# Galois orbits in the moduli space of all triangles

Curtis T. McMullen

13 May 2023

**Abstract**

Every $a$ in the torus $A = \mathbb{R}^3/2\mathbb{Z}^3$ determines a unique spherical, Euclidean or hyperbolic triangle $T(a)$ with angles $(\pi a_i)$. In this paper we study the Galois orbits $\operatorname{Gal}(a)$ of torsion points $a \in A$, focusing on the *ramification density*

$$\rho(a) = \frac{|\{b \in \operatorname{Gal}(a) \;:\; T(b) \text{ is spherical}\}|}{|\operatorname{Gal}(a)|}.$$

We show that the closure $\overline{R}$ of the set of values of $\rho(a)$ is a countable subset of $[0,1]$, with 0 and 1 as isolated points. The *spectral gaps* at 0 and 1 lead to general finiteness statements for the classical triangle groups $\Delta(p,q,r) \subset \mathrm{SL}_2(\mathbb{R})$. For example, we obtain a conceptual proof, based on equidistribution, that the set of arithmetic triangle groups is finite.

## Contents



Research supported in part by the NSF. Typeset March 25, 2024.

# 1 Introduction

In this paper we study the orbits of the group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the moduli space of all triangles, and in particular how often the Galois conjugates of a given triangle are spherical. This analysis leads to finiteness statements for the classical hyperbolic triangle groups

$$\Delta(p, q, r) \subset \mathrm{SL}_2(\mathbb{R}),$$

including as special cases results of Takeuchi and Waterman–Maclachlan. It also illuminates the significance of certain Galois–invariant flats in moduli space, where all triangles are of the same type, and provides a geometric test for arithmeticity.

**Triangles.** Consider the torus $A = (\mathbb{R}/2\mathbb{Z})^3$. Every $a = (a_1, a_2, a_3) \in A$ determines a unique spherical, hyperbolic or Euclidean triangle $T(a)$ with angles $(\pi a_i)$, and every triangle arises for some $a$. Thus one can regard $A$ as the space of all triangles (with ordered, oriented edges; see §4).

**Galois orbits.** Let $a \in A$ be a torsion point, let $\langle a\rangle \subset A$ be the finite group it generates, and let

$$\mathrm{Gal}(a) = \{b \in A \,:\, \langle b\rangle = \langle a\rangle\}.$$

If we identify $A$ with a torus in $(\mathbb{C}^*)^3$, then the coordinates of $a$ become roots of unity, and $\mathrm{Gal}(a)$ gives its orbit under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

**Spectral gaps and ramification.** The *ramification density* of a torsion point $a \in A$ is defined by

$$\rho(a) = \frac{|\{b \in \mathrm{Gal}(a) \,:\, T(b) \text{ is spherical}\}|}{|\,\mathrm{Gal}(a)|}.$$

Let $R \subset [0, 1]$ be the set of all possible values of $\rho(a)$, and $\overline{R}$ its closure. We refer to $\overline{R}$ as the *ramification spectrum*. Let

$$\begin{aligned}\rho_H &= \inf\{\rho(a) \,:\, \rho(a) > 0\} \quad \text{and}\\ \rho_S &= \sup\{\rho(a) \,:\, \rho(a) < 1\}.\end{aligned}$$

In §5 we will show:

**Theorem 1.1** *The ramification spectrum $\overline{R}$ is a closed, countable subset of $[0, 1]$. In fact $\overline{R} \subset \mathbb{Q}$, and $0$ and $1$ are isolated points of $\overline{R}$.*

**Corollary 1.2** *We have $0 < \rho_H < \rho_S < 1$.*

The finiteness results that follow depend crucially on these *spectral gaps* at 0 and 1. As far as typical behavior is concerned, we have:

**Theorem 1.3** *Let $a_n \in A$ be a sequence of torsion points. Then either*

$$\rho(a_n) \to 1/3,$$

*or there is a proper closed subgroup of $A$ containing $a_n$ for infinitely many $n$.*

**Corollary 1.4** *The value of $\rho(a)$ is close to $1/3$ for most $a$. More precisely, for any $\epsilon > 0$, the set where $|\rho(a) - 1/3| > \epsilon$ is contained in a finite union of proper closed subgroups of $A$.*

The value of $1/3$ is to be expected: if $a \in A$ is chosen at random, the probability that $T(a)$ is spherical is $1/3$ (see Figure 1).

**Equidistribution.** The proofs of the results above rely on a general equidistribution theorem (§2); in the case at hand, it can be stated as follows. Let $\overline{a}$ denote the uniform probability measure on $\mathrm{Gal}(a) \subset A$, and let $\overline{B}$ denote normalized Haar measure on a subtorus $B \subset A$. These measures combine to give a natural measure $\overline{a} + \overline{B}$ on $\mathrm{Gal}(a) + B$.

**Theorem 1.5** *Let $a_n$ be a sequence of torsion points in $A$. Then after passing to a subsequence, we have*

$$\overline{a}_n \to \overline{a} + \overline{B}$$

*for some torsion point $a$ and subtorus $B$. Moreover, $a_n \in \mathrm{Gal}(a) + B$ for all $n \gg 0$.*

For variants of this result which include points of small height, see e.g. [BG, §4]. The existence of a spectral gap follows from the final statement above, which is special to roots of unity (points of height zero).

**Triangle groups.** Next we discuss applications to classical triangle groups. We will use the shorthand $1/(p, q, r)$ for $(1/p, 1/q, 1/r)$.

Let $p, q, r > 0$ be integers such that $1/p + 1/q + 1/r < 1$, and let $a = 1/(p, q, r)$. Then $T(a)$ is a hyperbolic triangle. The associated reflection group acting on $\mathbb{H}$ contains, with index two, the cocompact *triangle group*

$$\Delta = \Delta(p, q, r) \subset \mathrm{SL}_2(\mathbb{R}),$$

which is well–defined up to conjugation. The group $\Delta$ naturally lives inside the unit group of the ring

$$B = \mathbb{Q}[\Delta] \subset \mathrm{M}_2(\mathbb{R}),$$

which is a quaternion algebra over the trace field

$$K = \mathbb{Q}(\operatorname{tr} g \ : \ g \in \Delta) = \mathbb{Q}(\cos(\pi/p), \cos(\pi/q), \cos(\pi/r)).$$

Note that $K$ is totally real.

At each infinite place $v$ of $K$, $B \otimes_K K_v$ is either a division algebra or a matrix algebra. In the former case, $B$ is *ramified* at $v$, and $\Delta$ is realized as a subgroup of SU(2); while in the latter case, $B$ *splits* at $v$, and $\Delta$ is realized as a subgroup of $\mathrm{SL}_2(\mathbb{R})$. As we will see in §7, the ramified places of $B$ correspond to the spherical triangles in the Galois orbit of $T(a)$, and hence

$$\rho(1/(p,q,r)) = \frac{|\text{infinite places } v \text{ where } B \text{ is ramified}|}{|\text{all infinite places } v \text{ of } K|}. \tag{1.1}$$

By Corollary 1.14 below, there are only finitely many $(p, q, r)$ such that $\rho(1/(p, q, r)) = 0$. Thus Corollaries 1.2 and 1.4 imply:

**Corollary 1.6** *For all but finitely many $(p, q, r)$, the number of infinite places of $K$ where $B$ is ramified is comparable to the number of places where it splits. More precisely, we have:*

$$0 < \rho_H \leq \rho(1/(p, q, r)) \leq \rho_S < 1.$$

**Corollary 1.7** *For most values of $(p, q, r)$, the quaternion algebra $B$ is ramified at approximately $1/3$ of the infinite places of $K$.*

**Example.** Letting $a = 1/(p, q, r)$, we have verified that $|\rho(a) - 1/3| < 1/10$ for more than 89% of the triangle groups $\Delta(p, q, r)$ with $p, q, r \leq 100$.

**Finiteness results.** Note that $[K : \mathbb{Q}] \to \infty$ as $\max(p, q, r) \to \infty$. By Corollary 1.6, any condition that forces the proportion of ramified places to go to zero or one as $[K : \mathbb{Q}] \to \infty$ defines a finite set of triangle groups. In particular we have (§7):

**Corollary 1.8 (Takeuchi)** *There are only finitely many arithmetic triangle groups.*

**Corollary 1.9 (Waterman – Maclachlan)** *There are only finitely many totally hyperbolic triangle groups.*

Here $\Delta$ is *totally hyperbolic* if $B$ is unramified at all infinite places of $K$.

In the first case, all 76 cocompact arithmetic triangles groups are known; see [Tak, Theorem 3], [MR, §13.3]. In the second case, covered by [WM, Theorem 4], the corresponding problem is open; see Conjecture 1.15 below. In both cases the original proofs of finiteness are more intricate, and different in spirit, from the proof based on equidistribution we present here.

**Geometry of moduli space.** We now describe the moduli space of all triangles in more detail, and analyze the locus where $\rho(a) = 0$ or 1. The main results are Theorems 1.10 and 1.12 below. The first plays an important role in the proofs of Theorems 1.1 and 1.3, and the second is critical for applications to triangle groups such as Corollary 1.6. Theorem 1.11 gives a criterion for arithmeticity in terms of moduli space.

To state these results, it is useful to regard $\mathcal{T} = \mathbb{R}^3$, the universal cover of $A$, as an analogue of Teichmüller space. Let

$$L = \left\{ a \in \mathbb{Z}^3 \ : \ \sum_1^3 a_i = 0 \bmod 2 \right\} \tag{1.2}$$

denote the $D_3$ or checkerboard lattice in $\mathbb{R}^3$, and let

$$\|a\|_1 = \sum_1^3 |a_i|$$

denote the $L^1$–norm. In §4 we will show:

**Theorem 1.10** *For any $a \in \mathbb{R}^3$, the triangle $T(a)$ is hyperbolic, Euclidean or spherical according to whether the $L^1$ distance from $a$ to $L$, given by*

$$\|a - L\|_1 = \inf_{b \in L} \|a - b\|_1,$$

*is $< 1$, $= 1$ or $> 1$.*

Using the result above, the partition of $\mathcal{T}$ into spherical, Euclidean and hyperbolic triangles,

$$\mathcal{T} = \mathbb{R}^3 = S \cup E \cup H, \tag{1.3}$$

can be readily described as follows.

