EISENSTEIN INTEGERS AND EQUILATERAL IDEAL TRIANGLES

GREG MCSHANE

ABSTRACT. We discuss the relationship between Penner's λ-length and the norms of Eisenstein integers. This leads to a geometric proof of the fact, attributed to Fermat, that every prime p of the form $3k+1$ is the norm of an Eisenstein integer that is can be written as $a^2 - ab + b^2$ for some $a, b \in \mathbb{Z}$.

1. INTRODUCTION

A celebrated result of Fermat characterises those primes which can be written as the sum of two squares

Theorem 1.1 (Fermat)**.** Let p be a prime then the equation

$$x^2 + y^2 = p$$

has a solution in integers iff $p = 2$ or $p - 1$ is a multiple of 4.

An alternative formulation of this result is that a prime p is the norm of a Gaussian integer if and only if $p = 2$ or p is of the form $4k + 1$.

There are many proofs of this result. In particular, about 1984 Heath-Brown published a proof of Theorem 1.1 in the journal of the Oxford University undergraduate mathematics society. His proof arose from a study of the account of Liouville's papers on identities for parity functions. Zagier's celebrated one line proof [11] is a clever reformulation of this argument but probably the most elegant formulation is the recent proof given by Dolan [2]. Heath-Brown studies the action of a Klein four group on a finite set and uses considerations of parity to show that there is a fixed point for one of the group elements which is a solution to $x^2+y^2 = p$. In a previous work [6] we gave a geometric proof which, in some sense, mimics Heath-Browns proof. Our proof was based on the analysis of the action of a Klein four group on complete geodesics with two ends terminating at cusps, often referred to as *arcs*, of λ-length p on the surface $\mathbb{H}/\Gamma(2)$. The interested reader might consult [9, 10] for a nice account of how arcs and their λ-lengths play a role in Diophantine approximation etc.

1.1. **Eisenstein integers.** The Eisenstein integers, $\mathbb{Z}[\omega]$, is the ring of integers of the cyclotomic field $\mathbb{Q}(\omega)$ where ω denotes an irrational cubic root of unity. If $a + b\omega \in \mathbb{Z}$ is an Eisenstein integer then its norm is

$$a^2 - ab + b^2.$$

There is an analogue of Theorem 1.1 in this setting:

Theorem 1.2. Let p be a prime then the equation

$$a^2 - ab + b^2 = p \tag{1}$$

has a solution in integers iff $p = 3$ or p is of the form $6k + 1$.

This work was partially funded by the Equipe Action ToFu part of Persyval-Lab.

It is easy to see why this condition is necessary since if we make a reduction modulo p then in $\mathbb{F}_p$ the equation becomes

$$\bar{a}^2 - \bar{a}\bar{b} + \bar{b}^2 = 0$$

and so , for $p > 3$, $\bar{a}/\bar{b}$ is a cubic root of -1 ie an element of order 6 in the group of units $\mathbb{F}_p^*$ and it follows that 6 divides the order of $\mathbb{F}_p^*$ that is $p - 1$.

In this note we show that this condition is sufficient. As before our proof uses the action of a group of automorphisms of the surface $\mathbb{H}/\Gamma(2)$ on arcs of λ-length p.

1.2. **Farey tessalation and λ-lengths.** The *Farey tessalation* is a fundamental object in the theory of Fuchsian groups. It is a tessalation of hyperbolic space by *ideal triangles*.

