$\mathrm{SL}(2,\mathbb{C})$ REPRESENTATIONS AND 3-DIMENSIONAL HYPERBOLIC GEOMETRY

STEVEN BOYER AND CHI CHEUK TSANG
FALL 2024

1. $\mathrm{SL}(2,\mathbb{C})$ representations

1.1. **The group $\mathrm{SL}(2,\mathbb{C})$.** The main character of this course is

$$\mathrm{SL}(2,\mathbb{C}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a,b,c,d \in \mathbb{C}, ad-bc=1 \right\}.$$

We recall some basic facts and terminology for this group.

Proposition 1.1. $\mathrm{SL}(2,\mathbb{C})$ *is homeomorphic to* $\mathbb{R}^3 \times S^3$*. In particular it has 3 complex dimensions and is simply connected.*

Proof. The map $\mathrm{SL}(2,\mathbb{C}) \to \mathbb{C}^2 \backslash \{0\}$ that takes each matrix to its first column is a fibration whose fiber over $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is $\left\{ \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \mid b \in \mathbb{C} \right\} \cong \mathbb{C}$. Thus $\mathrm{SL}(2,\mathbb{C}) \cong (\mathbb{C}^2 \backslash \{0\}) \times \mathbb{C} \cong \mathbb{R}^3 \times S^3$. $\square$

The **characteristic polynomial** of an element $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{C})$ is $x^2 - (a+d)x + 1$. Its eigenvalues are $\frac{a+d+((a+d)^2-4)^{\frac{1}{2}}}{2}$, with associated eigenvectors $\begin{bmatrix} \frac{a-d+((a+d)^2-4)^{\frac{1}{2}}}{2} \\ c \end{bmatrix}$.

The matrix A can be classified under a trichotomy depending on the value of $\mathrm{tr}(A) = a+d$:

- If $\mathrm{tr}(A) \in (-2,2)$, then A admits two distinct conjugate eigenvalues that lie on the unit circle. In this case we say that A is **elliptic**.
- If $\mathrm{tr}(A) \pm 2$, then A has an eigenvalue ± 1 with multiplicity two. Either $A = \pm I$, or the ± 1-eigenspace of A is 1-dimensional. In the latter case we say that A is **parabolic**.
- If $\mathrm{tr}(A) \notin [-2,2]$, then A admits two distinct eigenvalues that lie off the unit circle. In this case we say that A is **hyperbolic**.

If $A \neq \pm I$, the conjugancy class of A is determined by its trace. For example, if $A \neq \pm I$, then $A^n = I$ if and only if $\mathrm{tr}(A) = 2\cos(\frac{2k\pi}{n}) \neq \pm 2$ for some $k \in \mathbb{Z}$.

Proposition 1.2. *The center of* $\mathrm{SL}(2,\mathbb{C})$ *is* $\{\pm I\}$

Proof. Two matrices $A, B \in \mathrm{SL}(2,\mathbb{C})$ commute if and only if A preserves the eigenspaces of B. In particular if $A \in \mathrm{SL}(2,\mathbb{C})$ commutes with every element then A must be a multiple of the identity. $\square$

We define

$$\mathrm{PSL}(2,\mathbb{C}) = \mathrm{SL}(2,\mathbb{C})/\{\pm I\}$$

Finally, we record some trace identities

Proposition 1.3.

- *For $A \in \mathrm{SL}(2,\mathbb{C})$, $\mathrm{tr}(A^2) = \mathrm{tr}(A)^2 - 2$.*
- *For $A, B \in \mathrm{SL}(2,\mathbb{C})$, $\mathrm{tr}(AB) + \mathrm{tr}(AB^{-1}) = \mathrm{tr}(A)\mathrm{tr}(B)$.*

Corollary 1.4. *For any word w, $\mathrm{tr}(w(A_1, ..., A_n))$ is a polynomial function in $(\mathrm{tr}(A_{i_1}...A_{i_k}))_{1 \le i_1 < ... < i_k \le n}$.*

Proof. If there are any occurrences of A_i^p with $p > 1$ in w, we cyclically permute the terms until $w = w' A_i^p$. Then we apply the identity

$$\mathrm{tr}(w' A_i^p) + \mathrm{tr}(w' A_i^{p-2}) = \mathrm{tr}(w' A_i^{p-1})\mathrm{tr}(A_i)$$

to reduce p to 0 or 1.

Similarly, if there are any occurrences of A_i^p with $p < 0$, we can reduce to the case where $p = 0$ or 1.

Once all the exponents in the word are 1, if there are any repeat occurrences of A_i, we cyclically permute until $w = w' A_i w'' A_i$. Then we apply the identity

$$\mathrm{tr}(w' A_i w'' A_i) + \mathrm{tr}(w' w''^{-1}) = \mathrm{tr}(w' A_i)\mathrm{tr}(w'' A_i)$$

to reduce the length of the word.

Repeating this process, we eventually reduce to the case when $w = A_{i_1}...A_{i_k}$ for distinct $1 \le i_1, ..., i_k \le n$. At this point, we can first cyclically permute so that i_1 is the smallest of the i_r, then apply the identity

$$\mathrm{tr}(A_{i_1}...A_{i_r}A_{i_{r+1}}...A_{i_k}) + \mathrm{tr}(A_{i_1}...A_{i_r}A_{i_k}^{-1}...A_{i_{r+1}}^{-1}) = \mathrm{tr}(A_{i_1}...A_{i_r})\mathrm{tr}(A_{i_{r+1}}...A_{i_k})$$

to rearrange the terms and to reduce to shorter words, inductively. □

1.2. **Representation variety.** Let π be a finitely presented group, i.e. $\pi = \langle s_1, ..., s_m \mid r_1, ..., r_n \rangle$ for a finite set of generators $s_1, ..., s_m$ satisfying a finite set of relations $r_1, ..., r_n$. A $\mathrm{SL}(2,\mathbb{C})$**-representation** is a homomorphism $\rho : \pi \to \mathrm{SL}(2,\mathbb{C})$.

The set of $\mathrm{SL}(2,\mathbb{C})$-representations of π is called the **representation variety** of π and is denoted by $R(\pi)$. Its structure as a variety comes from considering the matrix entries of $\rho(s_1), ..., \rho(s_m)$ as coordinates and writing

$$R(\pi) = \left\{ a_1, b_1, c_1, d_1, ..., a_m, b_m, c_m, d_m \in \mathbb{C} \mid a_i d_i - b_i c_i = 1, r_j\left(\begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

and noting that the matrix entries of $r_j\left(\begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}\right)$ are polynomials in a_i, b_i, c_i, d_i.

Remark 1.5. To be more specific, $R(\pi)$ is a complex affine algebraic set. Some authors require varieties to be irreducible. We will not impose this requirement in these notes.

$\mathrm{SL}(2,\mathbb{C})$ acts on $R(\pi)$ via conjugation: $(A \cdot \rho)(g) = A\rho(g)A^{-1}$ for every $g \in \pi$.

A representation $\rho \in R(\pi)$ is **reducible** if $\rho(\pi)$ preserves a 1-dimensional subspace of $\mathbb{C}^2$. Equivalently, a representation $\rho \in R(\pi)$ is reducible if $\rho(\pi)$ can be conjugated into the subgroup of upper triangular matrices. We show two more characterizations of reducibility.

Lemma 1.6 ([CS83, Corollary 1.2.2])**.** *A representation $\rho \in R(\pi)$ is reducible if and only if $\mathrm{tr}(\rho(g)) = 2$ for every $g \in [\pi, \pi]$.*

Proof. Suppose ρ is reducible. Then up to a conjugation, we can assume that $\rho(\pi)$ is a subgroup of the upper triangular matrices. Every element in the commutator subgroup of the upper triangular matrices has trace 2.

Conversely, suppose $\mathrm{tr}(\rho(g)) = 2$ for every $g \in [\pi, \pi]$. If $\rho(\pi)$ is abelian, then the elements of the subgroup must be all of the same type and have the same fixed points. Otherwise, $\rho(g) \neq I$ for some $g \in [\pi, \pi]$. Since $\mathrm{tr}(\rho(g)) = 2$, $\rho(g)$ is parabolic and has a unique eigenspace L.

We claim that $\rho(h)$ preserves L for every $h \in [\pi, \pi]$. Suppose otherwise. Up to a conjugation, we can assume that $L = \left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle$ and $\rho(g) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then $\rho(h) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ where $a + d = 2$ and $c \neq 0$. But then $gh \in [\pi, \pi]$ and $\mathrm{tr}(\rho(gh)) = \mathrm{tr} \begin{bmatrix} a & a+b \\ c & c+d \end{bmatrix} = a + d + c \neq 2$. Contradiction.

This implies that $\rho(h)$ preserves L for every $h \in \pi$. Indeed, $\rho(hgh^{-1})$ is parabolic and preserves $(\rho(h))(L)$, and $hgh^{-1} \in [\pi, \pi]$, hence $(\rho(h))(L) = L$. This shows that $\rho(\pi)$ is reducible. $\square$

Lemma 1.7 ([CS83, Lemma 1.5.1]). *Let $\rho \in R(\pi)$ be an irreducible representation. Suppose $\rho(g) \neq \pm I$ for some $g \in \pi$. Then there exists $h \in \pi$ such that the restriction of ρ to $\langle g, h \rangle$ is irreducible and* $\mathrm{tr}(\rho(h)) \neq \pm 2$.

Proof. We first show that there exists $h \in H$ such that the restriction of ρ to $\langle g, h \rangle$ is irreducible. If $\rho(g)$ is parabolic then there exists h such that $\rho(h)$ does not preserve the eigenspace of $\rho(g)$, and we are done. Otherwise $\rho(g)$ has two distinct eigenspaces L_1 and L_2. There exists h_1, h_2 such that h_1 does not preserve L_1 and h_2 does not preserve L_2. If h_1 does not preserve L_2 we take $h = h_1$. If h_2 does not preserve L_1 we take $h = h_2$. Otherwise h_1 preserves L_2 and h_1 preserves L_1, and we take $h = h_1 h_2$.

If $\mathrm{tr}(\rho(h)) \neq \pm 2$ then we are done. Otherwise $\rho(h)$ is parabolic. For every n, $\rho(h^n)$ is also parabolic and shares the same eigenspace as $\rho(h)$. Thus

$$\begin{aligned} \text{the restriction of } \rho \text{ to } \langle g, h \rangle \text{ is irreducible} &\iff \rho(g) \text{ does not fix the eigenspace of } \rho(h) \\ &\iff \rho(g) \text{ does not fix the eigenspace of } \rho(h^n) \\ &\iff \text{the restriction of } \rho \text{ to } \langle g, h^n \rangle \text{ is irreducible.} \end{aligned}$$

Up to conjugation we can assume $\rho(h) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Meanwhile, suppose $\rho(g) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Since $\rho(g)$ cannot preserve $\left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle$, $c \neq 0$. Now $\mathrm{tr}(\rho(gh^n)) = \mathrm{tr} \begin{bmatrix} a & na+b \\ c & nc+d \end{bmatrix} = (a + d) + nc$, hence for suitable n, $\mathrm{tr}(\rho(gh^n)) \neq \pm 2$ and the restriction of ρ to $\langle g, gh^n \rangle = \langle g, h^n \rangle$ is irreducible. $\square$

Corollary 1.8. *A representation $\rho \in R(\pi)$ is reducible if and only if the restriction of ρ to every two-generator subgroup is reducible.*

1.3. **Character variety.** The **character** of a representation $\rho \in R(\pi)$ is the function $\chi_\rho : \pi \to \mathbb{C}$ defined by $\chi_\rho(g) = \mathrm{tr}(\rho(g))$.

Two conjugate representations have the same character. When restricted to irreducible representations, the converse is also true.

Proposition 1.9 ([CS83, Proposition 1.5.2]). *Let $\rho, \rho' \in R(\pi)$. Suppose ρ is irreducible and $\chi_\rho = \chi'_\rho$. Then ρ and ρ' are conjugate.*

Proof. It follows from Lemma 1.6 that ρ' is irreducible.

By Lemma 1.7, there exists g, h such that the restriction of ρ to $\langle g, h \rangle$ is irreducible and $\chi_\rho(h) \neq \pm 2$. By Lemma 1.6 again, the restriction of ρ' to $\langle g, h \rangle$ is also irreducible.

Up to individually conjugating ρ and ρ', we can assume that $\rho(h) = \rho(h') = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$ where $\lambda \neq \pm 1$. Meanwhile, let $\rho(g) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\rho'(g) = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$. Since the restriction of ρ and ρ' to $\langle g, h\rangle$ are irreducible, $b, c, b', c' \neq 0$. Thus up to further individual conjugation by diagonal matrices, we can assume that $b = b' = 1$.

We compute

$$a + d = \chi_\rho(g) = \chi_{\rho'}(g) = a' + d'$$
$$a\lambda + d\lambda^{-1} = \chi_\rho(gh) = \chi_{\rho'}(gh) = a'\lambda + d'\lambda^{-1}$$

Thus $a = a'$, $d = d'$. This implies that $c = ad + 1 = a'd' + 1 = c'$.

Now for every $k \in \pi$. Suppose $\rho'(k) = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ and $\rho'(k) = \begin{bmatrix} p' & q' \\ r' & s' \end{bmatrix}$. As above, the equalities $\chi_\rho(k) = \chi_{\rho'}(k)$ and $\chi_\rho(kh) = \chi_{\rho'}(kh)$ imply $p = p'$ and $s = s'$. We also compute

$$pa + qc + r + sd = \chi_\rho(kg) = \chi_{\rho'}(kg) = pa + q'c + r' + sd$$
$$(pa + qc)\lambda + (r + sd)\lambda^{-1} = \chi_\rho(kgh) = \chi_{\rho'}(kgh) = (pa + q'c)\lambda + (r' + sd)\lambda^{-1}$$

Thus $pa + qc = pq + q'c$ and $r + sd = r' + sd$, which implies that $q = q'$ and $r = r'$ since $c \neq 0$. This shows that $\rho = \rho'$. $\square$

The set of characters of $\mathrm{SL}(2, \mathbb{C})$-representations of π is called the **character variety** of π and is denoted by $X(\pi)$.

The structure of $X(\pi)$ as a variety comes from considering the regular map $t : R(\pi) \to \mathbb{C}^{2^m-1}$ defined by $t(\rho) = (\mathrm{tr}(\rho(s_{i_1} \ldots s_{i_k})))|_{1 \leq i_1 < \ldots < i_k \leq m}$. Chevalley's theorem in algebraic geometry ([Har92, Theorem 3.16]) shows that the image $t(R(\pi))$ is a variety.

More specifically, Chevalley's theorem states that $t(R(\pi))$ is a constructible set, i.e. the complement of an affine algebraic set in an affine algebraic set. But in fact, Culler and Shalen showed that $t(R(\pi))$ is an affine algebraic set ([CS83, Corollary 1.4.5]).

Corollary 1.4 implies that t descends to a bijection $X(\pi) \to t(R(\pi))$. We identify $X(\pi)$ with $t(R(\pi))$ via this bijection. In particular, $X(\pi)$ inherits a structure as a variety, for which $t : R(\pi) \to X(\pi)$ is a regular map.

Corollary 1.4 also implies that for every $g \in \pi$, the function $I_g(\chi_\rho) = \mathrm{tr}(\rho(g))$ is a well-defined regular function on $X(\pi)$.

Let $X^{\mathrm{red}}(\pi)$ be the subset consisting of characters of reducible representations.

Proposition 1.10. *$X^{\mathrm{red}}(\pi)$ is a closed algebraic subset of $X(\pi)$.*

Proof. By Lemma 1.6, $X^{\mathrm{red}}(\pi) = \cap_{g \in [\pi,\pi]} I_g^{-1}(2)$. $\square$

The following observation is useful for computing $X^{\mathrm{red}}(\pi)$.

Lemma 1.11. *$X^{\mathrm{red}}(\pi)$ is the subset of characters of representations into the subgroup of diagonal matrices* $\left\{ \begin{bmatrix} a & 0 \\ 0 & a^{-1} \end{bmatrix} \right\} \cong \mathbb{C}$.

Proof. Every reducible representation can be conjugated into a representation into the subgroup of upper triangular matrices. Given such a representation $\rho(g) = \begin{bmatrix} a(g) & b(g) \\ 0 & a(g)^{-1} \end{bmatrix}$, we can define another representation $\rho'(g) = \begin{bmatrix} a(g) & 0 \\ 0 & a(g)^{-1} \end{bmatrix}$ which has the same character. $\square$

Lemma 1.11 makes computing $X^{\text{red}}(\pi)$ very easy. We define $X^{\text{irr}}(\pi)$ to be $\overline{X(\pi)\backslash X^{\text{red}}(\pi)}$. This is the actual interesting part of $X(\pi)$.

1.4. **Representations via hyperbolic geometry.** The hyperbolic space $\mathbb{H}^3$ is the space $\{(x, y, z) \in \mathbb{R}^3 \mid z > 0\}$ with the Riemannian metric $g = \frac{1}{z^2}(dx^2 + dy^2 + dz^2)$.

The **hyperbolic space** $\mathbb{H}^3$ is the Riemannian manifold

$$\left(\left\{(x, y, z) \in \mathbb{R}^3 \mid z > 0\right\}, \frac{1}{z^2}(dx^2 + dy^2 + dz^2)\right).$$

The **geodesic lines** in $\mathbb{H}^3$ are straight lines and circles that are perpendicular to the $\{z = 0\}$ plane. The **geodesic planes** in $\mathbb{H}^3$ are the flat planes and spheres that are perpendicular to the $\{z = 0\}$ plane. The **horospheres** in $\mathbb{H}^3$ are the horizontal flat planes and the spheres that are tangent to the $\{z = 0\}$ plane.

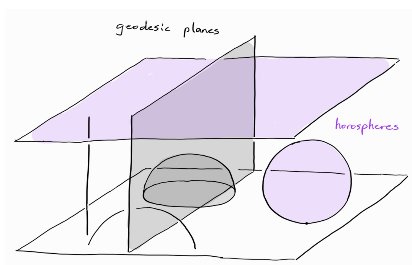

FIGURE 1.

The boundary at infinity of $\mathbb{H}^3$ is $S^2_\infty = \mathbb{C} \cup \{\infty\}$. Here we identify the $\{z = 0\}$ plane with $\mathbb{C}$ by $(x, y) \longleftrightarrow x + iy$.

We define $\overline{\mathbb{H}^3}$ to be the union of $\mathbb{H}^3 \cup S^2_\infty$, endowed with the topology where $\mathbb{H}^3$ is an open set, and regions enclosed by geodesic planes define basis of open neighborhoods for points on S^2_∞. In particular, $\overline{\mathbb{H}^3}$ is homeomorphic to the closed 3-ball.

Remark 1.12. Some authors prefer using the ball model for $\mathbb{H}^3$. This is the image of our half-space model under the map

$$(x, y, z) \mapsto \left(\frac{2x}{x^2 + y^2 + (z+1)^2}, \frac{2y}{x^2 + y^2 + (z+1)^2}, \frac{2(z+1)}{x^2 + y^2 + (z+1)^2} - 1\right).$$

Under this map, $\mathbb{H}^3$ is sent to the open unit ball and S^2_∞ is sent to the unit sphere, thus $\overline{\mathbb{H}^3}$ is exactly the closed unit ball. The Riemannian metric on $\mathbb{H}^3$ is transformed to $g = \frac{4}{(1-x^2-y^2-z^2)^2}(dx^2 + dy^2 + dz^2)$.

$\mathrm{SL}(2, \mathbb{C})$ acts on S^2_∞ by $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot w = \frac{aw+b}{cw+d}$. The maps $w \mapsto \frac{aw+b}{cw+d}$ in this action are called the **Möbius transformations**. They preserve the straight lines and circles on $\mathbb{C}$. The fixed points of $w \mapsto \frac{aw+b}{cw+d}$ are $\frac{(a-d)\pm((a+d)^2-4)^{\frac{1}{2}}}{2c}$. Note that they correspond exactly to the eigenspaces of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Each Möbius transformation extends uniquely into a homeomorphism of $\overline{\mathbb{H}^3}$ that restricts to an isometry of $\mathbb{H}^3$ by requiring that the extension preserves the geodesic planes.

The trichotomy of elements in $\mathrm{SL}(2,\mathbb{C})$ corresponds to a trichotomy of isometric actions.