1. The Euclidean locus $E \subset \mathbb{R}^3$ is the union of countably many planes of the form $a_1 \pm a_2 \pm a_3 = n$, where $n \in \mathbb{Z}$ is odd.

2. The hyperbolic locus $H \subset \mathbb{R}^3$ is the union of countable many open, regular octahedra, namely the unit $L^1$-balls centered at the points of $L$.

3. The spherical locus $S \subset \mathbb{R}^3$ is related to the standard tiling of $\mathbb{R}^3$ by unit cubes of the form $[0,1]^3 + p$, $p \in \mathbb{Z}^3$. It consists of countably many open tetrahedra with vertices in $\mathbb{Z}^3 - L$, one inscribed in each such cube (see Figure 1).

The closures of the components of $H$ and $S$ give a tiling of $\mathbb{R}^3$ by regular octahedra and tetrahedra, with edges of length $\sqrt{2}$. Two tiles of the same type ($S$ or $H$) can only meet along an edge or at a vertex.

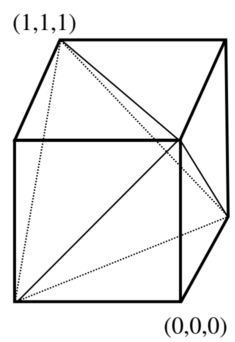


Figure 1. The spherical locus in $[0,1]^3$ is the interior of an inscribed tetrahedron.

**The mapping–class group.** Next we describe the analogue of the mapping–class group $\mathrm{Mod}(\mathcal{T})$ acting on $\mathcal{T} = \mathbb{R}^3$.

Let $U \subset \mathrm{O}(3)$ be the symmetry group of the octahedron defined by $\sum |a_i| \leq 1$. This group of order 48, isomorphic to $(\pm 1)^3 \rtimes S_3$, acts on $\mathbb{R}^3$ by sign changes and coordinate permutations. Let $L$ act on $\mathbb{R}^3$ by translations, and let

$$\mathrm{Mod}(\mathcal{T}) = L \rtimes U.$$

This is the largest group of affine transformations of $\mathbb{R}^3$ preserving the partition $H \cup E \cup S$.

**Moduli space.** The function $\rho(a)$ lifts from $A_{\mathrm{tor}}$ to a $\mathrm{Mod}(\mathcal{T})$–invariant function on $\mathbb{Q}^3$, which we denote by the same symbol. Similarly $T(a)$ lifts to $\mathbb{R}^3$, and for all $g \in \mathrm{Mod}(\mathcal{T})$, the triangles $T(a)$ and $T(g \cdot a)$ differ only in the ordering and orientation of their edges. Thus we can regard

$$\mathcal{M} = \mathcal{T} / \operatorname{Mod}(\mathcal{T}) \tag{1.4}$$

as the *moduli space* of all triangles.

**Normalized angle data.** We can also identify $\mathcal{M}$ with a compact convex subset of $\mathbb{R}^3$, namely the space of *normalized angles*:

$$\mathcal{M} \cong \{a \in \mathbb{R}^3 \,:\, 0 \leq a_1 \leq a_2 \leq a_3 \quad \text{and} \quad a_2 + a_3 \leq 1\}. \tag{1.5}$$

Every orbit of $\operatorname{Mod}(\mathcal{T})$ meets this copy of $\mathcal{M}$ in a unique point, so it also forms a natural fundamental domain for the mapping–class group.

**Arithmeticity.** We remark that the natural projection $A \to \mathcal{M}$ has degree

$$192 = |L/2\mathbb{Z}^3| \times |(\pm 1)^3| \times |S_3|.$$

In (§7) we will show:

**Theorem 1.11** *The group $\Delta(p, q, r)$ is arithmetic iff there is a unique hyperbolic triangle in the projection of* $\operatorname{Gal}(a)$ *to $\mathcal{M}$, where $a = 1/(p, q, r)$.*

For example, $\Delta(2, 3, 7)$ is arithmetic because its 'Galois conjugates' $\Delta(2, 3, 7/2)$ and $\Delta(2, 3, 7/3)$ are spherical. It is easy to recover the list of cocompact arithmetic triangle groups using Theorem 1.11, once one knows that $p, q, r \leq 30$ for all such groups. (The criterion also applies when $p, q$ or $r = \infty$.)

**Exceptional flats.** A *flat* $F = a + B \subset A$ is the translate of a closed, connected subgroup $B \subset A$. Similarly, a flat $F = a + V \subset \mathbb{R}^3$ is the translate a vector subspace $V$.

We can now finally describe the loci where $\rho(a) = 0$ and $\rho(a) = 1$. Consider the following flats in $\mathbb{R}^3$:

- $S_1 =$ the line defined by $a_2 = a_3 = 1/2$;
- $H_2 =$ the plane defined by $a_1 = 0$; and
- $H_1 =$ the line through $(0, 1/2, 1/2)$ and $(-1/2, 0, -1/2)$.

It is easy to see that $\rho(a) = 0$ for all $a \in E$ (all Galois conjugates remain Euclidean; cf. Theorem 6.2). The remaining cases are covered by:

**Theorem 1.12** *There exist finite sets $H_0$ and $S_0$ in $\mathcal{M} \subset \mathbb{R}^3$ such that for all $a \in \mathbb{Q}^3 - E$,*

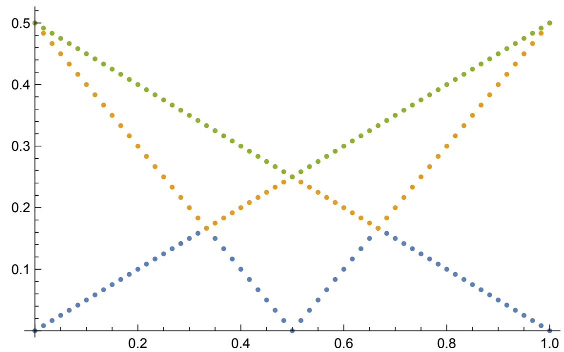


Figure 2. The projection of $H_1$ to moduli space $\mathcal{M}$.

1. *We have* $\rho(a) = 1$ *iff* $a \in \mathrm{Mod}(\mathcal{T}) \cdot (S_0 \cup S_1)$*; and*

2. *We have* $\rho(a) = 0$ *iff* $a \in \mathrm{Mod}(\mathcal{T}) \cdot (H_0 \cup H_1 \cup H_2)$*.*

**Remarks.**

1. The set $S_0$ is given explicitly in §6, but at present an explicit description of $H_0$ is unknown; cf. Conjecture 1.15 below.

2. We have $\rho(a) = 1$ for $a \in S_1 - E$ because a hyperbolic triangle (of finite area) can never have two right angles, and this property is Galois invariant.

3. Similarly, we have $\rho(a) = 0$ for $a \in H_2$ because a spherical triangle can never have an angle of zero.

4. The locus $H_1$ is more subtle. It corresponds, in $\mathbb{R}^3$, to a segment joining two points on skew edges of the octahedron $P$ defined by $\|a\|_1 \leq 1$. (See Figure 5 in §6.) The projection of this segment to $\mathcal{M}$ is given by a family of normalized triples $(a_1(t), a_2(t), a_3(t))$, $t \in [0, 1]$; the piecewise–linear functions $a_i(t)$ are graphed in Figure 2. Since $\sum a_i(t) < 1$ except at the endpoints, the line $H_1$ is essentially contained in $H$.

The flats $S_1, H_1$ and $H_2$ in $\mathbb{R}^3$ each cover a flat in $A = \mathbb{R}^3/2\mathbb{Z}^3$. Using the fact that $2\mathbb{Z}^3$ has finite index in $\mathrm{Mod}(\mathcal{T})$, we readily obtain:

**Corollary 1.13** *The locus in* $A$ *where* $\rho(a) = 0$ *or* $1$ *is a finite union of flats.*

**Corollary 1.14** *There are only finitely many* $(p,q,r)$ *such that* $\rho(1/(p,q,r)) = 0$.

**Proof.** By Theorem 1.12, any torsion point satisfying $\rho(a) = 0$ is accounted for by $H_0$, $H_1$ or $H_2$. Since $p,q,r$ are finite, their reciprocals are never 0, so we have no cases accounted for by $H_2$. Examining Figure 2, we find there is 1 case (up to permutation) accounted for by $H_1$, namely $(p,q,r) = (3,6,6)$; and there are finitely many additional cases accounted for by $H_0$. ∎

**Questions.**

1. What is the value of $\rho_H$? We note that $0 < \rho_H \leq \rho((9,11,19)/35) = 1/12$.
2. What is the value of $\rho_S$? We note that $1 > \rho_S \geq \rho(1/(2,3,11)) = 4/5$. The latter parameter corresponds to an arithmetic triangle group. A concrete value of $\rho_S$ would make our proof of Corollary 1.8 effective.
3. Is the ramification spectrum $\overline{R}$ homeomorphic to the ordinal $\omega^3 + 1$? See Corollary 3.4 for a result in this direction.

We conclude with a conjecture that motivated the present paper.

**Conjecture 1.15** *There are exactly 11 totally hyperbolic triangle groups. They are given by* $(p,q,r) =$

$$\begin{array}{cccccc} (2,4,6), & (2,6,6), & (3,4,4), & (3,6,6), & (2,6,10), & (3,10,10), \\ (5,6,6), & (6,10,15) & (4,6,12), & (6,9,18), & \textit{and} & (14,21,42). \end{array}$$

These groups are studied in detail in a sequel [Mc]. The first four examples are arithmetic and commensurable; the next three examples are also commensurable; and for the last three examples, $a = 1/(p,q,r)$ is proportional to $(1,2,3)$. (We note that the line $\mathbb{R} \cdot (1,2,3)$ meets $H$ in a set of unusually high density, namely $5/6$.)

It is readily verified, using Corollary 7.3 below, that all $(p,q,r)$ above give totally hyperbolic triangle groups, This list also appears in [WM].

**Acknowledgements.** I would like to thank Matt Baker and Alan Reid for many useful conversations. In particular Reid pointed out the reference [WM].

## 2 Galois orbits

In this section we prove an elementary equidistribution result for the Galois orbits of torsion points on a torus. Theorem 1.5 is a special case of Theorem 2.1 below.

**Compact groups and measures.** Let $A$ be a compact topological group. We say $A$ is a *torus* if it is isomorphic to $\mathbb{R}^N/\mathbb{Z}^N$ for some $N$. A element $a \in A$ is *torsion*, of order $n$, if it generates a finite subgroup $\langle a\rangle \cong \mathbb{Z}/n$ in $A$. We let $A_{\mathrm{tor}}$ denote the set of torsion points in $A$.