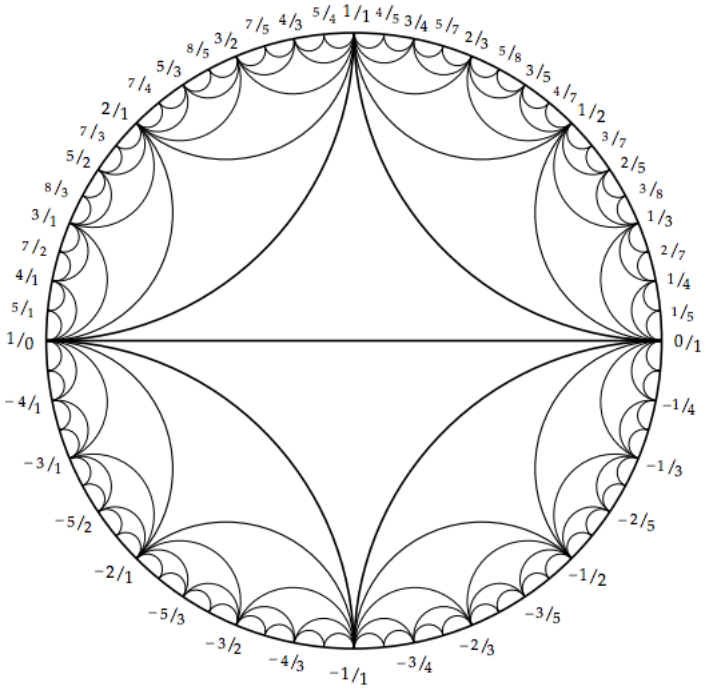

FIGURE 1. Farey diagram.

The tessalation is invariant under the action of the modular group $\Gamma = \mathrm{PSL}(2, \mathbb{Z})$. The quotient of the upper half space by the modular group is the *modular orbifold*. This is a non compact singular surface with a single cusp and two cone points. One of the cone points lifts to the orbit of i under Γ and the other to the orbit of ω. The principal congruence subgroup $\Gamma(2)$ is a torsion free, normal subgroups of index 6 and the modular orbifold admits a degree 6 cover corresponding to $\Gamma(2)$ namely, the *three punctured sphere* $\mathbb{H}/\Gamma(2)$. Whilst this surface is non compact it has finite hyperbolic area equal to 2π. The non compactness is due to the presence of *cusps* and in fact $\mathbb{H}/\Gamma(2)$ has exactly three cusps one for each orbit of $\Gamma(2)$ on the extended rationals $\mathbb{Q} \cup \{\infty\}$.

We will see (Lemma 2.3) that the reciprocal direction in Theorem 1.2 can be stated in terms of lengths of arcs on this surface. For the purposes of this paper an *arc* geodesic is the projection to $\mathbb{H}/\Gamma(2)$ of a Poincaré geodesic with both endpoints in the extended rationals $\mathbb{Q} \cup \{\infty\}$. The geodesic edges of the Farey tessalation yield a set of three such geodesic, in fact the shortest arcs, on the quotient surfaces. The geodesics obtained from projecting the Farey tessalation are *simple*, that is they have no self intersections, and

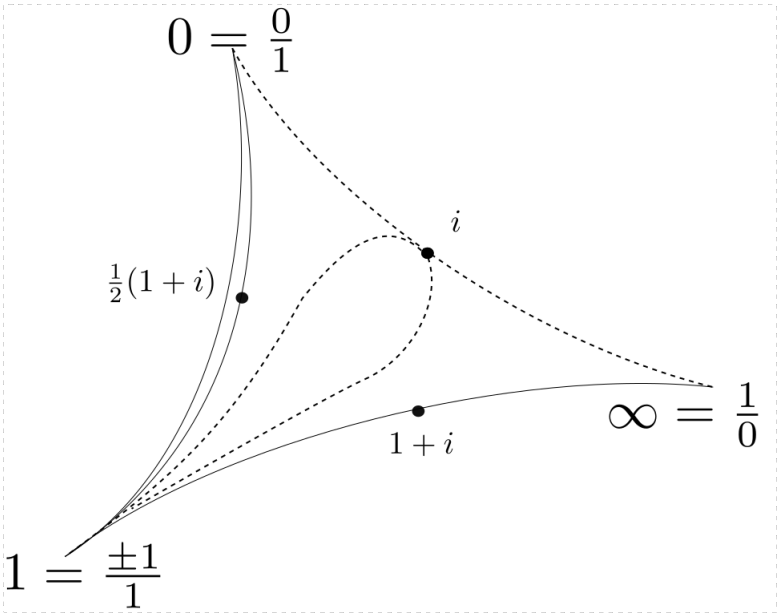

FIGURE 2. Three punctured sphere with cusps labelled by $\Gamma(2)$-orbits and some arcs. Note that the dotted geodesic in the middle is the edge of an ideal triangle which is embedded but not properly immersed as two spikes meet at the cusp ∞.