- If A is elliptic with eigenvalues $e^{\pm i\theta}$, then A rotates $\mathbb{H}^3$ around the geodesic line passing through the fixed points $\frac{(a-d)+((a+d)^2-4)^{\frac{1}{2}}}{2c}$ by an angle of 2θ. In this case, A can be conjugated to $\begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \cdot w = e^{2i\theta} w$ which rotates $\mathbb{H}^3$ around the vertical axis $\{x = y = 0\}$ by an angle of 2θ.
- If A is parabolic, then A shears $\mathbb{H}^3$ along the horospheres based at the fixed point $\frac{a-d}{2c}$. In this case, A can be conjugated to $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot w = w + 1$ which shears $\mathbb{H}^3$ along the horospheres $\{z = z_0\}$.
- If A is hyperbolic with eigenvalues $\lambda^{\pm 1}$, then A translates and rotates $\mathbb{H}^3$ along the geodesic line passing through the fixed points $\frac{(a-d)+((a+d)^2-4)^{\frac{1}{2}}}{2c}$. In this case, A can be conjugated to $\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \cdot w = \lambda^2 w$ which translates $\mathbb{H}^3$ along the vertical axis $\{x = y = 0\}$ by a distance of $2\log|\lambda|$ and an angle of $2\arg\lambda$.

Note that $\mathrm{SL}(2,\mathbb{C})$ acts transitively on the frame bundle of $\mathbb{H}^3$. Indeed, suppose (v_1, v_2, v_3) is a positively oriented orthonormal basis of vectors at a point $p \in \mathbb{H}^3$. There is a geodesic line tangent to v_1, going from $w_1 \in S^2_\infty$ to $w_2 \in S^2_\infty$. The map $w \mapsto \frac{w-w_1}{w-w_2}$ sends ℓ to the vertical axis $\{x = y = 0\}$. Hence p lies on this axis with v_1 pointing vertically upwards. By applying the maps $w \mapsto \lambda w$, we can translate p to $(0,0,1)$ and rotate (v_2, v_3) to be $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$.

Meanwhile, the kernel of the action is $\pm I$. Indeed, a matrix in the kernel cannot be elliptic, parabolic, nor hyperbolic by the above analysis.

This implies the following two propositions.

Proposition 1.13. $\mathrm{PSL}(2,\mathbb{C})$ *is isomorphic to the group of orientation-preserving isometries of* $\mathbb{H}^3$.

Proposition 1.14. $\mathrm{PSL}(2,\mathbb{C})$ *is homeomorphic to the frame bundle* $\mathrm{Fr}(\mathbb{H}^3)$.

Suppose M is an orientable 3-manifold. If M admits a complete hyperbolic metric g, then the universal cover $\widetilde{M}$ is isometric to $\mathbb{H}^3$. The fundamental group $\pi_1(M)$ acts by orientation-preserving isometries on $\widetilde{M} \cong \mathbb{H}^3$. Thus we have a discrete faithful representation $\rho_g : \pi_1(M) \to \mathrm{Isom}^+(\mathbb{H}^3) \cong \mathrm{PSL}(2,\mathbb{C})$. Note that the isometry $\widetilde{M} \cong \mathbb{H}^3$ is only determined up to an orientation-preserving isometry of $\mathbb{H}^3$. Consequently, ρ_g is only determined up to conjugation by elements of $\mathrm{PSL}(2,\mathbb{C})$.

Now, in general, a $\mathrm{PSL}(2,\mathbb{C})$ representation may or may not lift to a $\mathrm{SL}(2,\mathbb{C})$ representation. The obstruction is a second Stiefel Whitney class $w_2 \in H^2(\pi; \mathbb{Z}/2)$. When a lift exists, it is unique up to multiplication by homomorphisms $\pi \to \{\pm I\}$. In particular, there are $|H^1(\pi; \mathbb{Z}/2)|$ lifts.

An important fact is that $\mathrm{PSL}(2,\mathbb{C})$ representations coming from a hyperbolic orientable 3-manifold always lift.

Proposition 1.15 (Thurston)**.** *Let g be a complete hyperbolic metric on an orientable 3-manifold. The representation ρ_g admits a lift $\widehat{\rho_g} : \pi_1(M) \to \mathrm{SL}(2,\mathbb{C})$.*

Proof. We identify $\mathrm{PSL}(2,\mathbb{C})$ with the frame bundle $\mathrm{Fr}(\mathbb{H}^3)$ as in Proposition 1.14. Under this identification, the action of $\rho_g(\pi_1(M))$ on $\mathrm{PSL}(2,\mathbb{C})$ by left multiplication is identified with the action of $\pi_1(M)$ on $\mathrm{Fr}(\mathbb{H}^3)$.

Let $\rho_g(\widehat{\pi_1(M)})$ be the preimage of $\rho_g(\pi_1(M))$ under the projection $\mathrm{SL}(2,\mathbb{C}) \to \mathrm{PSL}(2,\mathbb{C})$. We have $\mathrm{SL}(2,\mathbb{C})/\rho_g(\widehat{\pi_1(M)}) \cong \mathrm{PSL}(2,\mathbb{C})/\rho_g(\pi_1(M)) \cong \mathrm{Fr}(\mathbb{H}^3)/\pi_1(M) = \mathrm{Fr}(M)$.

Since $\mathrm{SL}(2,\mathbb{C})$ is simply connected by Proposition 1.1, this shows that $\pi_1(\mathrm{Fr}(M)) \cong \rho_g(\widehat{\pi_1(M)}))$.

Meanwhile, every orientable 3-manifold is parallelizable. Thus $\pi_1(\mathrm{Fr}(M)) \cong \pi_1(M \times \mathrm{SO}(3)) \cong \pi_1(M) \times \mathbb{Z}/2$. We can thus pick a section $\pi_1(M) \to \pi_1(M) \times \mathbb{Z}/2 \cong \rho_g(\widehat{\pi_1(M)})) \subset \mathrm{SL}(2,\mathbb{C})$. □

The lift $\widehat{\rho_g} : \pi_1(M) \to \mathrm{SL}(2,\mathbb{C})$ is uniquely determined up to conjugation and multiplication by homomorphisms $\pi(M) \to \{\pm I\}$. Now, an element $A \in \mathrm{SL}(2,\mathbb{C})$ has the same trace as $-A$ if and only if A has order 4. But $\pi_1(M)$ is torsion-free, so lifts that differ by multiplication of a nontrivial homomorphism cannot have the same character. This shows that g uniquely determines $|H^1(M;\mathbb{Z}/2)|$ points in the character variety $X(\pi_1(M))$.

The first half of the class will focus on studying the character variety near one of these points.

1.5. **Examples.** We compute the character variety of some two-generator groups. For each of these groups, we will let s_1, s_2 be a set of generators, and embed $X(\pi)$ into $\mathbb{C}^3$ via the map $t(\chi) = (\chi(s_1), \chi(s_2), \chi(s_1 s_2))$ as in Section 1.3.

Example 1.16 (Free group F_2)**.** Consider $F_2 = \langle s_1, s_2 \rangle$. We first compute $X^{\mathrm{red}}(F_2)$. The representations of F_2 into the diagonal subgroup are

$$\left\{ \rho(s_1) = \begin{bmatrix} w_1 & 0 \\ 0 & w_1^{-1} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} w_2 & 0 \\ 0 & w_2^{-1} \end{bmatrix} \mid w_1, w_2 \in \mathbb{C} \right\}.$$

Thus, by Lemma 1.11, $X^{\mathrm{red}}(F_2)$ is the complex dimension 2 variety

$$\{(w_1 + w_1^{-1}, w_2 + w_2^{-1}, w_1 w_2 + w_1^{-1} w_2^{-1}) \mid w_1, w_2 \in \mathbb{C}\}.$$

Meanwhile, we claim that $X(F_2) = \mathbb{C}^3$. Indeed, for any $(w_1, w_2, w_3) \in \mathbb{C}$, the representation

$$\rho(s_1) = \begin{bmatrix} w_1 & 1 \\ 0 & w_1^{-1} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} w_2 & 0 \\ w_3 & w_2^{-1} \end{bmatrix}$$

maps to

$$(w_1 + w_1^{-1}, w_2 + w_2^{-1}, w_1 w_2 + w_1^{-1} w_2^{-1} + w_3).$$

These points fill up all of $\mathbb{C}^3$.

Thus $X^{\mathrm{irr}}(F_2) = \mathbb{C}^3$ as well.

Example 1.17 (Free product $\mathbb{Z}/p * \mathbb{Z}/q$)**.** Consider $\mathbb{Z}/p * \mathbb{Z}/q = \langle s_1, s_2 \mid s_1^p = s_2^q = 1 \rangle$. We first compute $X^{\mathrm{red}}(\mathbb{Z}/p * \mathbb{Z}/q)$. The representations of $\mathbb{Z}/p * \mathbb{Z}/q$ into the diagonal subgroup are

$$\left\{ \rho(s_1) = \begin{bmatrix} e^{\frac{2ik_1\pi}{p}} & 0 \\ 0 & e^{-\frac{2ik_1\pi}{p}} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} e^{\frac{2ik_2\pi}{q}} & 0 \\ 0 & e^{\frac{-2ik_2\pi}{q}} \end{bmatrix} \mid k_1, k_2 \in \mathbb{Z} \right\}.$$

Thus, by Lemma 1.11, $X^{\mathrm{red}}(\mathbb{Z}/p * \mathbb{Z}/q)$ is the union of finitely many points

$$\left\{ \left(2\cos(\frac{2k_1\pi}{p}), 2\cos(\frac{2k_2\pi}{q}), 2\cos(\frac{2k_1\pi}{p} + \frac{2k_2\pi}{q}) \right) \mid k_1, k_2 \in \mathbb{Z} \right\}.$$

Meanwhile, we claim that $X(\mathbb{Z}/p * \mathbb{Z}/q)$ is the union of finitely many lines

$$\left\{ \left(2\cos(\frac{2k_1\pi}{p}), 2\cos(\frac{2k_2\pi}{q}) \right) \mid k_1, k_2 \in \mathbb{Z} \right\} \times \mathbb{C}.$$

Indeed, for any representation ρ, since $\rho(s_1)^p = \rho(s_2)^q = 1$, $\mathrm{tr}(\rho(s_1)) = 2\cos(\frac{2k_1\pi}{p})$ and $\mathrm{tr}(\rho(s_2)) = 2\cos(\frac{2k_2\pi}{q})$ for some $k_1, k_2 \in \mathbb{Z}$.

Conversely, for any $k_1, k_2 \in \mathbb{Z}$, $w \in \mathbb{C}$, the representation

$$\rho(s_1) = \begin{bmatrix} e^{\frac{2ik_1\pi}{p}} & 1 \\ 0 & e^{-\frac{2ik_1\pi}{p}} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} e^{\frac{2ik_2\pi}{q}} & 0 \\ w & e^{-\frac{2ik_2\pi}{q}} \end{bmatrix}$$

maps to

$$\left(2\cos(\frac{2k_1\pi}{p}), 2\cos(\frac{2k_2\pi}{q}), 2\cos(\frac{2k_1\pi}{p} + \frac{2k_2\pi}{q}) + w\right).$$

Thus $X^{\mathrm{irr}}(\mathbb{Z}/p * \mathbb{Z}/q) = \left\{\left(2\cos(\frac{2k_1\pi}{p}), 2\cos(\frac{2k_2\pi}{q})\right) \mid k_1, k_2 \in \mathbb{Z}\right\} \times \mathbb{C}$ as well.

Example 1.18 (Torus knot complement)**.** Consider the group $\pi = \langle s_1, s_2 \mid s_1^p = s_2^q \rangle$, where p and q are relatively prime integers. This is the fundamental group of the (p, q)-torus knot complement.

We first compute $X^{\mathrm{red}}(\pi)$. The representations of π into the diagonal subgroup are

$$\left\{\rho(s_1) = \begin{bmatrix} w^q & 0 \\ 0 & w^{-q} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} w^p & 0 \\ 0 & w^{-p} \end{bmatrix} \mid w \in \mathbb{C}\right\}.$$

Thus, by Lemma 1.11,

$$X^{\mathrm{red}}(\pi) = \{(w^q + w^{-q}, w^p + w^{-p}, w^{p+q} + w^{-p-q} \mid w \in \mathbb{C}\} \cong \mathbb{C}.$$

Meanwhile, we claim that $X^{irr}(\mathbb{Z}/p * \mathbb{Z}/q)$ is the union of finitely many lines

$$\left\{\left(2\cos(\frac{k_1\pi}{p}), 2\cos(\frac{k_2\pi}{q})\right) \mid k_1, k_2 \in \mathbb{Z} \text{ , same parity, and not divisible by } p, q \text{ respectively}\right\} \times \mathbb{C}.$$

Indeed, for any irreducible representation ρ, since $\rho(s_1^p) = \rho(s_2^q)$ commutes with every element of $\rho(\pi)$, it must be $\pm I$. However, $\rho(s_1), \rho(s_2) \neq \pm I$. So, depending on whether $\rho(s_1^p) = \rho(s_2^q) = \pm I$, we have $\mathrm{tr}(\rho(s_1)) = 2\cos(\frac{k_1\pi}{p}), \mathrm{tr}(\rho(s_2)) = 2\cos(\frac{2k_2\pi}{q})$ for k_1, k_2 even/odd and not divisible by p, q respectively.

Conversely, for any $k_1, k_2 \in \mathbb{Z}$, same parity, and not divisible by (p, q) respectively, the representation

$$\rho(s_1) = \begin{bmatrix} e^{\frac{ik_1\pi}{p}} & 1 \\ 0 & e^{-\frac{ik_1\pi}{p}} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} e^{\frac{ik_2\pi}{q}} & 0 \\ w & e^{-\frac{ik_2\pi}{q}} \end{bmatrix}$$

maps to

$$\left(2\cos(\frac{k_1\pi}{p}), 2\cos(\frac{k_2\pi}{q}), 2\cos(\frac{k_1\pi}{p} + \frac{k_2\pi}{q}) + w\right).$$

Example 1.19 (Figure-eight knot complement)**.** Consider the group $\pi = \langle s_1, s_2 \mid s_1[s_1, s_2] = [s_1, s_2]s_2^{-1} \rangle$. This is the fundamental group of the figure-eight knot complement.

We first compute $X^{\mathrm{red}}(\pi)$. The representations of π into the diagonal subgroup are

$$\left\{\rho(s_1) = \begin{bmatrix} w & 0 \\ 0 & w^{-1} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} w^{-1} & 0 \\ 0 & w \end{bmatrix} \mid w \in \mathbb{C}\right\}.$$

Thus, by Lemma 1.11,

$$X^{\mathrm{red}}(\pi) = \{(w + w^{-1}, w + w^{-1}, 2) \mid w \in \mathbb{C}\} \cong \mathbb{C}.$$

Meanwhile, we claim that $X^{irr}(\pi) = \{(x, x, y) \mid x^2 - y = 1\}$. Indeed, for any irreducible representation ρ, since s_1 and s_2^{-1} are conjugate, they share the same eigenvalues $a^{\pm 1}$. Up to conjugation we can assume that the a-eigenspace of $\rho(s_1)$ and $\rho(s_2)$ are $\left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle$ and $\left\langle \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\rangle$ respectively. Correspondingly, we can conjugate so that $\rho(s_1) = \begin{bmatrix} a & 1 \\ 0 & a^{-1} \end{bmatrix}$ and $\rho(s_2) = \begin{bmatrix} a^{-1} & 0 \\ c & a \end{bmatrix}$ for some $c \in \mathbb{C}$.

A computation shows that

$$s_1[s_1,s_2] = \begin{bmatrix} a+2ac+ac^2-a^3c+a^{-1}c+a^{-1}c^2-ac & -c+a^2-a^{-2}c \\ a^{-2}c+a^{-2}c^2-c & -a^{-3}c+a^{-1} \end{bmatrix}$$

$$[s_1,s_2]s_2^{-1} = \begin{bmatrix} a+2ac+ac^2-a^3c+a^{-1}c+a^{-1}c^2-ac & -a^{-2}-a^{-2}c+1 \\ c^2-a^2c+a^{-2}c & -a^{-3}c+a^{-1} \end{bmatrix}$$

hence $s_1[s_1,s_2] = [s_1,s_2]s_2^{-1}$ if and only if $a^2+a^{-2} = c+1$.

Substituting $x = \mathrm{tr}(\rho(s_1)) = \mathrm{tr}(\rho(s_2)) = a+a^{-1}$ and $y = \mathrm{tr}(\rho(s_1s_2)) = c+2$, we have $x^2-y=1$.

Conversely, given x, y such that $x^2-y=1$, the representation

$$\rho(s_1) = \begin{bmatrix} a & 1 \\ 0 & a^{-1} \end{bmatrix}, \rho(s_2) = \begin{bmatrix} a^{-1} & 0 \\ c & a \end{bmatrix}$$

with $x = a+a^{-1}$ and $y = c+2$ is mapped to (x,x,y).

We will see later in the class that the points $(2,2,3)$ and $(-2,-2,3)$ correspond to the unique complete hyperbolic metric on the figure-eight knot complement.

2. Thick-thin decomposition

Throughout these notes, when we say a **hyperbolic 3-manifold**, we mean an orientable 3-manifold with a complete hyperbolic metric g.

Let M be a hyperbolic 3-manifold. The **injectivity radius** of a point $x \in M$ is the largest value of R such that the restriction of the universal cover $\widetilde{M} \to M$ to the open ball of radius R centered around a lift of x is injective.

Fix $\varepsilon > 0$. We define the **thick part** of M to be the set of points that have injectivity radius $\geq \frac{\varepsilon}{2}$ and denote it by $M_{[\varepsilon,\infty)}$. We define the **thin part** of M to be the set of points that have injectivity radius $\leq \frac{\varepsilon}{2}$ and denote it by $M_{(0,\varepsilon]}$.

Theorem 2.1 (Thick-thin decomposition)**.** *There exists a universal constant $\varepsilon_3 > 0$ such that for every hyperbolic 3-manifold M and for every $\varepsilon < \varepsilon_3$, the thin part $M_{(0,\varepsilon]}$ is a disjoint union of tubes around closed geodesic curves of length $< \varepsilon$ and truncated cusps.*

We will follow the presentation in [Mar22, Section 4].

2.1. Margulis lemma.

Lemma 2.2. *Let G be a Lie group. There is a neighbourhood U of $e \in G$ such that every discrete subgroup generated by elements in U is nilpotent.*

Proof. Fix some path metric on G. Consider the map $[\cdot,\cdot] : G \times G \to G$ that sends (g,h) to $[g,h]$. The map is smooth and sends $G \times \{e\}$ and $\{e\} \times G$ to e, so its differential at (e,e) is the zero map. Thus there exists $\varepsilon_0 > 0$ such that for all $g, h \in B_{\varepsilon_0}(e)$, $d([g,h],e) \leq \frac{1}{2}\min\{d(g,e), d(h,e)\}$.

Suppose Γ is a discrete subgroup of G that is generated by elements in $\Gamma \cap B_{\varepsilon_0}(e)$. There exists $\varepsilon_1 > 0$ such that $\Gamma \cap B_{\varepsilon_1}(e) = \{e\}$. Let k be a large enough integer such that $\varepsilon_1 > \frac{\varepsilon_0}{2^k}$. Then for all $g_1, ..., g_k \in \Gamma \cap B_{\varepsilon_0}(e)$, $[g_1,[g_2,...,[g_{k-1},g_k]...]] \in \Gamma \cap B_{\varepsilon_1}(e) = \{e\}$.

Hence $[G,[G,...,[G,G]...]] = \langle [g_1,[g_2,...,[g_{k-1},g_k]...]] \rangle = \{e\}$ and G is nilpotent of class k. $\square$

Lemma 2.3 (Margulis lemma)**.** *There exists a universal constant $\varepsilon_3 > 0$ such that, for every $x \in \mathbb{H}^3$, any discrete subgroup in $\mathrm{Isom}^+(\mathbb{H}^3) = \mathrm{PSL}(2,\mathbb{C})$ that is generated by elements that move x less than ε_3 is virtually nilpotent.*

Proof. Since $\text{Isom}^+(\mathbb{H}^3)$ acts transitively on $\mathbb{H}^3$, it suffices to prove the lemma for x being a fixed point.

Let K be the stabilizer of x in $\text{Isom}^+(\mathbb{H}^3)$. This is a compact subgroup that is isomorphic to $\text{SO}(3)$. By Lemma 2.2, there is a neighbourhood U of $e \in \text{Isom}^+(\mathbb{H}^3)$ such that every discrete subgroup generated by elements in U is nilpotent. The set UK is a neighborhood of K.