The *Galois orbit* of a torsion point is given by

$$\mathrm{Gal}(a) = \{b \in A \,:\, \langle b\rangle = \langle a\rangle\}.$$

It carries a natural uniform probability measure $\overline{a}$, satisfying

$$\overline{a}(E) = \frac{|E \cap \mathrm{Gal}(a)|}{|\,\mathrm{Gal}(a)|}$$

for any Borel set $E$.

Although we have defined $\mathrm{Gal}(a)$ purely in terms of group theory, it also coincides with the usual Galois orbit of $a$ if we regard $A$ as a subgroup of the complex torus $(\mathbb{C}^*)^N$. For example, when $a = \zeta_n = \exp(2\pi i/n)$ in $A = S^1 \subset \mathbb{C}$, the set $\mathrm{Gal}(a)$ consists of all the primitive $n$th roots of unity; these form a single orbit under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, by irreducibility of the cyclotomic polynomial $\Phi_n(z)$. The corresponding measure is given by

$$\overline{\zeta_n} = \frac{1}{\phi(n)} \sum_{\gcd(i,n)=1} \delta_{\zeta_n^i},$$

where $\phi(n) = |(\mathbb{Z}/n)^*|$ is the Euler $\phi$–function and $\delta_x$ is the measure of mass one supported on $x$.

**Pushforward.** The following useful fact is readily verified: if $\pi : A \to B$ is a homomorphism between tori, and $a \in A$ is torsion, then so is $b = \pi(a)$ and

$$\pi_*(\overline{a}) = \overline{\pi(a)} = \overline{b}. \tag{2.1}$$

**Galois flats.** A *Galois flat* is a subset of $A$ of the form

$$G = \mathrm{Gal}(a) + B,$$

where $B$ is a subtorus of $A$. We can always choose $a$ such that $\langle a\rangle \cap B = (e)$. We note that torsion points are dense in $G$, $G$ generates the subgroup $\langle a\rangle + B$, and $G$ is Galois invariant, in the sense that

$$a \in G \cap A_{\mathrm{tor}} \implies \mathrm{Gal}(a) \subset G.$$

**Primitive measures.** Every Galois flat $G = \mathrm{Gal}(a) + B$ carries a natural *primitive measure*, which we denote by

$$m = \overline{a} + \overline{B}.$$

Here $\overline{B}$ is the unique translation–invariant measure on $B$, and $m$ is the pushforward of $\overline{a} \times \overline{B}$ to $G$ under addition.

**Equidistribution.** The next result describes the distribution of Galois orbits in $A$. Recall that the space of probability measures on $A$ is compact (in the weak* topology), so it suffices to analyze limiting measures.

**Theorem 2.1** *Let $a_n \in A$ be a sequence of torsion points, and suppose the measures $\overline{a}_n$ converge. Then:*
*(i) There exists a primitive measure such that $\overline{a}_n \to \overline{a} + \overline{B}$, and*
*(ii) $a_n \in \mathrm{Gal}(a) + B$ for all $n \gg 0$.*

The existence of spectral gaps will follow from part (ii).

**Corollary 2.2** *Let $a_n \in A$ be a sequence of torsion points. Then either (i) there is a closed subgroup of $A$ that contains $a_n$ for infinitely many $n$; or (ii) the Galois orbit of $a_n$ becomes equidistributed in $A$ as $n \to \infty$.*

By similar reasoning one can show:

**Corollary 2.3** *The set of all primitive measures on $A$ is closed, and if*

$$\overline{a}_n + \overline{B}_n \to \overline{a} + \overline{B},$$

*then $\mathrm{Gal}(a_n) + B_n \subset \mathrm{Gal}(a) + B$ for all $n \gg 0$.*

**Corollary 2.4** *The set of primitive measures on $A$ is homeomorphic to the compact, countable ordinal $\omega^N + 1$, $N = \dim A$.*

**Proof.** Let $M_i$ denote the set of primitive measures $\overline{a} + \overline{B}$ with $\dim B \geq i$. In view of Theorem 2.1, the derived set $DM_0 = \overline{M_0} - M_0$ is equal to $M_1$. A similar argument shows that $DM_i = M_{i+1}$ for every $i < N$; and $M_N$ consists of the single point $\overline{A}$. ■

**Characters.** We now turn to the proof of Theorem 2.1. The following two statements are well-known.

**Proposition 2.5** *For any $n \geq 1$, the sum of the primitive $n$th roots of unity is equal to the Möbius function $\mu(n)$.*

**Proof.** Since for $n > 1$, the sum of all roots of $z^n = 1$ is zero, the sum $\sigma(n)$ of the primitive $n$th roots of unity satisfies $\sum_{d|n} \sigma(d) = 0$; and evidently $\sigma(1) = 1$. These two identities also hold for $\mu(n)$, and determine it uniquely. ■

**Corollary 2.6** *The primitive $n$th roots of unity become equidistributed in $S^1$ as $n \to \infty$.*

**Proof.** The functions $z^k$ span a dense subset of $C(S^1)$, so it suffices to show that for $k \neq 0$, the measure $\overline{\zeta}_n$ satisfies

$$I_k(n) = \langle z^k, \overline{\zeta}_n \rangle = \int_{S^1} z^k \, \overline{\zeta}_n \to 0$$

as $n \to \infty$. First consider the case $k = 1$. Then $I_1(n)$ is just the average of the primitive $n$th roots of unity, and by Proposition 2.5, we have

$$|I_1(n)| = |\mu(n)|/\phi(n) \leq 1/\phi(n) \to 0.$$

To treat the case of general $k$, observe that by (2.1), the pushforward or $\overline{\zeta}_n$ under $z^k$ is $\overline{\zeta}_m$, where $m = n/\gcd(k, n)$. Since $\gcd(k, n) \leq k$, we have

$$|I_k(n)| = |I_1(n/\gcd(k, n))| \to 0$$

in this case as well. ■

**Proof of Theorem 2.1.** Let $\widehat{A} \cong \mathbb{Z}^N$ denote the group of characters $\chi : A \to S^1$. Passing to a subsequence, we can assume that the order of the subgroup of $S^1$ generated by $\chi(a_n)$ converges to a limit $M(\chi)$ (possibly $\infty$) for all $\chi$. Let

$$F = \{\chi \in \widehat{A} \, : \, M(\chi) < \infty\}.$$

It is clear that $M(\chi_1\chi_2) \leq M(\chi_1)M(\chi_2)$ and $M(\chi^{-1}) = M(\chi)$; thus $F$ is a subgroup of $\widehat{A}$. Moreover $M(\chi) \leq dM(\chi^d)$, so $\widehat{A}/F$ is torsion–free. Thus

$$B = F^{\perp} = \bigcap \{\ker(\chi) \, : \, \chi \in F\}$$

is a connected subtorus of $A$. Choose a complementary torus $C \subset A$. We then have

$$A = B \times C, \quad \text{and} \quad \widehat{A} = \widehat{B} \times \widehat{C},$$

with $\widehat{C} = F$.

Write $a_n = (b_n, c_n) \in B \times C$. Choose a basis $(\chi_1, \ldots, \chi_c)$ for $\widehat{C}$. Since $M(\chi_i) < \infty$ for all $i$, the order of $c_n$ is uniformly bounded above; thus we can pass to a subsequence such that $c_n = c$ is constant. Then $a_n = (b_n, c)$.

To show $m = \overline{B} \times \overline{c}$, it suffices to show that for all $\chi \in \widehat{A}$, we have

$$\langle \chi, m \rangle = \langle \chi, \overline{B} \times \overline{c} \rangle = \begin{cases} \langle \chi, \overline{c} \rangle & \text{if } \chi \in \widehat{C}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

To check this, first suppose $\chi \in \widehat{C}$. By equation (2.1), the projection of $\overline{a}_n$ to $C$ is $\overline{c}$; thus

$$\langle \chi, m \rangle = \lim \langle \chi, \overline{a}_n \rangle = \langle \chi, \overline{c} \rangle.$$

On the other hand, if $\chi \in \widehat{A} - \widehat{C}$, then $M(\chi) = \infty$; thus the order of $\langle \chi(a_n) \rangle$ in $S^1$ tends to infinity. This implies, by Corollary 2.6, that the Galois orbit of $\chi(a_n)$ becomes equidistributed, and hence

$$\langle \chi, m \rangle = \lim \langle \chi, \overline{a}_n \rangle = 0.$$

Thus $m = \overline{B} \times \overline{c}$, which is the same as the primitive measure $\overline{c} + \overline{B}$.

The same argument shows that, after passing to a subsequence, we have

$$a_n = (b_n, c) \in B \times \mathrm{Gal}(c) = \mathrm{supp}(m)$$

for all $n$. The same must be true for the original sequence, at least for all $n \gg 0$, else we could have first passed to a subsequence with $a_n \notin \mathrm{supp}(m)$ for all $n$, and no passage to a further subsequence would rectify this situation. Finally, $B \times \mathrm{Gal}(c) = \mathrm{Gal}(c) + B$. ■

**Remark.** The behavior of Galois orbits described above is analogous to the behavior of unipotent orbits in homogeneous spaces [Rn], [MS]. See also [Ric] for related results.

# 3 Spectral gaps

In this section we establish the existence of spectral gaps in a general setting.

Let $A = \mathbb{R}^n / \mathbb{Z}^n$ be a torus as in §2. The following result underlies the proofs of Theorems 1.1 and 1.3, to be given in §5.

**Theorem 3.1** *Let $U \subset A$ be an open set such that $\partial U$ is a finite union of Galois flats. Let $\overline{R} \subset [0,1]$ be the closure of $R = \{\overline{a}(U) \,:\, a \in A_{\mathrm{tor}}\}$. Then:*

1. *We have $\overline{R} \subset [0,1] \cap \mathbb{Q}$;*
2. *The points $0$ and $1$ are isolated in $\overline{R}$ (if they belong to $\overline{R}$); and*
3. *There exists a finite set of Galois flats $G_i$ such that for all $a \in A_{\mathrm{tor}}$, we have $\overline{a}(U) = 0 \iff a \in \bigcup_1^n G_i$.*

We first establish a Noetherian property for Galois flats.

