class. The subgroup

$$\mathrm{K} = \mathrm{PGL}_2(\mathbf{Z}_p)$$

fixes the class of the standard lattice $\mathbf{Z}_p^2$. We claim it is precisely the stabilizer of it.

> Lemma 3.4. The matrices in $\mathrm{GL}_2(\mathbf{Q}_p)$ fixing the equivalence class of the standard lattice $\mathbf{Z}_p^2$ are precisely those in $\mathrm{GL}_2(\mathbf{Z}_p)$.

Proof. First, we need to understand which pairs of elements generate standard lattice. Take a choice of generators e_1, e_2. If we have another generating pair $ae_1 + be_2$ and $ce_1 + de_2$, to produce e_1 and e_2 as a $\mathbf{Z}_p$-linear combination asks that $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has an inverse with entries in $\mathbf{Z}_p$. Thus, other generators are just pairs $g \cdot e_1, g \cdot e_2$ for $g \in \mathrm{GL}_2(\mathbf{Z}_p)$. □

When we move to $\mathrm{PGL}_2(\mathbf{Q}_p)$, we get $\mathrm{PGL}_2(\mathbf{Z}_p)$ as the stabilizer of the class of the standard lattice. In $\mathrm{GL}_2(\mathbf{Q}_p)$, matrices besides those in $\mathrm{GL}_2(\mathbf{Z}_p)$ can stabilize the class of the standard lattice: we also have the matrices $\lambda \mathrm{I}$ for $\lambda \in \mathbf{Q}_p^\times$.

Now we further can note that $\mathrm{PGL}_2(\mathbf{Q}_p)$ acts transitively on lattice classes in $\mathbf{Q}_p^2$. As a result, equivalent lattices are in bijection with

$$\mathrm{PGL}_2(\mathbf{Q}_p)/\mathrm{PGL}_2(\mathbf{Z}_p).$$

These will form the vertices of the tree. For edges, we need to define a notion of distance.

> Lemma 3.5. Given two lattices Λ_1 and Λ_2, we may find a basis
>
> $$\Lambda_1 = \mathbf{Z}_p v \oplus \mathbf{Z}_p w$$
>
> such that $\Lambda_2 = \mathbf{Z}_p p^n v \oplus \mathbf{Z}_p p^m w$ for integers m and n. We may assume $n \geq m$.

Proof. This claim is equivalent to the Cartan decomposition

$$\mathrm{GL}_2(\mathbf{Q}_p) = \coprod_{a \in \mathrm{A}} \mathrm{GL}_2(\mathbf{Z}_p) \cdot a \cdot \mathrm{GL}_2(\mathbf{Z}_p),$$

where A consists of matrices $\begin{pmatrix} p^n & 0 \\ 0 & p^m \end{pmatrix}$ with $n \geq m$. This is often called a KAK decomposition, with $k \in \mathrm{K}$ corresponding to $\mathrm{GL}_2(\mathbf{Z}_p)$.

Assuming this, without loss of generality Λ_1 is the standard lattice and $g \cdot \Lambda_1 = \Lambda_2$. Writing g in this form $k_1 a k_2$, since $\mathrm{GL}_2(\mathbf{Z}_p)$ fixes the underlying lattice (just changes the basis) we can actually just apply a to the basis. $\square$

We define $d(\Lambda_1, \Lambda_2) = |n - m|$. This is constant in an equivalence class of lattices, and so is well-defined on the classes $[\Lambda_1], [\Lambda_2]$.

> Definition 3.6. The set $\mathrm{X} = \mathrm{PGL}_2(\mathbf{Q}_p)/\mathrm{PGL}_2(\mathbf{Z}_p)$ can be upgraded to a graph by defining an edge between lattice classes in $[\Lambda]$ and $[\Lambda']$ if $d([\Lambda], [\Lambda']) = 1$. This is called the *Bruhat-Tits tree*.

Thus, we have now defined a graph. The action of $\mathrm{GL}_2(\mathbf{Q}_p)$ on a lattice class sends neighboring classes to neighboring classes, since the values of n and m are preserved in our bases for $g\Lambda_1$ and $g\Lambda_2$. Thus, we get an action on the graph.

> Theorem 3.7. The graph X is a tree, and moreover it is $p+1$-regular.

Proof. First, we justify why X is a $(p+1)$-regular graph. The distance between classes $[\Lambda_1]$ and $[\Lambda_2]$ is 1 if and only if we can pick representative lattices Λ_1 and Λ_2 such that

$$p\Lambda_1 \subseteq \Lambda_2 \subseteq \Lambda_1.$$

If the distance is one, then we can write in some basis $\Lambda_1 = \mathbf{Z}_p v \oplus \mathbf{Z}_p w$ and $\Lambda_2 = p^{n+1}\mathbf{Z}_p v \oplus p^n \mathbf{Z}_p w$. Then by scaling, we can instead pick a representative $\Lambda_2' \in [\Lambda_2]$ given by $p\mathbf{Z}_p v \oplus \mathbf{Z}_p w$. In this case, the desired containments hold. Conversely, putting both lattices into normal form (v, w) for Λ_1 and $(p^n v, p^m w)$ for Λ_2, for $\Lambda_2 \subseteq \Lambda_1$ we need $m \geq 0$. If $p\Lambda_1 \subseteq \Lambda_2$, $n \leq 1$. The claim then follows.

All such classes $[\Lambda_2]$ can be found by enumerating lines in $\Lambda_1/p\Lambda_1 \simeq \mathbf{F}_p^2$. The reason we consider lines is simply that we want to consider the class and not the specific representative, so we must consider Λ_2 and $c\Lambda_2$ the same for $c \in \mathbf{Q}_p$. These lines are given by points of $\mathbf{P}^1(\mathbf{F}_p)$, of which there are $p+1$.

Next, we will show that X is a tree. Given a particular lattice Λ_1, each class $[\Lambda_2] \in \mathrm{X}$ has precisely one representative Λ_2 satisfying

$$\Lambda_2 \subseteq \Lambda_1$$

and Λ_1/Λ_2 is *monogenic*, or cyclic as a $\mathbf{Z}_p$-module. Before, we asked for this to be $\mathbf{Z}_p/p\mathbf{Z}_p \simeq \mathbf{F}_p$. The justification is similar to before, we just put both in the normal form. However, this applies for lattices of arbitrary distance $d(\Lambda_1, \Lambda_2)$. If the quotient is $\mathbf{Z}/p\mathbf{Z}$, then the two classes are neighbors: again use the normal form.

Now we show X is connected first. Given $[\Lambda_1]$ and $[\Lambda_2]$, pick representatives as done above to get

$$\Lambda_2 \subseteq \Lambda_1$$

and $\Lambda_1/\Lambda_2 \simeq \mathbf{Z}/p^d\mathbf{Z}$. Taking a Jordan-Holder sequence for this $\mathbf{Z}_p$-module we get lattices

$$\Lambda_2 = L_n \subset L_{n-1} \subset \ldots \subset L_0 = \Lambda_1.$$

Here, the successive quotients are $\mathbf{Z}/p\mathbf{Z}$ so we get from the first argument that each is distance one from the next. This produces a path.

Finally, we will see that we indeed have a tree. Suppose we have a path $[\Lambda_0], \ldots, [\Lambda_n]$ of classes of adjacent vertices in X, and that it has no backtracking. We want to show this is not a circuit.

First, pick representatives Λ_i so

$$\Lambda_n \subset \ldots \subset \Lambda_{i+1} \subset \Lambda_i \subset \ldots \Lambda_0$$

and $\Lambda_i/\Lambda_{i+1} \simeq \mathbf{Z}/p\mathbf{Z}$ (equivalently, $p\Lambda_i \subset \Lambda_{i+1} \subset \Lambda_i$). This chain can be turned into the standard chain

$$\Lambda_n = p^n\mathbf{Z}_p v \oplus \mathbf{Z}_p w \subset \ldots \subset p\mathbf{Z}_p v \oplus \mathbf{Z}_p w \subset \mathbf{Z}_p v \oplus \mathbf{Z}_p w = \Lambda_0$$

by applying an element of $\mathrm{GL}_2(\mathbf{Q}_p)$. This can be shown inductively: for a path of length one it's just our normal form. Assuming we can do it for a path of length n, do it for $n+1$ by taking a matrix

$$g \in \begin{pmatrix} 1 & p^n\mathbf{Z}_p \\ 0 & 1 \end{pmatrix}.$$

Let's see how to make this matrix. We know $p\Lambda_n \subset \Lambda_{n+1} \subset \Lambda_n$ corresponds to a line in $\mathbf{P}(\Lambda_n/p\Lambda_n)$, and also $\Lambda_{n+1} \neq p\Lambda_{n-1} = p^n\mathbf{Z}_p v \oplus p\mathbf{Z}_p w$ as this creates backtracking.

In $\mathbf{F}_p^2 \simeq \Lambda_n/p\Lambda_n$, we can take the basis given by images of $p^n v$ and w. The lattice $p\Lambda_{n-1}$ then maps to the line through $(1,0)$. Then Λ_{n+1}, being a different line, corresponds to a line through some $(y,1)$, and we use $\begin{pmatrix} 1 & -y \\ 0 & 1 \end{pmatrix}$ to map it to the line through $(0,1)$. Now use $p^n x$ for $x \in \mathbf{Z}_p$ in the matrix, where $\overline{x} = -y$ when reduced modulo p. This will definitely put Λ_{n+1} in the desired form, and a matrix of this form preserves $\Lambda_0, \ldots, \Lambda_n$ in the standard chain.

At this point, we have seen any path without backtracking can be turned into the standard chain via the $\mathrm{GL}_2(\mathbf{Q}_p)$ action. However, this action is an action on the graph X, so it follows that our original path was a circuit if and only if the standard chain is a circuit. It is not, so the claim follows. $\square$

Before we move to studying carefully the action of $\mathrm{GL}_2(\mathbf{Q}_p)$ on X, we will want to understand how X connects to the story of lattices over $\mathbf{R}$.

Before, we looked at the action of $\mathrm{PGL}_2(\mathbf{R})$ by linear fractional transformations on $\mathbf{C} \setminus \mathbf{R}$, the union of the upper and lower half planes. We can equivalently understand this as

$$\mathbf{P}^1(\mathbf{C}) \setminus \mathbf{P}^1(\mathbf{R}),$$

since this just adds points at infinity which we get rid of. A better way to think about this is that $\mathbf{P}^1(\mathbf{C})$ parameterizes equivalence classes of lattices $\mathbf{Z}^2 \to \mathbf{C}$ up to multiplication by a scalar. Note that this is not just classifying maps, since identifying $[v : w]$ with $\mathbf{Z}v \oplus \mathbf{Z}w$ picks a basis; no distinction is made between the same lattice with different bases.

However, we don't want all lattices. We want just the rank two ones. This corresponds to removing $\mathbf{P}^1(\mathbf{R})$, so $\mathbf{P}^1(\mathbf{C}) \setminus \mathbf{P}^1(\mathbf{R})$ classifies rank two $\mathbf{Z}$-lattices in $\mathbf{C}$.

> Remark 3.8. This is a really natural condition to think about lattices with respect to in this setting: $\mathbf{C}/\Lambda$ for a rank two $\mathbf{Z}$-lattice Λ is a complex torus. As a Riemann surface, these are isomorphic when $[\Lambda_1] = [\Lambda_2]$, the brackets denoting the equivalence class of the lattice up to multiplication by $c \in \mathbf{C}^\times$.

In the p-adic world, we have a thing called the *Drinfeld upper half plane*. This has its points given by

$$\Omega_{\mathbf{C}_p} := \mathbf{P}^1(\mathbf{C}_p) \setminus \mathbf{P}^1(\mathbf{Q}_p),$$

where $\mathbf{C}_p$ is the completion of the algebraic closure of $\mathbf{Q}_p$, or $\widehat{\overline{\mathbf{Q}}}_p$. This is a complete algebraically closed field, like $\mathbf{C}$, and is in fact abstractly isomorphic to $\mathbf{C}$. The group $\mathrm{PGL}_2(\mathbf{Q}_p)$ acts on $\Omega_{\mathbf{C}_p}$.

In particular, points of $\mathbf{P}^1(\mathbf{C}_p) \setminus \mathbf{P}^1(\mathbf{Q}_p)$ correspond to rank two $\mathbf{Q}_p$-lattices in $\mathbf{C}_p$.

There is a natural $\mathrm{PGL}_2(\mathbf{Q}_p)$-equivariant map

$$\Omega_{\mathbf{C}_p} \to \mathrm{X}$$

sending a lattice in $\mathbf{C}_p$ to a norm on $\mathbf{Q}_p^2$ given by restricting the valuation on $\mathbf{C}_p$ to the image of the injective map $\mathbf{Q}_p^2 \to \mathbf{C}_p$.

Remark 3.9. I skimmed over this in class because I didn't want to get too into the details (as $\mathbf{C}_p$ can be quite confusing, and understandably so). Here is a more in depth explanation of how to make the valuation

$$\nu_{\mathbf{C}_p} : \mathbf{C}_p \to \mathbf{R} \cup \infty.$$

First, we'll figure out how to do this for $\overline{\mathbf{Q}_p}$. It is a theorem there is a unique extension of $\nu_p : \mathbf{Q}_p \to \mathbf{Z} \cup \infty$ to any finite extension $K/\mathbf{Q}$. The unique extension is explicitly given by

$$\nu_K := \frac{1}{[K : \mathbf{Q}_p]}(\nu_p \circ \mathbf{N}_{K/\mathbf{Q}_p})$$

where $\mathbf{N}_{K/\mathbf{Q}_p}$ denotes the field norm. This is given by the determinant $\det(y)$ of the linear map $x \mapsto yx$ on K as a $\mathbf{Q}_p$-vector space for an element $y \in K$.

Now $\overline{\mathbf{Q}}_p$ is the union of all finite extensions (for pendants: direct limit with respect to inclusions). We extend to $\nu_{\overline{\mathbf{Q}}_p} : \overline{\mathbf{Q}}_p \to \mathbf{Q} \cup \infty$ accordingly: to evaluate on x, find an extension it lies in and use the norm there. The choice of extension does not matter.

Now to move to $\mathbf{C}_p$, we take the completion. This does not change the image of the valuation: it is still $\mathbf{Q}\cup\infty$, and not $\mathbf{R}$. This is because given a Cauchy sequence of elements x_i in $\overline{\mathbf{Q}}_p$, we define $\nu_{\mathbf{C}_p}$ to be the limit of the valuations $\nu_{\overline{\mathbf{Q}}_p}$ in $\mathbf{Q}\cup\infty$. This is a limit entirely within $\mathbf{Q}$ and the valuation is *non-Archimedean*, so it just becomes eventually constant and remains within $\mathbf{Q} \cup \infty$. It does not take values in all of $\mathbf{R} \cup \infty$, as you might initially believe. This defines $\nu_{\mathbf{C}_p}$, which is what we needed.

To see that our map $\Omega_{\mathbf{C}_p} \to$ X actually lands in X, we need to reinterpret X in terms of norms.

Theorem 3.10. The topological space of X has its points naturally in bijection with the set of equivalence classes of norms on $\mathbf{Q}_p^2$.

A norm on $\mathbf{Q}_p^2$ is a function $\gamma : \mathbf{Q}_p^2 \to \mathbf{R} \cup \infty$ such that $\gamma(x) = \infty$ iff $x = 0$, $\gamma(cx) = \nu(c) + \gamma(x)$ for $c \in \mathbf{Q}_p$ and $\gamma(x + y) \geq \inf(\gamma(x), \gamma(y))$. They are equivalent if $\gamma - \gamma' = C$ for a constant C.

Given a vertex $[\Lambda]$ set

$$\gamma(x) := -\inf\{n \in \mathbf{Z} : p^n x \in \Lambda\}$$

for a representative Λ. For edges, if our point is $(1-t)[\Lambda] + t[\Lambda']$, picking $\Lambda = \mathbf{Z}_p v \oplus \mathbf{Z}_p w$ and $\Lambda' = p\mathbf{Z}_p v \oplus \mathbf{Z}_p w$, set

$$\gamma(av + bw) = \inf(\nu(a) - t, \nu(b)).$$

This interpolates between what happens at the edges; for example, $t = 0$ gives an equivalent way to write the norm at $[\Lambda]$.