their complement consists of a pair of (ideal) triangles. Of course, a arc α has infinite length with respect to the Poincaré metric but one can define a useful geometric quantity by truncating the geodesic ie removing a portion of infinite length inside the cusp regions of the surface. This idea of associating a finite length to an *arc*, that is a simple arc, appears in Penner's work on moduli [7] (see paragraph 2.3.) on a punctured surface to be the exponential of half the length of the portion outside of some fixed system of cusp regions. We extend this notion to non simple arcs: Lemma 2.5 shows that in the context we consider a λ-length is always the determinant of an integer matrix.

1.3. **Properly immersed ideal triangles.** An *ideal triangle* in $\mathbb{H}$ is the convex hull of a triple of distinct points in $a, b, c \in \partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$. We say that a, b, c are the *vertices of the triangle* and that the triangle has *spikes* at a, b, c. We identify the group of orientation preserving automorphisms of $\mathbb{H}$ with $P\mathrm{SL}(2, \mathbb{R})$ acting by linear fractional transformations. This action is simply transitive on triples of distinct points in $(a, b, c) \in (\partial\mathbb{H})^3$ so there is only one ideal triangle up to isometry.

Any ideal triangles in the Farey tesselation, for example the one with vertices $0, 1, \infty$ projects to an embedded ideal triangle in $\mathbb{H}/\Gamma(2)$. We will see later that each of the edges of this triangle has λ-length 1 so it is *equilateral* for this notion of length. It is also invariant by the group generated by the map

$$\psi : z \mapsto \frac{1}{-z+1}.$$

The fixed points in $\mathbb{C}$ of this map satisfy

$$z^2 - z + 1 = 0,$$

and there is exactly one in $\mathbb{H}$ namely ω. By a standard argument ω is the *barycenter* of the ideal triangle $0, 1, \infty$. Since there is a unique ideal triangle up to isometry the notion of barycenter is well defined for any ideal triangle.

Each edge of the ideal triangle $0, 1/3, 2/3$ has λ-length 3 so again it is equilateral. However, it does not project to an embedded triangle but it is immersed on $\mathbb{H}/\Gamma(2)$ and further each of its spikes ends at a different cusp on $\mathbb{H}/\Gamma(2)$. We say that the ideal triangle with vertices a, b, c is *properly immersed* iff each of its spikes ends at a different cusp.

1.4. **Sketch of proof.** We study the action of a cyclic subgroup of the automorphisms of $\mathbb{H}/\Gamma(2)$ induced by ψ. More precisely we consider its action on the set of immersed equilateral ideal triangles in $\mathbb{H}/\Gamma(2)$ whose sides have λ-length p. Let F denote the cardinal of this set. By counting incidences (see Section 4) between arcs of λ-length p and the cusps of $\mathbb{H}/\Gamma(2)$ we have the following equation:

$$3 \times F = 2(p-2) \times 3. \tag{2}$$

The RHS is $3F$ precisely because all the triangles are properly immersed so that each spike ends at a different cusp on $\mathbb{H}/\Gamma(2)$ (see Lemma 4.2) and the LHS follows from our Lemma 4.1 which also appeared in our work [6]. So $F = 2(p-2)$ and if $p = 3k+1$ then $F = 2(3k-1)$ which is not a multiple of 3. It follows that there is are two ψ-orbits of length 1 that is ψ leaves two ideal triangles of this family invariant.

Then by considering the barycenter of a lift of one of these invariant ideal triangles we obtain p as the norm of an element of $\mathbb{Z}[\omega]$.