**Lemma 3.2** *There exist only finitely many maximal Galois flats contained in a given closed set $X \subset A$.*

**Proof.** Suppose to the contrary. Let $Gal(a_i) + B_i$ be an infinite sequence of distinct maximal flats in $X$. Passing to a subsequence we can assume that $\overline{a}_i + \overline{B}_i \to \overline{a} + \overline{B}$ for $a, B$. Since $X$ is closed, this limiting measure is still supported in $X$. But we also have $\mathrm{Gal}(a_i) + B_i \subset \mathrm{Gal}(a) + B \subset X$ for all $i \gg 0$, by Corollary 2.3. This contradicts maximality. ■

**Corollary 3.3** *Let $G_1, G_2, \ldots, G_n$ be the maximal Galois flats contained in $X$. Then $\mathrm{Gal}(a) \subset X \iff a \in \bigcup_1^n G_i$.*

**Proof of Theorem 3.1.** Let $\mathcal{P}$ be the compact, countable set of primitive measures on $A$ (defined in §2). Since $\partial U$ is a finite union of Galois flats, any $m \in \mathcal{P}$ satisfies $m(\partial U) = 0$ or $1$. In the latter case $m$ is supported in $\partial U$.

Define $\rho : \mathcal{P} \to [0,1]$ by $\rho(m) = m(U)$. We claim $\rho$ is continuous. Indeed, if $m_n \to m$ in $\mathcal{P}$ and $m(\partial U) = 0$, then $m_n(U) \to m(U)$ by general principles. Otherwise, as remarked above, $m$ is supported in $\partial U$, and hence, by Corollary 2.3, the same is true of $m_n$ for all $n \gg 0$. Thus $m_n(U) \to 0 = m(U)$, and we have continuity in this case as well.

Since $\rho$ is continuous, $\rho(\mathcal{P})$ is a compact, countable set. Moreover, measures of the form $\overline{a}$ are dense in $\mathcal{P}$, so $\overline{R} = \rho(\mathcal{P})$ and thus $\overline{R}$ is countable. Finally we observe $\rho(m) \in \mathbb{Q}$ for all $m \in \mathcal{P}$. This follows from the fact that the support $\mathrm{Gal}(a) + B$ of $m$, and the open set $U$, are both locally defined by rational linear equations on $A = \mathbb{R}^n/\mathbb{Z}^n$, and

$$\rho(\overline{a} + \overline{B}) = \frac{\mathrm{vol}(U \cap (\mathrm{Gal}(a) + B))}{\mathrm{vol}(\mathrm{Gal}(a) + B)}.$$

Next let us check that 0 is isolated in $\overline{R}$. If not, we can find a sequence find $m_n \in \mathcal{P}$ such that $m_n(U) \to 0$ but $m_n(U) > 0$ for all $n$. Pass to a subsequence such that $m_n \to m$; then $m(U) = 0$, so its support is disjoint from $U$. But the support of $m_n$ is contained in the support of $m$ for all $n \gg 0$, contradicting our assumption that $m_n(U) > 0$.

The proof that 1 is isolated is similar. Suppose $m_n \in \mathcal{P}$, $0 < m_n(U) < 1$, and $m_n(U) \to 1$. As remarked above, since $m_n(U) > 0$ we have $m_n(\partial U) = 0$ for all $n$. Passing to a subsequence, we can also assume that $m_n \to m \in \mathcal{P}$ with $m(U) = 1$. Since $m$ is supported in $\overline{U}$, so is $m_n$ for all $n \gg 0$. But $m_n(\partial U) = 0$, so $m_n(U) = 1$ for all $n \gg 0$, contrary to assumption.

For the final statement, apply Corollary 3.3 to $X = A - U$. ■

In view of Corollary 2.4, the proof also shows:

**Corollary 3.4** *The set $\overline{R}$ is the continuous image of the countable compact ordinal $\omega^n + 1$, $n = \dim A$.*

## 4 Triangles

In this section we formalize the notion of a triangle, and prove Theorem 1.10. That is, we show that for every $a = (a_1, a_2, a_3) \in \mathbb{R}^3$ there exists an essentially unique spherical, Euclidean or hyperbolic triangle $T(a)$ with angles $(\pi a_i)$, and that these three alternatives correspond to

$$\|a - L\|_1 < 1, = 1 \quad \text{or} \quad > 1.$$

Along the way we will show:

**Theorem 4.1** *Given $a, b \in \mathcal{T} = \mathbb{R}^3$, we have $b \in \operatorname{Mod}(\mathcal{T}) \cdot a$ if and only if $T(a)$ is isomorphic to $T(b)$.*

This justifies the statement that $\mathcal{M} = \mathcal{T} / \operatorname{Mod}(\mathcal{T})$ can be regarded as the moduli space of all triangles.

**Geometry of lines.** To set the stage, let $\mathbb{X}$ denote a complete, simply connected surface of constant curvature $-1$, 0 or 1. Then $\mathbb{X}$ is isomorphic to the hyperbolic plane $\mathbb{H}^2$, the Euclidean plane $\mathbb{E}^2$, or the unit sphere $\mathbb{S}^2$. The *automorphism group* of $\mathbb{X}$ refers to its isometry group when $\mathbb{X} = \mathbb{H}^2$ or $S^2$; for $\mathbb{E}^2$, we also include similarities.

A *line* is a complete, oriented geodesic $L \subset \mathbb{X}$. We say $L$ and $L'$ are *parallel* if they are disjoint and, in the hyperbolic case, they meet at a point at infinity.

Given a pair of unit tangent vectors $v, v' \in T_p\mathbb{X}$, and $L$ and $L'$ the oriented lines they determine, we define the dot product of these two lines by

$$L \cdot L' = \langle v, v' \rangle = \cos\theta. \tag{4.1}$$

Here $\theta$ is the angle of a rotation about $p$ that moves $L$ to $L'$. Note that

$$(-L) \cdot L' = -(L \cdot L').$$

This definition extends in a natural way to parallel lines, with $L \cdot L' = \pm 1$ depending on orientation.

**Triangles.** A *triangle* $T = (L_1, L_2, L_3)$ is an ordered triple of oriented lines $T = (L_1, L_2, L_3)$ in $\mathbb{H}^2$, $\mathbb{E}^2$ or $S^2$. This definition is chosen to deal uniformly with different geometries. We adopt the following conventions.

1. With one exception, all three lines are required to be distinct, and every pair of lines $L_i$ and $L_j$ must either be parallel or meet in $\mathbb{X}$.

2. The exception occurs in $\mathbb{E}^2$: here, we allow a degenerate triangle in which all three lines coincide (up to orientation). This case is characterized by the condition that $|L_i \cdot L_j| = 1$ for all $i, j$.

**Quadratic forms.** For the proof of Theorem 4.3 we will relate geometry and quadratic forms. This perspective also fits well with quaternion algebras (§7).

Motivated by equation (4.1), for each $a \in \mathbb{R}^3$ we define a quadratic form on $\mathbb{R}^3$ by

$$Q(a) = \langle e_i, e_j \rangle = \begin{pmatrix} 1 & -\cos\pi a_3 & -\cos\pi a_2 \\ -\cos\pi a_3 & 1 & -\cos\pi a_1 \\ -\cos\pi a_2 & -\cos\pi a_1 & 1 \end{pmatrix}. \tag{4.2}$$

**Theorem 4.2** *We have* $\det Q(a) = 0$ *if and only if*

$$a_1 \pm a_2 \pm a_3 = 1 \bmod 2 \tag{4.3}$$

*for some choice of signs. Otherwise, the form* $Q(a)$ *has signature* $(2,1)$ *or* $(3,0)$.

**Proof.** Choose $z_i \in S^1 \subset \mathbb{C}$ such that $2\cos\pi a_i = z_i + 1/z_i$ for $i = 1, 2, 3$. We then compute

$$\det Q(a) = -\frac{(1 + z_1 z_2 z_3)(z_1 z_2 + z_3)(z_2 z_3 + z_1)(z_3 z_1 + z_2)}{4 z_1^2 z_2^2 z_3^2}.$$

The numerator vanishes exactly when $z_1 z_2^{\pm 1} z_3^{\pm 1} = -1$, which translates into condition (4.3).

To check the signature, suppose $\det Q(a) \neq 0$. Let $V_i$ be the subspace of $\mathbb{R}^3$ spanned by $(e_j, e_k)$, where $(i, j, k)$ is a permutation of $(1, 2, 3)$. Then $\det Q(a)|V_i = 1 - \cos^2 \pi a_i$. Changing $a$ slightly (without changing the signature of $Q(a)$), we can assume that $Q(a)|V_i$ has signature $(2, 0)$ for all $i$. Since the diagonal entries of $Q(a)$ are also positive, it follows that $Q(a)$ has signature $(3, 0)$ or $(2, 1)$. ■

Note that signature $(3, 0)$ and $(2, 1)$ correspond to $\det Q(a) > 0$ and $\det Q(a) < 0$ respectively.

**Theorem 4.3** *For any $a \in \mathbb{R}^3$, there exists an essentially unique triangle $T(a) = (L_1, L_2, L_3)$ such that*

$$L_i \cdot L_j = -\cos \pi a_k \tag{4.4}$$

*for any permutation $(i, j, k)$ of $(1, 2, 3)$. This triangle is hyperbolic, Euclidean or spherical depending on whether $\det Q(a)$ is $< 0$, $= 0$ or $> 0$.*

By essentially unique, we mean $T(a) \subset \mathbb{X}$ is unique up to an automorphism of $\mathbb{X}$ and an overall change of sign, $(L_1, L_2, L_3) \mapsto (-L_1, -L_2, -L_3)$. These operations clearly preserve $L_i \cdot L_j$.

**Proof.** Let $\langle x, y\rangle_Q$ be the symmetric bilinear form on $\mathbb{R}^3$ with basis $(e_1, e_2, e_3)$ defined by $Q(a)$.

First suppose $\det Q(a) > 0$. Then $Q(a)$ has signature $(3, 0)$ by Theorem 4.2, the locus $\langle x, x\rangle_Q = 1$ is isometric to the standard sphere $S^2$, and the oriented great circles

$$L_i = S^2 \cap e_i^{\perp}$$

satisfy $L_i \cdot L_j = \langle e_i, e_j\rangle_Q = -\cos \pi a_k$ by the definition of $Q(a)$. Thus $T(a) = (L_1, L_2, L_3)$ satisfies equation (4.4). Conversely, any other triple of lines with these inner products determines a basis for $\mathbb{R}^3$ with quadratic form $Q(a)$, establishing uniqueness.