> Remark 3.11. It might be tempting to say there should be some sort of map from $\mathbf{P}^1(\mathbf{C})/\mathbf{P}^1(\mathbf{R})$ to a tree that one uses to study $\mathrm{PGL}_2(\mathbf{R})$. However, the existence of such a tree is very much a p-adic phenomenon. Note that in defining the norm on a lattice, we needed the non-Archimedean nature of $\mathbf{Q}_p$ (being able to use $\inf(\nu(a), \nu(b))$) to get exactly what the norm should be, rather than some inequality.
>
> One can think of X as describing what the reduction modulo p of $\Omega_{\mathbf{C}_p}$ looks like. Finding an analogous graph for the upper half plane would be asking for $\mathbf{R}$ to have a residue field.

The Drinfeld upper half space $\Omega_{\mathbf{C}_p}$ can be thought of as a "tubular neighborhood" of the tree X. This can be a useful perspective for combinatorially understanding quotients of $\Omega_{\mathbf{C}_p}$ in terms of the tree X.

3.2 Structure theory

Unfortunately, the group $\mathrm{GL}_2(\mathbf{Q}_p)$ acts by inversions on the Bruhat-Tits tree X. As an example, consider the path $[\mathbf{Z}_p \oplus \mathbf{Z}_p] \to [p\mathbf{Z}_p \oplus \mathbf{Z}_p]$. The matrix

$$\begin{pmatrix} 0 & p \\ 1 & 0 \end{pmatrix}$$

swaps these two lattice classes, therefore creating an inversion.

However, it is true that an index two subgroup $\mathrm{GL}_2(\mathbf{Q}_p)^+$ does act without inversions.

Definition 3.12. The group $\mathrm{GL}_2(\mathbf{Q}_p)^+$ is defined as the kernel of

$$\mathrm{GL}_2(\mathbf{Q}_p) \xrightarrow{\det} \mathbf{Q}_p \xrightarrow{\nu} \mathbf{Z} \longrightarrow \mathbf{Z}/2\mathbf{Z}.$$

We use $\mathrm{GL}_2(\mathbf{Q}_p)^0$ to denote the kernel of $\mathrm{GL}_2(\mathbf{Q}_p) \to \mathbf{Z}$. There are then inclusions

$$\mathrm{SL}_2(\mathbf{Q}_p) \subset \mathrm{GL}_2(\mathbf{Q}_p)^0 \subset \mathrm{GL}_2(\mathbf{Q}_p)^+.$$

We will show $\mathrm{GL}_2(\mathbf{Q}_p)^+$ acts without inversions, so that all of its subgroups also do.

For a finite $\mathbf{Z}_p$-module M, let $\ell(M)$ denote the length as a $\mathbf{Z}_p$-module (the maximal length of a chain of submodules). We will apply this to quotients of $\mathbf{Z}_p$-lattices.

Definition 3.13. Let Λ_1 and Λ_2 be lattices. We define

$$\chi(\Lambda_1, \Lambda_2) = \ell(\Lambda_1/\Lambda_3) - \ell(\Lambda_2/\Lambda_3).$$

This does not depend on the choice of $\Lambda_3 \subseteq \Lambda_1 \cap \Lambda_2$. You should think of this as measuring the difference in size between these lattices (which is literally true in the sense of Haar measures with respect to $\mu(\Lambda_3) = 1$).

Let ν denote the p-adic valuation. As it turns out, the following is true:

Lemma 3.14. We have

$$\chi(\Lambda, g\Lambda) = \nu(\det(g)).$$

In particular, this only depends on $g \in \mathrm{GL}_2(\mathbf{Q}_p)$.

Proof. Put $g\Lambda$ in the normal form $\mathbf{Z}_p p^n v \oplus \mathbf{Z}_p p^m w$ where $\Lambda = \mathbf{Z}_p v \oplus \mathbf{Z}_p w$. The claim follows by computing both as $m+n$: $\nu(\det(g))$ is clear, since in the chosen basis we have $g = ak$ where a is the diagonal matrix with entries p^n and p^m and $k \in \mathrm{GL}_2(\mathbf{Z}_p)$. The determinant of the latter is a p-adic unit in $\mathbf{Z}_p^\times$, and hence has valuation 0. We are then left with $m+n$.

Next, for $\chi(\Lambda, g\Lambda)$ we can take Λ_3 to be $p^{\min(0,a)}\mathbf{Z}_p \oplus p^{\min(0,b)}\mathbf{Z}_p$. A similar computation proves the result. $\square$

We saw $\chi(\Lambda, g\Lambda) = m+n$ if we put $g\Lambda$ in the normal form $\mathbf{Z}_p p^n v \oplus \mathbf{Z}_p p^m w$ where $\Lambda = \mathbf{Z}_p v \oplus \mathbf{Z}_p w$. Now note that

$$d(\Lambda, g\Lambda) = |a-b| \equiv a+b \pmod 2.$$

We then have $\nu(\det(g)) \equiv d(\Lambda, g\Lambda) \pmod 2$.

Because of this, if $g \in \mathrm{GL}_2(\mathbf{Q}_p)^+$ then the distance between Λ and $g\Lambda$ is always even. It follows that we cannot have an inversion.

Since we want to directly apply Bass-Serre theory eventually, we will now restrict ourselves to subgroups of $\mathrm{GL}_2(\mathbf{Q}_p)^+$, and further to subgroups of $\mathrm{GL}_2(\mathbf{Q}_p)^0$. In particular, all results apply to $\mathrm{SL}_2(\mathbf{Q}_p)$.

All of the subgroups we will study arise as various stabilizers in the Bruhat-Tits tree. The first we will look at are stabilizers of vertices, which we have already looked at in the case of $\mathrm{GL}_2(\mathbf{Q}_p)$. There, we found the stabilizer of the standard lattice is $\mathrm{GL}_2(\mathbf{Z}_p)$ (although note that the stabilizer of the lattice *class* is $\mathbf{Q}_p^\times \mathrm{GL}_2(\mathbf{Z}_p)$). For a particular lattice, the stabilizers are conjugates of $\mathrm{GL}_2(\mathbf{Z}_p)$, so we get all maximal compact subgroups of $\mathrm{GL}_2(\mathbf{Q}_p)$.

In X, instead of individual lattices we look at classes lattices. It turns out these have good behavior when we look at subgroups of $\mathrm{GL}_2(\mathbf{Q}_p)^0$.

> Lemma 3.15. If $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^0$, then $\mathrm{Stab}_\Lambda = \mathrm{Stab}_{[\Lambda]}$.

This says that the stabilizers behave differently from the entire action of $\mathrm{GL}_2(\mathbf{Q}_p)$, where it is different between the lattice and the lattice class.

Proof. Pick a representative Λ of the class $[\Lambda]$. If there is a group element $g \in \mathrm{G}$ such that $g\Lambda = c \cdot \Lambda$ for $c \in \mathbf{Q}_p^\times$, then from its definition

$$\chi(\Lambda, g\Lambda) = 2\nu(c).$$

As $g \in \mathrm{GL}_2(\mathbf{Q}_p)^0$ we see that $2\nu(c) = \nu(\det g) = 0$ and hence $\nu(c) = 0$.

Thus, $c \in \mathbf{Z}_p^\times$. It follows $c\Lambda = \Lambda$. $\square$

A nice consequence of this is the following. A bounded subgroup is one where all $g \in \mathrm{G}$ have matrix entries g_{ij} with $\nu(g_{ij}) \leq d$ for some fixed integer d.

> Corollary 3.16. The maximal bounded subgroups of $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^0$ are precisely the stabilizers of vertices in X when acted upon by G.

Proof. It suffices to show that $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^0$ is bounded if and only if G leaves a vertex $x \in \mathrm{X}$ fixed. Indeed, if this is the case then bounded subgroups of $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^0$ are precisely those which have nontrivial fixed points. The maximal subgroups with nontrivial fixed points are stabilizers, since any subgroup with a fixed point is contained within the stabilizer of that fixed point by definition.

Now we show the revised claim. First, if $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^0$ is bounded then we can find a lattice stable under G. Pick a lattice $\Lambda_0 \subseteq \mathbf{Q}_p^2$. Putting $\Lambda = \sum_{g \in \mathrm{G}} g\Lambda_0$, the fact that G is bounded means this remains a lattice as everything lands inside of $p^{-d}\Lambda_0$ for some d (so in particular, it is a finitely generated, as it lands inside of a finitely generated $\mathbf{Z}_p$-module).

By construction it is stable under G. It follows that for bounded G, it fixes a particular lattice. By the lemma, fixing a lattice is the same as fixing a lattice class. Hence, bounded G have fixed points.

On the other hand, suppose that G has a fixed point on X. That means there is a lattice Λ stable under G, by the lemma again. If there is a lattice stable under G, it is immediate that G is bounded, since it is contained within a conjugate of $\mathrm{GL}_2(\mathbf{Z}_p)$. $\square$

Thus, we see stabilizers of vertices give us important information about subgroups of general $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^0$. Next, we look at stabilizers of edges.

> Proposition 3.17. Let e be an edge of X with bounding vertices represented by
>
> $$p\Lambda \subset \Lambda' \subset \Lambda$$
>
> and $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^0$. Then
>
> $$\mathrm{Stab}(e) = \{g \in \mathrm{Stab}([\Lambda]) : \overline{g} \in \mathrm{GL}(\Lambda/p\Lambda) \text{ fixes } \Lambda'/\Lambda\}.$$

Since the stabilizer of the lattice classes agree with the stabilizers of the actual lattices, the stabilizer of the edge can be identified with $\mathrm{Stab}(\Lambda) \cap \mathrm{Stab}(\Lambda')$. Using the inclusions

$$p\Lambda \subset \Lambda' \subset \Lambda$$

we see additionally stabilizing Λ' just means the reduction stabilizes the corresponding line.

To give an explicit example, consider $\mathrm{G} = \mathrm{SL}_2(\mathbf{Q}_p)$. Then we get what's called an Iwahori subgroup, conjugate to those consisting of matrices

$$\begin{pmatrix} a & b \\ pc & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}_p).$$

These Iwahori subgroups are quite important for number-theoretic applications, and also play an important role in the structure of the group. With $\mathrm{GL}_n(\mathbf{C})$, we have a decomposition

$$\mathrm{GL}_n(\mathbf{C}) = \coprod_{\sigma \in S_n} \mathrm{B}\sigma\mathrm{B}$$

where σ denotes a permutation matrix, and B is the subgroup of upper triangular matrices. We have a decomposition for p-adic groups that is similar in nature, where B is replaced by an Iwahori subgroup. This is called the Iwahori decomposition.

Finally, we come to two additional types of subgroups that can be seen with stabilizers: Borel and Cartan subgroups.

> Definition 3.18. An *end* b of X is an infinite path with no backtracking.
>
> An element $g \in \mathrm{GL}_2(\mathbf{Q}_p)$ leaves an end $b = [\Lambda_0], [\Lambda_1], \ldots$ invariant if $g[\Lambda_i] = [\Lambda_{i+d}]$ for some fixed integer d and $i \gg 0$.

Ends of X can be neatly understood as elements of $\mathbf{P}^1(\mathbf{Z}_p) = \mathbf{P}^1(\mathbf{Q}_p)$. To see this, observe that we saw before that neighbors of a vertex are in bijection with $\mathbf{P}^1(\mathbf{Z}/p\mathbf{Z})$. Extending this, vertices which are distance d away are parameterized by $\mathbf{P}^1(\mathbf{Z}/p^d\mathbf{Z})$. This is because points distance d away are given by Λ_1 where $p^d\Lambda_0 \subset \Lambda_1 \subset \Lambda_0$ after picking a representative Λ_0 (this is precisely the monogenic quotient condition, just rephased). These correspond to lines in $\mathbf{Z}/p^d\mathbf{Z}$.

Now, what does it mean for these choices to be compatible? Well, we need an infinite chain of sublattices of Λ_0, and without backtracking. These are enumerated by elements of

$$\mathbf{P}^1(\mathbf{Z}_p) \simeq \varprojlim \mathbf{P}^1(\mathbf{Z}/p^d\mathbf{Z}) \simeq \mathbf{P}^1(\varprojlim \mathbf{Z}/p^d\mathbf{Z}).$$

I want to explain precisely what this notation means. An inverse limit of groups takes a diagram like

$$\ldots \to \mathbf{Z}/p^n\mathbf{Z} \to \mathbf{Z}/p^{n-1}\mathbf{Z} \to \ldots \to \mathbf{Z}/p\mathbf{Z}$$

where each map is reduction modulo p^{i-1} from $\mathbf{Z}/p^i\mathbf{Z}$ and produces a group G. The group G is characterized by a universal property:

Definition 3.19. We say

$$\mathrm{G} = \varprojlim \mathrm{G}_i$$

for a system of groups G_i with maps $f_i : \mathrm{G}_i \to \mathrm{G}_{i-1}$ is the *inverse limit* of the G_i if:

- It has maps $\pi_i : \mathrm{G} \to \mathrm{G}_i$ such that $f_i \circ \pi_i = \pi_{i-1}$
- It is universal with respect to this property.

This second condition means there is a *unique* homomorphism from G to any other G' equipped with such maps.

Remark 3.20. Confusingly, an inverse limit is actually a limit in Grp (and not a colimit) in the category theory sense.

Practically, $\varprojlim_i \mathrm{G}_i$ can be described as

$$\mathrm{G} = \{g \in \prod_i \mathrm{G}_i : g_{i-1} = f_i(g_i)\}.$$

This has maps $\pi_i : \mathrm{G} \to \mathrm{G}_i$ given by projection to each G_i. By definition, the desired diagram commutes: $f_i \circ \pi_i(g) = \pi_{i-1}(g) = g_{i-1}$.

In our case, we see this description precisely describes a p-adic integer, by asking that there are compatibilities in all the quotients $\mathbf{Z}_p/p^n\mathbf{Z}_p \simeq \mathbf{Z}/p^n\mathbf{Z}$. This just means that it has a well-defined expansion as a p-adic integer: the projection maps a p-adic integer to the truncated p-adic expansion $\sum_{i=0}^n c_i p^i$, and the compatibilities are checking that the coefficients for lower powers of p remain the same.

We get $\mathbf{P}^1(\mathbf{Z}_p) \simeq \varprojlim \mathbf{P}^1(\mathbf{Z}/p^d\mathbf{Z})$ by asking that there is a chain of adjacent vertices distance one apart; the element of $\mathbf{P}^1(\mathbf{Z}/p^d\mathbf{Z})$ enumerates the vertex at distance d. If reducing modulo p^{d-1} gives the previous vertex, they are adjacent. Thus, we arrive at the following result:

> LEMMA 3.21. Ends of X based at the standard lattice $[\mathbf{Z}_p^2]$ are naturally enumerated by elements of $\mathbf{P}^1(\mathbf{Z}_p) = \mathbf{P}^1(\mathbf{Q}_p)$.

Now we want to compute the stabilizers of ends under the full $\mathrm{GL}_2(\mathbf{Q}_p)$ action, which is much less difficult now given this description. Preserving an end is equivalent to preserving the line in $\mathbf{P}^1(\mathbf{Q}_p)$ under the natural action of $\mathrm{GL}_2(\mathbf{Q}_p)$. Such elements are enumerated by the Borel subgroup

$$\mathrm{B} = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$$

with entries in $\mathbf{Q}_p$ inside of $\mathrm{GL}_2(\mathbf{Q}_p)$. These enumerate the stabilizers of ends; they are all conjugate, depending on which line is stabilized.

3.3 Amalgams and SL_2

Let $\mathrm{G} \leq \mathrm{GL}_2(\mathbf{Q}_p)^+$ throughout this section. We will be using the fact that G acts on the Bruhat-Tits tree X without inversions to apply Bass-Serre theory.