1.5. **The case $p = 3$.** In fact it is not difficult to show that there are exactly two immersed equilateral with sides of λ-length 3 in $\mathbb{H}/\Gamma(2)$: the ideal triangles with vertices $0, 1/3, 2/3$ and $0, 4/3, 5/3$. Each of these is invariant under the automorphism induced by ψ so their barycenters are in the $\mathrm{SL}(2, \mathbb{Z})$ orbit of ω.

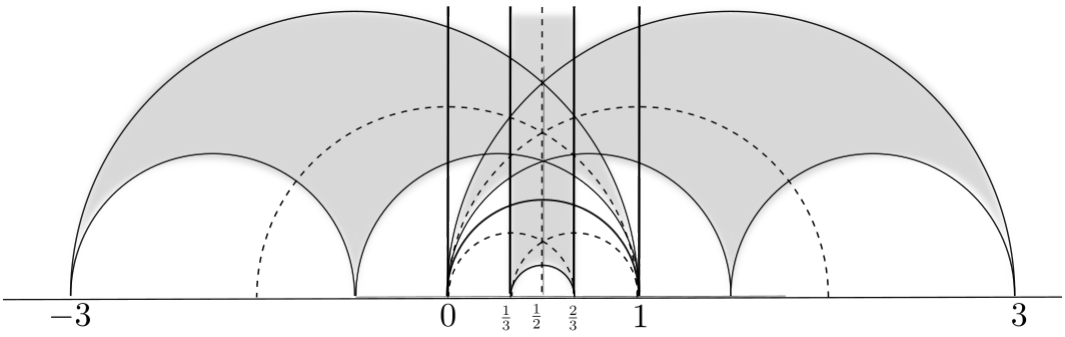

FIGURE 3. Three intersecting lifts of an equilateral ideal triangle. The dotted lines intersect in two points the lower of which is the barycenter of $1/3, 2/3, \infty$ which coincides with the barycenter of $0, 1/2, 1$. The upper point is the barycenter of $0, 1, \infty$ that is $1 + \omega$.

Let us compute the barycenter of $0, 1/3, 2/3$ explicitly to check this. One has

$$3 = 2^2 - 2 + 1 = |2\omega + 1|^2.$$

Let

$$f : z \mapsto \frac{1}{2z+1}$$

then the map

$$f \circ \psi \circ f^{-1} : z \mapsto \frac{2z-1}{3z-1}$$

is conjugate to ψ and permutes the vertices of the ideal triangle

$$\infty \mapsto 2/3 \mapsto 1/3 \mapsto \infty.$$

so it leaves our ideal triangle invariant. One also sees from this that the barycenter of $0, 1/3, 2/3$ is the fixed point of $f \circ \psi \circ f^{-1}$ which is just

$$\frac{\omega}{2\omega+1} = \frac{3/2+\sqrt{3}i/2}{|2\omega+1|^2} = \frac{3+\sqrt{3}i}{2|2\omega+1|^2} = \frac{1}{2} + \frac{\sqrt{3}}{6}i.$$

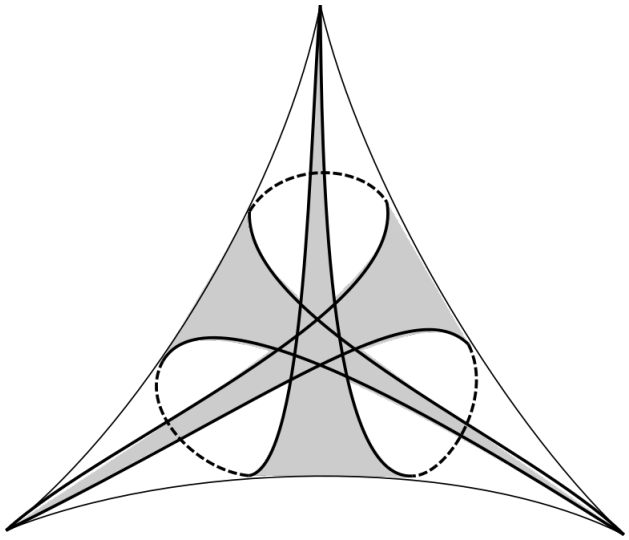

FIGURE 4. Projection of the ideal triangle $1/3, 2/3, \infty$ to $\mathbb{H}/\Gamma(2)$.