For $\varepsilon > 0$, let V_ε be the set of isometries that move x less than ε. Pick ε_0 so that the closure of V_{ε_0} lies in UK. By compactness of $\overline{V_{\varepsilon_0}}$, there exists finitely many elements $k_1, ..., k_m \in K$ such that $\overline{V_{\varepsilon_0}} \subset \bigcup_{i=1}^m Uk_i$. Pick ε_3 small enough so that $V_{\varepsilon_3}^m \subset V_{\varepsilon_0}$.

Now let Γ be a discrete subgroup that is generated by $\Gamma \cap V_{\varepsilon_3}$. Let Γ_U be the subgroup of Γ generated by elements in $\Gamma \cap U$. We claim that Γ_U has index $\leq m$ in Γ.

To show this, we let $r(i)$ be the number of cosets of Γ_U in Γ that can be represented by products of $\leq i$ elements in $\Gamma \cap V_{\varepsilon_3}$. Note that $1 = r(0) \leq r(1) \leq ...$, and if $r(i) = r(i+1)$ for some i, then $r(i+1) = r(i+2) = ...$ Thus it suffices to show that $r(m) \leq m$.

If we have more than m cosets that can be represented by products of $\leq m$ elements in $\Gamma \cap V_{\varepsilon_3}$, then since each of these representatives lie in $V_{\varepsilon_3}^m \subset V_{\varepsilon_0} \subset \bigcup_{i=1}^m Uk_i$, two of them can be represented by elements h, h' in the same Uk_i. But then $h'h^{-1} \in \Gamma \cap U \subset \Gamma_U$, so they represent the same coset. □

The optimal value of ε_3 is known as the **Margulis constant**. At the time of writing, it is known that the Margulis constant is between 0.104 and 0.616. The lower bound is due to Meyerhoff [Mey87] and the upper bound comes from the SnapPea census manifold `m027(-4,1)`.

Margulis lemma holds for hyperbolic spaces in all dimensions. In dimension 3, and in the torsion free setting, the virtually nilpotent groups can be described very explicitly.

Lemma 2.4. *Every virtually nilpotent, discrete, torsion free subgroup of* $\text{Isom}^+(\mathbb{H}^3)$ *is of one of the following types:*

- *The trivial subgroup.*
- $\mathbb{Z}$*, generated by a hyperbolic isometry.*
- $\mathbb{Z}$*, generated by a parabolic isometry.*
- $\mathbb{Z}^2$*, generated by two parabolic isometries.*

Proof. Let Γ be such a subgroup. Let H be a finite index nilpotent subgroup. If H is trivial then Γ is finite. But since Γ is torsion free, it must be trivial. If H is nontrivial, then any element g in the last nontrivial subgroup in the lower central series commutes with every element of H. Thus all the elements in H have the same fixed points.

Since H is discrete and torsion free, it cannot contain any elliptic elements, thus all nontrivial elements in H must be of the same type and share the same fixed points on S^2_∞.

More specifically, if H contains a hyperbolic element preserving a geodesic line ℓ, then every nontrivial element in H preserves ℓ. The subgroup of elements in $\text{Isom}^+(\mathbb{H}^3)$ that preserve ℓ is homeomorphic to $\mathbb{C}\backslash\{0\}$. Since H is discrete and torsion free, it is isomorphic to $\mathbb{Z}$.

If H contains a parabolic element preserving a point $w \in S^2_\infty$, then every nontrivial element in H preserves w. The subgroup of parabolic elements in $\text{Isom}^+(\mathbb{H}^3)$ that preserve w is homeomorphic to $\mathbb{C}$. Since H is discrete and torsion free, it is isomorphic to $\mathbb{Z}$ or $\mathbb{Z}^2$. □

We say that a nontrivial subgroup of $\text{Isom}^+(\mathbb{H}^3)$ is **elementary** if it is of one of the forms in Lemma 2.4.

2.2. **Tubes and truncated cusps.** Suppose Γ is an elementary subgroup. For $\varepsilon > 0$, we define $S_\Gamma(\varepsilon)$ to be the set

$$\{x \in \mathbb{H}^3 \mid \text{there exists } g \in \Gamma \text{ that moves } x \text{ for a distance of } \leq \varepsilon\}.$$

For any $h \in \mathrm{Isom}^+(\mathbb{H}^3)$, $S_{h\Gamma h^{-1}}(\varepsilon) = h(S_\Gamma(\varepsilon))$. In particular, for any h that centralizes Γ, we have $h(S_\Gamma(\varepsilon)) = S_\Gamma(\varepsilon)$.

Proposition 2.5. *If $\Gamma \cong \mathbb{Z}$ is generated by a hyperbolic element g that translates points by $\log|\lambda|$ along an axis ℓ, then $S_\Gamma(\varepsilon) = \begin{cases} \text{some closed } \delta \text{ neighborhood of } \ell & \text{if } \log|\lambda| \leq \varepsilon \\ \varnothing & \text{if } \log|\lambda| > \varepsilon \end{cases}$. In particular, $S_\Gamma(\varepsilon)/\Gamma$ is some closed tubular neighborhood of a closed geodesic curve of length $\leq \varepsilon$.*

Proof. Up to conjugation, we can assume that ℓ is the vertical axis $\{x = y = 0\}$ in $\mathbb{H}^3$.

Since any isometry that preserves ℓ commutes with g, $S_\Gamma(\varepsilon)$ is a cone $\{(x, y, z) \in \mathbb{H}^3 \mid x^2 + y^2 \leq r^2 z^2\}$.

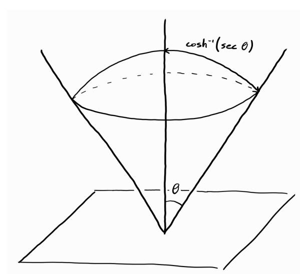

FIGURE 2.

The distance δ between $\partial S_\Gamma(\varepsilon)$ and ℓ is $\int_0^\theta \sec\theta d\theta = \cosh^{-1}(\sec\theta)$, where $r = \tan\theta$.

Also, we observe that

$$\begin{aligned} S_\Gamma(\varepsilon) \neq \varnothing &\iff \ell \subset S_\Gamma(\varepsilon) \\ &\iff \log|\lambda| \leq \varepsilon. \end{aligned}$$

□

Proposition 2.6. *If $\Gamma \cong \mathbb{Z}$ or $\mathbb{Z}^2$ is generated by parabolic elements that fix a point $w \in S^2_\infty$, then $S_\Gamma(\varepsilon) =$ some closed horoball based at w. In particular, $S_\Gamma(\varepsilon)/\Gamma \cong \begin{cases} S^1 \times \mathbb{R} \times [0, \infty) & \text{if } \Gamma \cong \mathbb{Z} \\ T^2 \times [0, \infty) & \text{if } \Gamma \cong \mathbb{Z}^2 \end{cases}$.*

Proof. Up to conjugation, we can assume that $w = \infty$. Since any parabolic isometry that preserves w centralizes Γ, $S_\Gamma(\varepsilon)$ is an upper half space $\{(x, y, z) \in \mathbb{H}^3 \mid z \geq z_0\}$. □

If a manifold M is the interior of a manifold with boundary $\overline{M}$, then a **cusp** of M is a boundary component of $\overline{M}$. In the context of Proposition 2.6, we say that the end of $S_\Gamma(\varepsilon)/\Gamma$ is a **rank one cusp** or **rank two cusp** depending on whether $\Gamma \cong \mathbb{Z}$ or $\mathbb{Z}^2$. Correspondingly, we refer to $S_\Gamma(\varepsilon)/\Gamma$ as a **rank one/two truncated cusp**.

Observe that a rank one truncated cusp has infinite volume while a rank two truncated cusp has finite volume. This follows from the fact that $\mathrm{vol}(K \times [z_0, \infty)) = \frac{1}{z_0^2}\mathrm{area}(K)$ for any $K \subset \mathbb{R}^2$.

2.3. **Proof and consequences.** We are now ready to prove thick-thin decomposition.

Proof of Theorem 2.1. Let M be a hyperbolic 3-manifold. A point x lies in the thin part $M_{(0,\varepsilon]}$ if and only if there exists $g \in \pi_1(M)$ such that g moves x by a distance of $\leq \varepsilon$. Thus the preimage of $M_{(0,\varepsilon]}$ in $\widetilde{M}$ can be written as

$$\widetilde{M_{(0,\varepsilon]}} = \bigcup_{g \in \pi_1(M)} S_g(\varepsilon)$$

where $S_g(\varepsilon) = \{x \in \mathbb{H}^3 \mid g \text{ moves } x \text{ for a distance of } \leq \varepsilon\}$.

By Lemma 2.3 and Lemma 2.4, as long as ε is smaller than the Margulis constant, then if two sets $S_{g_1}(\varepsilon)$ and $S_{g_2}(\varepsilon)$ intersect, they must lie in the same elementary subgroup. Thus we can further rewrite

$$\widetilde{M_{(0,\varepsilon]}} = \bigsqcup_{\Gamma \subset \pi_1(M)} S_\Gamma(\varepsilon)$$

where the disjoint union is taken over all elementary subgroups of $\pi_1(M)$.

Thus, modding out the action of $\pi_1(M)$,

$$M_{(0,\varepsilon]} = \bigsqcup_{\Gamma} S_\Gamma(\varepsilon)/\Gamma$$

where the disjoint union is now taken over all conjugacy classes of elementary subgroups of $\pi_1(M)$. □

Theorem 2.1 implies that whether a hyperbolic 3-manifold has finite volume is solely determined by its topology.

Corollary 2.7. *A hyperbolic 3-manifold M has finite volume if and only if it is the interior of a compact 3-manifold with torus boundary components.*

Proof. It is a general fact that a complete Riemannian manifold with boundary M with injectivity radius bounded from below by some $\varepsilon > 0$ has finite volume if and only if it is compact. Indeed, if M has finite volume, then it has finite diameter, otherwise the $\frac{\varepsilon}{2}$ balls along arbitrarily long distance-minimizing geodesics gives arbitrarily large volume.

Meanwhile, by the thick-thin decomposition, a hyperbolic 3-manifold is always homeomorphic to the interior of its thick part.

These two facts suffice to show the corollary: If M has finite volume, then its thick part has finite volume and so is compact. M is homeomorphic to the interior of this compact manifold with torus boundary components. Conversely, if M is the interior of a compact 3-manifold with torus boundary components, then M cannot have rank one cusps and its thick part is homeomorphic to this compact 3-manifold, thus is compact. □

3. Deforming hyperbolic representations

In this section, we set up some tools for understanding $\mathrm{SL}(2, \mathbb{C})$ representations that lie close to one that is induced by a complete hyperbolic metric.

3.1. **Dirichlet tessellation.** Let M be a hyperbolic 3-manifold. We identify $\widetilde{M} \cong \mathbb{H}^3$ and pick a base point $x_0 \in \mathbb{H}^3$.

The **Dirichlet domain** of $\pi_1(M)$ centered around x_0 is the set

$$D = \{x \in \mathbb{H}^3 \mid d(x, x_0) \leq d(x, gx_0) \text{ for all } g \in \pi_1(M)\}.$$

Since $\pi_1(M) \subset \mathrm{Isom}^+(\mathbb{H}^3)$ is discrete, D is a closed convex neighborhood of x_0. Its boundary can be written as a union of subsets of the form

$$F_h = \{x \in \mathbb{H}^3 \mid d(x, x_0) = d(x, hx_0) \leq d(x, gx_0) \text{ for all } g \in \pi_1(M)\}$$

as h varies over $\pi_1(M)$.

In fact, we claim that the union is a locally finite union, i.e. each compact subset only meets finitely many F_h. This follows from the discreteness of $\pi_1(M)$: The set of elements $h \in \pi_1(M)$ that can attain the minimum in $\min_{h\in\pi_1(M)} d(x, hx_0)$ for $x \in K$ is finite. This also implies that each F_h is a geodesic polygon.

D is a fundamental domain in the following sense: One can recover M by gluing the face F_h to the face $F_{h^{-1}}$ via h, for each pair of faces $(F_h, F_{h^{-1}})$ in the boundary of D.

Definition 3.1. A hyperbolic 3-manifold M is **geometrically finite** if it admits a Dirichlet domain with finitely many faces.

It can be shown that M admits a Dirichlet domain with finitely many faces for a particular basepoint x_0 if and only if it admits a Dirichlet domain with finitely many faces for any basepoint x_0. See [Rat19, Theorem 12.4.5].

Proposition 3.2. *A hyperbolic 3-manifold M with finite volume is geometrically finite.*

Proof. Choose ε small enough so that the thin part of M consists only of truncated cusps that do not contain the base point x_0.

Let $\widetilde{S}$ be the lift of a truncated cusp S, and let Γ be the elementary subgroup of $\pi_1(M)$ that stabilizes $\widetilde{S}$. We claim that the set of elements $h \in \pi_1(M)$ that can attain the minimum in $\min_{h\in\pi_1(M)} d(x, hx_0)$ for $x \in S$ lies in finitely many cosets of Γ.

Indeed, the truncated cusp S is naturally a subset of the quotient $\mathbb{H}^3/\Gamma$ that has compact boundary, so by discreteness of $\pi_1(M)$, there can only be finitely many images of points hx_0 in $\pi_1(M) \cdot x_0$ that can attain the minimum in $\min d(x, hx_0)$ for $x \in S$.

This implies that up to decreasing ε and shrinking $\widetilde{S}$, the Dirichlet domain D centered around x_0 meets $\widetilde{S}$ in a set that is combinatorially isomorphic to $P \times [0, \infty)$. In particular, D has finitely many sides in the thin part.

Meanwhile, the intersection of D with the lift of the thick part of M has finite diameter, since it is the fundamental domain for a compact manifold with boundary. Thus D has only finitely many sides in the lift of the thick part as well. □

The tessellation of $\mathbb{H}^3$ by the translates gD, $g \in \pi_1(M)$, is the **Dirichlet tessellation**.

A Dirichlet tessellation is **generic** if on every edge the points are equi-distant to exactly three hyperbolic translates or four parabolic translates of the basepoint. Since $\pi_1(M)$ is torsion free, we can always perturb the basepoint to get a generic Dirichlet tessellation. See [Bea95, Theorem 9.4.5] for a proof of the 2-dimensional analogue of this fact.

3.2. **Ford tessellation and Epstein-Penner decomposition.** The Dirichlet tessellation construction can be generalized as follows: Instead of taking a single base point x_0, we can take a finite set of points $x_1, ..., x_k$ and define

$$D_i = \{x \in \mathbb{H}^3 \mid d(x, x_i) \leq d(x, gx_j) \text{ for all } g \in \pi_1(M), j = 1, ..., k\}.$$

Suppose M has finite volume. Then it has a finite number of truncated cusps. A Ford tessellation is constructed by picking k basepoints as above and letting the basepoints go to infinity within lifts of the truncated cusps.

More precisely, let $S_1, ..., S_k$ be disjoint truncated cusps of M. We do not require the boundary of each S_i to have the same injectivitiy radius. Let $\widetilde{S}_i$ be a component of the preimage of S_i in $\mathbb{H}^3$. We define the **Ford domain** containing $\widetilde{S}_i$ to be

$$F_i = \{x \in \mathbb{H}^3 \mid d(x, \widetilde{S}_i) \leq d(x, g\widetilde{S_j}) \text{ for all } g \in \pi_1(M), j = 1, ..., k\}.$$

The same argument as before shows that each F_i is a convex closed neighborhood of $\widetilde{S}_i$ with boundary given by a locally finite union of geodesic polygons. The tessellation of $\mathbb{H}^3$ by the translates gF_i, $g \in \pi_1(M), i = 1, ..., k$, is the **Ford tessellation**.

Note however that a Ford domain is not a fundamental domain of M. If Γ_i is the elementary subgroup of $\pi_1(M)$ that stabilizes $\widetilde{S}_i$, then F_i is sent to itself under Γ_i. To formulate the appropriate fundamental domain statement, consider the hyperbolic polygons $\overline{F_i} = F_i/\Gamma_i$. Under our assumption that M has finite volume, each $\overline{F_i}$ has finitely many faces and can be glued along pairs of faces to recover M.

A retraction of the i^{th} cusp onto $\partial\overline{F_i}$ induces a retraction of M onto the cell complex $S = \bigcup_i \partial\overline{F_i}$ in M. In other words, S is a spine of M.

The dual ideal cellulation to S is called the Epstein-Penner decomposition. More precisely, on every 2-cell σ of S, the points are equi-distance from two truncated cusps S_{i_1} and S_{i_2}. We take the geodesic line between the i_1^{th} and i_2^{th} cusp as an edge in the Epstein-Penner decomposition.

Similarly, on every 1-cell ε of S, the points are equi-distance from a collection of truncated cusps S_{i_p}. We take the geodesic polygon between these cusps as a face in the Epstein-Penner decomposition.

This decomposes M into hyperbolic ideal polyhedra, i.e. polyhedra of $\mathbb{H}^3$ whose vertices lie on S^2_∞. We will make use of this decomposition in Section 5.

3.3. **Developing map.** Let M be a geometrically finite hyperbolic 3-manifold. Let $\rho : \pi_1(M) \to \mathrm{Isom}^+(\mathbb{H}^3)$ be the representation induced by the hyperbolic metric on M.

Choose a generic Dirichlet tessellation of M. Let D be one of the Dirichlet domains. Let $\mathcal{G}$ be the set $\{g \in \pi_1(M) \mid D \cap \rho(g)D \neq \varnothing\}$. We impose the Dirichlet tessellation $\bigcup_g gD$ on $\widetilde{M}$ and refer to the identification map $H : \widetilde{M} \to \mathbb{H}^3$ as the **developing map**.

H is ρ-equivariant by definition. For $\rho' : \pi_1(M) \to \mathrm{Isom}^+(\mathbb{H}^3)$ close to ρ, we will show that one can deform H to into a ρ'-equivariant map H.

Given ρ', we consider the subset

$$D' = \{x \in \mathbb{H}^3 \mid d(x, x_0) \leq d(x, \rho'(g)x_0) \text{ for all } g \in \mathcal{G}\}.$$

For ρ' close enough to ρ, the faces of D' lie close to those of D, hence D' lies close to D and D' is combinatorially isomorphic to D.

Because of this, we can pick a cellular homeomorphism $H' : D \to D'$ that is close to the identity map, and such that for every pair of faces $(F_{h^{-1}}, F_h)$ on D, we have $H'_{F_{h^{-1}}} = \rho'(h)H'|_{F_h}$. We can then extend H' to $\widetilde{M}$ by setting $H(gx) = \rho'(g)H(x)$ for every $g \in \pi_1(M)$.

Note that H' is a local homeomorphism, but is in general neither injective nor surjective. Correspondingly, $\rho'(\pi_1(M))$ is in general not discrete. Nevertheless, H' induces a hyperbolic metric on M by pulling back the metric on $\mathbb{H}^3$. This metric is in general not complete.

For future purposes, we also point out that the vertices of D depend algebraically on ρ'. Thus each vertex of the Dirichlet tessellation depend algebraically on ρ'.

Example 3.3. Consider the representation $\rho : \mathbb{Z}^2 \to \mathrm{Isom}^+(\mathbb{H}^3)$ defined by

$$\rho\left(\begin{bmatrix}1\\0\end{bmatrix}\right)(w) = w+1$$

$$\rho\left(\begin{bmatrix}0\\1\end{bmatrix}\right)(w) = w+i$$

where we identify an isometry of $\mathbb{H}^3$ with the corresponding Möbius transformation on S^2_∞.

For small $\varepsilon > 0$, we can deform ρ into $\rho' : \mathbb{Z}^2 \to \mathrm{Isom}^+(\mathbb{H}^3)$ defined by

$$\rho'\left(\begin{bmatrix}1\\0\end{bmatrix}\right)(w) = (1+\varepsilon)w+1$$

$$\rho'\left(\begin{bmatrix}0\\1\end{bmatrix}\right)(w) = (1+i\varepsilon)w+i$$

The deformed developing map spirals around the axis $\{x+iy = -\frac{1}{\varepsilon}\}$.

Let S be a truncated rank two cusp of M. Consider the subgroup $\Gamma = \pi_1(S) \subset \pi_1(M)$. Let g_1, g_2 be a set of generators for Γ. Under the initial representation ρ, g_1 and g_2 are sent to parabolic isometries that fix a point on S^2_∞.