> THEOREM 3.22. Suppose that the closure of G contains $\mathrm{SL}_2(\mathbf{Q}_p)$. Then it follows that the fundamental domain for the action of G on the Bruhat-Tits tree X is a single segment.

Proof. We'll prove this by trying to write down exactly what the action is on edges and vertices.

To begin, we look at vertices. Pick a basepoint $[\Lambda]$, say $\Lambda = \mathbf{Z}_p^2$. We partition X into X^+ and X^-, where X^+ consists of vertices at an even distance from $[\Lambda]$ and X^- those at an odd distance.

We saw previously that $g \in \mathrm{GL}_2(\mathbf{Q}_p)^+$ preserves the distance between vertices modulo 2, which we used to prove that it acts without inversions. This means that G preserves this partition of X. We claim that the fundamental domain has two vertices, so what we want to show is that for any $[\Lambda^+] \in \mathrm{X}^+$ there is $g \in \mathrm{G}$ so $g \cdot [\Lambda] = [\Lambda^+]$, and similarly for X^-. That is, G should act transitively on both components of the partition.

This is where we use the fact that G contains $\mathrm{SL}_2(\mathbf{Q}_p)$ in its closure, which forces G to be just large enough so that this occurs. Suppose that $[\Lambda^+]$ is at distance $2n$ from the standard lattice $[\mathbf{Z}_p^2] = [\Lambda]$. Using our normal form, pick a basis v, w of $\Lambda = \mathbf{Z}_p^2$ such that a representative Λ^+ can be chosen as

$$\Lambda^+ = p^n \mathbf{Z}_p v \oplus p^{-n} \mathbf{Z}_p w.$$

We do this by appropriately scaling the normal form for any given representative, so that the pair (n, m) in the normal form becomes $(n, -n)$. Now let

$$s = \begin{pmatrix} p^n & 0 \\ 0 & p^{-n} \end{pmatrix}.$$

Note that s is an element of the closure of G, since it lies in $\mathrm{SL}_2(\mathbf{Q}_p)$.

Observe that $s\Lambda = \Lambda^+$. As $\mathrm{GL}_2(\mathbf{Z}_p)$ is open in $\mathrm{GL}_2(\mathbf{Q}_p)$, we have

$$s\mathrm{GL}_2(\mathbf{Z}_p) \cap \mathrm{G} \neq \emptyset.$$

Indeed, if s has an open neighborhood U (such as $s\mathrm{GL}_2(\mathbf{Z}_p)$) that did *not* meet G, then it cannot lie in the closure of G: $\mathrm{GL}_2(\mathbf{Q}_p) \setminus \mathrm{U}$ is closed and contains G. Then s would not lie in a closed subset containing G, contradicting that it lies in the closure.

Thus, it cannot be the case that $s\mathrm{GL}_2(\mathbf{Z}_p)$ dotes not intersect G. It follows that we can write $g = su$ for $u \in \mathrm{GL}_2(\mathbf{Z}_p) = \mathrm{Aut}(\mathbf{Z}_p^2)$ and $g \in \mathrm{G}$, so this particular g will also send $[\Lambda]$ to $[\Lambda^+]$.

Now what about X^-? The argument here is entirely analogous, because once we pick a neighbor of $[\mathbf{Z}_p^2] = [\Lambda]$ all we have done is pick a new lattice instead of the standard lattice and the same argument applies.

Thus, the action has precisely two orbits consisting of X^+ and X^-. To deduce the claim, we just need to verify that the group G acts transitively on edges.

Knowing what we do about vertices, it suffices to prove that G acts transitively on the edges connecting to the standard lattice $[\mathbf{Z}_p^2]$. To see this, suppose we take one of these edges e and want to produce a group element sending it to some edge e'. Then e' has a bounding vertex $[\Lambda^+]$ which is in X^+, so take $g \in \mathrm{G}$ sending $[\mathbf{Z}_p^2]$ to that vertex. Then it must be the case that some edge connecting to $[\mathbf{Z}_p^2]$ is sent to e' by g: the action of g induces a bijection between edges connecting to $[\mathbf{Z}_p^2]$ and to $[\Lambda^+]$. Thus, to send $e \mapsto e'$ first pick a group element sending e to the appropriate edge $g^{-1}(e')$ connecting to $[\mathbf{Z}_p^2]$, and then act by g.

Recall that these adjacent edges are in bijection with $\mathbf{P}^1(\mathbf{F}_p)$, and the $\mathrm{Stab}_{[\Lambda]}$ action on them is via the element $\overline{g} \in \mathrm{GL}(\Lambda/p\Lambda)$. Because the closure of G contains $\mathrm{SL}_2(\mathbf{Q}_p)$, the image of the homomorphism

$$\mathrm{Stab}_{[\Lambda]} \to \mathrm{Aut}(\Lambda/p\Lambda) = \mathrm{PGL}_2(\mathbf{F}_p)$$

contains $\mathrm{PSL}_2(\mathbf{F}_p)$. But $\mathrm{PSL}_2(\mathbf{F}_p)$ still acts transitively on $\mathbf{P}^1(\mathbf{F}_p)$, so we are done. $\square$

Thus, we now know that X//G has graph of groups

$$\underset{\mathrm{Stab}_\Lambda}{\bullet} \xrightarrow{\mathrm{Stab}_e} \underset{\mathrm{Stab}_{\Lambda'}}{\bullet}.$$

where Λ and Λ' are representative lattices in X^+ and X^-, and e an edge connecting their classes.

Corollary 3.23. We have

$$\mathrm{G} \simeq \mathrm{Stab}_{\Lambda} *_{\mathrm{Stab}_e} \mathrm{Stab}_{\Lambda'}.$$

This follows from the main theorem of Bass-Serre theory, since G acts without inversions due to being a subgroup of $\mathrm{GL}_2(\mathbf{Q}_p)^+$ From the previous section, we know the stabilizers of these lattices are maximal compact subgroups of G, and can describe the stabilizer of the edge as an Iwahori subgroup.

In the specific case of $\mathrm{SL}_2(\mathbf{Q}_p)$, we can make this a bit more explicit.

Corollary 3.24. We have

$$\mathrm{SL}_2(\mathbf{Q}_p) \simeq \mathrm{SL}_2(\mathbf{Z}_p) *_{\mathrm{I}} \mathrm{SL}_2(\mathbf{Z}_p)$$

where

$$\mathrm{I} = \begin{pmatrix} \mathbf{Z}_p & \mathbf{Z}_p \\ p\mathbf{Z}_p & \mathbf{Z}_p \end{pmatrix} \subset \mathrm{SL}_2(\mathbf{Z}_p)$$

is an Iwahori subgroup.

The injections here are the identity map, and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & pb \\ p^{-1}c & d \end{pmatrix},$$

since the stabilizer of the vertex in X^- is obtained by conjugate with the appropriate element.

However, there are more clever ways that we can use this result. Note that we didn't explicitly ask that G contain $\mathrm{SL}_2(\mathbf{Q}_p)$, only that its closure does. This means our results apply to groups like

$$\mathrm{SL}_2(\mathbf{Z}[1/p]) \subseteq \mathrm{GL}_2(\mathbf{Q}_p)^+$$

which still contain $\mathrm{SL}_2(\mathbf{Q}_p)$ in their closure. There, we can also obtain an amalgamated product decomposition. We obtain in this case

$$\mathrm{SL}_2(\mathbf{Z}[1/p]) \simeq \mathrm{SL}_2(\mathbf{Z}) *_{\Gamma} \mathrm{SL}_2(\mathbf{Z})$$

where Γ is the subgroup of $\mathrm{SL}_2(\mathbf{Z})$ consisting of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where $p|c$.

3.4 Ihara's theorem and Mumford uniformization

We'll being with talking about uniformization over $\mathbf{C}$.

> Definition 3.25. A *Riemann surface* is a connected complex manifold of (complex) dimension one.
>
> Explicitly, we can take a Riemann surface to be a manifold equipped with charts $\mathrm{U} \to \mathrm{D}$ where D is the unit disk in $\mathbf{C}$. We ask that the transition maps are holomorphic.

For example, $\mathbf{C}$ is a Riemann surface, and so is a complex torus $\mathbf{C}/\mathbf{Z}^2$. Viewed as a real manifold, they have dimension two (think of the torus example), and so they are surfaces.

The genus of such a surface is the number of holes that it has. This can be computed topologically through the fundamental group: $\pi_1(\mathrm{X})^{\mathrm{ab}} \simeq \mathbf{Z}^{2g}$ if the Riemann surface has genus g. For example, a torus has genus one.

The usual uniformization uses the upper half plane. The idea is to find some sort of highly symmetric object, and then study its quotients to produce a classification of Riemann surfaces.

Let's talk about the usual way this is done first. In the case of genus 0, nothing much is happening. For genus one, these are precisely elliptic curves. They will be of the form $\mathbf{C}/(\mathbf{Z} \oplus \tau\mathbf{Z}) \simeq \mathbf{C}^\times/q^{\mathbf{Z}}$ via the exponential map $z \mapsto \exp(2\pi i z)$ and $q = \exp(2\pi i \tau)$. The main result is that this is isomorphic (as a Riemann surface) to the complex solutions of the elliptic curve

$$\mathrm{E}_q : y^2 + xy = x^3 + a_4(q)x + a_6(q)$$

where $a_4(q)$ and $a_6(q)$ are certain explicit power series. In summary, we have an explicit equation depending on q such that $\mathrm{E}_q(\mathbf{C}) \simeq \mathbf{C}^\times/q^{\mathbf{Z}}$, and varying q parameterizes all genus one Riemann surfaces so we have produced explicit equations in general. Note that we present this in a slightly different way than normal using $\mathbf{C}^\times/q^{\mathbf{Z}}$. The reason for this is that $\mathbf{C}/(\mathbf{Z}\oplus\tau\mathbf{Z})$ does not translate very well p-adically, while $\mathbf{C}_p^\times/q^{\mathbf{Z}}$ can make perfect sense.

For genus $g \geq 2$, we can obtain any Riemann surface of genus $g \geq 2$ as a quotient $\mathbb{H}/\Gamma$ for $\Gamma \leq \mathrm{PSL}_2(\mathbf{R})^+$ discrete, torsion free, and cocompact (the quotient is compact). The reason for this uniformization result is the following theorem.

Theorem 3.26. Every simply connected Riemann surface is isomorphic to either $\mathbf{C}$, $\mathbb{H}$ or $\mathbf{P}^1(\mathbf{C})$.

It follows that the universal cover of a Riemann surface, which can always be equipped with a Riemann surface structure, must fall into one of these cases. The genus 0 and 1 cases are covered by $\mathbf{P}^1(\mathbf{C})$ and $\mathbf{C}$ respectively (for example, all genus one Riemann surfaces are $\mathbf{C}/\Lambda$ for a lattice $\Lambda \subset \mathbf{C}$ and hence have universal cover $\mathbf{C}$). The genus $g \geq 2$ case always has universal cover $\mathbb{H}$, which is why we should expect this sort of uniformization.

However, this isn't perfect to adapt to the p-adic setting, the main reason being that the notion of simply connectedness is more complicated. In complex analysis, we also have *Schottky uniformization*. This allows us to give a method for constructing the same Riemann surfaces, but in a way that translates better to the p-adic setting because it bypasses using a classification of possible universal covers.

Definition 3.27. A Schottky group is a subgroup $\Gamma \leq \mathrm{PSL}_2(\mathbf{C})$ constructed in the following way. Take some point $p \in \mathbf{P}^1(\mathbf{C})$, the Riemann sphere. Given any Jordan curve (a non-intersecting continuous loop) not passing through p, it divides $\mathbf{P}^1(\mathbf{C})$ into two regions and we can define the exterior to be the region containing p and the interior to be the region not containing p.

Now pick $2g$ Jordan curves $A_1, \ldots, A_g, B_1, \ldots, B_g$ with disjoint interiors. There is a group Γ of elements in $\mathrm{PSL}_2(\mathbf{C})$ (acting by linear fractional transformations on $\mathbf{P}^1(\mathbf{C})$) consisting of the transformations which take the *outside* of A_i to the *inside* of B_i. Any group obtained in this way is called a *Schottky group*.

It is unfortunately not true that all Schottky groups are obtained by taking Jordan curves to be circles. One can show the following fact, which will lead us to the definition in the p-adic case.

Proposition 3.28. Any Schottky group $\Gamma \leq \mathrm{PSL}_2(\mathbf{C})$ is discrete, finitely generated, and free.

The converse is also possible with some technical additional requirements. Now, the real content of this is the following procedure for producing an arbitrary closed Riemann surface of genus g.

First, given a Schottky group Γ construct the *limit set* $\mathcal{L}_\Gamma$. This consists of the points in $\mathbf{P}^1(\mathbf{C})$ which are obtained as $\lim_{\gamma_n \in \Gamma} \gamma_n(q)$ for $q \in \mathbf{P}^1(\mathbf{C})$, and *distinct* γ_n.

Then, we can take $\mathbf{P}^1(\mathbf{C}) \setminus \mathcal{L}_\Gamma$. This will construct our desired Riemann srufaces.

> Theorem 3.29. Any compact Riemann surface of genus $g \geq 1$ admits a Schottky uniformization.

Proof. The argument is actually fairly short and neat, so I'll give it here. Fist, let S be a compact genus $g \geq 1$ Riemann surface. By taking g curves independent in π_1^{ab}, we get a connected Riemann surface.

After cutting away these curves, call the resulting surface $\tilde{\mathrm{S}}$. It is a theorem that this is biholomorphic (that is, there is a holomorphic map with holomorphic inverse) to a region in $\mathbf{P}^1(\mathbf{C})$ with $2g$ boundary components $\tilde{\Gamma}_i^{\pm}$. There is then a Möbius transformation γ_i sending $\tilde{\Gamma}_i^{+}$ to $\tilde{\Gamma}_i^{-}$ and also $\tilde{\mathrm{S}} \cap \gamma_i(\tilde{\mathrm{S}}) = \tilde{\Gamma}_i^{-}$. The subgroup Γ of $\mathrm{PSL}_2(\mathbf{C})$ generated by the γ_i is Schottky, and $\tilde{\mathrm{S}}$ is a fundamental domain for Γ. We recover S as

$$\widetilde{\mathrm{S}}/\Gamma = \Big(\bigcup_{\gamma \in \Gamma} \gamma(\widetilde{\mathrm{S}})\Big)/\Gamma.$$

Thus, any compact genus g Riemann surface can be produced as $(\mathbf{P}^1(\mathbf{C}) \setminus \mathcal{L}_\Gamma)/\Gamma$. $\square$

Let's give an explicit example in genus one showing how to make a torus. Pick D_0 and D_∞ to be small disks around 0 and ∞ in $\mathbf{P}^1(\mathbf{C})$. Then $\tilde{\mathrm{S}}$ corresponds to the annulus $\{z \in \mathbf{C} : |z| \in [r_1, r_2]\}$. The element γ_1 is $z \mapsto r_2/r_1 \cdot z$. The group $\Gamma = \langle \gamma_1 \rangle$ is just $\mathbf{Z}$, and the corresponding region is $\Omega = \mathbf{P}^1(\mathbf{C}) - \{0, \infty\}$. Then we get the torus as Ω/Γ, which effectively becomes $\tilde{\mathrm{S}}/\gamma_1$ identify the boundary components of an annulus to get a torus.

Going the other way, say we have a torus $\mathrm{S} = \mathbf{C}^2/\Lambda$. Take a side of the fundamental parallelogram for Λ, and view this as a closed loop on S. Say $\Lambda \simeq \mathbf{Z} \oplus \mathbf{Z}$, so we can say that we took the segment from 0 to 1. Cutting this segment gives a cylinder, given by gluing the two sides $[0, i]$ and $[1, 1+i]$ together. The quotient $\mathbf{C}/\mathbf{Z}$ is identified with $\mathbf{P}^1(\mathbf{C}) \setminus \{0, \infty\}$ via $z \mapsto \exp(2\pi i z)$; we can therefore place the cylinder we got by cutting S in this by applying this map. Indeed, the image of the cylinder under this map is going to be $\{z : |z| \in [e^{-2\pi}, 1]\}$, and the map $z \mapsto z+1$ on $\mathbf{C}$ induces the Möbius transformation γ_1 as before.