1.6. **Thanks.** It is a pleasure to thank Louis Funar, Hidetoshi Masai, Robert Penner, Vlad Sergiescu, and Xu Binbin for discussions and suggestions over the years.

Some of the text of this paper was suggested by GitHub Copilot[12, 13].

2. Reciprocals of norms and arcs

As we stated above an alternative formulation of Theorem 1.1 is that a prime p is the norm of a Gaussian integer if and only if $p = 2$ or $p - 1$ is a multiple of 4. We recall the result from [6] which relates $\mathrm{SL}(2,\mathbb{Z})$-orbits of $i \in \mathbb{C}$ to representations of a number as a sum of squares and then discuss the analogous result for $\mathrm{SL}(2,\mathbb{Z})$-orbits of ω.

2.1. **Gaussian integers.** The group of integer matrices $\mathrm{SL}(2,\mathbb{Z})$ acts on $\mathbb{H}$ by linear fractional transformations that is:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}),\ z \in \mathbb{H},\ \begin{pmatrix} a & b \\ c & d \end{pmatrix}.z = \frac{az+b}{cz+d}.$$

The key lemma that relates this $\mathrm{SL}(2,\mathbb{Z})$ action to sums of squares is:

Lemma 2.1. Let n be a positive integer. The number of ways of writing n as a sum of squares

$$n = c^2 + d^2$$

with c, d co prime positive integers is equal to the number of integers $0 < k < n-1$ co prime to n such that the line

$$\{k/n + it,\ t > 0\}$$

contains a point in the $\mathrm{SL}(2,\mathbb{Z})$ orbit of i.

Proof. Suppose there is such a point which we denote w. The point w is a fixed point of some element of order 2 in $\mathrm{SL}(2,\mathbb{Z})$. Since the Ford circles are $\mathrm{SL}(2,\mathbb{Z})$ invariant this element must permute F with the Ford circle tangent to the real line at the real part of w. So, in particular, w is the midpoint of the line that it lies on and by Lemma 2.5 one has:

$$\frac{1}{n} = \operatorname{Im} \frac{1}{n}(k+i) = \operatorname{Im} \frac{ai+b}{ci+d} = \frac{\operatorname{Im} i}{c^2+d^2}. \tag{3}$$

Conversely if c, d are co prime integers then there exists a, b such that

$$ad - bc = 1 \Rightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}).$$

By applying a suitable iterate of the parabolic transformation $z \mapsto z+1$, one can choose w such that $0 \le \operatorname{Re} w < 1$. So if $n = c^2 + d^2$ then $\frac{ai+b}{ci+d}$ is on one of the lines of the family in the statement.

□

2.2. **Eisenstein integers.** We replace i by ω in (3) above:

$$= \operatorname{Im} \frac{a\omega + b}{c\omega + d} = \frac{\operatorname{Im} \omega}{c^2 - cd + d^2} = \frac{\sqrt{3}/2}{c^2 - cd + d^2}. \tag{4}$$

This expression is not so tidy but we can still exploit it to prove Theorem 1.2. By mimicking the proof of the previous lemma one can easily show:

Lemma 2.2. Let n be a positive integer. The number of ways of writing n as

$$n = c^2 - cd + d^2$$

with c, d co prime positive integers is equal to the number of integers $0 < k < 2n-1$ co prime to $2n$ such that the vertical line $\{k/2n + it,\ t > 0\}$ contains a point in the $\mathrm{SL}(2,\mathbb{Z})$ orbit of ω.