The argument when $\det Q(a) < 0$ is similar. In this case $Q(a)$ has signature $(2, 1)$. Using the Minkowski model for hyperbolic space, we take $\mathbb{H}^2$ to be one component of the hyperboloid defined by $\langle x, x\rangle_Q = -1$, and let $L_i = \mathbb{H}^2 \cap e_i^{\perp}$ as before. Then $T(a) = (L_1, L_2, L_3)$ is the desired hyperbolic triangle.

Finally suppose $\det Q(a) = 0$. In this case, $R = \operatorname{Ker} Q(a)$ is typically one dimensional, and the induced form on $\mathbb{R}^3/R \cong \mathbb{R}^2$ has signature $(2, 0)$.

Then the lines $L_i = e_i^\perp \subset \mathbb{R}^2$, once displaced from the origin, give desired Euclidean triangle. When $\dim R = 2$, we obtain a degenerate Euclidean triangle. ■

**Isomorphic triangles.** Let us say triangles $T = (L_1, L_2, L_3)$ and $T' = (L'_1, L'_2, L'_3)$ in $\mathbb{X}$ are *isomorphic* if there a $g \in \operatorname{Aut} \mathbb{X}$, a permutation $\sigma \in S_3$, and a choice of signs, such that

$$(g(L'_i)) = (\pm g(L_{\sigma i})) \quad \text{for } i = 1, 2, 3.$$

We can now give the:

**Proof of Theorem 4.1.** Clearly the isomorphism class of $T(a)$ depends only on $Q(a)$, which itself is invariant under the action of the subgroup of finite index $(2\mathbb{Z}^3) \rtimes (\pm 1)^3$ in $\operatorname{Mod}(\mathcal{T})$. Note that reversing the orientation of $L_1$ has the effect of negating $L_1 \cdot L_2$ and $L_1 \cdot L_3$, which is the same as adding $(0, 1, 1)$ to $a$, since $\cos(\pi + x) = -\cos(x)$. Thus $\mathbb{R}^3/(L \rtimes (\pm 1))$ gives the moduli space of triangles with unoriented edges, and taking the further quotient by $S_3$ removes their ordering, leaving only the isomorphism class of $T(a)$. ■

**The cube model.** Let $P \subset [0, 1]^3$ be the inscribed tetrahedron with vertices $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ and $(1, 1, 1)$. To help visualize the partition $\mathbb{R}^3 = H \cup E \cup S$ (see Figure 1), we will show:

**Theorem 4.4** *Given $a \in [0, 1]^3$, the triangle $T(a)$ is Euclidean iff $a \in \partial P$. It is spherical if $a$ lies in the interior of $P$, and otherwise hyperbolic.*

**Proof.** The Euclidean case follows from equation (4.3), and the remaining statements can be verified by computing $\det Q(a)$ at the center and vertices of the cube. ■

**Proof of Theorem 1.10.** In view of the fact that the partition $S \cup E \cup H$, the $L^1$ metric on $\mathbb{R}^3$, and the lattice $L$ are all invariant under $\operatorname{Mod}(\mathcal{T})$, its suffices to verify that the description of $S \cup E \cup H$ in terms of $\|a - L\|_1$ is correct on the unit cube, which we have just done. ■

**Example: Spherical isosceles triangles.** Here is an example that illustrates the transition from hyperbolic to spherical geometry and back again. Fix a small number $s > 0$, and consider the family of isosceles triangles $T(t, t, s)$, $t \in [0, 1]$. Let $t_0 = (1 - s)/2$ and $t_1 = (1 + s)/2$. The triangle

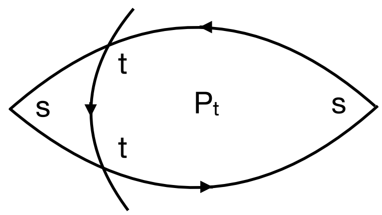


Figure 3. A family of spherical isosceles triangles in the lune $L_s$.

$T(t,t,s)$ is hyperbolic in the range $[0,t_0)$ and $(t_1,1]$; it is spherical in the range $(t_0,t_1)$; and it is Euclidean at the endpoints $t_0$ and $t_1$.

The most interesting behavior occurs in the spherical range. For $t \in (t_0,t_1)$, $T(t,t,s)$ determines a spherical triangle $P_t$, cut out by a great circle crossing a lune $L_s$ of angle $s$; see Figure 3. As $t$ approaches $t_1$ from below, the triangle $P_t$ expands to fill the whole lune; as it does, the shrinking complementary triangle $L_s - P_t$ can be rescaled to yield a Euclidean limit. This limit $T(t_1,t_1,s)$ serves to connect the spherical and hyperbolic regimes.

Note that for $t > t_1$, there is no embedded spherical triangle with internal angles $(t,t,s)$, since (by Gauss–Bonnet) its area would exceed that of the lune $L_s$. Instead, $T(t,t,s)$ gives a hyperbolic triangle with internal angles $(s,1-t,1-t)$, and the usual orientation of one edge flipped.

**Moduli space, *reprise*.** We remark that the embedding $\mathcal{M} \subset \mathbb{R}^3$ given by equation (1.5) provides each isomorphism class of triangle with a canonical representative $T(a) = (L_1,L_2,L_3)$. This triangle has *internal* angles $(\pi a_1, \pi a_2, \pi a_3)$, and it is hyperbolic, Euclidean or spherical depending on whether $\sum a_i < 1$, $= 1$ or $> 1$. In the spherical case, the great circles $(L_1,L_2,L_3)$ cut $S^2$ into 8 triangular regions, and among these $T(a)$ has the smallest area.

**Remark: pairs of pants.** One can similarly prove that there exists a unique hyperbolic pair of pants with boundary geodesics of given lengths $\ell_i \geq 0$, $i = 1,2,3$, by replacing the off–diagonal entries of $Q(a)$ with $-\cosh(\ell_i/2)$; cf. [Th, p.83, p.263].

# 5 Orbits of triangles

In this section we prove our main results on Galois orbits and spectral gaps, namely Theorems 1.1 and 1.3 and their corollaries. The proofs combine the equidistribution results from §2 and §3 with the geometric properties of the partition

$$\mathcal{T} = \mathbb{R}^3 = H \cup E \cup S \tag{5.1}$$

established in §4.

**Partition of $\boldsymbol{A}$.** Since the partition (5.1) is invariant under $\mathrm{Mod}(\mathcal{T})$, it descends to a partition

$$A = A(H) \cup A(E) \cup A(S).$$

Here $A(H)$ and $A(S)$ are open, and $A(E)$ is the closed set defined by

$$a_1 \pm a_2 \pm a_3 = 1 \bmod 2.$$

We observe that

$$\frac{\operatorname{vol} S}{\operatorname{vol} A} = \frac{1}{3}, \tag{5.2}$$

as can be seen geometrically in Figure 1. Clearly $A(E)$ is a finite union of Galois flats; in fact $A(E) = (1,1,1) + B'$, where

$$B' = \{a \ : \ a_1 \pm a_2 \pm a_3 = 0 \bmod 2\}$$

is the union of four subtori in $A$.

**Proof of Theorem 1.1 and Corollary 1.2.** Note that $\rho(a) = \overline{a}(A(S))$ for all $a \in A_{\mathrm{tor}}$. By Theorem 1.10 we have $\partial A(S) = \partial A(H) = A(E)$, and as just remarked, $A(E)$ is a finite union of Galois flats. Thus we can apply Theorem 3.1 with $U = A(S)$ to deduce that the ramification spectrum $\overline{R}$ is a closed subset of $\mathbb{Q} \cap [0,1]$ with 0 as an isolated point. To show that 1 is also an isolated point, apply the same argument with $U = A(H)$.

Corollary 1.2 is immediate. ∎

**Proof of Theorem 1.3 and Corollary 1.4.** Let $a_n$ be a sequence of torsion points in $A$. Assume that no subsequence lies in a proper, closed subgroup of $A$. Then $\overline{a}_n \to \overline{A}$ by Corollary 2.2. Since $\overline{A}(\partial A(S)) = 0$, we also have $\rho(a_n) = \overline{a}_n(A(S)) \to \overline{A}(A(S))$, and $\overline{A}(A(S)) = 1/3$ by equation (5.2) above.

To prove Corollary 1.4 we use the fact that the proper closed subgroups of $A$ form a countable set, say $\{A_1, A_2, A_3, \ldots\}$. Let $X = \{a \in A_{\mathrm{tor}} : |\rho(a) - 1/3| > \epsilon\}$, and suppose $X$ is *not* contained in $\bigcup_1^n A_i$ for any $n$. Choose $a_n \in X - \bigcup_1^n A_i$. Then $a_n \notin A_i$ for all $n \gg 1$, so $\rho(a_n) \to 1/3$, as we have just seen. This contradicts the definition of $X$. Thus $X$ is contained in $\bigcup_1^n A_i$ for some $n$. ∎

## 6 Exceptional flats

In this section we prove Theorem 1.12, characterizing the locus in $A$ where $\rho(a) = 0$ or 1.

By Theorem 1.10, the partition $\mathbb{R}^3 = H \cup E \cup S$ is invariant under $\operatorname{Mod}(\mathcal{T})$. We say a flat $F = a + V \subset \mathbb{R}^3$ of dimension $d$ is *essentially* contained in $S$ if $F - S$ has $d$–dimensional measure zero, and similarly for $H$.

**Theorem 6.1** *Let $F \subset \mathbb{R}^3$ be a flat of dimension* 1 *or more. Suppose that $F$ is maximal among flats essentially contained in $S$ or among flats essentially contained in $H$. Then up to the action of* $\operatorname{Mod}(\mathcal{T})$*, $F$ is either:*

1. *The line $S_1$ defined by $a_2 = a_3 = 1/2$;*
2. *The plane $H_2$ defined by $a_1 = 0$, or*
3. *The line $H_1$ passing through* $(0, 1/2, 1/2)$ *and* $(-1/2, 0, -1/2)$ *in* $\mathbb{R}^3$.

This classification is the main step in the proof of Theorem 1.12. The notation $S_1, H_1$ and $H_2$ is the same as in §1.