This turns out to be the correct one to consider in the p-adic world. We can define a p-adic Schottky group as follows:

Definition 3.30. A p-adic Schottky group is a free finitely generated discrete subgroup Γ of $\mathrm{PGL}_2(\mathbf{Q}_p)$.

We will try to do the same exact thing. As we saw with the Drinfeld upper half plane, the correct analogue of $\mathbf{P}^1(\mathbf{C})$ should be $\mathbf{P}^1(\mathbf{C}_p)$.

Definition 3.31. Set $\Omega_\Gamma := \mathbf{P}^1(\mathbf{C}_p) \setminus \mathcal{L}_\Gamma$, where $\mathcal{L}_\Gamma$ is the set of limit points of the p-adic Schottky group Γ, defined in precisely the same way. Then Ω_Γ still admits a Γ action, and we can form a space Ω_Γ/Γ.

There is additional p-adic analytic structure on this space beyond being a topological space (like for a Riemann surface, where we have analytic structure), but it is beyond the scope of the course. However, without going into this, we can still discuss the meaningful consequences of this construction.

A *projective curve* over $\mathbf{Q}_p$ is the zero locus inside $\mathbf{P}^n(\mathbf{Q}_p)$ of some homogeneous polynomials with coefficients in $\mathbf{Q}_p$. We ask that it be irreducible (can't break it up into multiple components cut out by polynomials) and dimension one to be considered a projective curve.

A *smooth projective curve* refers to the curve not having singularities; roughly, that we don't have any additional tangent lines. There is again a notion of genus for smooth projective curves over an arbitrary field, although we cannot get away with a topological definition anymore (the algebraic one will agree).

For example, consider $y^2 = x^3$ over $\mathbf{C}$. This can be projectivized as $y^2z - x^3 = 0$, so that the solutions make sense in $\mathbf{P}^2(\mathbf{C})$. This has dimension one, since we've used a single equation to cut the dimension down from two to one. However, at the point $(0, 0, 1)$, it is not smooth. This is because we can draw multiple tangent lines: the curve looks like a cusp.

The uniformization for genus one Riemann surfaces we previously discussed shows that over $\mathbf{C}$ they can be realized as projective curves (usually smooth). In particular, we saw that any genus one Riemann surface can be realized as $\mathbf{C}/(\mathbf{Z} \oplus \tau\mathbf{Z})$. Then this was isomorphic to the complex points of $y^2 + xy = x^3 + a_4(\tau)x + a_6(\tau)$, which is going to be a projective curve once we add appropriate powers of z so that it cuts out a curve in $\mathbf{P}^2$. In fact, we can always do such a thing in the compact case:

THEOREM 3.32. There is an equivalence of categories between compact Riemann surfaces and smooth projective curves.

Thus, p-adically what we would like to see is that Ω_Γ/Γ recovers the $\mathbf{C}_p$ solutions of some p-adic smooth projective curve. This is remarkably indeed the case!

THEOREM 3.33 (Mumford uniformization, paraphrased). Let Γ be a p-adic Schottky group of rank g. Then there exists a curve X_Γ (which is smooth projective and of genus g over $\mathbf{Q}_p$) such that there is a p-adic analytic isomorphism

$$\Omega_\Gamma/\Gamma \simeq \mathrm{X}_\Gamma(\mathbf{C}_p).$$

The genus of the curve is the rank of Γ.

In particular, the construction again produces smooth projective curves just like it did in the previous case.

EXAMPLE 3.34. Suppose we are working in the genus one case. Then for $|q| < 1$, we have

$$\mathbf{C}_p^\times/q^{\mathbf{Z}} \simeq \mathrm{E}_q(\mathbf{C}_p),$$

where E_q is again cut out by $y^2 + xy = x^3 + a_4(q)x + a_6(q)$, as the power series make sense over $\mathbf{Q}_p$.

Note that I am again hiding a bit of extra structure in this isomorphism: it is not just talking about the topology but rather the analytic structure as well. This is just like we had with genus one Riemann surfaces $\mathbf{C}/(\mathbf{Z}\oplus\tau\mathbf{Z})$: they are all topologically equivalent to tori, but the homeomorphism need not be a *biholomorphism*, that is, respecting the analytic structure. The space $\mathbf{C}_p^\times/q^{\mathbf{Z}}$ is called the Tate curve.

Our work on the Bruhat-Tits tree can help us simplify the definition of a p-adic Schottky group.

THEOREM 3.35 (Ihara). Every discrete torsion-free subgroup Γ of $\mathrm{PGL}_2(\mathbf{Q}_p)$ acts freely on X and Γ is a free group.

Proof. We'll show the action of Γ on the Bruhat-Tits tree X must actually be a free action, since after that Bass-Serre theory finishes the theorem off.

Indeed, if we had a nontrivial stabilizer then it would need to be a finite group since Γ is discrete, and the stabilizer in $\mathrm{PGL}_2(\mathbf{Q}_p)$ of a vertex is $\mathrm{PGL}_2(\mathbf{Z}_p)$ which is compact. A discrete subgroup of a compact group is finite, just for topological reasons.

Thus, the stabilizers of the action are at least finite. But a finite group G always has torsion, which contradicts Γ being torsion-free since the stabilizer is a subgroup. It follows that all stabilizers of vertices are trivial. There are no inversions, so the action is free and by the subcase of Bass-Serre theory for free actions we conclude that Γ is also free. □

Thus, we can equivalently define a p-adic Schottky group as a discrete torsion-free subgroup of $\mathrm{PGL}_2(\mathbf{Q}_p)$.

The tree X helps us understand the curve X_Γ quite explicitly when the genus is ≥ 2 (we need no help when $g = 1$ by the previous example).

> Theorem 3.36. Suppose Γ has rank $g \geq 2$. Then the curve X_Γ has split degenerate stable reduction, and conversely any $\mathbf{Q}_p$-curve which has split degenerate stable reduction is some X_Γ up to isomorphism.

We call that curves X_Γ we can produce *Mumford curves*. The takeaway here is that we don't produce all curves like we do over $\mathbf{C}$, but we can tell exactly which curves we get.

Understanding the exact meaning of the theorem statement requires some algebraic geometry, but we can explain the gist of what its implications are in a more elementary way. By reduction, we mean that it is possible to choose equations over $\mathbf{Z}_p$ for the curve (so that over $\mathbf{Q}_p$ it behaves the same) and then reduce those coefficients modulo $p\mathbf{Z}_p$ to get equations over $\mathbf{F}_p$ to define $\overline{X_\Gamma}$ as a curve over $\mathbf{F}_p$. Asking that the reduction is "split degenerate stable" means we put some geometric conditions on what it looks like.

Specifically, we ask for some control over singular points (points where it is not smooth) on each irreducible component and what those look like, and that the irreducible components after "desingularizing" them become copies of $\mathbf{P}^1$. Then, we ask that each of these $\mathbf{P}^1$'s meet other components at at least 3 points.

These $\mathbf{P}^1$'s produce what is called the *reduction graph*. This is simply defined as the dual graph of the $\mathbf{P}^1$'s: for each projective line we make that into a vertex of the graph. A point of intersection between $\mathbf{P}^1$'s turns into an edge connecting those vertices. This graph can be quite explicitly understood using the combinatorics of the Bruhat-Tits tree X.

It turns out that for $\mathbf{P}^1(\mathbf{C}_p) \setminus \mathbf{P}^1(\mathbf{Q}_p)$, the reduction graph is X *exactly*. This is why I said before that we can think of the Drinfeld upper half plane $\Omega_{\mathbf{C}_p}$ as a tubular neighborhood of X.

For $\mathbf{P}^1(\mathbf{C}_p) \setminus \mathcal{L}_\Gamma$, the reduction graph is some subtree X_Γ of X which still has an action by Γ. The quotient of this graph X_Γ by Γ gives the reduction graph. It is quite often that we do not need to modify the Drinfeld upper half space, because the limit points are precisely $\mathbf{P}^1(\mathbf{Q}_p)$. So, we can produce many examples where the reduction graph is literally a quotient of the the Bruhat-Tits tree X by some p-adic Schottky group.

4 Ramanujan graphs

In this section, we'll study a little bit of spectral graph theory and use the Bruhat-Tits tree X to construct some interesting graphs as X/Γ, for $\Gamma \leq \mathrm{PGL}_2(\mathbf{Q}_p)$.

These graphs will be optimal expanders, or Ramanujan graphs. Roughly, this means that they approximate random graphs extremely well; we can read off such properties from the eigenvalues of the adjacency matrix.

While this will not be the way we prove that the graphs we construct are Ramanujan, Mumford uniformization can also provide inspiration for an alternative construction of Ramanujan graphs. The idea is roughly that X/Γ is a good potential source for Ramanujan graphs since the graphs we get are $(p+1)$-regular and an infinite regular tree has exactly the right eigenvalues to be Ramanujan.

We saw that the graphs X/Γ appear as reduction graphs of Mumford curves. It is possible to pick a curve over $\mathbf{Q}$ appearing in number theory and relate it to a Mumford curve, and therefore matching up the reduction graph modulo p. We can produce operators from the geometry of this curve, and these will induce the adjacency operator on the graph. However, as they come from number theory there are then methods to control the eigenvalues. This approach actually reflects more of Ihara's original approach, which was to match up eigenvalues of the adjacency operator on X/Γ with eigenvalues of Brandt matrices. Eichler's trace formula then allows us to reinterpret the nontrivial (not equal to $p+1$) eigenvalues as eigenvalues of T_p on cusp forms in $S_2(\Gamma_0(q))$, just like we'll do.

The approach we will take also uses number theory techniques, but will be done in a much simpler way. The strategy I follow is outlined by Wen-Ching Winnie Li in https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0441, §5. This has the disadvantage missing some of the beautiful geometry behind the construction of the graph, but also has the advantage of being much simpler to write down and requiring less number theory to appreciate.

4.1 Spectra of graphs

The idea of this last part of the course will be to give constructions of particular graphs which are good spectral expanders.

To begin, we'll discuss the notion of a spectrum of a finite graph. Given a finite graph G, we have an associated vector space

$$L^2(\mathrm{G}) := \{\text{functions } \mathrm{G} \to \mathbf{C}\} = \bigoplus_{v \in \mathrm{G}} \mathbf{C} \cdot \delta_v$$

of functions on the graph. Here, the function δ_v outputs 1 when evaluated on v and 0 otherwise.

Remark 4.1. For an infinite graph, we ask that the absolute values $|f(v)|$ are square summable; we no longer include all functions.

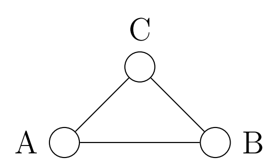
There is a natural operator $\mathrm{A_G}$ on $L^2(\mathrm{G})$, sending

$$f \mapsto \mathrm{A_G}(f)(v) := \sum_{v' \in N_v} f(v')$$

where N_v is the set of neighbors of v. That is, every value is replaced with the average of those around it.

This linear operator $\mathrm{A_G}$ is called the *adjacency* operator, since it uses vertices adjacent to v in the definition. The corresponding matrix in the usual basis of delta functions δ_v is the adjacency matrix, whose ij entry just indicates whether or not there is an edge from vertex i to vertex j.

Example 4.2. Consider the following graph K_3:

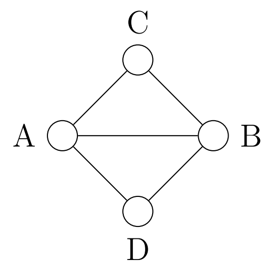


The adjacency matrix of this graph is a 3 by 3 matrix, given by

$$\mathrm{A}_{K_3} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

This is symmetric, because our graph is y default undirected.

As a more interesting example, consider this graph with four vertices:



The adjacency matrix is given by

$$\mathrm{A_G} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

Here, we make the matrix by ordering vertices as A, B, C, D. Note that our previous adjacency matrix appears as a submatrix, since it is a subgraph.

The adjacency matrix tells us a great deal about the graph. We will motivate the notion of the spectrum of a graph, or the set of eigenvalues of the adjacency matrix, by making some simple observations about its properties.

LEMMA 4.3. The ij entry of of A_G^n counts the number of length n paths from v_i to v_j.

Proof. This can be seen inductively from the definition of $\mathrm{A_G}(f)$. Namely, we count the length $n-1$ paths from v_i to all neighbors of v_j, and compute the number of length n paths as the sum of these. □

Let us now make the assumption that G is a d-regular graph. This means that the degree of each vertex is d. Then the matrix

$$\mathrm{W_G} := \frac{1}{d}\mathrm{A_G}$$

is called the *weighted* adjacency matrix. The corresponding operator is given by

$$\frac{1}{|N_v|}\sum_{v' \in N_v} f(v')$$

since $|N_v| = d$ by definition. The weighted adjacency operator now has the following nice interpretation, similar to the previous lemma.

LEMMA 4.4. Let $f \in L^2(\mathrm{G})$ be a probability distribution. Then $\mathrm{W}_{\mathrm{G}}^n(f)$ gives the probability we land on a vertex v after sampling our starting location according to f and taking a length n random walk.

Proof. We normalized A_{G}, so that it is now an averaging operator. Given probability assignments after $n-1$ step walks, we compute the probability for the length n walks by averaging. □

Thus, we can interpret W_{G} as a sort of random walk matrix. If we think about the stationary distribution, if we can diagonalize it's clear that the rate of convergence of $\mathrm{W}_{\mathrm{G}}^n f$ to a stationary distribution for $n \gg 0$ is governed by the magnitude of its eigenvalues: these determine how quickly in n it becomes like a projection to the stationary distribution.

First, let's assume that our d-regular graph G is connected. We would like for random walks to converge to a stationary distribution over time, or for $\mathrm{W}_{\mathrm{G}}^n f$ to converge to some probability distribution. What we will precisely mean by a stationary distribution is that we have $\pi \in L^2(\mathrm{G})$ a probability distribution (so $\sum_{v\in\mathrm{G}} \pi(g) = 1$), and

$$\mathrm{W}_{\mathrm{G}}\pi = \pi.$$

There is an obvious candidate: $\pi = (1/|\mathrm{G}|, \ldots, 1/|\mathrm{G}|)$. This is an eigenvector with eigenvalue one.

We would like there to always be a unique stationary distribution on a connected d-regular graph. Unfortunately, this is not the case. Consider the graph with

$$\mathrm{W}_{\mathrm{G}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

A random walk here just oscillates: after diagonalizing, the eigenvalues are 1 and -1 so we never really get convergence. It turns out this happens if and only if G is not bipartite.

DEFINITION 4.5. Set $\pi = (1/|\mathrm{G}|, \ldots, 1/|\mathrm{G}|) \in L^2(\mathrm{G})$ on a connected d-regular graph G. For such a G, define

$$\lambda(\mathrm{G}) = \max(|\lambda_2|, |\lambda_n|) = \max_{v \perp \pi} \frac{\|\mathrm{W}_{\mathrm{G}} v\|}{\|v\|}.$$

Proposition 4.6. Assuming G is connected and d-regular. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of W_{G}. Then:

- $\lambda_1 = 1 > \lambda_2 \geq \ldots \geq \lambda_n \geq -1$.
- For any probability distribution $f \in L^2(\mathrm{G})$ we have

$$\|\mathrm{W}_{\mathrm{G}}^k f - \pi\| \leq \lambda(\mathrm{G})^k \|f - \pi\|.$$

- Assume G is not bipartite. Then $\mathrm{W}_{\mathrm{G}}^k f$ always converges to π.