The fact that the denominator is $2n$ and not n is perhaps what is most disturbing but the geodesic is paired to an equilateral ideal triangle which suits our approach:

Lemma 2.3. Let $p > 2$ be prime and $0 < k < p$ an integer co prime to $2p$ such that the line $\{k/2n + it,\ t > 0\}$ contains a point in the $\mathrm{SL}(2,\mathbb{Z})$-orbit of ω. Then the ideal triangle with vertices $\infty, (k-1)/2p, (k+1)/2p$:

- has barycenter at $\frac{k+i\sqrt{3}}{2n}$.
- is equilateral with sides of λ-length p.

The first part follows from the discussion above and the second part is a corollary of Lemma 2.5 below.

2.3. **Ford circles, lengths.** We now recall some standard ideas from hyperbolic geometry necessary to define λ-length. We define an *arc* to be a Poincare geodesic with endpoints in $\partial\mathbb{H}$ a pair of extended rationals, that is elements of $\mathbb{Q} \cup \infty$.

We denote by F the set $\{z, \operatorname{Im} z > 1\}$ this is a *horoball in* $\mathbb{H}$ centered at ∞. The image of F under the $\mathrm{SL}(2,\mathbb{Z})$ action consists of F and infinitely many disjoint discs, which we will refer to as *Ford circles*, each tangent to the real line at some rational m/n. We adopt the convention that F is also a Ford circle of infinite radius tangent to the extended real line at $\infty = 1/0$.

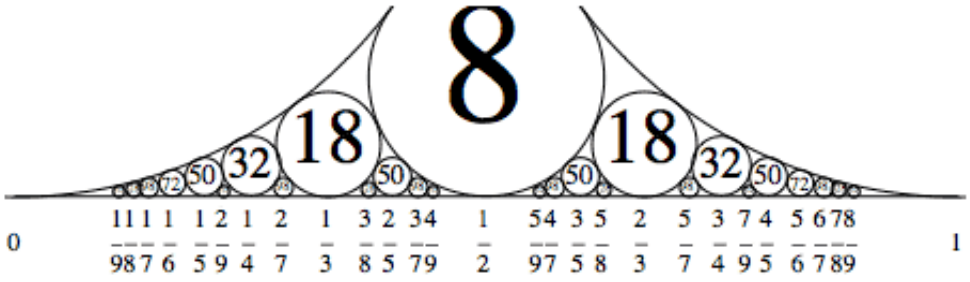

FIGURE 5. Ford circles with tangent points and curvatures. Recall that the curvature of a euclidean circle is the reciprocal of the square of its radius.

The following is well known and is easily checked:

Lemma 2.4. .

The Ford circle tangent to the real line at m/n has Euclidean diameter $1/n^2$.

2.3.1. λ-*length.* Let $a/c, b/d$ be a pair of distinct rationals. We define the *length* of the arc joining these rationals to be half the length, with respect to the Poinacre metric on $\mathbb{H}$, of the portion outside of the Ford circles tangent at $a/c, b/d$.

Following Penner [7] we define the λ-length of an arc to be the exponential of this length. It is a consequence of Lemma 2.5 below that the arcs of λ-length 1 are the edges in the so-called *Farey diagram* (see Figure 1). The lemma is a simple exercise left to the reader:

Lemma 2.5. Let $a/c, b/d$ be a pair of distinct extended rationals. Then the λ-length of the arc joining them is the determinant of the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Further if $a/b = 1/0$ then the arc is a vertical line whose midpoint has imaginary part equal to $1/d$.

Proof. By transivity of $P\mathrm{SL}(2,\mathbb{R})$ on the extended rationals we may assume $a/c = 1/0$ and so the determinant of the matrix is d. The Ford circle at $1/0$ is $\{z, \operatorname{Im} z > 1\}$ and at b/d it is a circle of euclidean diameter $1/d^2$. The top of this circle is tangent to the line $\{z, \operatorname{Im} z = 1/d^2\}$ and the $z \mapsto d^2 z$ maps this to the boundary of the Ford circle $\{z, \operatorname{Im} z \geq 1\}$. It is easy to see that the distance between these sets, and so the length of the segment outside the Ford circles is $2 \log d$. Thus the λ-length is equal to the determinant as required.