In the deformed representation ρ', the images of g_1 and g_2 still have to commute. Hence one of the following two cases is true:

- g_1 and g_2 are parabolic, and fix a common point on S^2_∞.
- g_1 and g_2 are elliptic and/or hyperbolic, and fix a common axis ℓ.

The **generalized Dehn filling invariant** of ρ' is a point on $S^2 = \mathbb{R}^2 \cup \{\infty\}$ defined as follows:

If g_1 and g_2 are parabolic, then the generalized Dehn filling invariant of ρ' is ∞.

If g_1 and g_2 are elliptic and/or hyperbolic, we identify $\mathrm{Isom}^+(\mathbb{H}^3) = \mathrm{PSL}(2,\mathbb{C})$ and consider the values of $\mathrm{tr}(\rho'(g_i))$. Since ρ' is close to ρ, $\mathrm{tr}(\rho'(g_i))$ is close to ± 2. Hence there exists r_i close to 1 and θ_i close to 0 such that $\mathrm{tr}(\rho'(g_i)) = \pm(r_i e^{i\theta_i} + r_i^{-1}e^{-i\theta_i})$. There exists unique $p_1, p_2 \in \mathbb{R}$ such that

$$\begin{cases} p_1 \log r_1 + p_2 \log r_2 = 0 \\ p_1\theta_1 + p_2\theta_2 = 2\pi \end{cases}$$

The generalized Dehn filling invariant of ρ' is $(p_1, p_2) \in \mathbb{R}^2$.

The motivation for this terminology comes from the fact that if (p_1, p_2) are a pair of relatively prime integers, then the representation $\rho' : \pi_1(M) \to \mathrm{Isom}^+(\mathbb{H}^3)$ descends to a representation $\rho' : \pi_1(M_{(p_1,p_2)}) \to \mathrm{Isom}^+(\mathbb{H}^3)$ where $M_{(p_1.p_2)}$ is the 3-manifold obtained by Dehn filling M along the slope $p_1g_1 + p_2g_2$, i.e. removing the truncated cusp S and gluing in a solid torus with the meridian being glued onto a curve of homotopy class $p_1g_1 + p_2g_2$.

Let $\widehat{M}$ be the cover of M corresponding to the kernel of $\pi_1(M) \to \pi_1(M_{(p_1,p_2)})$. Then the developing map H' descends to a map $\widehat{M} \to \mathbb{H}^3$. The universal cover $\widetilde{M_{(p_1,p_2)}}$ can be obtained from $\widehat{M}$ by filling each cusp along the slope $p_1g_1 + p_2g_2$. Correspondingly, we can complete the map $\widehat{M} \to \mathbb{H}^3$ into a map $\widetilde{M_{(p_1,p_2)}} \to \mathbb{H}^3$. Thus we have a hyperbolic metric on $M_{(p_1,p_2)}$.

4. Closed manifolds

The goal of this section is to show that $\mathrm{SL}(2,\mathbb{C})$ representations arising from hyperbolic metrics on *closed* 3-manifolds are rigid.

Theorem 4.1 (Mostow rigidity)**.** *Let M_1 and M_2 be two closed hyperbolic 3-manifolds. Every isomorphism $\pi_1(M_1) \to \pi_1(M_2)$ is induced by a unique isometry $M_1 \to M_2$.*

We will prove Theorem 4.1 via the approach in [Mar22, Sections 5 and 13]. See also [Thu97, Chapters 5 and 6]. Note that this is not Mostow's original approach via quasiconformal extensions. For an exposition of that approach, see [Thu97, Chapter 5.9].

4.1. **Boundary extension of pseudo-isometries.** Let X and Y be two metric spaces. A map $F : X \to Y$ is a **pseudo-isometry** if there exists $C > 0$ such that $\frac{1}{C}d(x_1, x_2) - C \leq d(F(x_1), F(x_2)) \leq Cd(x_1, x_2)$ for all $x_1, x_2 \in X$.

The first ingredient to proving Mostow's rigidity is the following theorem.

Theorem 4.2. *Every quasi-isometry $F : \mathbb{H}^3 \to \mathbb{H}^3$ extends to a continuous map $\overline{F} : \overline{\mathbb{H}^3} \to \overline{\mathbb{H}^3}$.*

Note that if $\overline{F}$ exists, then it is unique. In fact, for every point $w \in S^2_\infty$, if ℓ is a geodesic line ending at w, then $\overline{F}(w)$ must be the endpoint of $F(\ell)$. Conversely, the strategy of the proof is to define $\overline{F}$ in this way then check that it is well-defined and continuous.

To show that this definition is well-defined, we prove the following sequence of lemmas.

Lemma 4.3. *Let ℓ be a geodesic segment in $\mathbb{H}^3$. Let $\pi : \mathbb{H}^3 \to \ell$ be the map defined by sending each point to its closest point on ℓ. Then for every path γ that lies at a distance of $\geq R$ away from ℓ, we have* $\mathrm{length}(F(\gamma)) \leq \frac{1}{\cosh R}\mathrm{length}(\gamma)$.

Proof. It suffices to show this for segments of γ that get mapped to the interior of ℓ, since all other segments get mapped to the endpoints of ℓ and do not contribute to the length of $F(\gamma)$.

The restriction of π to a point x mapping to the interior of ℓ is smooth, so in turn it suffices to show that $||d\pi(v)|| \leq \frac{1}{\cosh R}||v||$ for every $v \in T_x\mathbb{H}^3$. Finally, it suffices to show this inequality when v is orthogonal to $\ker d\pi$.

In this case v is tangent to the cone $\{(x, y, z) \in \mathbb{H}^3 \mid x^2 + y^2 = r^2z^2\}$ where $d(x, \ell) = \cosh^{-1}(\sec\theta) \leq R$ and $r = \tan\theta$ as in Proposition 2.5. We compute $||d(\pi(v))|| = (\cos\theta)||v|| \leq \frac{1}{\cosh R}||v||$. □

Lemma 4.4. *Let $F : \mathbb{H}^3 \to \mathbb{H}^3$ be a pseudo-isometry. There exists R such that for every $x, y \in \mathbb{H}^3$, the image of the geodesic segment $[x, y]$ between x and y lies within a R-neighborhood of the geodesic segment $[F(x), F(y)]$ between $F(x)$ and $F(y)$.*

Proof. Let $[p, q]$ be a maximal segment of $[x, y]$ that gets mapped to outside the $\cosh^{-1}(2C^2)$-neighborhood of $[F(x), F(y)]$. Consider the concatenated path $[F(p), \pi(F(p))] * \pi(F([p, q])) * [\pi(F(q)), F(q)]$. By Lemma 4.3, its length is bounded from above by $\cosh^{-1}(2C^2) + \frac{1}{2C^2}\mathrm{length}(F([p, q])) + \cosh^{-1}(2C^2) \leq \frac{1}{2C}d(p, q) + 2\cosh^{-1}(2C^2)$. On the other hand, this is a path from $F(p)$ to $F(q)$, so its length is bounded from below by $d(F(p), F(q)) \geq \frac{1}{C}d(p, q) - C$.

Thus

$$\frac{1}{C}d(p, q) - C \leq \frac{1}{2C}d(p, q) + 2\cosh^{-1}(2C^2)$$
$$d(p, q) \leq 2C(2\cosh^{-1}(2C^2) + C)$$

This implies that $\text{length}(F([p,q])) \le Cd(p,q) \le 2C^2(2\cosh^{-1}(2C^2)+C)$, thus $F([x,y])$ lies within the $\cosh^{-1}(2C^2) + C^2(2\cosh^{-1}(2C^2)+C)$ neighborhood of $[F(x), F(y)]$. $\square$

We can now check that our definition of $\overline{F}$ is well-defined.

Lemma 4.5. *Let $F : \mathbb{H}^3 \to \mathbb{H}^3$ be a pseudo-isometry. There exists R such that for every geodesic ray ℓ, there is a geodesic ray ℓ' such that $F(\ell)$ lies within the R neighborhood of ℓ'.*

Proof. Suppose ℓ starts at x. Take a sequence of points y_i on ℓ that converge to infinity. For each i, let v_i be the unit tangent vector of $[F(x), F(y_i)]$. Up to passing to a subsequence, v_i converges to some unit vector v. Let ℓ' be the geodesic ray starting at $F(x)$ with tangent vector v.

Take R as in Lemma 4.4. Then for each i, $F([x,y_i]) \subset N_R([F(x), F(y_i)])$. For each $L > 0$, since v_i converges to v and $d(F(x), F(y_i)) \ge \frac{1}{C} d(x, y_i) - C \to \infty$, the sets $N_R([F(x), F(y_i)])$ converge to $N_R(\ell')$ within $B_L(x)$. More precisely, this means that $\bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} N_R([F(x), F(y_i)]) \cap B_L(x) = N_R(\ell') \cap B_L(x)$. This implies that $F([x,y])$ is contained within $N_R(\ell')$ within $B_L(x)$. Taking $L \to \infty$, this proves the lemma. $\square$

It remains to show that the extended map is continuous.

Lemma 4.6. *Let $F : \mathbb{H}^3 \to \mathbb{H}^3$ be a pseudo-isometry. There exists R such that for any geodesic ray ℓ and geodesic plane H orthogonal to ℓ, if ℓ' is a geodesic ray such that $F(\ell) \subset N_R(\ell')$ as in Lemma 4.5, then the image of the projection of $F(H)$ to ℓ' has diameter $\le R$.*

Proof. Let s be a geodesic ray on H starting at $x = \ell \cap H$. Let s_n be the segment of s_n of distance between $n-1$ and n from x. Then $\text{length}(\pi(F(s))) = \sum_n \text{length}(\pi(F(s_n))) \le \sum_n \frac{1}{\cosh n} \text{length}(F(s_n)) \le \sum_n \frac{C}{\cosh n}$. Since H is the union of all such geodesic rays s, we have $\text{diam}(\pi(F(H))) \le \sum_n \frac{C}{\cosh n} < \infty$. $\square$

Lemma 4.7. *The extended map $\overline{F} : \overline{\mathbb{H}^3} \to \overline{\mathbb{H}^3}$ is continuous.*

Proof. It suffices to show that $\overline{F}$ is continuous on S^2_∞. This follows from Lemma 4.6 since it implies that $\overline{F}$ sends a basis of open neighborhoods of $w \in S^2_\infty$ into a basis of open neighborhoods of $\overline{F}(w)$. $\square$

In fact, one can show that $\overline{F}$ is **quasiconformal** on S^2_∞. This means that under the round metric on S^2_∞ (see Remark 1.12), the function

$$K(w) = \lim_{r \to 0} \frac{\sup_{p,q \in B_r(w)} d(\overline{F}(p), \overline{F}(q))}{\inf_{p,q \in B_r(w)} d(\overline{F}(p), \overline{F}(q))}$$

is uniformly bounded.

4.2. **Volume of hyperbolic tetrahedra.** Let Δ be a hyperbolic ideal tetrahedron. Suppose e_1 and e_2 are a pair of opposite edges. Let ℓ be the shortest geodesic segment between e_1 and e_2. Since ℓ meets e_1 and e_2 perpendicularly, the isometry that rotates π around ℓ sends e_1 and e_2 to themselves, thus sends Δ to itself. This shows that the dihedral angles of each opposite pair of edges of Δ agree.

We let the three pairs of dihedral angles be α, β, γ. Then $\alpha + \beta + \gamma = \pi$. Indeed, up to an isometry, we can place one vertex of Δ at ∞, then the three faces of Δ that meet ∞ must be vertical planes, thus bound a region of the form (Euclidean triangle) $\times (0, \infty)$. Since α, β, γ are the angles of this Euclidean triangle, they add up to π.

Conversely, given $\alpha, \beta, \gamma > 0$ that add up to π, we can construct a hyperbolic ideal tetrahedron with those dihedral angles by first constructing an Euclidean triangle with those angles on the $\{z = 0\}$ plane, then taking ∞ as the fourth vertex. From this argument one can also see that α, β, γ uniquely determine Δ up to isometry.

Proposition 4.8. *The volume of the hyperbolic ideal tetrahedron with dihedral angles is* $\Lambda(\alpha)+\Lambda(\beta)+\Lambda(\gamma)$ *where* $\Lambda(\theta) = -\int_0^\theta \log|2\sin\theta| d\theta$ *is the Lobachevsky function.*

Proof. A routine calculus exercise. See [Mar22, Section 13.1] or [Thu97, Chapter 7.2]. ☐

A crucial observation is that the Lobachevsky function is strictly concave on $[0,\pi]$, since $\Lambda''(\theta) = -\cot\theta < 0$ for $\theta \in (0,\pi)$. This implies that the volume functional is strictly concave on the space $\{(\alpha,\beta,\gamma) \in [0,\pi]^3 \mid \alpha+\beta+\gamma = \pi\}$. Thus the volume functional attains a unique maximum of $3\Lambda(\frac{\pi}{3}) \approx 1.015$ at the regular ideal hyperbolic tetrahedron.

Corollary 4.9. *The volume of any hyperbolic tetrahedron, ideal or not, is* $\leq v_3$.

Proof. Any hyperbolic tetrahedron is contained in a hyperbolic ideal tetrahedron, which we just reasoned has volume $\leq v_3$. ☐

4.3. **Simplicial volume.** Let X be a topological space. Recall that a k-dimensional simplicial chain in X is a formal linear combination $a = \lambda_1 a_1 + \cdots + \lambda_n a_n$ where each a_i is a map from the k-dimensional simplex Δ^k to X. We define the norm of a to be $||a|| = \sum |\lambda_i|$.

Given a class $\alpha \in H_k(X;\mathbb{R})$, we define the **norm** of α to be the infimum of the norm of its representatives, i.e.

$$||\alpha|| = \inf\{||a|| \mid \alpha = [a]\}.$$

Definition 4.10 (Gromov)**.** Let M be an oriented closed 3-manifold. The **simplicial volume** of M, which we denote by $||M||$, is the norm of its fundamental class $[M] \in H_3(M;\mathbb{R})$.

Definition 4.10 works for oriented closed manifolds of all dimensions. Also, note that since $||-[M]|| = ||[M]||$, the simplicial volume doesn't actually depend on the choice of orientation.

To give some motivation for the terminology 'simplicial *volume*', we state the following properties. We will only use property (1) in the sequel.

Proposition 4.11.

(1) If $f : M_1 \to M_2$ *is a map of degree* d*, then* $||M_1|| \geq d||M_2||$.

(2) If $f : M_1 \to M_2$ *is a covering map of degree* d*, then* $||M_1|| = d||M_2||$.

(3) $||M_1 \# M_2|| = ||M_1|| + ||M_2||$.

Proof. For (1), let $a = \lambda_1 a_1 + \cdots + \lambda_n a_n$ be a cycle representing $[M_1]$. Then $f_*(a) = \lambda_1(f\circ a_1)+\cdots+\lambda_n(f\circ a_n)$ represents $d[M_2]$. Thus $d||M_2|| \leq ||f_*(a)|| = \sum_i |\lambda_i| = ||a||$.

For (2), let $a = \lambda_1 a_1 + \cdots + \lambda_n a_n$ be a cycle representing $[M_2]$. Then $f^*(a) = \lambda_1 \sum_{\widehat{a_1} \in f^{-1}(a_1)} \widehat{a_1} + \cdots + \lambda_n \sum_{\widehat{a_n} \in f^{-1}(a_n)} \widehat{a_n}$ represents $[M_1]$, where $f^{-1}(a_i)$ denotes the set of d lifts of a_i. Thus $||M_1|| \leq ||f^*(a)|| = \sum_i d|\lambda_i| = d||a||$. Together with (1), this implies (2).

(3) is nontrivial, in particular it is not true in dimension 2. See [Gro82, Section 3.5] for a proof. ☐

Proposition 4.11(1) implies that the simplicial volume of any closed orientable 3-manifold of the form $S \times S^1$, where S is a surface, has zero simplicial volume. Indeed, for each $d \geq 2$, S^1 has a degree d map to itself, which induces a degree d map of $S \times S^1$ to itself. More generally, one can show that any Seifert fibered manifold has zero simplicial volume. See [Mar22, Proposition 13.2.10].

On the other hand, closed hyperbolic 3-manifolds always have positive simplicial volume. In fact, the simplicial volume of a closed hyperbolic 3-manifold is essentially its hyperbolic volume.

Theorem 4.12. *Let M be a closed hyperbolic 3-manifold, then* $\mathrm{vol}(M) = v_3||M||$.

We will prove Theorem 4.12 by showing the two directions of inequalities in Proposition 4.13 and Proposition 4.14.

To show that $\mathrm{vol}(M) \leq v_3||M||$, we use the **cycle straightening operation**: Let $a = \lambda_1 a_1 + \cdots + \lambda_n a_n$ be a k-cycle. Let $\Delta = \bigsqcup \Delta_i^k$ be the disjoint union of the domain simplices of a_i, and abusing notation, let a be the disjoint union of the maps a_i. Inductively, for $i = 1, ..., n$, we homotope a on each collection of i-cells on which a restricts to the same map, so that the image of these i-cells is the geodesic i-simplex spanned by the image of its $i+1$ vertices. The straightened cycle represents the same homology class and has the same norm as a.

Proposition 4.13. *Let M be a closed hyperbolic 3-manifold, then* $\mathrm{vol}(M) \leq v_3||M||$.

Proof. Suppose $a = \lambda_1 a_1 + \cdots + \lambda_n a_n$ is a cycle that represents $[M]$. We straighten a. If ω is the volume form of M, then $\mathrm{vol}(M) = \int_M \omega = \sum \lambda_i \int_{\Delta_i} a_i^* \omega \leq \sum |\lambda_i| \mathrm{vol}(a_i(\Delta_i)) \leq \sum \lambda_i v_3 = v_3||a||$. Here Corollary 4.9 is used for the last inequality. Note that in the second to last inequality, $\int_{\Delta_i} a_i^*\omega = \pm \mathrm{vol}(a_i(\Delta_i))$ depending on whether the straightened map a_i is orientation-preserving. □

The proof that $\mathrm{vol}(M) \leq v_3||M||$ is more involved.

Recall the notion of a Dirichlet tessellation. When M is closed, we claim that each Dirichlet domain has finite diameter. Indeed, otherwise there are points inside the domain D that are arbitrarily far away from x_0, and these project down to points that are arbitrarily far away from a fixed point in M.

Proposition 4.14. *Let M be a closed hyperbolic 3-manifold, then* $\mathrm{vol}(M) \geq v_3||M||$.

Proof. We define a **t-simplex** to be a regular hyperbolic tetrahedron with labelled vertices, where all the vertices are at distance t with a center point. Let $S(t)$ be the set of t-simplices in $\mathbb{H}^3$. $\mathrm{Isom}^+(\mathbb{H}^3)$ acts freely on $S(t)$ with two orbits: one orbit $S^+(t)$ contains all tetrahedron with vertices labelled positively, and one orbit $S^-(t)$ contains all tetrahedron with vertices labelled negatively. Thus $S^\pm(t) \cong \mathrm{Isom}^+(\mathbb{H}^3)$, and we can pick a measure μ on $S^\pm(t)$ that is invariant under the action of $\mathrm{Isom}^+(\mathbb{H}^3)$.

Since $S^\pm(t) \cong \mathrm{Isom}^+(\mathbb{H}^3)$, the quotient $S^\pm(t)/\pi_1(M) \cong \mathrm{Isom}^+(\mathbb{H}^3)/\pi_1(M) \cong \mathrm{Fr}(M)$ is compact. Thus up to rescaling the measure μ, we can assume $\mu(S^\pm(t)/\pi_1(M)) = 1$.

Let Σ be the quotient of $(\pi_1(M))^4$ under the diagonal action of $\pi_1(M)$. For each $\sigma = [g_0, g_1, g_2, g_3] \in \Sigma$, we define a map $a_\sigma : \Delta \to M$ by sending the i^{th} vertex of the 3-simplex Δ to the image of $g_i x_0$, sending the edges and faces to geodesic segments and triangles bounded by these points, then projecting the whole tetrahedron to M.

Consider some Dirichlet tessellation $\mathbb{H}^3 = \bigcup_{g \in \pi_1(M)} gD$ for $\pi_1(M)$. Let d be the diameter of each Dirichlet domain. We define $S^\pm_\sigma(t) \subset S^\pm(t)$ to be the set of all t-simplices whose i^{th} vertex lies in $g_i D$. Finally, we define $\lambda_\sigma(t) = \mu(S^+_\sigma(t)) - \mu(S^-_\sigma(t))$.