Sketch. We sketch the argument briefly. Firstly, writing down W_{G} as a matrix shows immediately that $\|\mathrm{W}_{\mathrm{G}} f\| \leq \|f\|$ (here we regard f as a vector, and take the usual norm). Note also that it is real symmetric, which shows $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ since by the by the spectral theorem it is diagonalizable with real eigenvalues.

We have constructed an eigenvector $\pi = (1/|\mathrm{G}|, \ldots, 1/|\mathrm{G}|)$ for 1, so we know that $\lambda_1 = 1$. To see that $\lambda_1 > \lambda_2$, suppose that there existed an additional eigenvector with eigenvalue one. It turns out the dimension of the 1-eigenspace computes the number of connected components of G, so this is ruled out by the assumption that G is connected.

Now for the second point, we again use the spectral theorem but appeal to the fact that we can choose an orthonormal basis consisting of eigenvectors. Let $v_1, \ldots, v_n$ be the corresponding orthonormal eigenvectors, and for a probability distribution f let $f = \sum_i f_i v_i$. Then

$$\|\mathrm{W}_{\mathrm{G}} f - \pi\|^2 = \sum_{i=2}^{n} \lambda_i^2 f_i^2 \leq \lambda(\mathrm{G})^2 (f_2^2 + \ldots + f_n^2) = \lambda(\mathrm{G})^2 \|f - \pi\|^2.$$

Thus, $\|\mathrm{W}_{\mathrm{G}}^k f - \pi\| \leq \lambda(\mathrm{G})^k \|f - \pi\|$.

If G is not bipartite, then $\lambda_n \neq -1$ and hence $\lambda(\mathrm{G}) < 1$. □

It is possible to slightly modify the notion of a random walk so that we always get convergence to a stationary distribution. Regardless, what the previous proposition tells us is that $\lambda(\mathrm{G})$ controls the "mixing rate" of G. A large value of λ intuitively corresponds to a graph that is bottlenecked in some way, which causes random walks to converge slowly. A low value of λ causes rapid convergence. Intuitively, these behave close to a random graph. This means they are well-connected, as random walks converge very quickly. This is formalized by the *expander mixing lemma*.

LEMMA 4.7. Let G be a d-regular connected graph on n vertices, and take λ as before. Let S, T be subsets of the vertex set of G and let $e(S, T)$ denote the number of edges connecting members of S to T. Then we have

$$\left| e(S, T) - \frac{d|S||T|}{n} \right| \leq \lambda\sqrt{|S||T|}.$$

Note that this quantity being 0 is the expected value for a random graph.

4.2 Motivation for Ramanujan graphs

A natural question to ask is whether or not we can make graphs where λ is very small.

Definition 4.8. Let G be a d-regular connected graph. We say that G is an ε-expander if $\lambda(G) = \max_{v \perp \pi} \frac{\|W_G v\|}{\|v\|} < \varepsilon$.

Small values of ε make better expansion, since for example random walks converge faster or the mixing lemma tells us we have closer behavior to a random graph. The advantage of such graphs is that they have properties of a random graph, but are psuedorandom: we can give concrete constructions of them.

The following result tells us that there is a sort of upper bound on how good of an expander G can be.

Theorem 4.9 (weak version of Alon-Boppana). Fix d and $\varepsilon > 0$. Then there exists n such that all connected d-regular graphs G with n vertices have

$$\lambda(G) > \frac{2\sqrt{d-1}}{d} - \varepsilon.$$

Proof. We are equivalently claiming that for A_G, $\max(|\lambda_2|, |\lambda_n|) \leq 2\sqrt{d-1}$. We actually achieve equality for an infinite d-regular tree. This looks something like the following:

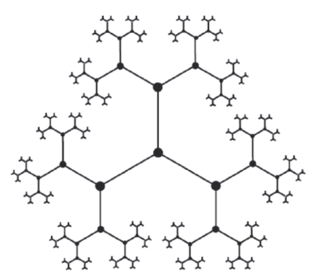


Figure 4: An infinite 3-regular tree you might have seen before in connection to $SL_2(\mathbf{Q}_2)$. Credit to Wikipedia.

Regarding the infinite d-regular tree $\widetilde{G}$ as the universal cover of an arbitrary finite d-regular graph G of size n, the intuition is that the universal cover has the best case spectral behavior. Consider a closed random walk γ starting at $v \in \mathrm{G}$. Then there is a unique lift to $\widetilde{\mathrm{G}}$. The lifts which are again closed random walks are precisely those which have a trivial class in $[\gamma] \in \pi_1(\mathrm{G}, v)$. Thus, the number of closed walks is at least as many as in the infinite tree $\widetilde{\mathrm{G}}$.

Less closed walks starting from v corresponds to a smaller value of $\lambda(\mathrm{G})$. Indeed, the number of such walks is the some diagonal entry of $\mathrm{A}_{\mathrm{G}}^k$, and for k large this is controlled by $\lambda(\mathrm{G})^k$ (we take $\lambda(\mathrm{G})$ for the adjacency matrix, so everything is scaled up by d). By adding these up for all $v \in \mathrm{G}$ we get the trace, and $\operatorname{tr} \mathrm{A}_{\mathrm{G}}^k \leq d^k + n\lambda(\mathrm{G})^k$. Thus, given that we have at least as many walks as in the universal cover which meets the spectral bound exactly, we see intuitively why the result should hold.

Let us make this precise now. The number of closed walks of length $2k$ for a vertex on the infinite d-regular tree can be explicitly calculated as $\frac{1}{k+1}\binom{2k}{k}d(d-1)^k$. It follows that

$$\frac{n}{k+1}\binom{2k}{k}d(d-1)^k \leq \operatorname{tr} \mathrm{A}_{\mathrm{G}}^{2k} \leq d^k + n\lambda(\mathrm{G})^{2k},$$

The first inequality comes from the previous covering space theory argument. We then obtain a concrete lower bound

$$\lambda(\mathrm{G})^{2k} \geq \frac{1}{k+1}\binom{2k}{k}d(d-1)^k - \frac{d^{2k}}{n}$$

Now take $n \to \infty$, and also k but keep k small relative to n. The second term $\frac{d^{2k}}{n}$ goes away, and the first term looks like $C_d \cdot 2^{2k}(d-1)^k$ (C_d is a constant depending on d) using the Catalan number asymptotic, so taking $2k$th roots we get the bound. $\square$

Thus, if we wish to construct a family of d-regular graphs the best we can really hope for is that they all have $\lambda(\mathrm{G}) \leq \frac{2\sqrt{d-1}}{d}$.

> Definition 4.10. A connected d-regular graph G is Ramanujan if $\lambda(\mathrm{G}) \leq \frac{2\sqrt{d-1}}{d}$.

If we use the adjacency matrix for $\lambda(\mathrm{G})$, the bound is just $2\sqrt{d-1}$ which is the definition you'll usually see.

Due the properties that are immediate from this definition, such as optimal convergence to a uniform distribution for random walks or optimal approximation of a random graph, Ramanujan graphs are extremely useful in computer science. Explicit, practical constructions of these are needed.

4.3 Connecting back to the tree

We saw in the proof of the Alon-Boppana bound that an infinite d-regular tree is an optimal expander, so from the beginning our strategy should be looking at quotients of this tree and trying to understand what happens to the spectrum of the adjacency operator.

In the case $d = p+1$, we can use the Bruhat-Tits tree X. This has the distinct advantage that we already know an action of $\mathrm{PGL}_2(\mathbf{Q}_p)$ on it, and so we can easily make quotients of X. The question is then which quotients are the right ones.

In this section, I'll give a representation theoretic perspective for how to identify the right subgroups. In the final section of the notes, I'll give a different proof that takes the perspective of X/Γ describing the reduction of a curve modulo p and trying to give a geometric interpretation for everything.

The subgroups we are interested in come from quaternion algebras.

> DEFINITION 4.11. A quaternion algebra over $\mathbf{Q}$ is a ring B which is a 4-dimensional $\mathbf{Q}$-vector space, such that there exist $\alpha, \beta \in \mathrm{B}$ such that
> $$\alpha^2 = a, \beta^2 = b, \beta\alpha = -\alpha\beta$$
> for $a, b \in \mathbf{Q}^\times$. We denote this algebra by $\left(\frac{a,b}{\mathbf{Q}}\right)$.

The algebra $\mathrm{B} = \left(\frac{a,b}{\mathbf{Q}}\right)$ comes equipped with a norm. Indeed, a general element θ can be written as
$$\theta = x + y\alpha + z\beta + w\alpha\beta,$$
with conjugate $\overline{\theta} = x - y\alpha - z\beta - w\alpha\beta$. The norm is $\mathbf{N}(\theta) = \theta\overline{\theta}$, which lies in $\mathbf{Q}$.

If $\mathrm{B}_p := \mathrm{B} \otimes \mathbf{Q}_p$ is a division ring (i.e. every nonzero element has a multipliciative inverse), we say p ramifies in B. We say ∞ ramifies if $\mathrm{B} \otimes \mathbf{R}$ is a division ring as well. Then we define the discriminant
$$\mathrm{disc(B)} := \pm \prod_{p \text{ ramified}} p$$
with a positive sign if $\mathrm{B} \otimes \mathbf{R} \simeq \mathrm{M}_2(\mathbf{R})$, the ring of 2 by 2 real matrices. Note that over any field F, a quaternion algebra is either a division algebra or $\mathrm{M}_2(F)$.

Now take any quaternion algebra over $\mathbf{Q}$ of discriminant d which is ramified at ∞, and unramified at p. Call this B, and set
$$\mathrm{D} := \mathrm{B}^\times/\mathrm{Z},$$

where Z is the center of the quaternion algebra.

> EXAMPLE 4.12. Take the standard Hamilton quaternions, $\mathbf{Q} \oplus i\mathbf{Q} \oplus j\mathbf{Q} \oplus k\mathbf{Q}$ with $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k$. This is ramified at ∞ because we get the quaternions, which are division ring. This is unramified at any odd prime p, so this suffices.

An *order* of B is a $\mathbf{Z}$-submodule $\mathcal{O}$ where $\mathcal{O} \otimes \mathbf{Q} \simeq \mathrm{B}$ which is a subring. Such a subring necessarily consists of integral elements in B. An order is maximal if there is no other order containing it; maximal orders are not unique for B. For example, in $\mathrm{M}_2(\mathbf{Q})$ a maximal order is $\mathrm{M}_2(\mathbf{Z})$. We can write maximal orders as

$$\mathcal{O} = \mathbf{Z} \oplus \mathbf{Z}\omega_1 \oplus \mathbf{Z}\omega_2 \oplus \mathbf{Z}\omega_3,$$

for certain elements ω_i.

We want to define $\mathrm{D}(\mathrm{A})$ and $\mathrm{B}^\times(\mathrm{A})$ for a ring A; there are several choices of defining the integral structure, and so for the sake of concreteness we pick a maximal order.

> DEFINITION 4.13. Pick a maximal order $\mathcal{O}$. Define $\mathrm{D}(\mathrm{A}) := (\mathrm{A} \otimes \mathcal{O})^\times/\mathrm{A}^\times$, and $\mathrm{B}^\times(\mathrm{A}) := (\mathrm{A} \otimes \mathcal{O})^\times$.

> REMARK 4.14. In class, I wrote down how to define this in terms of solutions to equations over A. This is equivalent to just picking a particular order (not necessarily maximal), which is what we used for $\mathrm{D}(\mathbf{Z})$ and $\mathrm{B}^\times(\mathbf{Z})$.

Now we're ready to define Γ.

> DEFINITION 4.15. Set $\Gamma := \mathrm{D}(\mathbf{Z}[1/p])$. Quite explicitly, this is
>
> $$\{x \in \mathcal{O} : \mathbf{N}(x) = \pm p^k\}/\{\pm p^k\}$$
>
> as $\mathbf{Z}[1/p]^\times = \{\pm p^k\}$.
>
> This is a subgroup of $\mathrm{PGL}_2(\mathbf{Q}_p)$, using the fact that B is unramified at p (and so $\mathrm{B} \otimes \mathbf{Q}_p \simeq \mathrm{M}_2(\mathbf{Q}_p)$, the algebra of 2x2 matrices with entries in $\mathbf{Q}_p$).

Remark 4.16. Over $\mathbf{Q}$, this is almost p-adic Schottky group as a subgroup of $\mathrm{PGL}_2(\mathbf{Q}_p)$. The only issue is that it may have elements of finite order. However, it always contains a Schottky group of finite index.

We will show that X/Γ is a Ramanujan graph. Define

$$\mathbf{A_Q} := \mathbf{Q} \otimes_{\mathbf{Z}} (\mathbf{R} \times \prod_p \mathbf{Z}_p)$$

where the product is taken over all primes p. Equivalently, we set

$$\mathbf{A_Q} \subseteq \mathbf{R} \times \prod_p \mathbf{Q}_p$$

to be the subset of elements such that for almost all p the $\mathbf{Q}_p$ component lies in $\mathbf{Z}_p$. These are equivalent, because looking at the first definition when we interpret it as a subset of $\mathbf{R} \times \prod_p \mathbf{Q}_p$ what we obtain are elements $q \cdot (x_\infty, x_2, x_3, x_5, \ldots)$ and the effect of q is to multiply each component by q. However, $q \in \mathbf{Z}_p$ for only finitely many p. This means all but finitely many components must be within $\mathbf{Z}_p$. Conversely, given an element $x \in \mathbf{R} \times \prod_p \mathbf{Q}_p$ such that for almost all p the projection to $\mathbf{Q}_p$ lies in $\mathbf{Z}_p$, take the finitely many p-adic places where $x_p \in p^\nu \mathbf{Z}_p$ for $\nu < 0$, and set $q \in \mathbf{Q}$ to be the product of these powers p^ν. Then we can write x as element of $\mathbf{Q} \otimes_{\mathbf{Z}} (\mathbf{R} \times \prod_p \mathbf{Z}_p)$, since $q\mathbf{Z}_p = p^\nu \mathbf{Z}_p$ and hence we can choose $x'_p \in \mathbf{Z}_p$ so $x_p = qx'_p$.

This brings up a general construction.

Definition 4.17. Let X_i be topological spaces, and U_i open sets in X_i. The restricted product $\prod_{i \in \mathrm{I}} (\mathrm{X}_i, \mathrm{U}_i)$ consists of elements of $\prod_{i \in \mathrm{I}} \mathrm{X}_i$ such that for all but finitely many i the projection lies in U_i.

The ring $\mathbf{A_Q}$ is given as a restricted product

$$\prod_v (\mathbf{Q}_v, \mathcal{O}_v)$$

where v is a prime or ∞. At a prime p, we set $\mathcal{O}_v = \mathbf{Z}_p$ and at ∞ we set $\mathcal{O}_v = \mathbf{R}$.

One might remark that we don't have a ring of integers for $\mathbf{R}$, so the definition might seem a bit off. Surely the real place messes things up? Well, because we define the restricted product to be elements of $\mathbf{R} \times \prod_p \mathbf{Q}_p$ such that for *almost all* v the projection lies in $\mathcal{O}_v$, changing a finite number of $\mathcal{O}_v$ does not matter. Thus, we can put whatever we want for $\mathbf{R}$, and it will not even change the result as a set.

Another natural question you might have is why we don't just take $\prod_v \mathbf{Q}_v$. After all, the idea of $\mathbf{A}_{\mathbf{Q}}$ seems to be to try to put together all of the completions in one place. The problem is actually just a topological one:

> Lemma 4.18. Let X_i be locally compact topological spaces. Then $\prod_i X_i$ need not be locally compact, but a restricted product where almost all U_i are compact gives a locally compact topological space.

The property of local compactness is quite important: all completions of $\mathbf{Q}$ have this topological property, and it is crucial in many arguments in number theory so we want to preserve it. For example, this topological property allows one to define a Fourier transform on any completion of $\mathbf{Q}$, and due to using a restricted product it is also possible on $\mathbf{A}_{\mathbf{Q}}$.

The key result in this is the following equality.