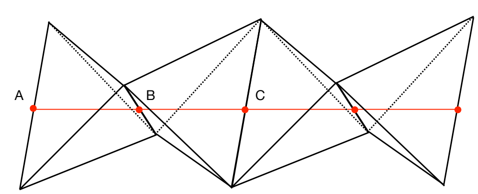


Figure 4. A chain of adjacent tetrahedra in $\overline{S}$.

**Proof.** We begin with the spherical case. Let $F$ be a maximal flat essentially contained in $S$. Let $P$ be the tetrahedron inscribed in $[0, 1]^3$ with vertices $(1, 1, 1)$, $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. By Theorem 1.10, $\overline{S} = \operatorname{Mod}(\mathcal{T}) \cdot P$, so up to the action of $\operatorname{Mod}(\mathcal{T})$, we can assume that $F$ meets the interior of $P$. Since the faces of $P$ are all adjacent to cells of $H$ (see Figure 1), $F \cap \partial P$ must be contained in the edges of $P$. This is impossible if $\dim(F) = 2$, so $\dim(F) = 1$.

Since the line $F$ meets the interior of $P$, it cannot pass through any vertex of $P$ — if it did, it would also pass through the opposite face. Similarly, $F$ cannot meet two adjacent edges of $P$ — if it did, then it would lie in the face they span and not meet the interior of $P$.

Thus $F$ must be the unique line joining a pair of points $A$ and $B$ on opposite edges of $P$. There is a unique tetrahedron $P_1$ in $\overline{S}$ sharing the edge of $P_0 = P$ containing $B$. By the same reasoning, $F$ joins $B$ to a point on the opposite edge of $P_1$. Continuing this way, we find that $F$ passes through chain of adjacent tetrahedra, $P = P_0, P_1, P_2, \ldots$, meeting edge to edge. Such a chain must be parallel to one of the coordinate axis in $\mathbb{R}^3$, as can be seen by considering the chain of circumscribed cubes. The same is true of $F$, and hence of the vector $A - B$ (see Figure 4). Thus up to the action of the stabilizer of $P$ in $\mathrm{Mod}(\mathcal{T})$, we may assume that $A - B = (1, 0, 0)$, and hence $A = (0, 1/2, 1/2, )$, $B = (1, 1/2, 1/2)$, and $F = S_1$.

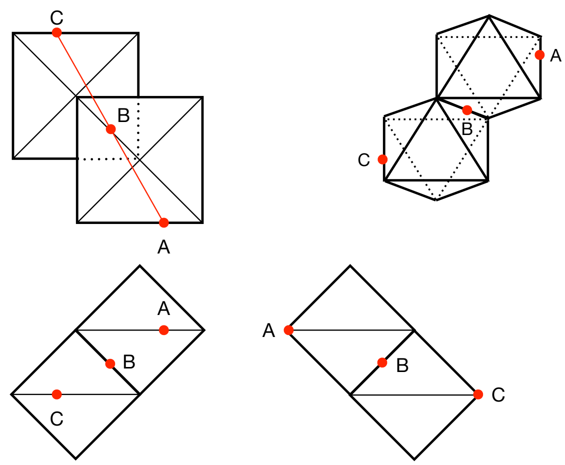


Figure 5. A line segment passing through two adjacent octahedra. Top, front, side and perspective views are shown.

We now turn to the hyperbolic case. Let $F$ be a maximal flat essentially contained in $H$, and let $P$ be the octahedron defined by $\|a\|_1 \leq 1$. By Theorem 1.10, $\overline{H} = \mathrm{Mod}(\mathcal{T}) \cdot P$. We may assume that $F$ meets the interior of $P$, and, by the same reasoning as in the spherical case, that $F$ only meets the edges of $\partial P$, not the faces.

If $\dim F = 2$, then $F$ must contain any edge of $P$ that it meets, and it follows quickly that $F$ coincides with one of the coordinate planes. Up to the action of $\mathrm{Mod}(\mathcal{T})$, we can assume this plane is $H_2$.

Now assume $\dim F = 1$. Since $F$ is maximal, it does not lie in any coordinate plane. Thus it does not pass through any of the vertices of $P$.

Therefore it joins a pair of points $A$ and $B$ in the interiors of different edges of $P$. These edges cannot lie in a common plane. Thus we can assume, up to the action of $\mathrm{Mod}(\mathcal{T})$, that $A$ and $B$ lie in the configuration shown in Figure 5.

There is then a unique octahedron $P_1$, adjacent to $P_0$ at $B$, such that $F$ joins a pair of points $B$ and $C$ on disjoint, skew edges of $\partial P_1$. As can be seen from the diagram, the edge of $P_1$ on which $C$ lies is uniquely determined by the edges of $P_0$ containing $A$ and $B$. By the same reasoning, these first two edges uniquely determine an infinite chain of octahedra $P_0, P_1, P_2, \ldots$ containing $F$, which in turn uniquely determines the direction of $F$ and hence the vector $A - B$. Up to the symmetries of $P_0$, we find that $A - B = (1/2, 1/2, 1)$ and $F = H_1$. ■

**From flats in $\mathbb{R}^3$ to flats in $A$.** Next we verify that $S_1$, $H_1$ and $H_2$ give rise to finite unions of Galois flats in $A$. To make this precise, let

$$\pi : \mathcal{T} = \mathbb{R}^3 \to A = \mathbb{R}^3 / 2\mathbb{Z}^3$$

be the natural projection, and let

$$A(X) = \pi(\mathrm{Mod}(\mathcal{T}) \cdot X) \subset A$$

for any subset $X$ in $\mathbb{R}^3$.

**Theorem 6.2** *For $X = E, H_1, H_2$ or $S_1$, the locus $A(X) \subset A$ is a finite union of Galois flats.*

**Proof.** It will be convenient to describe flats in $A$ using the coordinates $z = (z_1, z_2, z_3)$ defined by $z_j = \exp(\pi i a_j)$. Note that for any integers $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, the locus in $A$ defined by

$$z^\alpha = z_1^{\alpha_1} z_2^{\alpha_2} z_3^{\alpha_3} = -1$$

is a finite union of Galois flats. Indeed, it is the preimage of the Galois flat $\mathrm{Gal}(-1) \subset \mathbb{C}^*$ under the homomorphism $z \mapsto z^\alpha$. We also note that the intersection of two Galois flats is again a Galois flat, provided it is nonempty.

We begin with $A(E)$. Since $E$ is characterized by the condition $a_1 \pm a_2 \pm a_3 = 1 \bmod 2$, $A(E) \subset A$ is the union of the four Galois flats defined by

$$z_1 z_2^{\pm 1} z_3^{\pm 1} = -1.$$

(This case was also treated in §5.)

Similarly, we claim that $A(H_2)$ is the union of the Galois flats defined by $z_j^2 = 1$, $j = 1, 2, 3$. Indeed, $H_2$ itself is defined by the condition $a_1 = 0$, so $\pi(H_2)$ is the subgroup of $A$ defined by $z_1 = 1$. The locus $\pi(L + H_2)$ includes the additional Galois flat defined by $z_1 = -1$. Indeed, adding a nonzero element of the lattice $L$ just changes $z_j$ to $-z_j$ for two values of $j$. Thus $\pi(L + H_2)$ is defined by $z_1^2 = 1$. Permuting coordinates, we obtain the stated description of $A(H_2)$.

Next, we show that $A(S_1)$ is the union of Galois flats defined by the condition that $z_j^2 = z_k^2 = -1$ for some pair of indices $j \neq k$. Indeed, $S_1$ is defined by $a_2 = a_3 = 1/2$, so $\pi(S_1)$ is the torus translate defined by $z_2 = z_3 = i$. The locus $\pi(S_1 + L)$ is defined by $z_2 = \pm z_3 = \pm i$, or equivalently $z_2^2 = z_3^2 = -1$, and the action of $S_3 \subset \mathrm{Mod}(\mathcal{T})$ permutes these coordinates to give the stated description of $A(S_1)$.

Finally, $H_1$ is the line defined by $(0, 1/2, 1/2) + t \cdot (1, 1, 2)$, $t \in \mathbb{R}$, and thus

$$\pi(H_1) = \{z = (s, is, is^2) \ : \ s \in S^1\}.$$

As in the preceding cases, we find that $\pi(H_1 + L)$ is defined by the two conditions $z_1^2 = -z_2^2$ and $z_1 z_2 = z_3$, and hence it, along with $A(H_1)$, is a finite union of Galois flats as well. ■

**Proof of Theorem 1.12.** Let us first describe the locus $\rho(a) = 0$. Let $G_i \subset A$ be the union of the maximal Galois flats of dimension $i$ contained in the closed set $A - A(S)$, for $i = 0, 1, 2$. By Lemma 3.2 and Corollary 3.3, each $G_i$ is a finite union of Galois flats, and $\rho(a) = 0$ iff $a \in A_{\mathrm{tor}} \cap (\bigcup G_i)$. By the preceding result, we have $A(E) \cup A(H_1) \cup A(H_2) \subset G_1 \cup G_2$. On the other hand, $\pi^{-1}(G_1 \cup G_2)$ is a union of flats of dimension $\geq 1$ contained in $\mathbb{R}^3 - S$, which are in turn contained in $\mathrm{Mod}(\mathcal{T}) \cdot (H_1 \cup H_2)$ by Theorem 6.1. Thus we have

$$\bigcup G_i = G_0 \cup A(H_1) \cup A(H_2).$$

Let $H_0 = \pi^{-1}(G_0) \cap \mathcal{M} \subset \mathbb{R}^3$. Then $H_0$ is finite since $G_0$ is finite, and $A(H_0) = G_0$. Thus for $a \in \mathbb{Q}^3$ we have

$$\rho(a) = 0 \iff \pi(a) \in A(H_0 \cup H_1 \cup H_2) \iff a \in \mathrm{Mod}(T) \cdot (H_0 \cup H_1 \cup H_2),$$

as stated in Theorem 1.12. The proof for $\rho(a) = 1$ is similar, taking into account the fact that for torsion points a in the closed set $A - A(H)$, either $\rho(a) = 1$ or $a \in A(E)$. ■

**The totally spherical locus.** It is straightforward to check that in the statement of Theorem 1.12, we can take

$$S_0 = \left\{ \begin{array}{lll} (1/5,1/5,2/3) & (1/5,2/5,1/2) & (1/3,1/3,2/3) \\ (1/5,1/5,4/5) & (1/4,1/4,2/3) & (1/3,2/5,1/2) \\ (1/5,1/3,1/2) & (1/4,1/3,1/2) & (1/3,2/5,3/5) \\ (1/5,1/3,3/5) & (1/3,1/3,2/5) & (2/5,2/5,2/5) \\ (1/5,1/3,2/3) & (1/3,1/3,1/2) & \end{array} \right\}.$$

To compute this list, it suffices to enumerate those $a \in \mathcal{M}$ such that $T(a)$ is totally spherical ($\rho(a) = 1$), and $(a_2, a_3) \neq (1/2, 1/2)$. When $T(a)$ is totally spherical, the reflection group it generates must be finite, since the corresponding quaternion algebra $B$ is definite. Therefore the edges of $T(a)$ must arise from lines in the tetrahedral, octahedral, or icosahedral tiling of $S^2$ (Figure 6). This shows the denominators occurring in $a$ must be 5 or less, leading to the list above. (The infinite family of dihedral tilings is covered by $S_1$.)