We claim that $\lambda_\sigma(t) = 0$ for all but finitely many σ. This follows from the observation that if $\lambda_\sigma(t) \neq 0$, then the Dirichlet domains $g_i D$ have to be within $2t$ distance of each other, but the Dirichlet domains are locally finite. We also observe that

$$\begin{aligned} \sum_{\sigma \in \Sigma} |\lambda_\sigma(t)| &\leq \sum_{\sigma \in \Sigma} (\mu(S^+_\sigma(t)) + \mu(S^-_\sigma(t))) \\ &\leq \mu(S^+(t)/\pi_1(M)) + \mu(S^-(t)/\pi_1(M)) = 2. \end{aligned}$$

Consider the chain $a(t) = \sum_\sigma \lambda_\sigma(t) a_\sigma$. The two claims above show that $a(t)$ is well-defined and $||a(t)|| \leq 2$.

Furthermore, we claim that $a(t)$ is a cycle. Indeed, every term in $\partial a(t)$ is a geodesic triangle with vertices at g_0x_0, g_1x_0, g_2x_0, for some (g_0, g_1, g_2). The coefficient of such a term is $\sum_g \lambda_{[g_0,g_1,g_2,g]}(t) - \sum_g \lambda_{[g_0,g_1,g,g_2]}(t) + \sum_g \lambda_{[g_0,g,g_1,g_2]}(t) - \sum_g \lambda_{[g,g_0,g_1,g_2]}(t)$.

For the first summand, $\sum_g \lambda_{[g_0,g_1,g_2,g]}(t) = \sum_g \mu(S^+_{[g_0,g_1,g_2,g]}(t)) - \sum_g \mu(S^-_{[g_0,g_1,g_2,g]}(t)) = 0$, since the sets $\bigcup_g S^+_{[g_0,g_1,g_2,g]}(t)$ and $\bigcup_g S^-_{[g_0,g_1,g_2,g]}(t)$ are canonically isomorphic via reflecting in the geodesic plane spanned by g_0x_0, g_1x_0, g_2x_0. Similarly, the rest of the summands disappear.

Now fix a large t_0 such that for all $t > t_0$, we cannot have a positively oriented t-simplex and a negatively oriented t-simplex with vertices within d of each other. Then for each σ where $\lambda_\sigma(t) > 0$, a_σ is an orientation preserving map onto a hyperbolic tetrahedron whose vertices lie within a d-neighborhood of a t-simplex. Hence as $t \to \infty$, $\int_\Delta a_\sigma^*\omega = \mathrm{vol}(a_\sigma(\Delta)) \nearrow v_3$ uniformly. Similarly, for each σ where $\lambda_\sigma(t) < 0$, $\int_\Delta a_\sigma^*\omega = -\mathrm{vol}(a_\sigma(\Delta)) \searrow -v_3$ uniformly. Thus $\left\langle \frac{a(t)}{||a(t)||}, \omega \right\rangle = \frac{\sum_\sigma \lambda_\sigma(t) \int_\Delta a_\sigma^*\omega}{\sum_\sigma |\lambda_\sigma(t)|} \nearrow v_3$.

In particular, for large enough t, $\langle a(t), \omega\rangle > 0$, thus $a(t)$ represents a positive multiple of $[M]$, say $a(t) = k(t)[M]$. Here we note that $\langle a(t), \omega\rangle \leq v_3||a(t)|| \leq 2v_3$, thus $k(t) \leq \frac{2v_3}{\mathrm{vol}(M_1)}$.

Finally, we compute

$$\frac{||M||}{\langle [M], \omega\rangle} \leq \frac{||a(t)||}{\langle a(t), \omega\rangle} \to \frac{1}{v_3}.$$

□

4.4. **Proof of Mostow rigidity.** Since hyperbolic 3-manifolds are $K(\pi, 1)$, there is a homotopy equivalence $F : M_1 \to M_2$ that induces the given isomorphism $\pi_1(M_1) \to \pi_1(M_2)$. Up to a perturbation, we can assume that F is smooth. Lift F to $\widetilde{F} : \mathbb{H}^3 \to \mathbb{H}^3$.

We claim that $\widetilde{F}$ is a pseudo-isometry. Since M_1 and M_2 are compact, there exists C such that $||dF|| \leq C$, thus $||d\widetilde{F}|| \leq C$ as well. Then $d(\widetilde{F}(x), \widetilde{F}(y)) \leq Cd(x, y)$.

Let $G : M_2 \to M_1$ be a homotopy inverse to F. There exists a homotopy H_t from GF to identity. Let $\widetilde{G}$ be a lift of G such that the homotopy H_t lifts to a homotopy $\widetilde{H_t}$ from $\widetilde{G}\widetilde{F}$ to identity.

Up to increasing the value of C, the above argument shows that $d(\widetilde{G}(x), \widetilde{G}(y)) \leq Cd(x, y)$. Then we have $d(x, y) \leq d(x, \widetilde{G}\widetilde{F}(x)) + d(\widetilde{G}\widetilde{F}(x), \widetilde{G}\widetilde{F}(y)) + d(\widetilde{G}\widetilde{F}(y), y) \leq 2\max_{x,t} d(H_t(x), x) + Cd(\widetilde{F}(x), \widetilde{F}(y))$. Thus up to increasing the value of C again, we have $\frac{1}{C}d(x, y) - C \leq d(\widetilde{F}(x), \widetilde{F}(y))$.

By Theorem 4.2, we can extend $\widetilde{F}$ to a map $\overline{F}$ on $\overline{\mathbb{H}^3}$. We claim that $\overline{F}$ is a Möbius transformation on S^2_∞.

To show this, it suffices to show that whenever (w_0, w_1, w_2, w_3) are the vertices of a positively oriented regular ideal hyperbolic tetrahedron, then $(\overline{F}(w_0), \overline{F}(w_1), \overline{F}(w_2), \overline{F}(w_3))$ are the vertices of a positively oriented regular ideal hyperbolic tetrahedron as well. Indeed, once we show this, up to composing by an isometry, we can assume that $\overline{F}$ fixes the vertices of a regular ideal hyperbolic tetrahedron Δ. But then $\overline{F}$ has to fix the vertices of the regular ideal tetrahedra that share a face with Δ, and then it has to fix the vertices of regular ideal tetrahedra that share a face with such tetrahedra, and so on. The set of such vertices is dense in S^2_∞, thus $\overline{F}$ must be identity on S^2_∞.

For the sake of contradiction, we suppose (w_0, w_1, w_2, w_3) are the vertices of a positively oriented regular ideal hyperbolic tetrahedron, but the ideal tetrahedron with vertices at $(\overline{F}(w_0), \overline{F}(w_1), \overline{F}(w_2), \overline{F}(w_3))$ has volume $< v_3 - 2\delta$. We pick neighborhoods U_i of w_i in $\overline{\mathbb{H}^3}$ such that any tetrahedron with vertices within $(\overline{F}(U_0), \overline{F}(U_1), \overline{F}(U_2), \overline{F}(U_3))$ has volume $< v_3 - \delta$.

Recall the cycles $a(t) = \sum_{\sigma\in\Sigma} \lambda_\sigma(t)a_\sigma$ defined in the proof of Proposition 4.14. We say that $\sigma \in \Sigma$ is **bad** if the i^{th} vertex of a_σ is contained in U_i for each i. For large t, $\sum_{\text{bad } \sigma} |\lambda_\sigma(t)|$ is uniformly bounded from below by some $\eta > 0$, since we can always find Dirchlet domains g_iD inside U_i.

Also recall that $a(t)$ represents $k(t)[M_1]$ for large enough t, where $k(t) \leq \frac{2v_3}{\mathrm{vol}(M_1)}$. Now consider the cycle $F_* a(t) = \sum_{\sigma\in\Sigma} \lambda_\sigma(t)(F\circ a_\sigma)$. Since F is a homotopy equivalence, $F_*a(t)$ represents $k(t)[M_2]$. We straighten this cycle as in the proof of Proposition 4.13 and compute

$$\begin{aligned}
k(t)\mathrm{vol}(M_2) &= \sum_{\sigma\in\Sigma} \lambda_\sigma(t) \int_\Delta (F\circ a_\sigma)^*\omega \\
&= \sum_{\text{good } \sigma} \lambda_\sigma(t) \int_\Delta (F\circ a_\sigma)^*\omega + \sum_{\text{bad } \sigma} \lambda_\sigma(t) \int_\Delta (F\circ a_\sigma)^*\omega \\
&\leq v_3 \sum_{\text{good } \sigma} |\lambda_\sigma(t)| + (v_3 - \delta) \sum_{\text{bad } \sigma} |\lambda_\sigma(t)| \\
&< v_3 \sum_{\sigma\in\Sigma} |\lambda_\sigma(t)| - \delta\eta \\
&= k(t)\mathrm{vol}(M_1) - \delta\eta.
\end{aligned}$$

But then we have

$$\mathrm{vol}(M_1) - \mathrm{vol}(M_2) \geq \frac{1}{k(t)}\delta\eta \geq \frac{\mathrm{vol}(M_1)}{2v_3}\delta\eta > 0$$

which contradicts Proposition 4.11.

4.5. **Consequences of Mostow rigidity.** An immediate consequence of Mostow's rigidity is that the hyperbolic metric on a closed hyperbolic 3-manifold is unique up to homeomorphism.

Corollary 4.15. *Let M be a closed 3-manifold. Let g_1 and g_2 be two complete hyperbolic metrics on M. Then there exists a homeomorphism $h : M \to M$ such that $g_1 = h^* g_2$.*

Proof. Let $M_i = (M, g_i)$ and apply Theorem 4.1 to the identity map $\pi_1(M) \to \pi_1(M)$. □

In fact, in Corollary 4.15, we know that h induces the identity map on $\pi_1(M)$. Since every hyperbolic 3-manifold is a $K(\pi, 1)$, this implies that h is homotopic to identity. By a result of Gabai-Meyerhoff-Thurston [GMT03] we can conclude that h is isotopic to identity. In other words, the hyperbolic metric on a closed hyperbolic 3-manifold is unique up to *isotopy*.

Next, as promised at the start of the section, we deduce a result on the rigidity of $\mathrm{SL}(2,\mathbb{C})$ representations.

Corollary 4.16. *Let M be a closed hyperbolic 3-manifold. The component of the character variety that contains the character of a representation induced by the hyperbolic metric is zero dimensional.*

Proof. Let ρ be the $\mathrm{Isom}^+(\mathbb{H}^3) = \mathrm{PSL}(2,\mathbb{C})$ representation induced by the hyperbolic metric. Let $\widetilde{\rho}$ be a $\mathrm{SL}(2,\mathbb{C})$ representation obtained by lifting ρ. Suppose $\widetilde{\rho'}$ is a representation close to $\widetilde{\rho}$. Then there is an induced $\mathrm{PSL}(2,\mathbb{C})$ representation ρ' that lies close to ρ.

From the discussion in Section 3.3, we know that ρ' is induced from a hyperbolic metric g' on M. Here g' is automatically complete since M is closed. Hence by Theorem 4.1 ρ' and ρ are conjugate. □

For a closed hyperbolic 3-manifold M, the **trace field** of $\pi_1(M)$, denoted by $\mathbb{Q}(\mathrm{tr}(\pi_1(M)))$, is the field extension of $\mathbb{Q}$ generated by $\mathrm{tr}(g)$ for all $g \in \pi_1(M) \subset \mathrm{PSL}_2(\mathbb{C})$.

Corollary 4.17. *The trace field of a closed hyperbolic 3-manifold is a finite field extension of $\mathbb{Q}$.*

Proof. The character variety of $\pi_1(M)$ is a variety defined over $\mathbb{Z}$, hence the element of any zero dimensional component has algebraic number entries. See [MR03, Lemma 3.1.5]. □

Corollary 4.17 is the starting point of the study of the arithmetic of hyperbolic 3-manifolds. We refer to [MR03] (and Connor's course!) for more about this.

5. Gluing equations

For the next few sections, we turn our attention to studying finite volume hyperbolic 3-manifolds with a nonempty set of cusps.

5.1. **Geometric triangulations.** A **(topological) ideal triangulation** of a 3-manifold M is a decomposition of M into ideal tetrahedra glued along pairs of faces by homeomorphisms.

A **geometric triangulation** of a hyperbolic 3-manifold M is a decomposition of M into hyperbolic ideal tetrahedra glued along pairs of faces by isometries.

A geometric triangulation carries the data of its underlying ideal triangulation and the hyperbolic structure of each ideal tetrahedron. We explain how the latter data can be encoded by a complex number for each tetrahedron.

Given an ideal hyperbolic tetrahedron Δ, we fix a positively oriented labelling (v_0, v_1, v_2, v_3) of its vertices. Up to an isometry we can position v_0 at 0, v_1 at 1, and v_2 at ∞. Then the fact that the labelling is positively oriented means that $\Im v_3 > 0$. We define the **modulus** of the edge $[v_0, v_2]$ to be v_3.

In short, the modulus of an edge is the ratio between the complex lengths of the sides adjacent to its dihedral angle. In particular, the dihedral angle of an edge is the argument of its modulus.

The modulus of one edge determines the moduli of all other edges. Specifically, in the notation above,

- the moduli of the edges $[v_0, v_2]$ and $[v_1, v_3]$ are v_3,
- the moduli of the edges $[v_1, v_2]$ and $[v_0, v_3]$ are $\frac{1}{1-v_3}$,
- the moduli of the edges $[v_2, v_3]$ and $[v_0, v_1]$ are $\frac{v_3-1}{v_3}$.

The moduli of the edges in turn determines the dihedral angles, thus determines the isometry type of the tetrahedron.

Example 5.1. We construct a hyperbolic 3-manifold M by gluing the faces of two regular ideal hyperbolic tetrahedra as follows.

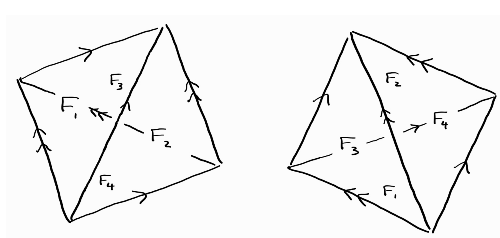

FIGURE 3.

By definition, M has a geometric triangulation. The modulus of every edge in the triangulation is $\frac{1+i\sqrt{3}}{2}$.

One can check that M is the figure eight knot complement. Indeed, $F_1 \cup F_2$ is a genus one fiber surface with monodromy $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. One can construct a homeomorphism between M and the figure eight knot complement by sending this surface to the Seifert surface of the figure eight knot.

Geometric triangulations are convenient for both conceptual and computational reasons. For example, the proof of the hyperbolic Dehn surgery theorem simplifies when the unfilled manifold admits a geometric triangulation.

However, it is currently unknown whether every hyperbolic 3-manifold (with a nonempty set of cusps) admits a geometric triangulation. In these notes, we will follow the approach of [Mar22, Section 15.3.2] and work around this by dealing with triangulations that are close to being geometric.

5.2. **Consistency and completeness equations.** In this subsection, we show that an ideal triangulation is the underlying triangulation of a geometric triangulation if and only if it solves two sets of equations.

Let τ be a geometric triangulation of a finite volume hyperbolic 3-manifold M. For each edge e and tetrahedron Δ for which e is incident to Δ, we denote by $z(e, \Delta)$ the modulus of e as an edge of Δ.

By the properties of the modulus, we always have

$$\begin{aligned} z(e,\Delta) &= z(e',\Delta) \text{ if } e \text{ and } e' \text{ are opposite edges in } \Delta \\ z(e,\Delta) &= \frac{1}{1-z(e',\Delta)} \text{ if } (e,e') \text{ determine a positive basis of a face at a vertex of } \Delta \end{aligned} \tag{5.1}$$

Let e be an edge of τ. Let $\widetilde{e}$ be a lift of e in the universal cover and let v be one of the vertices of $\widetilde{e}$. Up to an isometry, we can position v at ∞.

As we go around $\widetilde{e}$, the dihedral angles have to add up to 2π. In fact, if we multiply up the ratios between the complex lengths of adjacent sides, the terms will cancel each other out, leaving us with 1. In symbols,

$$\begin{aligned} \prod_{e \text{ incident to } \Delta} z(e,\Delta) &= 1 \\ \sum_{e \text{ incident to } \Delta} \arg z(e,\Delta) &= 2\pi \end{aligned} \tag{5.2}$$

We refer to Equation (5.1), as (e, Δ) ranges over all dihedral angles, and Equation (5.2), as e ranges over all edges, as the **consistency equations**.

Let v be a vertex of the triangulation. The **link** of v, which we denote by $L(v)$, is a torus that is topologically triangulated by the links of v in each tetrahedron. Let $\widetilde{v}$ be a lift of v in the universal cover. Up to an isometry, we can position $\widetilde{v}$ at ∞. Under this positioning, the links of $\widetilde{v}$ in each tetrahedron can be considered as Euclidean triangles.

Each $\gamma \in \pi_1(L(v))$ determines a deck transformation of $\widetilde{L(v)}$. Since this transformation has to preserve the angles of the Euclidean triangles, it is an affine transformation $z \mapsto az + b$. The number $a \in \mathbb{C}$ is the **derivative** of the affine transformation.

Let γ be an edge loop on $L(v)$. Observe that the product $(-1)^{|\gamma|} \prod_{(e,\Delta) \text{ right of } \gamma} z(e,\Delta)$ is the derivative of γ as a deck transformation of $\widetilde{L(v)}$. In particular, the value of the product only depends on the homotopy class of γ.

If γ is homotopically nontrivial, then since M is complete, it acts as a parabolic isometry on $\widetilde{M} \cong \mathbb{H}^3$. Thus the derivative of γ is 1. In symbols,

$$(-1)^{|\gamma|} \prod_{(e,\Delta) \text{ right of } \gamma} z(e,\Delta) = 1 \tag{5.3}$$

We refer to Equation (5.3), as v ranges over all cusps and γ ranges over all homotopically nontrivial curves on $L(v)$, as the **completeness equations**.

In fact, it suffices to consider the completeness equation for two curves on each cusp, namely a set of two curves that generate $\pi_1 L(v)$.

To conclude, we have shown the following proposition.

Proposition 5.2. *Given a geometric triangulation, the moduli of its edges satisfy both the consistency and the completeness equation.*

Conversely, suppose we are given an ideal triangulation τ of a 3-manifold M, where M has finitely many ends, each homeomorphic to $T^2 \times [0, \infty)$. Suppose we are given a number $z(e, \Delta) \in \{z \in \mathbb{C} \mid \Im z > 0\}$ for an edge e in each tetrahedron Δ.

First suppose that $z(e, \Delta)$ satisfy the consistency equations. Then we can construct M by gluing up pairs of faces of the relevant ideal hyperbolic tetrahedra. This endows M with a hyperbolic metric. However, this metric may not be complete.

To investigate completeness, we define a developing map $H : \widetilde{M} \to \mathbb{H}^3$ by requiring that H be a local isometry. Specifically, H can be constructed by first mapping an ideal hyperbolic tetrahedron in the lifted triangulation on $\widetilde{M}$ to $\mathbb{H}^3$ via an isometry, then extend the map to tetrahedra that meet the initial tetrahedron across faces, and continue inductively.

H is determined up to post-composition by an isometry of $\mathbb{H}^3$, depending on the choice of the isometry on the initial tetrahedron. In particular, for every $g \in \pi_1(M)$, $H \circ g = \rho(g) \circ H$ for some $\rho(g) \in \mathrm{Isom}^+(\mathbb{H}^3)$. This means that we can associate a representation $\rho : g \to \mathrm{Isom}^+(\mathbb{H}^3)$ to $z(e, \Delta)$, where ρ is determined up to conjugation of an element of $\mathrm{Isom}^+(\mathbb{H}^3)$.

Now suppose that $z(e, \Delta)$ satisfy the completeness equation as well. Then we claim that the lift of each end $T \times [0, \infty)$ of M is sent to a horoball under the developing map H. Indeed, up to an isometry we can assume that H sends the corresponding vertex v to ∞. Then from the completeness equations, we know that $\rho(g)$ is a parabolic isometry fixing ∞ for each $g \in \pi_1 T$. Moreover, since the moduli have positive imaginary part, $\rho(\pi_1 T)$ is a rank two discrete subgroup. Thus H sends $T \times [0, \infty)$ to a horoball based at ∞ as claimed. From this it follows that the hyperbolic metric on M is complete.

To conclude, we have shown the following proposition.