> Theorem 4.19. As sets, we have
>
> $$D(\mathbf{Q})\backslash D(\mathbf{A}_{\mathbf{Q}})/D(\mathbf{R})\prod_q D(\mathbf{Z}_q) \simeq X/\Gamma.$$
>
> Here, $H\backslash G/K$ means the set of cosets HgK.

Proof. The strong approximation theorem says that $D(\mathbf{Q})D(\mathbf{A}_S)$ is dense in $D(\mathbf{A}_{\mathbf{Q}})$, where $\mathbf{A}_S$ denotes the subring of $\mathbf{A}_{\mathbf{Q}}$ given by

$$\prod_{v\in S}\mathbf{Q}_v \times \prod_{v\notin S}\mathbf{Z}_v$$

for a finite set of absolute values in S (among the p-adic ones, or the normal one $|\cdot|$ giving $\mathbf{R}$). In our case, take S to be the Archimedean absolute value and $|\cdot|_p$. Then in the double coset, we get

$$D(\mathbf{Q})\backslash D(\mathbf{Q})\cdot(D(\mathbf{R})\times D(\mathbf{Q}_p)\times\prod_{q\neq p}D(\mathbf{Z}_q))/D(\mathbf{R})\prod_q D(\mathbf{Z}_q).$$

Now we use that

$$D(\mathbf{Q})\cap(D(\mathbf{R})\times D(\mathbf{Q}_p)\times\prod_{q\neq p}D(\mathbf{Z}_q)) = D(\mathbf{Z}[1/p]).$$

The reason is that $\mathbf{Q} \cap \mathbf{R} \times \mathbf{Q}_p \times \prod_{q \neq p} \mathbf{Z}_q$ is $\mathbf{Z}[1/p]$, using the embedding $q \mapsto (q, q, q, \ldots)$ to put $\mathbf{Q}$ inside of $\mathbf{A}_{\mathbf{Q}}$. Each of $\mathbf{R}, \mathbf{Q}_p, \mathbf{Z}_q$ are simply putting 0 in other components. Thus, saying that a rational number lies in the intersection is to say that it lies in $\mathbf{Z}_q$ for $q \neq p$; this means it lies in $\mathbf{Z}[1/p]$. We conclude that the desired intersection is $\mathrm{D}(\mathbf{Z}[1/p])$.

Now, use a result from group theory if H and G are both subgroups of the same group: $\mathrm{H} \backslash \mathrm{H} \cdot \mathrm{G}$ is $(\mathrm{H} \cap \mathrm{G}) \backslash \mathrm{G}$ as a set. The reason is

$$(\mathrm{H} \cap \mathrm{G}) \backslash \mathrm{G} \to \mathrm{H} \backslash \mathrm{H} \cdot \mathrm{G}$$

sending $(\mathrm{H} \cap \mathrm{G})g \to \mathrm{H}(\mathrm{H} \cap \mathrm{G})g = \mathrm{H}g$ is a well-defined map of sets, and is a bijection. It is certainly surjective, as elements $\mathrm{H} \cdot \mathrm{G}$ are by definition of the form $h \cdot g$, and $\mathrm{H}h \cdot g = \mathrm{H}g$. It is also injective: if g and g' are sent to the same coset, they differ by an element of H as $\mathrm{H}g = \mathrm{H}g'$. But then they differ by an element of $\mathrm{H} \cap \mathrm{G}$, being in the same subgroup G. If they differ by an element of $\mathrm{H} \cap \mathrm{G}$, they are the same coset in $(\mathrm{H} \cap \mathrm{G}) \backslash \mathrm{G}$.

Now apply this in our setting. We obtain

$$\mathrm{D}(\mathbf{Z}[1/p]) \backslash (\mathrm{D}(\mathbf{R}) \times \mathrm{D}(\mathbf{Q}_p) \times \prod_{q \neq p} \mathrm{D}(\mathbf{Z}_q)) / \mathrm{D}(\mathbf{R}) \prod_{q} \mathrm{D}(\mathbf{Z}_q).$$

Cancelling, we get

$$\mathrm{D}(\mathbf{Z}[1/p]) \backslash \mathrm{D}(\mathbf{Q}_p) / \mathrm{D}(\mathbf{Z}_p).$$

Now we see why we reduce D modulo its center: when we do this, because $\mathrm{B}(\mathbf{Q}_p) \simeq \mathrm{M}_2(\mathbf{Q}_p)^\times = \mathrm{GL}_2(\mathbf{Q}_p)$, we get

$$\mathrm{D}(\mathbf{Z}[1/p]) \backslash \mathrm{PGL}_2(\mathbf{Q}_p) / \mathrm{PGL}_2(\mathbf{Z}_p).$$

This is naturally X/Γ, since $\mathrm{PGL}_2(\mathbf{Q}_p)/\mathrm{PGL}_2(\mathbf{Z}_p)$ has a natural graph structure. $\square$

The ability to rewrite X/Γ in this way may seem inconsequential, but this is very important once you know what this coset means. For this, we will need to delve into modular forms.

> Remark 4.20. For people concerned about whether or not strong approximation applies, if we take a place which is unramified in our set S then it will apply for an arbitrary quaternion algebra $\mathrm{B}^\times$. We can deduce strong approximation for D from $\mathrm{B}^\times$.

> It is important to be careful about applying this: given a definite quaternion algebra (like the one we are using), it is possible for strong approximation to fail.

4.4 Modular forms

Recall the upper half plane comes equipped with an action of SL_2 by Möbius transformations. Namely, given $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$, this acts on $\mathfrak{h}$ via

$$z \mapsto \gamma \cdot z := \frac{az+b}{cz+d}.$$

We now want to identify functions which respect this action on $\mathbb{H}$.

> Definition 4.21. A modular function of weight k for $\mathrm{SL}_2(\mathbf{Z})$ is a function $f : \mathbb{H} \to \mathbf{C}$ which is holomorphic and $f(\gamma \cdot z) = (cz+d)^k f(z)$ for $\gamma \in \mathrm{SL}_2(\mathbf{Z})$.

One can define a modular form for a *congruence subgroup* as well.

> Definition 4.22. Define $\Gamma(N) := \ker(\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}))$. A congruence subgroup Γ is a subgroup such that $\Gamma(N) \subseteq \Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ for some N. The minimal N is called the level.

> Lemma 4.23. Congruence subgroups are precisely $\mathrm{SL}_2(\mathbf{Q}) \cap \mathrm{K}$, where K is a compact open subgroup in $\mathrm{SL}_2(\mathbf{A}_f)$.

Then, we modular function for Γ of weight k has the same definition but we only apply the condition for $\gamma \in \Gamma$.

It is not at all obvious that such functions exist. We can actually give an example, however. Consider the Eisenstein series for $k \in \mathbf{Z}_{\geq 4}$ defined by

$$G_k(z) = \sum_{(m,n) \in \mathbf{Z}^2 \setminus \{(0,0\}} \frac{1}{(mz+n)^k}$$

for $z \in \mathbb{H}$.

> Lemma 4.24. The function $G_k(z)$ is modular of weight k. It is 0 if k is odd, and converges absolutely whenever $k \geq 2$.

Proof. We'll assume that you can already show it converges by checking absolute convergence for $k \geq 2$.

First, it's clear that G_{2k+1} is identically zero because we can pair up (n, m) and $(-n, -m)$. This is fine to do, as we can check absolute convergence.

For G_{2k}, we after acting by γ we get

$$G_{2k}(\gamma \cdot z) = \sum_{(m,n)\in \mathbf{Z}^2\setminus(0,0)} \frac{1}{(m\frac{az+b}{cz+d} + n)^{2k}} = \sum_{(m,n)\in \mathbf{Z}^2\setminus(0,0)} \frac{(cz+d)^{2k}}{(n(az+b)+m(cz+d))^{2k}}.$$

But, this is just the same as

$$(cz+d)^{2k} \sum_{(m,n)\in \mathbf{Z}^2\setminus(0,0)} \frac{1}{(m'z+n')^{2k}}$$

where

$$\begin{pmatrix} m' \\ n' \end{pmatrix} = \begin{pmatrix} c & a \\ d & b \end{pmatrix} \begin{pmatrix} m \\ n \end{pmatrix}.$$

But, this induces a bijection on $\mathbf{Z}^2$ as the matrix is still in $\mathrm{SL}_2(\mathbf{Z})$. It also sends 0 to 0. So, this is actually a re-arrangement of $G_{2k}(z)$ and we are done. □

Thus, we see the reason this example works boils down to $\mathrm{SL}_2(\mathbf{Z})$ inducing a change of basis on a $\mathbf{Z}$-lattice.

Assume that $\Gamma \supseteq \left\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\rangle \simeq \mathbf{Z}$. Call this matrix T. The action of T on $\mathbb{H}$ shifts z, so this condition implies that our modular function with respect to Γ satisfies $f(z+1) = f(z)$. Then it makes sense to take a Fourier expansion of this function as since it is periodic. We write

$$f(z) = \sum_{n\in\mathbf{Z}} a_n(y) e^{2\pi i n x}.$$

We write this as $\sum_{n\in\mathbf{Z}} a_n q^n$ for $q = e^{2\pi i z}$, where the a_n must be constant due to f being holomorphic.

Not every modular function is so well behaved, so we impose extra conditions on the Fourier expansion. Consider the map $z \mapsto q = e^{2\pi i z}$ on $\mathbb{H}$. This sends $\mathbb{H}$ to the punctured unit disk $\mathrm{D} - \{0\}$. In particular, we have an isomorphism

$$\mathbb{H}/\langle \mathrm{T} \rangle \xrightarrow{\simeq} \mathrm{D} - \{0\}$$

via this map, and as q^n is a conformal map from $\mathbb{H}$ to the punctured unit disk we can view the modular function as coming from a holomorphic function on the unit disc $q \mapsto \sum_{n\in\mathbf{Z}} a_n q^n$, as precomposing with our isomorphism gives back the modular function. We would like this to be extendable to the origin (corresponding to $i\infty$), and in particular this means it should be bounded in a neighborhood of 0.

Definition 4.25. Let $f : \mathbb{H} \to \mathbf{C}$ be a modular function for Γ of weight k. This is a modular form of weight k if it is holomorphic at ∞. That is, $a_n = 0$ if $n < 0$. We denote the set of modular forms of weight k by $M_k(\Gamma)$.

We call it a cusp form if additionally $a_0 = 0$. The set of cusp forms is denoted $S_k(\Gamma)$.

Example 4.26. For $\mathrm{SL}_2(\mathbf{Z})$, $\bigoplus_k M_k(\mathrm{SL}_2(\mathbf{Z})) = \mathbf{C}[G_4, G_6]$. One usually normalizes G_4 and G_6 so they have Fourier expansions in $\mathbf{Z}[1/6][[q]]$; we call these E_4 and E_6.

Another specific case which is worth looking at is the case of the j-invariant, which helps give a more down to earth example of why people are interested in these sorts of functions at all.

Theorem 4.27. The weight 0 meromorphic modular functions are given by $\mathbf{C}(j)$, where

$$j(z) := \frac{\mathrm{E}_4^3}{\Delta} = 1728\frac{\mathrm{E}_4^3}{\mathrm{E}_4^3 - 27\mathrm{E}_6^2}.$$

The function j has a simple pole at ∞.

We can see this by looking at $\mathbb{H}/\mathrm{SL}_2(\mathbf{Z})$. A modular function of weight 0 descends to a function on this quotient, because the weight 0 condition means we are actually just asking for it to be fully invariant. If we don't add ∞, then looking at the usual fundamental domain

$$\mathcal{D} = \{z : \mathrm{Re}(z) \in [-1/2, 1/2]\} \cap \{z : |z| \geq 1\}$$

after we identify the $\mathrm{Re}(z) = -1/2$ and $1/2$ as well as the two segments $[\omega, i]$ and $[i, -\overline{\omega}]$ (T and S identify these) we get $\mathbf{P}^1(\mathbf{C}) \setminus \infty$. Thus, adding ∞ gives us $\mathbf{P}^1(\mathbf{C})$.

If we ask that a weight zero modular function is holomorphic at ∞, this is asking for a holomorphic function on $\mathbf{P}^1(\mathbf{C})$. This must then be constant as its image in $\mathbf{C}$ would be bounded (as $\mathbf{P}^1(\mathbf{C})$ is compact, hence the image is) and then we have a bounded entire function which must be constant.

If we remove this condition, we get $\mathbf{C}(z)$ as the meromorphic functions. Thus, meromorphic modular functions are given by

$$\mathbb{H} \cup \infty/\mathrm{SL}_2(\mathbf{Z}) \simeq \mathbf{P}^1(\mathbf{C}) \to \mathbf{C}$$

where the map $\mathbf{P}^1(\mathbf{C}) \to \mathbf{C}$ is some rational function in $\mathbf{C}(z)$. The isomorphism is called the j-function, and the just using z gives us a map to $\mathbf{C}$. This gives us the function $j(z)$ as in the theorem, and we already know all meromorphic modular functions are rational functions in j. It does not, however, give the explicit formula. We can verify the explicit formula works: it is certainly a meromorphic modular function, and we just need to check that it is injective on the fundamental domain. We can check the image is closed and open in $\mathbf{C}$, so it gives an isomorphism to $\mathbf{CP}^1$ once we add in the point and infinity and check that the pole is simple (which can be checked from q expansions).

The function $j(z)$ is very relevant for classifying elliptic curves, or genus one Riemann surfaces, over $\mathbf{C}$.

> Theorem 4.28. Genus one Riemann surfaces are all of the form $\mathbf{C}/\Lambda$ for a lattice Λ. We have $\mathbf{C}/\Lambda \simeq \mathbf{C}/\Lambda'$ as Riemann surfaces precisely if $\Lambda = c\Lambda'$, where $c \in \mathbf{C}^\times$.

We say Λ and Λ' are *homothetic* in this case. The upper half plane $\mathbb{H}$ enumerates lattices of the form $\mathbf{Z} \oplus \tau\mathbf{Z}$ for $\tau \in \mathbb{H}$ up to homothety. We can see this by revisiting our interpretation of $\mathbf{P}^1(\mathbf{C})\backslash\mathbf{P}^1(\mathbf{R})$ as moduli of lattices up to homothety with a chosen basis. Note that any lattice $\mathbf{Z}\omega_1 \oplus \mathbf{Z}\omega_2$ can be put in the form $\mathbf{Z} \oplus \omega_1/\omega_2\mathbf{Z}$, and with an appropriate choice of basis this lies in $\mathbb{H}$. Because we normalized, these lattices enumerated by $\mathbb{H}$ are not homothetic unless they are equal.

However, enumerating lattices with $\mathbb{H}$ still picks a basis. And when we pick a basis, it is possible we enumerate the same homothety class of lattice multiple times in $\mathbb{H}$. Using the $\mathrm{SL}_2(\mathbf{Z})$ action, the orbit of $z \in \mathbb{H}$ gives all the ways we can write the same lattice up to a change of basis. Thus, we arrive at the conclusion that $\mathbb{H}/\mathrm{SL}_2(\mathbf{Z})$ enumerates lattices up to homothety, without picking a basis.

> Corollary 4.29. Genus one Riemann surfaces up to isomorphism are enumerated by $\mathbb{H}/\mathrm{SL}_2(\mathbf{Z})$.

Note that j induces an isomorphism from $\mathbb{H} \cup \infty/\mathrm{SL}_2(\mathbf{Z})$ to $\mathbf{P}^1(\mathbf{C})$. In particular, it is injective on $\mathbb{H}/\mathrm{SL}_2(\mathbf{Z})$, and hence completely classifies genus one Riemann surfaces! Additionally, since there is an equivalence of categories between genus one Riemann surfaces and elliptic curves over $\mathbf{C}$, this also provides a complete isomorphism invariant on elliptic curves. The situation is even better than this: evaluating j on an elliptic curve can be explicitly calculated in terms of the coefficients of the equation for the elliptic curve. Take an elliptic curve $E/\mathbf{C}$ and write it in Weierstrass form $y^2 = x^3 + ax + b$. Then

$$j(E) = 1728\frac{4a^3}{4a^3 + 27b^2},$$

which is now simple to calculate.