Note that many different spherical triangles give the same reflection group, since $T(a)$ need not be a fundamental domain for that group.

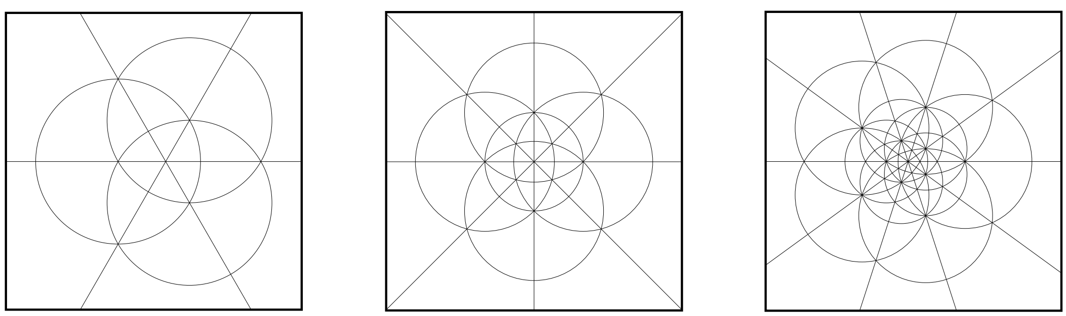

Figure 6. Triangular tilings of $S^2$.

**The totally hyperbolic locus.** On the other hand, we do not even know an upper bound for $|H_0|$. By searching $a \in \mathcal{M} \cap \mathbb{Q}^3$ with denominators $\leq 200$, one can verify the lower bound $|H_0| \geq 294$. To give just one example, $a = (1, 65, 131)/198 \in H_0$; the projection of its Galois orbit to $\mathcal{M}$ gives 20 distinct hyperbolic triangles, the most 'nearly Euclidean' of which is $T(a)$ itself, with $\sum a_i = 1 - 1/198$.

## 7 Quaternion algebras

In this section we prove the results used in §1 for applications to triangle groups, namely equation (1.1), Theorem 1.11 and Corollary 1.8.

Let $a = (a_1, a_2, a_3) \in A$ be a torsion point such that $T(a)$ is not Euclidean. Let $K(a)$ denote the totally real field

$$K(a) = \mathbb{Q}(\cos \pi a_1, \cos \pi a_2, \cos \pi a_3) \subset \mathbb{R}, \tag{7.1}$$

and let $B(a)$ be the quaternion algebra over $K(a)$ generated by three elements of norm one, $g_1, g_2, g_3$, satisfying

$$\mathrm{Tr}(g_i) = 2 \cos \pi a_i,\ i = 1, 2, 3, \quad \text{and} \quad g_1 g_2 g_3 = -1. \tag{7.2}$$

We first prove the following generalization of equation (1.1):

**Theorem 7.1** *For all $a \in A_{\mathrm{tor}}$ such that $T(a)$ is not Euclidean,*

$$\rho(a) = \frac{|\textit{infinite places } v \textit{ of } K(a) \textit{ where } B(a) \textit{ is ramified}|}{|\textit{all infinite places of } K(a)|}. \tag{7.3}$$

Corollary 1.9 on totally hyperbolic triangle groups follows immediately. Next, we turn to arithmeticity and prove Theorem 1.11; and finally, using the spectral gap $\rho_S < 1$, we show the number of arithmetic triangle groups is finite (Corollary 1.8). For more effective proofs of finiteness along quite different lines, see [Tak] and [MR, §11.3].

Note that we allow $a_i = 0$, so the analysis which follows can be applied to triangle groups $\Delta(p, q, r)$ with $p$, $q$ or $r = \infty$.

**Quaternion algebras.** We begin with some background material; for more details, see [MR] and [Voi].

Let $K$ be a field with $\mathrm{char}\, K \neq 2$, and let $B$ be a quaternion algebra over $K$, i.e. a central simple $K$–algebra of rank 4. There is a natural $K$–linear involution $x \mapsto x'$ on $B$ such that $(xy)' = y'x'$ and $x = x'$ if and only if $x \in K$. The trace and norm from $B$ to $K$ are defined by $\mathrm{Tr}(x) = x + x'$ and $N(x) = xx'$. If $N(x) = 1$ then $x' = x^{-1}$.

We can write $B = K \oplus B^0$, where $B^0$ is the set of elements of trace zero. Then $x^2 = -N(x)$ for all $x \in B^0$. The bracket $[x, y] = xy - yx$ on $B$ takes values in $B^0$, and satisfies $[x, y] = [x', y']$. Thus for all $x, y \in B$ we have:

$$[x, y]^2 = [x, y][x', y'] = (xy - yx)(x'y' - y'x') = \mathrm{Tr}(xyx'y') - 2N(xy). \tag{7.4}$$

For $K = \mathbb{R}$, the symmetric bilinear form $\mathrm{Tr}(xy)$ on $B^0$ is definite if $B$ is a division algebra, and indefinite if $B$ is split.

**Geometry of brackets.** The argument below is inspired by following observation. Let $B = M_2(\mathbb{R})$, and let $g, h \in SL_2(\mathbb{R})$ be a pair of elliptic elements fixing distinct points $p, q \in \mathbb{H}$. Then the bracket $r = [g, h] \in PGL_2(\mathbb{R})$ acts on $\mathbb{H}$ by reflection through the line $L = \overline{pq}$. This observation allows one to pass algebraically between triangle groups and reflection groups.

**Lemma 7.2** *The quaternion algebra $B(a) \otimes_{K(a)} \mathbb{R}$ is a division algebra iff $T(a)$ is spherical.*

**Proof.** Recall that $B(a)$ is generated by three elements of norm 1 with $\mathrm{Tr}(g_i) = 2\cos \pi a_i$, $i = 1, 2, 3$, and $g_1 g_2 g_3 = -1$. Let $(i, j, k)$ denote an arbitrary cyclic permutation of $(1, 2, 3)$, and let $r_i = [g_j, g_k]$. It is easy to see that $(r_1, r_2, r_3)$ gives a basis for $B^0$ over $K$. We will show there exists a $\lambda \in K$ such that

$$\mathrm{Tr}(r_i r_j) = 2\lambda Q_{ij}(a), \tag{7.5}$$

where $Q(a)$ is defined by equation (4.2). By Theorem 4.3, the form $Q(a)$ is definite when $T(a)$ is spherical an indefinite when it is hyperbolic, so the Lemma follows.

It remains to establish equation (7.5). Let

$$\lambda = -\mathrm{Tr}(g_k g_j g_i) - 2.$$

Since $g_i g_j g_k = -1$, we have $g_j g_k = -g_i'$; thus equation (7.4) gives

$$r_i^2 = \mathrm{Tr}(g_j g_k g_j' g_k') - 2 = -\mathrm{Tr}(g_i' g_j' g_k') - 2 = \lambda.$$

Similarly we have

$$r_j = [g_k, g_i] = [g_k, -g_k' g_j'] = g_k' g_j' g_k - g_j',$$

and thus

$$\begin{aligned} r_i r_j &= (g_j g_k - g_k g_j)(g_k' g_j' g_k - g_j') \\ &= 2g_k - g_k g_j g_k' g_j' g_k - g_k g_k' g_j g_k g_j' \\ &= g_k(2 - \mathrm{Tr}(g_j g_k' g_j' g_k)) = -\lambda g_k, \end{aligned}$$

using the fact that $g_k' g_j' = -g_i$. Thus

$$\mathrm{Tr}(r_i r_j) = -\lambda \,\mathrm{Tr}(g_k) = 2\lambda \cdot (-\cos \pi a_k)$$

and $\mathrm{Tr}(r_i^2) = 2\lambda \cdot 1$, which gives equation (7.5). ∎

**Galois groups.** Given $a \in A_{\mathrm{tor}}$, let $n$ be the order of the cyclic group $\langle a \rangle$. To relate $\mathrm{Gal}(a)$ to the infinite places of $K(a)$, we introduce the complex field:

$$L(a) = \mathbb{Q}(\exp(\pi i a_1), \exp(\pi i a_2), \exp(\pi i a_3)) = \mathbb{Q}(\zeta_n).$$

We have

$$\mathbb{Q} \subset K(a) \subset L(a) \subset \mathbb{C},$$

and all extensions are Galois. There is a canonical identification of groups,

$$G = \mathrm{Gal}(L(a)/\mathbb{Q}) = (\mathbb{Z}/n)^*,$$

under which $k \in (\mathbb{Z}/n)^*$ corresponds to the unique automorphism $\sigma_k$ of $L(a)$ satisfying $\sigma_k(\zeta_n) = \zeta_n^k$.