Proposition 5.3. *Given an ideal triangulation τ, a solution to the consistency and the completeness equations that has positive imaginary parts endows τ with the structure of a geometric triangulation with the prescribed moduli on its edges.*

In general, if $z(e, \Delta)$ does not satisfy the completeness equation, then the image of $\pi_1 T$ under ρ is a discrete rank two subgroup consisting of hyperbolic isometries preserving a geodesic line ℓ. The developing map H sends the end $T \times [0, \infty)$ to a neighborhood of ℓ.

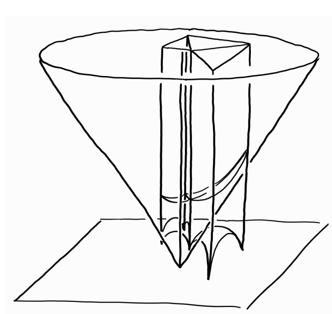

FIGURE 4.

Example 5.4. We work through the consistency and completeness equations of the ideal triangulation in Example 5.1.

We set the indeterminates for the moduli of the edges as shown in the figure below. In the figure, we also illustrate the link of the unique vertex.

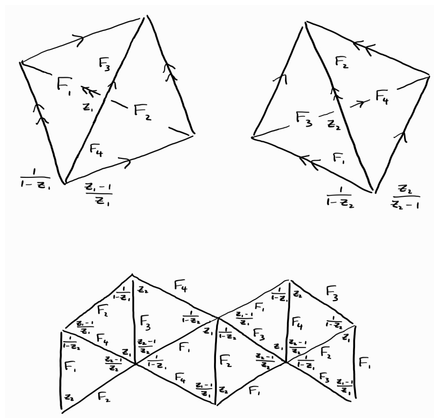

FIGURE 5.

The triangulation has two edges. The consistency equations for the two edges coincide and read

$$\begin{cases} \frac{(z_1-1)^2}{z_1}\frac{z_2}{(1-z_2)^2} = 1 \\ \arg z_1 + 2\arg\frac{z_1-1}{z_1} + \arg z_2 + 2\arg\frac{1}{z_2-1} = 2\pi \end{cases}$$

The completeness equations for the vertical and horizontal loops read

$$\begin{cases} -\frac{z_1-1}{z_1}\frac{z_2}{1-z_2} = 1 \\ \frac{z_2^2}{(1-z_2)^4} = 1 \end{cases}$$

If we require that $\Im z_1 > 0$ and $\Im z_2 > 0$, then the only solution to both the consistency and completeness equations is $z_1 = z_2 = \frac{1+i\sqrt{3}}{2}$. This recovers the geometric triangulation in Example 5.1.

5.3. **Deformation variety.** In this subsection, we explain how the consistency equations can be used to describe a local picture of the character variety.

Definition 5.5. Let τ be a finite ideal triangulation of a 3-manifold M. We define the **deformation variety** $\mathrm{Def}(M,\tau)$ to be the set of solutions to the consistency equations.

Note that here we consider all solutions, regardless of whether they have positive imaginary parts or not. In particular, given a solution that has negative imaginary part, it is a priori not clear whether it has any geometric meaning.

We describe one instance where solutions with negative imaginary part carry geometric meaning. Let M be a finite volume hyperbolic 3-manifold with a nonempty set of cusps. Recall the Epstein-Penner decomposition from Section 3.2. This decomposes M into convex ideal hyperbolic polyhedra. Each polyhedron can be decomposed into ideal tetrahedron. This does not immediately give a triangulation of M, since the triangulation of the faces of the polyhedron can differ. However, it is a fact that any two triangulations of a 2-dimensional polygon differ by diagonal switches. By inserting suitable flat ideal tetrahedra inbetween the faces, we get a **geometric triangulation with flat tetrahedra** on M.

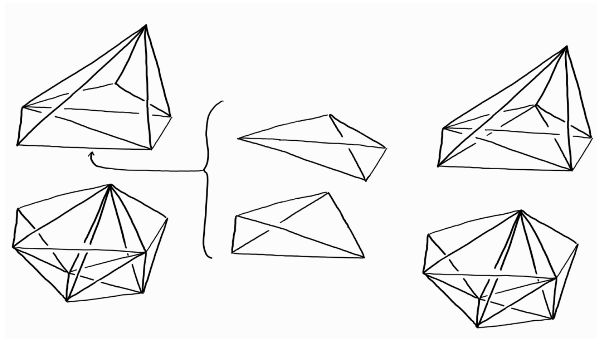

FIGURE 6.

A geometric triangulation with flat tetrahedra gives rise to a solution $z_0(e,\Delta)$ to the consistency and completeness equations as before. The only difference here is that $z_0(e,\Delta)$ is a real number different from $0,1,\infty$ when Δ is a flat tetrahedron.

To each face of identification between two polyhedra, we assign the face to one of the polyhedron. We then consider the union of tetrahedra that lie in the interior of a polyhedron P, as well as those associated to faces assigned to P, as being part of P.

Suppose $z(e,\Delta)$ is a solution lying close to $z_0(e,\Delta)$. Then the nondegenerate tetrahedra remain nondegenerate, i.e. if $\Im z_0(e,\Delta) > 0$ then $\Im z(e,\Delta) > 0$. The flat tetrahedra can become nondegenerate, remain flat, or (slightly) flip inside out, i.e. if $\Im z_0(e,\Delta) = 0$ then $\Im z(e,\Delta) > 0$, $= 0$, or < 0, respectively.

The union of tetrahedra lying in P determines a hyperbolic structure on P: One first takes the union of nondegenerate tetrahedra lying in the interior of P, then one tacks on the tetrahedra associated to the faces: a nondegenerate tetrahedron adds to P, making it more convex, a flat tetrahedron doesn't change the geometry of P, while a flipped tetrahedron subtracts from P, making it concave. Since any flipped tetrahedron is only slightly flipped, P remains homeomorphic to a 3-ball.

The fact that $z(e,\Delta)$ satisfies the consistency equation means that the faces of the polyhedron can be identified in the same manner to define a hyperbolic metric on M. As before, this allows us to define a developing map $H : \widetilde{M} \to \mathbb{H}^3$ and define an associated representation ρ.

Recall that ρ is only determined up to conjugation. Lifting from $\mathrm{Isom}^+(\mathbb{H}^3) \cong \mathrm{PSL}(2,\mathbb{C})$ to $\mathrm{SL}(2,\mathbb{C})$, we have a map from a neighborhood U_0 of $z_0(e,\Delta) \in \mathrm{Def}(M,\tau)$ to the character variety $X(\pi_1 M)$.

Since vertices of τ can be expressed as an algebraic function of the moduli, the value of $\rho(g)$ for each $g \in \pi_1 M$ is also an algebraic function of the moduli. This means that the map from $\mathrm{Def}(M,\tau)$ to $X(\pi_1 M)$ is algebraic.

Conversely, given a representation ρ close to the one ρ_0 associated to $z_0(e,\Delta)$, one can define a ρ-equivariant developing map $\widetilde{M} \to \mathbb{H}^3$ that is close to identity. We can read off a solution of the consistency equations by straightening the images of the lifted tetrahedra: If the straightened tetrahedron is nondegenerate, which must be the case if the tetrahedron is originally nondegenerate, then one uses the modulus of the tetrahedron. If the straightened tetrahedron is flat or flipped inside out, then one uses the conjugate of the modulus.

This gives an inverse map from a small neighborhood of ρ_0 in $X(\pi_1 M)$ to $\mathrm{Def}(M,\tau)$. Since the vertices of the tetrahedra depend algebraically on ρ, this map is also algebraic.

We have thus proved the following proposition.

Proposition 5.6. *A small neighborhood of $z_0(e,\Delta)$ in $\mathrm{Def}(M,\tau)$ is isomorphic to a small neighborhood of ρ_0 in $X(\pi_1 M)$.*

6. Hyperbolic Dehn surgery theorem

Let M be a finite volume hyperbolic 3-manifold with a nonempty set of cusps. Let $v \cong T^2 \times [0,\infty)$ be a cusp of M. Let $(g_{v,1}, g_{v,2})$ be a set of generators for $\pi_1 v$. Recall from Section 3.3 that we can associate to each $\rho \in X(\pi_1 M)$ close to the representation ρ_0 associated to the hyperbolic metric its generalized Dehn filling invariant $d_v(\rho) \in S^2 = \mathbb{R}^2 \cup \{\infty\}$. Putting together the maps d_v for each cusp v, we have a map $d : U \to (S^2)^{|\partial M|}$ where $U \subset X(\pi_1 M)$ is a neighborhood of ρ_0, where $d(\rho_0) = (\infty, \cdots, \infty)$.

The main theorem that we wish to show in this section is the following.

Theorem 6.1. *Let M be a finite volume hyperbolic 3-manifold with a nonempty set of cusps. The generalized Dehn filling invariant map $d : U \to (S^2)^{|\partial M|}$ is a local homeomorphism at ρ_0.*

Theorem 6.1 has the following important consequence, which is often refer to as the hyperbolic Dehn surgery theorem.

Corollary 6.2. *Let M be a finite volume hyperbolic 3-manifold with a nonempty set of cusps. There exists a neighborhood V of $(\infty, \cdots, \infty)$ in $(S^2)^{|\partial M|}$ such that for every $p = (p_{1,S}, p_{2,S})_S \in V$ with $(p_{1,S}, p_{2,S})$ a relatively prime pair of integers or ∞ for each cusp S, the Dehn filled manifold M_p admits a complete hyperbolic metric.*

In the sequel, we will refer to a Dehn filling as in Corollary 6.2 as a hyperbolic Dehn filling.

The strategy of the proof of Theorem 6.1 is to use Proposition 5.6 to consider the generalized Dehn filling invariant as a map on the deformation variety, then show that the map is a local diffeomorphism at the solution corresponding to the complete hyperbolic metric. We will follow [Mar22, Section 15].

6.1. Local smoothness.

For the rest of this section, we fix a geometric triangulation with flat tetrahedra τ as in Section 5.3.

In this subsection, we show that $\mathrm{Def}(M,\tau)$ is locally a smooth complex manifold of dimension equal to the number of cusps of M near the solution z_0 corresponding to the complete hyperbolic metric.

Observe that the number of edges equals the number of tetrahedra in τ. Indeed, this follows from $0 = \chi(M) = -|E(\tau)| + |F(\tau)| - |T(\tau)|$ and $|T(\tau)| = 2|F(\tau)|$. We denote $|E(\tau)| = |T(\tau)|$ by n and denote $|V(\tau)|$ by c.

We fix a labelling of the vertices of each tetrahedron in τ by v_0, v_1, v_2, v_3. We define a matrix $A \in \mathrm{Hom}(\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}, \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{E(\tau)})$ by

$$A = \begin{bmatrix} I & I & I \\ A_1 & A_2 & A_3 \end{bmatrix}$$

where I is the $n \times n$ identity matrix and

$$\begin{aligned}
(A_1)_{e,\Delta} &= \text{number of times } e \text{ is the edge labelled } [v_0, v_2] \text{ or } [v_1, v_3] \text{ in } \Delta \\
(A_2)_{e,\Delta} &= \text{number of times } e \text{ is the edge labelled } [v_1, v_2] \text{ or } [v_0, v_3] \text{ in } \Delta \\
(A_3)_{e,\Delta} &= \text{number of times } e \text{ is the edge labelled } [v_2, v_3] \text{ or } [v_0, v_1] \text{ in } \Delta
\end{aligned}$$

We define a matrix $B \in \mathrm{Hom}(\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{E(\tau)}, \mathbb{R}^{V(\tau)})$ by

$$B = \begin{bmatrix} B_T & B_E \end{bmatrix}$$

where

$$(B_T)_{v,\Delta} = -\text{number of times } v \text{ is incident to } \Delta$$
$$(B_E)_{v,e} = \text{number of times } v \text{ is incident to } e$$

Proposition 6.3. *We have an exact sequence*

$$\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \xrightarrow{\ A\ } \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{E(\tau)} \xrightarrow{\ B\ } \mathbb{R}^{V(\tau)} \longrightarrow 0$$

In particular the rank of A is $2n - c$.

Proof. It is straightforward to check that $BA = 0$.

We show that B_E is surjective: For any $v \in V(\tau)$, let f be a face of τ with a vertex at v. Let e_1 and e_2 be the two edges of f incident to v and e_3 be the remaining edges, then $B_E(\frac{1}{2}(e_1 + e_2 - e_3)) = v$.

B_E being surjective implies that B is surjective.

To show that $\ker B \subset \operatorname{im} A$, it suffices to show that $\ker A^T \subset \operatorname{im} B^T$. Suppose $y \in \ker A^T$. For each $v \in V(\tau)$, let f be a face of τ with a vertex at v. Let e_1 and e_2 be the two edges of f incident to v, and e_3 be the remaining edges. Let $w_v = \frac{1}{2}(y_{e_1} + y_{e_2} - y_{e_3})$.

Note that w_v only depends on v and not on f. This is because if f and f' are two faces of a tetrahedron Δ with vertices at v, then up to relabelling, we have $e_1 = e_1'$, and $(A^T y)_\Delta = 0$ implies that $y_\Delta + y_{e_2} + y_{e_3'} = 0$ and $y_\Delta + y_{e_2'} + y_{e_3} = 0$. In general, any two choices for f are related by a chain of faces of common tetrahedra.

It remains to check that $B^T w = y$. If e is an edge with vertices v and v', then we can take a face f that is incident to e. Up to relabelling, we have $e_1 = e_1' = e$, $e_2 = e_3'$, and $e_3 = e_2'$, hence $(B^T x)_e = w_v + w_{v'} = \frac{1}{2}(y_e + y_{e_2} - y_{e_3}) + \frac{1}{2}(y_e + y_{e_3} - y_{e_2}) = y_e$. If Δ is a tetrahedron, let e and e' be a pair of opposite edges. $(A^T y)_\Delta = 0$ implies that $y_\Delta + y_e + y_{e'} = 0$, thus $(B^T w)_\Delta = -\sum_{v \subset \Delta} w_v = -y_e - y_{e'} = y_\Delta$. ☐

We define a bilinear form ω on $\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ by

$$\omega(x, x') = \sum_\Delta \begin{bmatrix} x_{1,\Delta} & x_{2,\Delta} & x_{3,\Delta} \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x'_{1,\Delta} \\ x'_{2,\Delta} \\ x'_{3,\Delta} \end{bmatrix}$$

Proposition 6.4. *The rows of A generate a Legendrian subspace, i.e. $\omega(A_i, A_j) = 0$ for any pair of rows A_i, A_j of A, that has dimension $2n - c$.*

Proof. Proposition 6.3 implies that the rows of A generate a subspace of dimension $2n - c$, so it remains to show the Legendrian property.

It is clear that A_Δ lies in the radical of ω for each $\Delta \in T(\tau)$.

For $e, e' \in E(\tau)$, $\omega(A_e, A_{e'})$ is a signed count of tetrahedra for which e and e' are non-opposite edges. This signed count is zero since the tetrahedra can be paired together by shared faces. ☐

For each $e \in E(\tau)$, we define a function $G_e : \mathbb{C}^{T(\tau)} \to \mathbb{C}$ by

$$\begin{aligned} G_e(z) &= \sum_\Delta (A_1)_{e,\Delta} \log(z_\Delta) + \sum_\Delta (A_2)_{e,\Delta} \log(\frac{1}{1 - z_\Delta}) + \sum_\Delta (A_3)_{e,\Delta} \log(\frac{z_\Delta - 1}{z_\Delta}) \\ &= \sum_\Delta (A_1')_{e,\Delta} \log(z_\Delta) + \sum_\Delta (A_2')_{e,\Delta} \log(1 - z_\Delta) + N_e \pi i \end{aligned}$$

where $A_1' = A_1 - A_3$, $A_2' = A_3 - A_2$, and

$$N_e = \sum_{\Delta} (A_3)_{e,\Delta} = |\{\Delta \mid e \text{ is the edge labelled } [v_2, v_3] \text{ or } [v_0, v_1] \text{ in } \Delta\}|$$

is a constant depending only on e and τ.

Putting together the G_e, we have a function $G : \mathbb{C}^{T(\tau)} \to \mathbb{C}^{E(\tau)}$. The deformation variety $\mathrm{Def}(M, \tau)$ can be identified with the preimage $G^{-1}(2\pi i, \cdots, 2\pi i)$ near z_0.

Up to postcomposing by $N\pi i$, G can be factorized into

$$\mathbb{C}^{T(\tau)} \xrightarrow{L} \mathbb{C}^{T(\tau)} \oplus \mathbb{C}^{T(\tau)} \xrightarrow{A'} \mathbb{C}^{E(\tau)}$$

where $L : \mathbb{C}^{T(\tau)} \to \mathbb{C}^{T(\tau)} \oplus \mathbb{C}^{T(\tau)}$ is given by $L(z) = (\log(z), \log(1-z))$ and $A' = \begin{bmatrix} A_1' & A_2' \end{bmatrix}$. Hence at any point z, the differential $dG|_z$ can be factorized into

$$\mathbb{C}^{T(\tau)} \xrightarrow{dL|_z} \mathbb{C}^{T(\tau)} \oplus \mathbb{C}^{T(\tau)} \xrightarrow{A'} \mathbb{C}^{E(\tau)}$$

where $dL|_z = \begin{bmatrix} \frac{1}{z} \\ \frac{1}{1-z} \end{bmatrix}$.

Changing the coordinates in $\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ by $x \mapsto (x_{1,\Delta} - x_{3,\Delta}, x_{3,\Delta} - x_{2,\Delta}, x_{3,\Delta})$ transforms A to $\begin{bmatrix} 0 & 0 & I \\ A_1 - A_3 & A_3 - A_2 & A_3 \end{bmatrix} = \begin{bmatrix} 0 & I \\ A' & A_3 \end{bmatrix}$, so the row space of A is the direct sum of the row space of A' and the last $\mathbb{R}^{T(\tau)}$ summand. In particular, the rank of A' is $n - c$.

Changing the coordinates in $\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ by $x \mapsto (x_{1,\Delta} - x_{3,\Delta}, x_{3,\Delta} - x_{2,\Delta}, x_{3,\Delta})$ also transforms ω to

$$\omega(x, x') = \sum_{\Delta} \begin{bmatrix} x_{1,\Delta} - x_{3,\Delta} & x_{3,\Delta} - x_{2,\Delta} & x_{3,\Delta} \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x'_{1,\Delta} - x'_{3,\Delta} \\ x'_{3,\Delta} - x'_{2,\Delta} \\ x'_{3,\Delta} \end{bmatrix}$$

which restricts to a symplectic form on the first two summands $\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$. This is sometimes referred to as the **Neumann-Zagier symplectic form** in the literature.

Since the last $\mathbb{R}^{T(\tau)}$ summand lies in the row space of A, Proposition 6.4 implies the following proposition.

Proposition 6.5. *The rows of A' generate a Legendrian subspace of dimension $n - c$.*

Proposition 6.6. *At the solution z_0 corresponding to the complete hyperbolic metric, $dG|_{z_0} : \mathbb{C}^{T(\tau)} \to \ker B_E$ is surjective.*

Proof. Since we already know that $\operatorname{im} dG|_{z_0} \subset \ker B_E$, it suffices to show that the rank of $dG|_{z_0}$ is $n - c$. This is equivalent to the statement that the rank of $(dG|_{z_0})^T = (dL|_{z_0})^T (A')^T$ is $n - c$.

Throughout this proof, we will consider $(A')^T$ as a map from $\mathbb{R}^{E(\tau)}$ to $\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$, and $(dL|_{z_0})^T = \begin{bmatrix} \frac{1}{z} & \frac{1}{1-z} \end{bmatrix}$ as a map from $\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ to $\mathbb{C}^{T(\tau)}$, unless otherwise specified.