We'll now see an adelic enhancement of this theory, which will resemble the double coset we encountered before. Our first hint that something adelic is going on is the reinterpetation of congruence subgroups. For example, take $\Gamma_0(N)$, consisting of matrices in $\mathrm{SL}_2(\mathbf{Z})$ where the bottom left entry c is 0 modulo N. We can rewrite this as

$$\mathrm{SL}_2(\mathbf{Q}) \cap \mathrm{SL}_2(\mathbf{R}) \times \mathrm{K}^\infty$$

where $\mathrm{K}^\infty \leq \prod_p \mathrm{SL}_2(\mathbf{Z}_p)$ consists of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where after reducing modulo N the bottom left entry c becomes 0. Here, we embed $\mathrm{SL}_2(\mathbf{Q})$ into $\mathrm{SL}_2(\mathbf{A_Q})$ diagonally.

To see this, just note that we're asking for $q \in \mathrm{SL}_2(\mathbf{Q})$ such that $q \in \mathrm{SL}_2(\mathbf{Z}_p)$ for every p, so this already forces $q \in \mathrm{SL}_2(\mathbf{Z})$. The congruence condition means $N|c$, so we get the desired subgroup.

Theorem 4.30. We have

$$\Gamma_0(N)\backslash\mathrm{SL}_2(\mathbf{R}) \simeq \mathrm{SL}_2(\mathbf{Q})\backslash\mathrm{SL}_2(\mathbf{A_Q})/\mathrm{K}^\infty.$$

Proof. This is not as hard as it looks. We have

$$\mathrm{SL}_2(\mathbf{A_Q}) = \mathrm{SL}_2(\mathbf{Q}) \cdot (\mathrm{SL}_2(\mathbf{R}) \times \mathrm{K}^\infty),$$

for our particular choice. This is strong approximation. Now substitute this in. We get

$$\mathrm{SL}_2(\mathbf{Q})\backslash\mathrm{SL}_2(\mathbf{A_Q})/\mathrm{K}^\infty = \mathrm{SL}_2(\mathbf{Q})\backslash\mathrm{SL}_2(\mathbf{Q}) \cdot (\mathrm{SL}_2(\mathbf{R}) \times \mathrm{K}^\infty)/\mathrm{K}^\infty.$$

Now use $\mathrm{H} \cap \mathrm{G}\backslash\mathrm{G} \simeq \mathrm{H}\backslash\mathrm{H} \cdot \mathrm{G}$, like we did before. This gives us, for $\mathrm{H} = \mathrm{SL}_2(\mathbf{Q})$ and $\mathrm{G} = \mathrm{SL}_2(\mathbf{R}) \times \mathrm{K}^\infty$,

$$\Gamma_0(N)\backslash\mathrm{SL}_2(\mathbf{R}) \times \mathrm{K}^\infty/\mathrm{K}^\infty.$$

This is of course $\Gamma_0(N)\backslash\mathrm{SL}_2(\mathbf{R})$. □

It follows that $\mathrm{SL}_2(\mathbf{Q})\backslash\mathrm{SL}_2(\mathbf{A_Q})/\mathrm{K}_\infty\mathrm{K}^\infty$ where $\mathrm{K}_\infty = \mathrm{SO}_2(\mathbf{R})$ is $\Gamma_0(N)\backslash\mathbb{H}$. This is because $\mathrm{SL}_2(\mathbf{R})$ acts transitively on $\mathbb{H}$, with stabilizer SO_2 (at i).

Thus,

$$\Gamma_0(N)\backslash\mathbb{H} \simeq \mathrm{SL}_2(\mathbf{Q})\backslash\mathrm{SL}_2(\mathbf{A_Q})/\mathrm{K}_\infty\mathrm{K}^\infty,$$

which now looks awfully familiar when $N = 1$. We can also do this for GL_2.

> Corollary 4.31. Now set K^∞ to be $\mathrm{K}^\infty \leq \prod_p \mathrm{GL}_2(\mathbf{Z}_p)$, again with $c \equiv 0 \pmod N$ as the extra condition. Then
>
> $$\Gamma_0(N)\backslash\mathrm{GL}_2(\mathbf{R})^+ \simeq \mathrm{GL}_2(\mathbf{Q})\backslash\mathrm{GL}_2(\mathbf{A_Q})/\mathrm{K}^\infty.$$

Proof. It is essentially the same idea, we just check what happens with determinants. We have all but finitely many components of K^∞ equal to $\mathrm{GL}_2(\mathbf{Z}_p)$, and we always get all of $\mathbf{Z}_p^\times$ as the determinants. This produces enough determinants to allow us to modify the decomposition for $\mathrm{SL}_2(\mathbf{A_Q})$ to get

$$\mathrm{GL}_2(\mathbf{A_Q}) = \mathrm{GL}_2(\mathbf{Q}) \cdot (\mathrm{GL}_2(\mathbf{R})^+ \times \mathrm{K}^\infty).$$

Now use the exact same idea, this time using $\mathrm{GL}_2(\mathbf{Q}) \cap \mathrm{GL}_2(\mathbf{R})^+ \times \mathrm{K}^\infty = \Gamma_0(N)$. The determinant has to be positive, and moreover since it lies in $\mathrm{GL}_2(\mathbf{Z})$ after taking the intersection we see it is in fact one since determinants in $\mathrm{GL}_2(\mathbf{Z})$ are ± 1. This is why we get a subgroup of $\mathrm{SL}_2(\mathbf{Z})$ still. We obtain

$$\Gamma_0(N)\backslash\mathrm{GL}_2(\mathbf{R})^+ \times \mathrm{K}^\infty/\mathrm{K}^\infty,$$

which shows the result. $\square$

Moreover, if we use $\mathbf{R}_{>0}\mathrm{SO}_2(\mathbf{R}) = \mathrm{K}_\infty$, we get

$$\Gamma_0(N)\backslash\mathbb{H} = \mathrm{GL}_2(\mathbf{Q})\backslash\mathrm{GL}_2(\mathbf{A_Q})/\mathrm{K}_\infty\mathrm{K}^\infty.$$

This whole argument also eventually shows that we can lift modular forms to functions on

$$\mathrm{GL}_2(\mathbf{Q})\backslash\mathrm{GL}_2(\mathbf{A_Q}).$$

The reason is simply that we can lift the function to $\mathrm{GL}_2(\mathbf{R})^+$ by $f(r) := f(r \cdot i)$. Then use the previous result to try to interpret this adelically. We won't get into this since it's not needed. Instead, we'll look a slightly different interpretation that more directly uses the previous result.

For more general K^∞, including the previous example, we can realize modular forms as differential forms [2] $f(z)\mathrm{d}z^k$ on $\Gamma\backslash\mathbb{H}$ (say $\Gamma = \Gamma_0(N)$, matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathrm{SL}_2(\mathbf{Z})$ with $N|c$).

This can be explained informally as follows for weight $k = 2$. Suppose we have $f(z)\mathrm{d}z$, and we want it to make sense on $\mathbb{H}/\Gamma$. Then we need $f(z)\mathrm{d}z$ to be invariant under the action of $\gamma \in \Gamma_0(N)$. This means

$$f(z)\mathrm{d}z = f(\gamma \cdot z)\mathrm{d}(\gamma \cdot z).$$

But $\mathrm{d}(\gamma \cdot z) = \mathrm{d}(\frac{az+b}{cz+d}) = \frac{ad-bc}{(cz+d)^2}\mathrm{d}z$. Thus, the condition is really asking that f have weight 2.

These curves $\Gamma\backslash\mathbb{H}$ are called modular curves. The double coset we saw using D is exactly the analogue of this modular curve, but for division algebras. Note that these aren't too different, as GL_2 actually comes from a division algebra too: $\mathrm{M}_2(\mathbf{Q})$.

However, it is a lot simpler: instead of being a Riemann surface, it is now just a finite set. This means modular forms are extremely simple objects in terms of this double coset: there's no reason to worry about differentials anymore, we just look at the analogue of the modular curve.

Definition 4.32. A modular form of weight two of $\mathrm{B}^\times$ (for our desired congruence subgroup) is given by a function

$$\mathrm{B}^\times(\mathbf{Q})\backslash\mathrm{B}^\times(\mathbf{A_Q})/\mathrm{B}^\times(\mathbf{R})\prod_q \mathrm{B}^\times(\mathbf{Z}_q) \to \mathbf{C}.$$

Here, we use that this is a finite set. This vector space is denoted by $S_2(\mathrm{B}^\times)$.

Unlike the case of GL_2, we can non-canonically identify modular forms with functions on the modular curve (the finite double coset) because it is dimension 0.

Note that for GL_2, K_∞ was a proper subgroup of $\mathrm{GL}_2(\mathbf{R})$ when we were writing a modular curve adelically. The reason for this distinction is that $\mathrm{B}^\times(\mathbf{R})$ is compact, and we make the subgroup for GL_2 out of a maximal compact subgroup ($\mathrm{O}_2(\mathbf{R})$).

This double coset looks *almost* like $\mathbf{C}[X/\Gamma]$; we have just removed D, so it is no longer obviously a quotient of the Bruhat-Tits tree. However, it turns out that this finite set is still in bijection with X/Γ, so we will not need to define modular forms on D at all.

[2] These are *symmetric* differential forms: they are global sections of symmetric powers of the cotangent bundle, and not exterior powers.

In fact, this vector space is closely related to modular forms, which correspond to $\mathrm{B} = \mathrm{M}_2(\mathbf{Q})$: there is a nontrivial theorem which says we have an injection $S_2(\mathrm{B}^\times)/\mathrm{const} \subseteq S_2(\Gamma_0(d))$, where $d = \mathrm{disc}(\mathrm{B})$. We'll see how to upgrade this correspondence later to translate what happens with the adjacency operator. For now, we just want to observe the following:

> THEOREM 4.33. There is a natural identification $S_2(\mathrm{B}^\times) \simeq \mathbf{C}[\mathrm{X}/\Gamma]$.

Proof. This amounts to computing the magnitude of the cosets, since these are both just $\mathbf{C}$-valued functions on a finite set.

We have, using strong approximation,

$$\mathrm{B}^\times(\mathbf{Q})\backslash\mathrm{B}^\times(\mathbf{A}_\mathbf{Q})/\mathrm{B}^\times(\mathbf{R})\prod_q \mathrm{B}^\times(\mathbf{Z}_q) \simeq \mathrm{B}^\times(\mathbf{Z}[1/p])\backslash\mathrm{B}^\times(\mathbf{Q}_p)/\mathrm{B}^\times(\mathbf{Z}_p).$$

Now $\mathrm{B}^\times(\mathbf{Q}_p) \simeq \mathrm{GL}_2(\mathbf{Q}_p)$. We therefore end up with

$$\mathrm{B}^\times(\mathbf{Z}[1/p])\backslash\mathrm{GL}_2(\mathbf{Q}_p)/\mathrm{GL}_2(\mathbf{Z}_p).$$

Now $\mathrm{B}^\times(\mathbf{Z}[1/p])$ is just $\{x \in \mathcal{O} : \mathbf{N}(x) = p^k\}$. Viewing $\mathrm{GL}_2(\mathbf{Q}_p)/\mathrm{GL}_2(\mathbf{Z}_p)$ as the set of rank two $\mathbf{Z}_p$-lattices in $\mathbf{Q}_p^2$, note that $\mathrm{X} = \mathrm{GL}_2(\mathbf{Q}_p)/p^\mathbf{Z}\mathrm{GL}_2(\mathbf{Z}_p)$. That is, there is a unique power of p taking Λ to any representative lattice for $[\Lambda]$; this comes from $\mathbf{Q}_p^\times/\mathbf{Z}_p^\times \simeq p^\mathbf{Z}$.

With this in mind, we want to count orbits of $\mathrm{B}^\times(\mathbf{Z}[1/p])$ on lattices. Looking at a lattice class Λ, since $p^\mathbf{Z} \in \mathrm{B}^\times(\mathbf{Z}[1/p])$ the orbit at least contains all of $[\Lambda]$. In fact, the orbit is then precisely all lattices in $\mathrm{D}(\mathbf{Z}[1/p]) \cdot [\Lambda]$, since $\mathrm{D}(\mathbf{Z}[1/p])$ is just what we get after taking a quotient by $\pm p^\mathbf{Z}$. Note that ± 1 does nothing, as it stabilizes any lattice.

It follows that orbits of $\mathrm{B}^\times(\mathbf{Z}[1/p])$ are in bijection with orbits of $\mathrm{D}(\mathbf{Z}[1/p])$, by

$$\mathrm{B}^\times(\mathbf{Z}[1/p]) \cdot \Lambda \mapsto \mathrm{D}(\mathbf{Z}[1/p]) \cdot [\Lambda].$$

We can interpret what is happening as just enumerating the same orbits on X, but instead expanding each $[\Lambda]$ to all of its equivalent lattices. $\square$

We will now forget D now that we have made this identification; the use of D is purely to make the original quotient obviously a $(p+1)$-regular graph arising from a quotient of X, as the result is a bit trickier to deal with when using $\mathrm{B}^\times$.

4.5 Hecke operators on modular forms

We'll first give the classical Hecke operators, and then move on to explaining what the Hecke operators are for modular forms on D.

We will focus on the subgroup $\Gamma_0(N)$, since this is what we will look at later.

> Definition 4.34. The subgroup $\Gamma_0(N) \subseteq \mathrm{SL}_2(\mathbf{Z})$ consists of the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $c \equiv 0 \pmod N$.

We are particularly interested in the operator T_p, so I will focus on this. This operator can be given quite explicitly, just to convince you that it is directly computable.

> Definition 4.35. We can define T_p on $M_k(\Gamma_0(N))$ where $p \nmid N$ as
>
> $$\mathrm{T}_p f(z) := p^{k-1} f(pz) + \frac{1}{p} \sum_{a=0}^{p-1} f\left(\frac{z+a}{p}\right).$$
>
> If $f = \sum_n a_n q^n$ is the Fourier expansion,
>
> $$\mathrm{T}_p f := p^{k-1} \sum_{n \geq 1} a_n q^{pn} + \sum_{n \geq 1} a_{pn} q^n.$$
>
> Define T_n via $\mathrm{T}_{nm} = \mathrm{T}_n \mathrm{T}_m$ if $(n, m) = 1$, and $\mathrm{T}_{p^r} = \mathrm{T}_p \mathrm{T}_{p^{r-1}} - p \mathrm{T}_p^{r-1}$.

Let's see the utility of these operators in the special case of $S_2(\Gamma_0(N))$. Recall we interpreted these as differentials $f(z)\mathrm{d}z$ on $\mathbb{H}/\Gamma_0(N)$. Then, we define the inner product by just integrating them:

$$\langle f, g \rangle := \frac{i}{2} \int_{\mathbb{H}/\Gamma_0(N)} f(z)\overline{g}(z) \mathrm{d}z \mathrm{d}\overline{z} = \int_{\mathbb{H}/\Gamma_0(N)} f(z)\overline{g}(z) \mathrm{d}x \mathrm{d}y$$

where $z = x + iy$. What's happening in this computation is $\mathrm{d}z\overline{\mathrm{d}z} = (\mathrm{d}x + i\mathrm{d}y)(\mathrm{d}x - i\mathrm{d}y) = i\mathrm{d}y\mathrm{d}x - i\mathrm{d}x\mathrm{d}y = -2i\mathrm{d}x\mathrm{d}y$.

> Proposition 4.36. We have $\langle \mathrm{T}_p f, g \rangle = \langle f, \mathrm{T}_p g \rangle$ when $p \nmid N$.

We can define $\mathrm{T}_{mn} = \mathrm{T}_n \mathrm{T}_m$ for $(m, n) = 1$, and $\mathrm{T}_{p^r} = \mathrm{T}_{p^{r-1}} \mathrm{T}_p - p^{k-1} \mathrm{T}_{p^{r-2}}$

for $p \nmid N$. When $p|N$, we do not include the second term, and also define $\mathrm{T}_p f = \sum_{n\geq 1} a_{pn}q^n$. This defines operators T_n for all n.