**Proof of Theorem 7.1.** Given $k \in G$, let $v$ be the infinite place of $K(a)$ defined by $|x|_v = |\sigma_k(x)|$. Then we have

$$B_k = B(k \cdot a) \otimes_{K(k \cdot a)} \mathbb{R} \cong B(a) \otimes_{K(a)} K(a)_v = B(a)_v.$$

Every infinite place $v$ arises in this way, and the number of such places is $[K(a) : \mathbb{Q}]$. Two different $\sigma_k$ determine the same place $v$ exactly when their difference lies $H = \mathrm{Gal}(L(a)/K(a))$. Since $|H| = [L(a) : K(a)]$, using Lemma 7.2 we have:

$$\begin{aligned} \rho(a) &= \frac{|b \in \mathrm{Gal}(a) \ : \ T(b) \text{ is spherical}|}{|\mathrm{Gal}(a)|} = \frac{|k \in G \ : \ B_k \text{ is a division algebra}|}{[L(a) : \mathbb{Q}]} \\ &= \frac{|H| \cdot |v \ : \ B(a)_v \text{ is a division algebra}|}{|H| \cdot [K(a) : \mathbb{Q}]}. \end{aligned}$$

Canceling the factors of $|H|$ gives (7.3). ∎

**Total hyperbolicity.** In view of Theorem 1.10, we have:

**Corollary 7.3** *The group $\Delta(p, q, r)$ is totally hyperbolic if and only if*

$$\|ka - L\|_1 < 1$$

*for all $k \in (\mathbb{Z}/n)^*$, where $a = 1/(p, q, r)$ and $n = 2\,\mathrm{lcm}(p, q, r)$.*

For example, when $(p, q, r) = (14, 21, 42)$, $k = 31$ is relatively prime to $n = 84$, and we have

$$\|ka - L\|_1 = \|31/(14, 21, 42) - (2, 1, 1)\|_1 = 20/21 < 1.$$

The same inequality holds for all $k \in (\mathbb{Z}/n)^*$, and hence $\Delta(14, 21, 42)$ is totally hyperbolic. A similar check can easily be carried out for all $(p, q, r)$ appearing in Conjecture 1.15.

**Proof of Corollary 1.9.** By Theorem 7.1, $\Delta(p, q, r)$ is totally hyperbolic iff $\rho(1/(p, q, r) = 0$; thus the set of such groups is finite by Corollary 1.6. ■

**Arithmeticity.** We now recall material related to arithmeticity of the group

$$\Delta = \langle g_1, g_2, g_3 \rangle \subset B(a)^\times$$

determined by (7.2). For more details, see [MR, §8].

Since arithmeticity is a commensurability invariant, we pass to the subgroup of finite index $\Delta_0 = \langle g^2 \,:\, g \in \Delta \rangle$, whose quaternion algebra

$$B_0(a) = \mathbb{Q}[\Delta_0] \subset B(a)$$

is defined over the invariant trace field

$$K_0(a) = \mathbb{Q}(\cos^2(\pi a_1), \cos^2(\pi a_2), \cos^2(\pi a_3), \cos(\pi a_1)\cos(\pi a_2)\cos(\pi a_3)). \tag{7.6}$$

Note that $K_0(a)$ is equipped with an embedding into $\mathbb{R}$, and we have

$$B(a) = B_0(a) \otimes_{K_0(a)} K(a). \tag{7.7}$$

The choice of a maximal order $\mathcal{O} \subset B_0(a)$ determines an arithmetic group

$$\Gamma = \{g \in \mathcal{O} \,:\, N(g) = 1\} \subset B_0(a)^\times,$$

with a natural discrete embedding

$$\Gamma \subset (\mathrm{SU}_2)^r \times (\mathrm{SL}_2(\mathbb{R}))^s \subset \prod_{v|\infty} B_0(a)_v. \tag{7.8}$$

Here $v$ ranges over the infinite places of $K_0(a)$, and $r$ and $s$ are the number of ramified and split places for $B_0(a)$. Then as is well known [MR, Thm. 8.3.10],

*The group $\Delta$ is arithmetic iff $s = 1$.*

When $s = 1$, $\Delta$ is commensurable to $\Gamma$ since both project to lattices in the unique $\mathrm{SL}_2(\mathbb{R})$ factor in (7.8).

**The marked moduli space $\boldsymbol{\mathcal{M}_0}$.** To determine the split places of $B_0(a)$ geometrically, we introduce the projection

$$\mu : A = \mathbb{R}^3/(2\mathbb{Z}^3) \to \mathcal{M}_0 = \mathbb{R}^3/(L \rtimes (\pm 1)^3).$$

Here $\mathcal{M}_0$ is the moduli space of triangles with ordered, but unoriented, sides. There is a natural action of $S_3$ on $\mathcal{M}_0$, with quotient the moduli space $\mathcal{M}$ of equation (1.4).

**Theorem 7.4** *The places $v$ of $K_0(a)$ where $B_0(a)$ is split correspond bijectively to the hyperbolic triangles in $\mu(\mathrm{Gal}(a))$.*

**Proof.** We again use the identification $\mathrm{Gal}(L(a)/\mathbb{Q}) = (\mathbb{Z}/n)^*$. Every valuation $v$ on $K_0(a)$ has the form $|x|_v = |\sigma_k x|$, for some $k \in (\mathbb{Z}/n)^*$, and $B_0(a)_v$ is split iff $T(k \cdot a)$ is hyperbolic, in view of Lemma 7.2 and equation (7.7). Thus we need only determine when two different $k \in (\mathbb{Z}/n)^*$ give the same valuation $v$ on $K_0(a)$, or equivalently when $\sigma_k$ acts trivially on $K_0(a)$. Now it is clear from the definition of $K_0(a)$ that $\sigma_k$ acts trivially iff there exist $\epsilon_i = \pm 1$ such that

$$\cos \pi k a_i = \epsilon_i \cos \pi a_i, i = 1, 2, 3, \quad \text{and} \quad \epsilon_1 \epsilon_2 \epsilon_3 = 1.$$

But these conditions hold iff $\mu(a) = \mu(ka)$. Indeed, the first condition holds iff $a$ and $ka$ are equivalent under the action $\mathbb{Z}^3 \rtimes (\pm 1)^3$, and the second condition replaces $\mathbb{Z}^3$ with $L$. ∎

**Complement.** The proof shows that $[K_0(a) : \mathbb{Q}] = |\mu(\mathrm{Gal}(a))|$.

**Corollary 7.5** *The group $\Delta$ is arithmetic iff there is a unique hyperbolic triangle in $\mu(\mathrm{Gal}(a)) \subset \mathcal{M}_0$.*

**Example.** The group $\Delta$ associated to $a = 1/(5, 5, 5/2)$ is arithmetic, even though $T(a) \subset \mathbb{H}$ is not the fundamental domain for a reflection group. In fact $\Delta = \Delta(2, 5, 5)$ (see Figure 7).

**Proof of Theorem 1.11.** For convenience, we have expressed the criterion for arithmeticity of $\Delta(p, q, r)$ in terms of $\mathcal{M}$ instead of $\mathcal{M}_0$. To justify this criterion, it suffices to show that for $a = 1/(p, q, r)$, the projection map

$$p : \mathcal{M}_0 \to \mathcal{M}/S_3$$

is injective on $\mu(\mathrm{Gal}(a))$.

Injectivity is clear when the denominators $(p, q, r)$ of $a$ are distinct: a given point in the projection of $\mathrm{Gal}(a)$ to $\mathcal{M}$ has a unique lift to $\mathcal{M}_0$ whose denominators are in the right order. On the other hand, if $a = 1/(p, p, r)$, $p \neq r$, then every point in $\mu(\mathrm{Gal}(a))$ is fixed by $(12) \in S_3$, so the lift is unique in this case as well. Finally for $a = 1/(p, p, p)$, $S_3$ acts by the identity on $\mu(\mathrm{Gal}(a))$, so projection to $\mathcal{M}$ introduces no new identifications, and the proof is complete. ∎

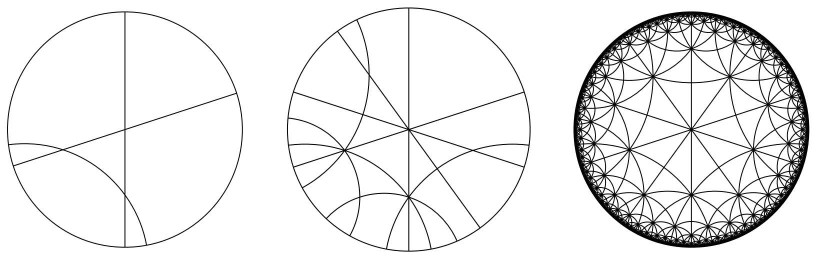

Figure 7. Unfolding the $(5,5,5/2)$ triangle.

**Remarks.** The criterion for arithmeticity just proved is special to $a$ of the form $1/(p,q,r)$. When $a=(1/5,2/5,1/4)$, the projection of $\mathrm{Gal}(a)$ to $\mathcal{M}$ contains a unique hyperbolic triangle, but its projection to $\mathcal{M}_0$ contains two, so $\Delta$ is not arithmetic.

**Finiteness: Proof of Corollary 1.8.** We conclude by the proving that the set of arithmetic triangle groups is finite.

Suppose $\Delta(p,q,r) \subset \mathrm{SL}_2(\mathbb{R})$ is arithmetic, and let $a=1/(p,q,r)$. By virtue of the spectral gap at 1 (Corollary 1.2), we have:

$$\rho(a) = 1 - [K_0(a):\mathbb{Q}]^{-1} \le \rho_S < 1.$$

Equivalently, we have $[K_0(a):\mathbb{Q}] \le (1-\rho_S)^{-1}$. There are only finitely many $a \in A_{\mathrm{tor}}$ satisfying this bound, so there are only finitely many arithmetic triangle groups. ■

## References


[BG] E. Bombieri and W. Gubler. *Heights in Diophantine Geometry.* Cambridge University Press, 2006.

[MR] C. Maclachlan and A. W. Reid. *The Arithmetic of Hyperbolic 3-Manifolds.* Springer-Verlag, 2003.

[Mc] C. McMullen. Triangle groups and Hilbert modular varieties. *Preprint, 2023.*

[MS] S. Mozes and N. Shah. On the space of ergodic invariant measures of unipotent flows. *Ergodic Theory Dynam. Systems* **15** (1995), 149–159.

[Rn] M. Ratner. Interactions between ergodic theory, Lie groups and number theory. In *Proceedings of the International Congress of Mathematicians (Zürich, 1994)*, pages 156–182. Birkhaüser, Basel, 1995.

[Ric] R. Richard. Manin–Mumford par le critère de Weyl. *J. Number Theory* **239** (2022), 137–150.

[Tak] K. Takeuchi. Arithmetic triangle groups. *J. Math. Soc. Japan* **29** (1977), 91–106.

[Th] W. P. Thurston. *Three-Dimensional Geometry and Topology*, volume 1. Princeton University Press, 1997.

[Voi] J. Voight. *Quaternion Algebras*. Springer, 2021.

[WM] P. L. Waterman and C. Maclachlan. Fuchsian groups and algebraic number fields. *Trans. AMS* **287** (1985), 353–364.

Mathematics Department, Harvard University, 1 Oxford St, Cambridge, MA 02138-2901