We partition $T(\tau)$ into $T_{\text{nondeg}}(\tau)$ and $T_{\text{flat}}(\tau)$. Since $\frac{1}{z_\Delta}, \frac{1}{1-z_\Delta} \in \mathbb{C}$ are linearly independent over $\mathbb{R}$ when $z_\Delta \notin \mathbb{R} \Leftrightarrow \Delta \in T_{\text{nondeg}}(\tau)$, $(dL|_{z_0})^T$ bijects $\mathbb{R}^{T_{\text{nondeg}}(\tau)} \oplus \mathbb{R}^{T_{\text{nondeg}}(\tau)}$ into $\mathbb{C}^{T_{\text{nondeg}}(\tau)}$. Meanwhile, $\ker(dL|_{z_0})^T \subset \mathbb{R}^{T_{\text{flat}}(\tau)} \oplus \mathbb{R}^{T_{\text{flat}}(\tau)}$ equals

$$\left\{ x' \in \mathbb{R}^{T_{\text{flat}}(\tau)} \oplus \mathbb{R}^{T_{\text{flat}}(\tau)} \mid x'_{1,\Delta} \frac{1}{z_\Delta} + x'_{2,\Delta} \frac{1}{1 - z_\Delta} = 0 \text{ for all } \Delta \in T_{\text{flat}}(\tau) \right\}$$

Claim 6.7. $\operatorname{im}(A')^T$ *intersects* $\ker(dL|_{z_0})^T$ *trivially.*

Proof. Suppose otherwise that there is a nonzero vector $x' \in \operatorname{im}(A')^T$ for which $(dL|_{z_0})^T x = 0$. Then

$$(x'_{1,\Delta}, x'_{2,\Delta}) = \begin{cases} (0,0) & \text{if } \Delta \in T_{\text{nondeg}}(\tau) \\ k_\Delta(z_\Delta, z_\Delta - 1) & \text{if } \Delta \in T_{\text{flat}}(\tau), \text{ for some } k_\Delta \in \mathbb{R} \end{cases}$$

We define a vector $x \in \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ by $(x_{1,\Delta}, x_{2,\Delta}, x_{3,\Delta}) = (x'_{1,\Delta}, -x'_{2,\Delta}, 0)$ for every $\Delta \in T(\tau)$. x can be obtained from $x' \in \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ by reversing the coordinate change $x \mapsto (x_{1,\Delta} - x_{3,\Delta}, x_{3,\Delta} - x_{2,\Delta}, x_{3,\Delta})$, thus $x \in \operatorname{im} A^T$.

Observe that $x_{i,\Delta} = 0$ for $\Delta \in T_{\text{nondeg}}(\tau)$, and $x_{i,\Delta}$ is either zero or a triple of distinct real numbers for $\Delta \in T_{\text{flat}}(\tau)$.

Now by Proposition 6.4, $\operatorname{im} A^T$ is Legendrian, so $\omega(x, A_e) = 0$ for any edge e. We take e to be an edge lying at the top of a stack of flat tetrahedra. Then e is incident to only one flat tetrahedron Δ, and $\omega(x, A_e) = x_{i,\Delta} - x_{i',\Delta}$ for values of i, i' depending on which edge e is labelled as in Δ. This means that at least two of $x_{i,\Delta}$ coincide, thus all of $x_{i,\Delta}$ are zero.

We repeat this argument on the next top edge of the stack, and proceed inductively to see that $x = 0$, which implies that $x' = 0$. $\square$

The claim implies that $\operatorname{im}(dG|_{z_0})^T = \operatorname{im}(dL|_{z_0})^T(A')^T$ has real dimension equals to that of $\operatorname{im}(A')^T$, which is $n - c$ by Proposition 6.5. We have to show that upon complexifying, $\operatorname{im}(dG|_{z_0})^T$ retains the same complex dimension.

To this end, we push forward ω under $(dL|_{z_0})^T|_{\mathbb{R}^{T_{\text{nondeg}}(\tau)} \oplus \mathbb{R}^{T_{\text{nondeg}}(\tau)}}$ into a symplectic form Ω on $\mathbb{C}^{T_{\text{nondeg}}(\tau)}$. We extend Ω into a bilinear form on $\mathbb{C}^{T(\tau)}$ by declaring $\mathbb{C}^{T_{\text{flat}}(\tau)}$ to be its radical.

For $\Delta \in T_{\text{nondeg}}(\tau)$, since $\Im z_\Delta > 0$, $\left(\frac{1}{z_\Delta}, \frac{1}{1-z_\Delta}\right)$ form a negatively oriented basis of $\mathbb{C}$ over $\mathbb{R}$. Thus $\Omega(v, iv) < 0$ for all $v \in \mathbb{C}^{T_{\text{nondeg}}(\tau)}$. We define a positive definite bilinear form on $\mathbb{C}^{T_{\text{nondeg}}(\tau)}$ by $\langle v, w \rangle = -\Omega(v, iw)$.

Since there is only one symplectic form on $\mathbb{R}^2$ up to scaling, Ω must be a real multiple of the standard symplectic form on each summand of $\mathbb{C}^{T_{\text{nondeg}}(\tau)}$. In particular, we have $\Omega(iv, iw) = \Omega(v, w)$, thus $\langle iv, iw \rangle = \langle v, w \rangle$.

Suppose the rank of $(dG|_{z_0})^T$ is less than $n - c$, then there exists vectors $y_1, y_2 \in \mathbb{R}^{E(\tau)}$ such that $(dG|_{z_0})^T(y_1 + iy_2) = 0$ while $(dG|_{z_0})^T(y_1) \neq 0$ and $(dG|_{z_0})^T(y_2) \neq 0$.

Since $(dG|_{z_0})^T(y_i)|_{\mathbb{C}^{T_{\text{flat}}(\tau)}} \in \mathbb{R}^{T_{\text{flat}}(\tau)}$, we must have $(dG|_{z_0})^T(y_i)|_{\mathbb{C}^{T_{\text{flat}}(\tau)}} = 0$, i.e. $(dG|_{z_0})^T(y_i) \in \mathbb{C}^{T_{\text{nondeg}}(\tau)}$.

But then

$$\begin{aligned} \Omega((dG|_{z_0})^T(y_1), (dG|_{z_0})^T(y_2)) &= \omega((dA')^T(y_1), (dA')^T(y_2)) \\ &= 0 \text{ by Proposition 6.5} \end{aligned}$$

which implies

$$\begin{aligned} 0 = ||(dG|_{z_0})^T(y_1 + iy_2)||^2 &= ||(dG|_{z_0})^T(y_1)||^2 + ||i(dG|_{z_0})^T(y_2)||^2 + \langle (dG|_{z_0})^T(y_1), i(dG|_{z_0})^T(y_2) \rangle \\ &= ||(dG|_{z_0})^T(y_1)||^2 + ||(dG|_{z_0})^T(y_2)||^2 + \Omega((dG|_{z_0})^T(y_1), (dG|_{z_0})^T(y_2)) \\ &= ||(dG|_{z_0})^T(y_1)||^2 + ||(dG|_{z_0})^T(y_2)||^2 \end{aligned}$$

hence $(dG|_{z_0})^T(y_1) = (dG|_{z_0})^T(y_2) = 0$, contradiction. $\square$

6.2. **Local coordinates.** In this subsection, we set up a system of local coordinates of $\mathrm{Def}(M,\tau)$ near z_0. These local coordinates are essentially the derivatives of curves on the cusps, and will be utilized to show Theorem 6.1.

Let v be a cusp and $L(v)$ be the link of v in τ. Let γ be an edge loop on $L(v)$. We can define a (row) vector $x_\gamma \in \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ by $x_\gamma = \begin{bmatrix}(x_1)_\gamma & (x_2)_\gamma & (x_3)_\gamma\end{bmatrix}$ where

$$\begin{aligned}
(x_1)_{\gamma,\Delta} &= \text{number of times the edges labelled } [v_0,v_2] \text{ and } [v_1,v_3] \text{ in } \Delta \text{ lies on the right of } \gamma\\
(x_2)_{\gamma,\Delta} &= \text{number of times the edges labelled } [v_1,v_2] \text{ and } [v_0,v_3] \text{ in } \Delta \text{ lies on the right of } \gamma\\
(x_3)_{\gamma,\Delta} &= \text{number of times the edges labelled } [v_2,v_3] \text{ and } [v_0,v_1] \text{ in } \Delta \text{ lies on the right of } \gamma
\end{aligned}$$

Proposition 6.8. *The vectors x_γ satisfy the following properties.*

(1) $\omega(x_\gamma, x_{\gamma'}) = \begin{cases} 0 & \textit{if } \gamma \textit{ and } \gamma' \textit{ lie on different links} \\ 2\langle \gamma, \gamma' \rangle & \textit{if } \gamma \textit{ and } \gamma' \textit{ lie on the same link} \end{cases}$

(2) $\omega(x_\gamma, A_i) = 0$ *for every row A_i of A.*

Proof. As pointed out in Proposition 6.4, A_Δ lies in the radical of ω for each $\Delta \in T(\tau)$. This implies half of (2). We claim that (1) implies the other half of (2). Indeed, given an edge e, let v be one of its vertices, and let γ be the degenerate edge loop γ at e. Then $x_\gamma = A_e$.

It remains to show (1). For each tetrahedron Δ, $\begin{bmatrix}(x_1)_{\gamma,\Delta} & (x_2)_{\gamma,\Delta} & (x_3)_{\gamma,\Delta}\end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix}(x_1)_{\gamma',\Delta} \\ (x_2)_{\gamma',\Delta} \\ (x_3)_{\gamma',\Delta}\end{bmatrix}$ splits into terms that come from segments of γ and γ' lying on the same vertex of Δ and those that come from segments lying on different vertices of Δ.

We first show that the terms that come from segments lying on the same vertex add up to give $2\langle \gamma, \gamma' \rangle$. Indeed, at every intersection point of γ and γ', there is a stack of tetrahedra lying on the right of γ and γ'. The sum of the terms for a single tetrahedron Δ equals the number of times a vertex of Δ is the outermost triangle in a wedge for a positive intersection point minus the number of times a vertex of Δ is the outermost triangle in a wedge for a negative intersection point. Since every intersection point has two outermost triangles, we end up with $2\langle \gamma, \gamma' \rangle$.

We then show that the terms that come from segments lying on different vertices cancel each other out. Indeed, such a term must arise from a pair of non-opposite edges (e, e') of Δ where γ passes through a vertex of e and γ' passes through a vertex of e'. Such a pair of edges span a face, which is adjacent to two tetrahedra. The two corresponding terms contributed by the two tetrahedra cancel each other out.

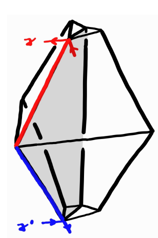

Figure 7.

$\square$

For each cusp v, we pick a homotopically nontrivial edge loop γ_v in $L(v)$. Putting the row vectors x_{γ_v} together, we obtain a matrix $x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \in \mathrm{Hom}(\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}, \mathbb{R}^{V(\tau)})$. Putting A and x together, we obtain $\overline{A} = \begin{bmatrix} A \\ x \end{bmatrix} \in \mathrm{Hom}(\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}, \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{E(\tau)} \oplus \mathbb{R}^{V(\tau)})$.

Proposition 6.9. *The rows of $\overline{A}$ generate a Legendrian subspace of dimension $2n$.*

Proof. Proposition 6.8 implies that the rows of $\overline{A}$ generate a Legendrian subspace.

Proposition 6.3 says that first $2n$ rows of $\overline{A}$ generate a $2n - c$ dimensional subspace. Thus it remains to show that the last c rows are linearly independent from the first $2n$ rows and from each other: For each cusp v, pick an edge loop μ_v on $L(v)$ such that $\langle \mu_v, \gamma_v \rangle = 1$. Then by Proposition 6.8, $\omega(x_{\mu_v}, A_i) = 0$ for each of the first $2n$ rows A_i and $\omega(x_{\mu_v}, x_{\gamma_{v'}}) = \langle \mu_v, \gamma_{v'} \rangle = \delta_{v,v'}$. □

As reasoned in Section 5.2, the derivative of γ_v as a deck transformation can be computed as

$$h_v(z) = (-1)^{|\gamma_v|} \prod_{(e,\Delta) \text{ right of } \gamma_v} z(e, \Delta)$$

We consider the function

$$\begin{aligned} H_v(z) &= \log h_v(z) \\ &= \sum_{(e,\Delta) \text{ right of } \gamma_v} \log z(e,\Delta) + |\gamma_v| \pi i \\ &= \sum_{\Delta} (x_1)_{v,\Delta} \log(z_\Delta) + \sum_{\Delta} (x_2)_{v,\Delta} \log(\frac{1}{1 - z_\Delta}) + \sum_{\Delta} (x_3)_{v,\Delta} \log(\frac{z_\Delta - 1}{z_\Delta}) + |\gamma_v| \pi i \\ &= \sum_{\Delta} (x_1')_{v,\Delta} \log(z_\Delta) + \sum_{\Delta} (x_2')_{v,\Delta} \log(1 - z_\Delta) + N_v \pi i \end{aligned}$$

where $x_1' = x_1 - x_3$, $x_2' = x_3 - x_2$, and

$$N_v = |\gamma_v| + \sum_{\Delta} (x_3)_{v,\Delta} = |\{\Delta \mid \text{the edge labelled } [v_2, v_3] \text{ or } [v_0, v_1] \text{ in } \Delta \text{ lies on the right of } \gamma_v\}|$$

is a constant depending only on v, γ_v, and τ.

Putting together G and the H_v, we have a function $G \times H : \mathbb{C}^{T(\tau)} \oplus \mathbb{C}^{T(\tau)} \to \mathbb{C}^{E(\tau)} \oplus \mathbb{C}^{V(\tau)}$, which maps the solution z_0 corresponding to the complete metric to $(2\pi i, \cdots, 2\pi i, 0, \cdots, 0)$.

Up to postcomposing by $N\pi i$, $G \times H$ can be factorized into

$$\mathbb{C}^{T(\tau)} \xrightarrow{L} \mathbb{C}^{T(\tau)} \oplus \mathbb{C}^{T(\tau)} \xrightarrow{\overline{A}'} \mathbb{C}^{E(\tau)} \oplus \mathbb{C}^{V(\tau)}$$

where $\overline{A}' = \begin{bmatrix} A' \\ x' \end{bmatrix}$. Hence at any point z, the differential $d(G \times H)|_z$ can be factorized into

$$\mathbb{C}^{T(\tau)} \xrightarrow{dL|_z} \mathbb{C}^{T(\tau)} \oplus \mathbb{C}^{T(\tau)} \xrightarrow{\overline{A}'} \mathbb{C}^{E(\tau)} \oplus \mathbb{C}^{V(\tau)}$$

The same argument used to deduce Proposition 6.5 from Proposition 6.4 allows us to deduce the following proposition from Proposition 6.9

Proposition 6.10. *The rows of $\overline{A}'$ generate a Legendrian subspace of dimension n.*

Proposition 6.11. *At the solution z_0 corresponding to the complete hyperbolic metric, $d(G \times H)|_{z_0} : \mathbb{C}^{T(\tau)} \to \ker B_E \oplus \mathbb{C}^{V(\tau)}$ is bijective.*

Proof. We have to show that the rank of $d(G \times H)|_{z_0}$ is n. This is equivalent to the statement that the rank of $(d(G \times H)|_{z_0})^T = (dL|_{z_0})^T (\overline{A}')^T$ is n.

To do this, we can use the same argument as in Proposition 6.6, using Proposition 6.9 instead of Proposition 6.4 and Proposition 6.10 instead of Proposition 6.5. $\square$

Corollary 6.12 (Choi [Cho04]). *The restriction of H to $\mathrm{Def}(M, \tau) \subset \mathbb{C}^{T(\tau)}$ is a local algebraic isomorphism at z_0.*

6.3. **Proof of hyperbolic Dehn surgery theorem.** We are now ready to prove the following theorem, which by Proposition 5.6 would imply Theorem 6.1.

Theorem 6.13. *Let M be a finite volume hyperbolic 3-manifold with a nonempty set of cusps. The generalized Dehn filling invariant map $d : U \to (S^2)^{|\partial M|}$ is a local biholomorphism on a neighborhood of z_0 on $\mathrm{Def}(M, \tau)$.*

Proof. For each cusp v and each $i = 1, 2$, let $\gamma_{v,i}$ be an edge loop on $L(v)$ with homotopy class $g_{v,i}$ Let $u_{v,i}(z) = \sum_{(e,\,\Delta)\text{ right of }\gamma_{v,i}} \log z(e, \Delta) + |\gamma_{v,i}|\pi i$. Then Corollary 6.12 states that $(u_{v,1})_v$ and $(u_{v,2})_v$ are holomorphic coordinates systems of $\mathrm{Def}(M, \tau)$ near z_0 that are 0 at z_0.

Fix a cusp v, and fix a lift $\widetilde{v}$ of v. Up to an isometry, we can assume that $\widetilde{v}$ lies at ∞. We have

$$\begin{aligned} u_{v,1} = 0 &\Leftrightarrow g_{v,1} \text{ is parabolic} \\ &\Leftrightarrow g_{v,2} \text{ is parabolic} \\ &\Leftrightarrow u_{v,2} = 0 \end{aligned}$$

This implies that the function $\frac{u_{v,2}}{u_{v,1}}$ is well-defined and analytic near z_0.

For $u_{v,1}, u_{v,2} = 0$, up to an isometry, we can assume that $\rho_0(g_{v,1})(z) = z + 1$ and $\rho_0(g_{v,2})(z) = z + w_v$. Thus there is a fundamental domain of the action on $\pi_1 L(v)$ on $\mathbb{R}^2$ that is a rectangle with vertices at 0, 1, w_v, $w_v + 1$. For $u_{v,1}, u_{v,2} \neq 0$, up to translation, we can assume that $\rho_0(g_{v,1})(z) = \exp(u_{v,1})z$ and $\rho_0(g_{v,2})(z) = \exp(u_{v,2})z$. Thus there is a fundamental domain of the action on $\pi_1 L(v)$ on $\mathbb{R}^2$ that is a curved rectangle with vertices at 1, $\exp(u_{v,1})$, $\exp(u_{v,2})$, $\exp(u_{v,1} + u_{v,2})$.

Take a sequence of points approaching z_0, then since the developing maps have to converge, the curved rectangles $[1, \exp(u_{v,1}), \exp(u_{v,2}), \exp(u_{v,1} + u_{v,2})]$ have to converge to $[0, 1, w_v, w_v + 1]$ up to translation and rescaling. We compute

$$\begin{aligned} &\lim_{z \to z_0} [1, \exp(u_{v,1}), \exp(u_{v,2}), \exp(u_{v,1} + u_{v,2})] \\ &= \lim_{z \to z_0} [0, 1, \frac{\exp(u_{v,2}) - 1}{\exp(u_{v,1}) - 1}, \frac{\exp(u_{v,1} + u_{v,2})}{\exp(u_{v,1}) - 1}] \text{ up to translation and rescaling} \\ &= [0, 1, \lim_{z \to z_0} \frac{u_{v,2}}{u_{v,1}}, \lim_{z \to z_0} \frac{u_{v,2}}{u_{v,1}} + 1] \end{aligned}$$

Thus $w_v = \lim_{z \to z_0} \frac{u_{v,2}}{u_{v,1}}$.

We define Φ to be the composition of the generalized Dehn filling coefficient map with the map $\phi(p) = (\frac{2\pi i}{p_{v,1} + p_{v,2} w_v})_v$. We claim that the derivative $d\Phi|_{z_0}$ is identity under the coordinates $(u_{v,1})_v$.

Indeed, recall that $p_{v,1}, p_{v,2}$ are defined as the solution to $p_{v,1} u_{v,1} + p_{v,2} u_{v,2} = 2\pi i$. Thus

$$\begin{aligned} p_{v,1} &= -2\pi \frac{\Re u_{v,2}}{\Im(\overline{u_{v,1}} u_{v,2})} \\ p_{v,2} &= 2\pi \frac{\Re u_{v,1}}{\Im(\overline{u_{v,1}} u_{v,2})} \end{aligned}$$

Near z_0, using the fact that $u_{v,2} \sim w_v u_{v,1}$, we have

$$p_{v,1} \sim -2\pi \frac{\Re w_v \Re u_{v,1} - \Im w_v \Im u_{v,1}}{|u_{v,1}|^2 \Im w_v}$$

$$p_{v,2} \sim 2\pi \frac{\Re u_{v,1}}{|u_{v,1}|^2 \Im w_v}$$

Thus

$$\frac{2\pi i}{p_{v,1} + p_{v,2} w_v} = u_{v,1}$$

Since ϕ is a local biholomorphism, we conclude that the generalized Dehn filling coefficient map is also a local biholomorphism. □

6.4. **Core geodesics and Mostow-Prasad rigidity.** In this section, we record an observation from the proof of Theorem 6.13.

Let v be a cusp of M. Whenever the generalized Dehn filling coefficient $(p_{v,1}, p_{v,2})$ for v are a pair of relatively prime inetegers, we have a complete hyperbolic metric on the filled manifold M_p, where v is one of the filled cusps. We can thus talk about the length of the core curve of the filling solid torus.