As a corollary, $S_2(\Gamma_0(N))$ admits a basis consisting of Hecke eigenforms, or f where $\mathrm{T}_n f = \lambda_n f$ for all $(n, N) = 1$. Note that a_1 in $\mathrm{T}_n f$ is a_n when $(n, N) = 1$ (we can see this already for primes). This means $\lambda_n = a_n$ when we normalize to get $a_1 = 1$ in the eigenforms, so the eigenvalues have meaning. For example, with $N = 1$ we get the following:

> Corollary 4.37. There is an orthonormal eigenbasis of $M_k(\mathrm{SL}_2(\mathbf{Z}))$ consisting of simultaneous eigenforms f for all Hecke operators T_n.

Proof. We can apply the spectral theorem! □

The very classical way of thinking about these operators would be that we try to produce operators T_n on $S_k(\Gamma_0(N))$ which respect the natural inner product, and also have a nice expression in terms of the q-expansions so that we can learn about the q-expansion from their eigenvalues.

We can make the origin of Hecke operators slightly clearer by explaining a uniform construction that works for all congruence subgroups. Let Γ_1 and Γ_2 be congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$, and let $\mathrm{GL}_2^+(\mathbf{Q})$ be the group of 2 by 2 matrices with positive determinant and entries in $\mathbf{Q}$.

For $\alpha \in \mathrm{GL}_2^+(\mathbf{Q})$, we can take a double coset

$$\Gamma_1 \alpha \Gamma_2 := \{\gamma_1 \alpha \gamma_2 : \gamma_i \in \Gamma_i\}.$$

There is an action on the left by Γ_1. This decomposes the double coset into orbit spaces $\Gamma_1 \beta_j$, so that

$$\Gamma_1 \alpha \Gamma_2 = \coprod_j \Gamma_1 \beta_j.$$

> Theorem 4.38. There are finitely many representatives β_j.

This allows us to define the Hecke operators.

Definition 4.39. Let $\alpha \in \mathrm{GL}_2^+(\mathbf{Q})$, and let Γ_1 and Γ_2 be congruence subgroups as before. Define the operator $\mathrm{T}_\alpha : M_k(\Gamma_1) \to M_k(\Gamma_2)$ as sending

$$f \mapsto \sum_{\gamma \in \Gamma_1 \backslash \Gamma_1 \alpha \Gamma_2} f|_\gamma = \sum_j f|_{\beta_j},$$

where $|_{\beta_j}$ denotes the action of the matrix coset representative β_j. We define this as

$$f|_\gamma(z) := (\det \gamma)^{k/2}(cz+d)^{-k} f(\gamma \cdot z).$$

For example, modularity for $\mathrm{SL}_2(\mathbf{Z})$ can be expressed as invariance under this action.

This takes in something in $M_k(\Gamma_1)$ and produces a function, as there are finitely many coset representatives and by modularity we can restrict to coset representatives. However, we would like for this to be in $M_k(\Gamma_2)$.

Lemma 4.40. The operator T_α sends $M_k(\Gamma_1) \to M_k(\Gamma_2)$.

Proof. What we want is $\mathrm{T}_\alpha f|_\gamma = \mathrm{T}_\alpha f$ for $\gamma \in \Gamma_2$. We have

$$\mathrm{T}_\alpha f|_\gamma = \sum_j f|_{\beta_j \gamma}.$$

Up to Γ_1, this just permutes the β_j. Indeed, $\beta_j\gamma \in \Gamma_1\alpha\Gamma_2$, as $\gamma \in \Gamma_2$ and $\beta_j \in \Gamma_1\alpha\Gamma_2$. Then when we decompose $\Gamma_1\alpha\Gamma_2$ into just cosets $\Gamma_1\beta_j$, $\beta_j\gamma$ lies in some $\Gamma_1\beta_{\gamma(j)}$. As f is invariant for Γ_1, we get $\sum_j f|_{\beta_j\gamma} = \sum_j f|_{\beta_{\gamma(j)}}$. The map $j \mapsto \gamma(j)$ is a permutation, as if $\beta_j\gamma$ and $\beta_{j'}\gamma$ lie in the same coset $\Gamma_1\beta_{\gamma(j)}$, then we can deduce β_j and $\beta_{j'}$ lie in the same coset. □

Lemma 4.41. Cusp forms in $S_k(\Gamma_1)$ get sent to cusp forms in $S_k(\Gamma_2)$ by T_α.

Let us think about the special case where $\Gamma_1 = \Gamma_2 = \mathrm{SL}_2(\mathbf{Z})$, and try to recover T_p. We claim that this is given by T_α, where $\alpha = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$.

Lemma 4.42. We have

$$\mathrm{SL}_2(\mathbf{Z})\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\mathrm{SL}_2(\mathbf{Z}) = \coprod_{0\le a\le p-1} \mathrm{SL}_2(\mathbf{Z})\begin{pmatrix} 1 & a \\ 0 & p \end{pmatrix} \cup \mathrm{SL}_2(\mathbf{Z})\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

We can interpret the left double coset as the determinant p integer matrices.

This gives us explicit coset representatives to work with.

Corollary 4.43. The corresponding Hecke operator for T_α is T_p up to a constant.

Proof. Using the explicit coset representatives from the lemma, applying $f|_\gamma$ we obtain the first definition of the Hecke operator. □

I want to explain what is happening here a bit more geometrically, just in the case of $\mathrm{SL}_2(\mathbf{Z})$. Of course, it works just as well for arbitrary congruence subgroups. If we use modular curves, we'll see how this readily generalizes to $S_2(\mathrm{B}^\times)$, where we only have access to a modular curve.

We get a diagram

$$\begin{tikzcd} & \Gamma_0(p)\backslash\mathbb{H} \arrow[dl, "\pi_2"'] \arrow[dr, "\pi_1"] & \\ \mathrm{SL}_2(\mathbf{Z})\backslash\mathbb{H} & & \mathrm{SL}_2(\mathbf{Z})\backslash\mathbb{H} \end{tikzcd}$$

sending an orbit $[\Gamma_0(p)z] \mapsto [\mathrm{SL}_2(\mathbf{Z})z]$ for π_2. For π_1, we apply the isomorphism

$$\Gamma_0(p)\backslash\mathbb{H} \simeq \alpha\Gamma_0(p)\alpha^{-1}\backslash\mathbb{H}$$

and then the map induced by the inclusion of $\alpha\Gamma_0(p)\alpha^{-1}$ into $\mathrm{SL}_2(\mathbf{Z})$.

Then if we are just interested in weight 0, we can define $\mathrm{T}_p f := (\pi_1)_* \circ \pi_2^* f$. The map π_2 pulls back f to a function on $\Gamma_0(p)\backslash\mathbb{H}$, and π_1 pushes it forward.

This works for weight two modular forms as well, by applying the same diagram but using the operations on $f(z)\mathrm{d}z$ (so the differential also transforms). We note that such forms get identified with $S_2(\mathrm{SL}_2(\mathbf{Z}))$.

> THEOREM 4.44. This also recovers the Hecke operator T_p for $S_2(\mathrm{SL}_2(\mathbf{Z}))$, up to a constant.

Proof. The initial pullback $\pi_2^* f$ does nothing particularly interesting on weight two cusp forms: it is simply telling us that a cusp form for $\mathrm{SL}_2(\mathbf{Z})$ of weight two also makes sense for $\Gamma_0(p)$. Thus, we can view it as the very same function $f(z)$ but in $S_2(\Gamma_0(p))$.

The final map $(\pi_1)_*$ requires a bit more thought. Observe that if $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ then $\alpha^{-1}\Gamma\alpha \cap \Gamma = \Gamma_0(p)$. Then, π_1 is explicitly given by the composition

$$(\alpha^{-1}\Gamma\alpha \cap \Gamma)\backslash\mathbb{H} \longrightarrow (\alpha\Gamma\alpha^{-1} \cap \Gamma)\backslash\mathbb{H} \xrightarrow{\tilde{\pi}_1} \Gamma\backslash\mathbb{H}.$$

The first map sends $\Gamma_0(p) \cdot z \mapsto (\alpha\Gamma_0(p)\alpha^{-1}) \cdot \alpha z$. In effect, we send f to $f|_{\alpha^{-1}}$ when we push forward: you pull back by the inverse in this case.

We aim to understand preimages for the second map. In the case of second map $\tilde{\pi}_1$ making up π_1, explicitly $((\tilde{\pi}_1)_* f)(z) := \sum_{y \in \pi_1^{-1}(z)} f(y)$. The preimages, for given $[\Gamma \cdot z] \in \Gamma\backslash\mathbb{H}$, are given by acting on z by a coset in $(\alpha\Gamma\alpha^{-1} \cap \Gamma)\backslash\Gamma$.

Now note that
$$(\alpha\Gamma\alpha^{-1} \cap \Gamma)\backslash\Gamma \simeq \Gamma\backslash\Gamma\alpha^{-1}\Gamma.$$

The bijection here takes a coset representative β_j in the right and writes $\beta_j = \alpha^{-1}\gamma_j$ where $\gamma_j \in \mathrm{SL}_2(\mathbf{Z}) = \Gamma$. These γ_j are the coset representatives on the left.

We therefore see what is happening now: we obtain $\sum_j f(z)|_{\alpha^{-1}\gamma_j}$ from $(\pi_1)_*$, by first applying α (the isomorphism) and then enumerating coset representatives to find preimages. But as $\beta_j = \alpha^{-1}\gamma_j$, we get the Hecke operator for α^{-1}. This only differs as a coset by a constant, and so it gives the same operator. □

A very similar definition can work for $S_2(\mathrm{B}^\times)$, thinking of these as also being functions on the modular curve and essentially using the same diagram. If one translates the previous definitions into adelic language, we see $\Gamma_0(p)$ only makes a change at p from $\mathrm{SL}_2(\mathbf{Z})$ when we look at corresponding subgroups of $\mathrm{GL}_2(\mathbf{A}_f)$. Thus, the double coset ends up being able to be reformulated in a way which is also local at p. Translating to $\mathrm{B}^\times$ we would get the following definition, which I have simplified using strong approximation.

Definition 4.45. Let $f \in S_2(\mathrm{B}^\times)$, thought of as a function on $\mathrm{B}^\times(\mathbf{Z}[1/p]) \setminus \mathrm{GL}_2(\mathbf{Q}_p)/\mathrm{GL}_2(\mathbf{Z}_p)$. Then

$$\mathrm{T}_p f(\gamma) := \sum_j f(\gamma\beta_j)$$

where

$$\mathrm{GL}_2(\mathbf{Z}_p)\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\mathrm{GL}_2(\mathbf{Z}_p) = \coprod_j \mathrm{GL}_2(\mathbf{Z}_p)\beta_j.$$

Now for the main point:

Theorem 4.46. Under the identification

$$\mathrm{B}^\times(\mathbf{Q})\backslash \mathrm{B}^\times(\mathbf{A_Q})/\mathrm{B}^\times(\mathbf{R})\prod_q \mathrm{B}^\times(\mathbf{Z}_q) \simeq \mathrm{X}/\Gamma$$

and hence function spaces $S_2(\mathrm{B}^\times) \simeq \mathbf{C}[\mathrm{X}/\Gamma]$, the Hecke operator T_p corresponds to the adjacency operator.

Proof. This amounts to unraveling definitions at this point: we just need to push the definition of Hecke operators through this identification.

Consider a function f on X/Γ. We know this is the same as the double coset

$$\mathrm{B}^\times(\mathbf{Z}[1/p]) \setminus \mathrm{GL}_2(\mathbf{Q}_p)/\mathrm{GL}_2(\mathbf{Z}_p).$$

Viewing f as a function there, we now take $\mathrm{GL}_2(\mathbf{Z}_p)\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\mathrm{GL}_2(\mathbf{Z}_p)$ and decompose it as cosets $\coprod_j \mathrm{GL}_2(\mathbf{Z}_p)\beta_j$, and we know $\mathrm{T}_p f(\gamma) = \sum_j f(\gamma\beta_j)$.

We now view f as a $\mathrm{B}^\times(\mathbf{Z}[1/p])$-invariant function on $\mathrm{GL}_2(\mathbf{Q}_p)/\mathrm{GL}_2(\mathbf{Z}_p)$, or equivalently a $\mathrm{D}^\times(\mathbf{Z}[1/p])$-invariant function on X after sending $\Lambda \mapsto [\Lambda]$; we've already checked that such functions are constant on all lattices in a lattice class, by identifying $S_2(\mathrm{B}^\times)$ with $\mathbf{C}[\mathrm{X}/\Gamma]$.

The action of $\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$ as an element of $\mathrm{B}^\times(\mathbf{Q}_p) = \mathrm{GL}_2(\mathbf{Q}_p)$ sends $[\mathbf{Z}_p^2]$ to a particular neighbor: the distance between $\mathbf{Z}_p \oplus \mathbf{Z}_p$ and $p\mathbf{Z}_p \oplus \mathbf{Z}_p$ is 1. The action of an element in $\mathrm{GL}_2(\mathbf{Z}_p)\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\mathrm{GL}_2(\mathbf{Z}_p)$ first fixes $[\mathbf{Z}_p^2]$, then sends it to a neighbor, and then sends it to another neighbor; the coset representatives β_j simply go through all the neighbors. It follows that the Hecke operator sends $f(x) \mapsto \mathrm{T}_p f(x) := \sum_{d(x,y)=1} f(y)$, or is just the adjacency operator. $\square$

Thus, we are reduced to computing the eigenvalues of T_p on $S_2(\mathrm{B}^\times)$.

4.6 Putting it all together

We have now seen that the adjacency matrix on X/Γ can be interpreted as the Hecke operator T_p on $S_2(B^\times)$. As I mentioned before, I slightly lied to you about what $S_2(B^\times)$ officially is: if we want to be consistent with the more general theory of automorphic forms, we want to remove the constant functions and not consider them as cusp forms. Our definition of a Hecke operator does make sense on the larger vector space.

Instead, what this amounts to is adding a one dimensional $(p+1)$-eigenspace consisting of constant functions for T_p on $S_2(B^\times)$, or equivalently the adjacency operator on $\mathbf{C}[X/\Gamma]$.

The following theorem relates the usual definition of cusp forms on $B^\times$ to classical cusp forms.

THEOREM 4.47 (Old fashioned Jacquet-Langlands). There is a Hecke-equivariant isomorphism

$$\varphi : S_2(B)/\text{const} \simeq S \subseteq S_2(\Gamma_0(d)).$$

By equivariant, we mean in particular that

$$\begin{tikzcd} S_2(B)/\text{const} \arrow[r, "\varphi"] \arrow[d, "T_p"] & S_2(\Gamma_0(d)) \arrow[d, "T_p"] \\ S_2(B)/\text{const} \arrow[r, "\varphi"] & S_2(\Gamma_0(d)) \end{tikzcd}$$

commutes.

In particular, the eigenvalues are the same. This is because if we have an eigenform f in $S_2(B)/\text{const}$, then $T_p f = \lambda f$. But we have a linear isomorphism, so $\varphi(T_p f) = T_p\varphi(f)$ is also $\varphi(\lambda f) = \lambda\varphi(f)$, and so $T_p\varphi(f) = \lambda\varphi(f)$. It follows that if eigenvalues of T_p on $S_2(\Gamma_0(d))$ satisfy the desired bounds, then so do eigenvalues of T_p on $S_2(B)/\text{const}$.

We can compute eigenvalues for T_p via classical theory on $S_2(\Gamma_0(d))$, which gives us access to a wealth of new tools. The proof of the following claim uses some more difficult algebraic geometry (and is complicated as well), so we'll skip the full argument. If you are interested, you can take a look at Diamond and Shurman's *A first course in modular forms*. A good chunk of the book is essentially dedicated to this result.

THEOREM 4.48. The eigenvalues of T_p are bounded by $2\sqrt{p}$ on $S_2(\Gamma_0(N))$.

Thus, we have shown X/Γ is Ramanujan!