This core curve is homotopic to $q_{v,1}g_1 + q_{v,2}g_2$, where $(p_{v,1}g_1 + p_{v,2}g_2, q_{v,1}g_1 + q_{v,2}g_2)$ is a basis for $L(v)$. The length of this core curve is the logarithm of the derivative of $\rho_p(q_{v,1}g_1 + q_{v,2}g_2)$, where ρ_p is the representative corresponding to the complete hyperbolic metric on M_p.

Theorem 6.14. *Let M be a finite volume hyperbolic 3-manifold with a nonempty set of cusps. Let v be a cusp. Suppose we have a sequence of hyperbolic Dehn fillings $\overline{M}_i$ where the filling coefficient for v converges to ∞, then the length of the core curve at v converges to 0.*

Proof. We use the notation as in the proof of Theorem 6.13. The length of the core curve is $\Re(q_{v,1}u_{v,1} + q_{v,2}u_{v,2}) = q_{v,1}\Re u_{v,1} + q_{v,2}\Re u_{v,2}$.

Meanwhile, we have $p_{v,1}q_{v,2} - p_{v,2}q_{v,1} = 1$, thus using the formulas we have for $p_{v,1}$ and $p_{v,2}$, we have $q_{v,1}\Re u_{v,1} + q_{v,2}\Re u_{v,2} = -\frac{1}{2\pi}\Im(\overline{u_{v,1}}u_{v,2}) \to 0$. □

Theorem 6.14 is very handy for translating results about closed hyperbolic 3-manifolds to cusped hyperbolic 3-manifolds. We demonstrate an instance of this by generalizing Mostow rigidity to cusped hyperbolic 3-manifolds.

Theorem 6.15 (Mostow-Prasad rigidity). *Let M_1 and M_2 be two finite volume hyperbolic 3-manifolds. Every isomorphism $\pi_1(M_1) \to \pi_1(M_2)$ is induced by a unique isometry $M_1 \to M_2$.*

Proof. For a finite volume hyperbolic 3-manifold, the fundamental groups of the cusps can be characterized algebraically as (the conjugacy classes of) the maximal $\mathbb{Z}^2$ subgroups. Thus an isomorphism $\pi_1(M_1) \cong \pi_1(M_2)$ induces a bijection between the cusps of M_1 and M_2, and an isomorphism on the fundamental groups of the cusps.

By filling the cusps with large enough coefficients, we have an induced isomorphism $\pi_1(\overline{M_1}) \cong \pi_1(\overline{M_2})$ between closed hyperbolic 3-manifolds. By Mostow rigidity, the isomorphism $\pi_1(M_1) \cong \pi_1(M_2)$ is induced by an isometry $\overline{M_1} \cong \overline{M_2}$.

Moreover, for large enough coefficients, we can assume that the core curves of the filling are the shortest curves in $\overline{M_i}$. Thus we can drill out these curves to get an isometry $M_1 \cong M_2$ between the incomplete metrics that are the restricted complete metrics. By taking a sequence of coefficients that converge to ∞, the incomplete metrics converge towards the complete hyperbolic metrics on M_i, thus we have an isometry $M_1 \cong M_2$. □

As for Mostow rigidity, an important consequence of Theorem 6.15 is that the hyperbolic metric on a finite volume hyperbolic 3-manifold is unique up to homeomorphism.

Corollary 6.16. *Let M be a finite volume hyperbolic 3-manifold. Let g_1 and g_2 be two complete hyperbolic metrics on M. Then there exists a homeomorphism $h : M \to M$ such that $g_1 = h^* g_2$.*

7. Angle structures and volume

The main goal of this section is to prove the following theorem.

Theorem 7.1. *Let M be a finite volume hyperbolic 3-manifold with a nonempty set of cusps. Suppose we have a sequence of hyperbolic Dehn fillings $\overline{M}_i$ where the filling coefficient converges to ∞, then* $\mathrm{vol}(\overline{M}_i) \nearrow \mathrm{vol}(M)$.

The strategy of the proof is to use a variation of the deformation variety. Instead of considering the moduli of edges, we only keep track of the dihedral angles. In a sense, we are only keeping track of the linear parts of the consistency equations. However, we disallow flat tetrahedra.

We then define a volume functional on the linear space of solutions, which agrees with the usual volume when the solution comes from a solution of full consistency equations. It will follow from Proposition 4.8 that the volume functional is strictly convex.

The crucial fact, first observed by Rivin, is that a point in the space of solutions is a critical point of the volume functional if and only if it comes from a solution of both the full consistency and completeness equations.

Thus if M admits a geometric triangulation, then Theorem 7.1 follows from the convexity of the volume functional. In general, we have to appeal to a result of Luo-Schleimer-Tillman that every hyperbolic 3-manifold has a finite cover that admits a geometric triangulation.

We follow [Mar22, Section 15.3].

7.1. **Strict angle structures.** Let M be a 3-manifold with finitely many ends, each homeomorphic to $T^2 \times [0, \infty)$. Let τ be an ideal triangulation of M.

Definition 7.2. A **strict angle structure** on τ is an assignment of an angle $\alpha(e, \Delta) \in (0, \pi)$ for each edge e in each tetrahedron Δ which satisfies:

(1) $\alpha(e, \Delta) = \alpha(e', \Delta)$ if e and e' are opposite edges in Δ.
(2) $\alpha(e, \Delta) + \alpha(e', \Delta) + \alpha(e'', \Delta) = \pi$ if e, e', e'' are the edges in Δ that are incident to a vertex.
(3) For any edge e, $\sum_\Delta \alpha(e, \Delta) = 2\pi$ where the sum is taken over all tetrahedra Δ that are incident to e.

For example, given a solution z to the consistency equations with $\Im z(e, \Delta) > 0$, the arguments $\alpha(e, \Delta) = \arg z(e, \Delta)$ is a strict angle structure.

Conversely, given a strict angle structure, we can endow each tetrahedron of τ the structure of a hyperbolic ideal tetrahedron. In fact, the modulus of a dihedral edge (e, Δ) can be recovered via the formula

$$z(e, \Delta) = \frac{\sin \alpha(e', \Delta)}{\sin \alpha(e'', \Delta)} \exp(i\alpha(e, \Delta))$$

where e, e', e'' are positively oriented at a vertex.

However, these hyperbolic ideal tetrahedra may not glue up into a hyperbolic metric on M. Indeed, Definition 7.2(3) is only the argument part of the consistency equation Equation (5.2). Without the modulus part, there might be shearing as one goes around an edge.

As in Section 6.1, we fix a labelling of the vertices of each tetrahedron and define the matrix $A \in \mathrm{Hom}(\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}, \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{E(\tau)})$ defined in Section 6.1. Then the space of all strict angle structures can be identified with $A^{-1}(\pi, ..., \pi, 2\pi, ..., 2\pi) \cap (\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)})_+$. We denote this space by $\mathcal{A}$.

Note that $T\mathcal{A} = \ker A$. We wish to describe this tangent space. To this end, we set up some notation.

Recall that bilinear form ω defined on $\mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$. Let $\langle \cdot, \cdot \rangle$ be the usual inner product. For $v \in \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$, we define $v^* \in \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)} \oplus \mathbb{R}^{T(\tau)}$ to be the vector satisfying $\omega(v, w) = \langle v^*, w \rangle$ for all w.

For each tetrahedron Δ, the row A_Δ lies in the radical of ω, thus $A^*_\Delta = 0$. For each edge e, A^*_e is the vector with entries

$$\begin{aligned}
(A^*_e)_{1,\Delta} &= \text{number of times } e \text{ is the edge labelled } [v_2, v_3] \text{ or } [v_0, v_1] \text{ in } \Delta \\
&\quad - \text{number of times } e \text{ is the edge labelled } [v_1, v_2] \text{ or } [v_0, v_3] \text{ in } \Delta \\
(A^*_e)_{2,\Delta} &= \text{number of times } e \text{ is the edge labelled } [v_0, v_2] \text{ or } [v_1, v_3] \text{ in } \Delta \\
&\quad - \text{number of times } e \text{ is the edge labelled } [v_2, v_3] \text{ or } [v_0, v_1] \text{ in } \Delta \\
(A^*_e)_{3,\Delta} &= \text{number of times } e \text{ is the edge labelled } [v_1, v_2] \text{ or } [v_0, v_3] \text{ in } \Delta \\
&\quad - \text{number of times } e \text{ is the edge labelled } [v_0, v_2] \text{ or } [v_1, v_3] \text{ in } \Delta
\end{aligned}$$

For each cusp v, let $(\gamma_{v,1}, \gamma_{v,2})$ be a pair of edge paths in $L(v)$ that generate $\pi_1 L(v)$. Recall the vectors $x_{v,i} = x_{\gamma_{v,i}}$ defined in Section 6.2. $x^*_{v,i}$ is the vector with entries

$$\begin{aligned}
(x^*_{v,i})_{1,\Delta} &= \text{number of times the edges labelled } [v_2, v_3] \text{ and } [v_0, v_1] \text{ in } \Delta \text{ lies on the right of } \gamma \\
&\quad - \text{number of times the edges labelled } [v_1, v_2] \text{ and } [v_0, v_3] \text{ in } \Delta \text{ lies on the right of } \gamma \\
(x^*_{v,i})_{2,\Delta} &= \text{number of times the edges labelled } [v_0, v_2] \text{ and } [v_1, v_3] \text{ in } \Delta \text{ lies on the right of } \gamma \\
&\quad - \text{number of times the edges labelled } [v_2, v_3] \text{ and } [v_0, v_1] \text{ in } \Delta \text{ lies on the right of } \gamma \\
(x^*_{v,i})_{3,\Delta} &= \text{number of times the edges labelled } [v_1, v_2] \text{ and } [v_0, v_3] \text{ in } \Delta \text{ lies on the right of } \gamma \\
&\quad - \text{number of times the edges labelled } [v_0, v_2] \text{ and } [v_1, v_3] \text{ in } \Delta \text{ lies on the right of } \gamma
\end{aligned}$$

Proposition 7.3. *$T\mathcal{A} = \ker A$ is generated by the $2n + c$ vectors A^*_e and $x^*_{v,i}$.*

Proof. The same argument as in Proposition 6.9 shows that $A^*_e, x^*_{v,i} \subset \ker A$. The dimension of $\ker A$ is $3|T(\tau)| - \dim \operatorname{im} A = 3n - (2n - c) = n + c$. Thus it remains to show that A^*_e, $x^*_{v,i}$ generate a space of dimension $\geq n + c$.

To do so, we use the same argument as in Proposition 6.9 to show that the vectors A_Δ, A_e, $x_{v,i}$ generate a space of dimension $2n + c$. Since the radical of ω has dimension n, the dual vectors $A^*_\Delta = 0$, A^*_e, $x^*_{v,i}$ generate a space of dimension $\geq n + c$. ∎

7.2. **Volume functional.** Recall the Lobachevsky function $\Lambda(\theta) = -\int_0^\theta \log|2\sin\theta| d\theta$. We define the volume functional $\mathrm{vol} : \mathcal{A} \to (0, \infty)$ by $\mathrm{vol}(\alpha) = \sum_{i,\Delta} \Lambda(\alpha_{i,\Delta})$.

It follows from Proposition 4.8 that if α comes from a solution z to the consistency equations with $\Im z(e, \Delta) > 0$, then $\mathrm{vol}(\alpha)$ is the volume of the glued up hyperbolic metric on M.

Proposition 7.4. *Given $\alpha \in \mathcal{A}$ and $x \in T\mathcal{A}$, we have $\frac{\partial \mathrm{vol}}{\partial x} = -\sum_{i,\Delta} x_{i,\Delta} \log(\sin\alpha_{i,\Delta})$ and $\frac{\partial^2 \mathrm{vol}}{\partial x^2} < 0$.*

Proof. Routine calculus exercise. ∎

Theorem 7.5 (Rivin)**.** *A point $\alpha \in \mathcal{A}$ is a critical point of* vol *if and only if it comes from a solution of the consistency and completeness equations.*

Proof. For each edge e,

$$\begin{aligned}
\frac{\partial \mathrm{vol}}{\partial A_e^*} &= -\sum_{i,\Delta}(A_e^*)_{i,\Delta}\log(\sin\alpha_{i,\Delta}) \\
&= \sum_{\Delta}(\text{number of times } e \text{ is the edge labelled } [v_0,v_2] \text{ or } [v_1,v_3] \text{ in } \Delta)\log(\frac{\sin\alpha_{2,\Delta}}{\sin\alpha_{3,\Delta}}) \\
&\quad+\sum_{\Delta}(\text{number of times } e \text{ is the edge labelled } [v_1,v_2] \text{ or } [v_0,v_3] \text{ in } \Delta)\log(\frac{\sin\alpha_{3,\Delta}}{\sin\alpha_{1,\Delta}}) \\
&\quad+\sum_{\Delta}(\text{number of times } e \text{ is the edge labelled } [v_2,v_3] \text{ or } [v_0,v_1] \text{ in } \Delta)\log(\frac{\sin\alpha_{1,\Delta}}{\sin\alpha_{2,\Delta}}) \\
&= \sum_{i,\Delta}(A_i)_{e,\Delta}\Re\log z_{i,\Delta}
\end{aligned}$$

Thus $\frac{\partial \mathrm{vol}}{\partial A_e^*}=0$ if and only if the modulus part of the consistency equation at e is satisfied.

Similarly, $\frac{\partial \mathrm{vol}}{\partial x_{v,i}^*}=0$ if and only if the modulus part of the completeness equation for $\gamma_{v,i}$ is satisfied.

Hence if $\alpha\in\mathcal{A}$ is a critical point of vol, then the consistency equations are satisfied, and we can construct a hyperbolic metric on M. The derivatives of every $\gamma_{v,i}$ have modulus 1. In fact, they must be 1, for otherwise the representation corresponding to the hyperbolic metric will not be discrete torsion free on $\pi_1 L(v)$. Hence the completeness equations are satisfied.

Conversely, if α comes from a solution of the consistency and completeness equations, then $\frac{\partial \mathrm{vol}}{\partial A_e^*}=0$ for all e and $\frac{\partial \mathrm{vol}}{\partial x_{v,i}^*}=0$ for all v,i. By Proposition 7.3, α is a critical point of vol. $\square$

As explained at the start of the section, Theorem 7.5 implies Theorem 7.1 when M admits a geometric triangulation.

7.3. **Geometric triangulations exist virtually.** The following theorem completes the proof of Theorem 7.1 in the general case.

Theorem 7.6 (Luo-Schleimer-Tillman)**.** *Every finite volume hyperbolic 3-manifold with a nonempty set of cusps has a finite sheeted cover $\widehat{M}$ that has a geometric triangulation.*

We refer to [LST08] or [Mar22, Theorem 15.4.11] for a proof.

7.4. **The set of volumes of hyperbolic 3-manifolds.** Let Ω be the set

$$\{\mathrm{vol}(M)\mid M \text{ is a finite volume hyperbolic 3-manifold}\}.$$

Theorem 7.7. *Ω is a well-ordered set, i.e. every subset of Ω has a minimum element.*

Theorem 7.7 follows from Theorem 7.1 and the following proposition.

Proposition 7.8. *For every fixed value $V>0$, there exists a finite collection of finite volume hyperbolic 3-manifold $M_1,...,M_N$ such that if M is a hyperbolic 3-manifold with $\mathrm{vol}(M)\le V$, then M can be obtained as some Dehn filling of some M_i.*

Proof. It is a general fact in Riemannian geometry that given ε,κ,V, the set of diffeomorphism types of complete Riemannian manifolds that have injectivity radius $\ge\varepsilon$, (norm of sectional) curvature $\le\kappa$, and volume $\le V$ is finite. A sketch of the proof goes as follows: If M is such a manifold, we can cover M with a uniformly (depending on κ and V) finite amount of geodesic ε balls. The value of κ bounds the local complexity of the dual cellulation, hence there are only finitely many diffeomorphism types that one can build out of the finitely many balls.

We now fix ε to be a Margulis constant. A generalization of the fact above, now allowing boundary, implies that there are finitely many diffeomorphism types $M_1, ..., M_N$ of thick parts of hyperbolic 3-manifolds with volume $\leq V$ is finite. Hence if M is a hyperbolic 3-manifold with volume $\leq V$, then M is diffeomorphic to some M_i union some tubes, i.e. some filling of M_i. □

Proof of Theorem 7.1. Given a subset Ω_0 of Ω, let V be an element of Ω_0. We obtain a finite collection of finite volume hyperbolic 3-manifold $M_1, ..., M_N$ from Proposition 7.8.

For each i, the set $\Omega_i = \{\mathrm{vol}(M) \mid M \text{ is a hyperbolic filling of } M_i\}$ is well-ordered: $\mathrm{vol}(M_i)$ is accumulated from above by $\mathrm{vol}(M)$ with M being the fillings of M_i with one cusp, those in turn are accumulated from above by $\mathrm{vol}(M)$ with M being the fillings of M_i with two cusps, and so on.

Thus $\bigcup_i \Omega_i$ is well-ordered and $\Omega_0 \subset \bigcup_i \Omega_i$ is well-ordered. □

From the proof of Theorem 7.7, the minimum c-fold accumulation point ω_c of Ω is the minimum volume among hyperbolic 3-manifolds with c-cusps. The values of these are known for some small values of c:

Theorem 7.9.

- *(Gabai-Meyerhoff-Milley [GMM09]) $\omega_0 \approx 0.94$, attained by the Fomenko-Matveev-Weeks manifold.*
- *(Cao-Meyerhoff [CM01]) $\omega_1 = 2v_3 \approx 2.02$, attained by the figure-eight knot complement.*
- *(Agol [Ago10]) $\omega_2 = v_8 \approx 3.66$, attained by the Whitehead link complement and the $(-2, 3, 8)$ pretzel link complement.*
- *(Yoshida [Yos13]) $\omega_4 = 2v_8 \approx 7.32$, attained by the 8^4_2 link complement.*

It is conjectured that $\omega_3 \approx 5.33$ attained by the L6a5 link complement (also known as the magic manifold) and the L8n4 link complement.

References

[Ago10] Ian Agol. The minimal volume orientable hyperbolic 2-cusped 3-manifolds. *Proc. Amer. Math. Soc.*, 138(10):3723–3732, 2010.

[Bea95] Alan F. Beardon. *The geometry of discrete groups*, volume 91 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. Corrected reprint of the 1983 original.

[Cho04] Young-Eun Choi. Positively oriented ideal triangulations on hyperbolic three-manifolds. *Topology*, 43(6):1345–1371, 2004.

[CM01] Chun Cao and G. Robert Meyerhoff. The orientable cusped hyperbolic 3-manifolds of minimum volume. *Invent. Math.*, 146(3):451–478, 2001.

[CS83] Marc Culler and Peter B. Shalen. Varieties of group representations and splittings of 3-manifolds. *Ann. of Math. (2)*, 117(1):109–146, 1983.

[GMM09] David Gabai, Robert Meyerhoff, and Peter Milley. Minimum volume cusped hyperbolic three-manifolds. *J. Amer. Math. Soc.*, 22(4):1157–1215, 2009.

[GMT03] David Gabai, G. Robert Meyerhoff, and Nathaniel Thurston. Homotopy hyperbolic 3-manifolds are hyperbolic. *Ann. of Math. (2)*, 157(2):335–431, 2003.

[Gro82] Michael Gromov. Volume and bounded cohomology. *Inst. Hautes Études Sci. Publ. Math.*, (56):5–99 (1983), 1982.

[Har92] Joe Harris. *Algebraic geometry*, volume 133 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1992. A first course.

[LST08] Feng Luo, Saul Schleimer, and Stephan Tillmann. Geodesic ideal triangulations exist virtually. *Proc. Amer. Math. Soc.*, 136(7):2625–2630, 2008.

[Mar22] Bruno Martelli. An introduction to geometric topology, 2022.

[Mey87] Robert Meyerhoff. A lower bound for the volume of hyperbolic 3-manifolds. *Canad. J. Math.*, 39(5):1038–1056, 1987.

[MR03] Colin Maclachlan and Alan W. Reid. *The arithmetic of hyperbolic 3-manifolds*, volume 219 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003.

[Rat19] John G. Ratcliffe. *Foundations of hyperbolic manifolds*, volume 149 of *Graduate Texts in Mathematics*. Springer, Cham, [2019] ©2019. Third edition [of 1299730].

[Thu97] William P. Thurston. *Three-dimensional geometry and topology. Vol. 1*, volume 35 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1997. Edited by Silvio Levy.

[Yos13] Ken'ichi Yoshida. The minimal volume orientable hyperbolic 3-manifold with 4 cusps. *Pacific J. Math.*, 266(2):457–476, 2013.