□

3. The three punctured sphere

We consider $\Gamma(2)$, the principal level 2 congruence subgroup of $\mathrm{SL}(2,\mathbb{Z})$. This group acts on $\mathbb{Z}^2$, that is pairs of integers, preserving parity.

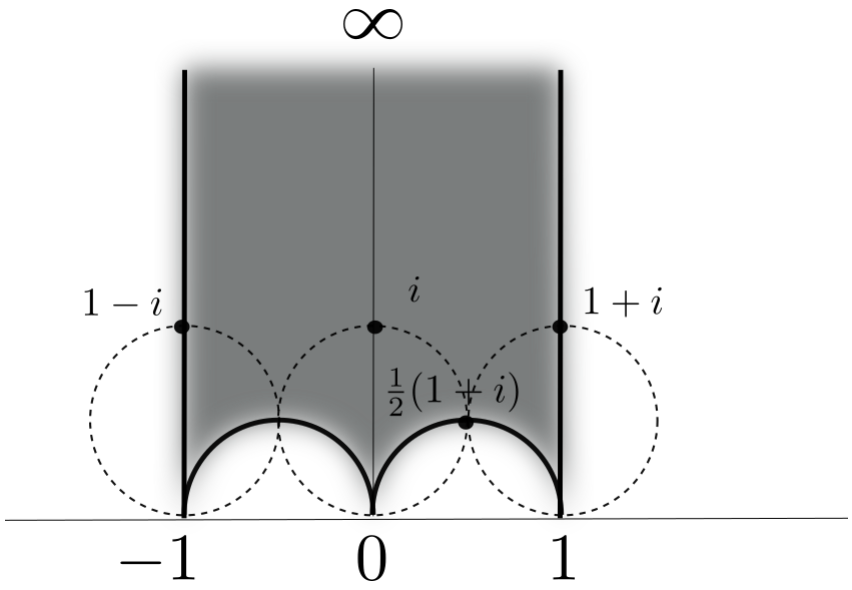

FIGURE 6. Standard fundamental domain for $\Gamma(2)$ and its decomposition into ideal triangles.

It also acts on $\mathbb{H}$ by linear fractional transformations and the quotient $\mathbb{H}/\Gamma(2)$ is conformally equivalent to the Riemann sphere minus three points which we will refer to as *cusps* (see Figure 2). Following convention we label these cusps $0, 1, \infty$ respectively corresponding to the three $\Gamma(2)$ orbits of $\mathbb{Q} \cup \infty$. Finally, the *standard fundamental domain* for $\Gamma(2)$ is the convex hull of the points $\infty, -1, 0, 1$. This region can be decomposed into two ideal triangles $\infty, -1, 0$ and $0, 1, \infty$ as in Figure 6. The edges of the ideal triangles project to three disjoint simple geodesics on $\mathbb{H}/\Gamma(2)$ and each edge has a *midpoint* which is a point of the $\mathrm{SL}(2,\mathbb{Z})$ orbit of i (see Figure 2).

3.0.1. *Cusp regions.* The image of a Ford circle on $\mathbb{H}/\Gamma(2)$ is a *cusp region* around one of the three cusps $0, 1, \infty$. Pairs of these cusp regions are tangent at one of the midpoints labelled $i, 1+i, \frac{1}{2}(1+i)$. It is not difficult to see that these cusp regions are permuted by the automorphisms of $\mathbb{H}/\Gamma(2)$. It follows that if an automorphism preserves a geodesic joining cusps on $\mathbb{H}/\Gamma(2)$ then it must permute the Ford regions at each end of a lift to $\mathbb{H}$.

3.1. **Automorphism groups of $\mathbb{H}/\Gamma(2)$.** From covering theory an isometry of $\mathbb{H}$ induces an automorphism of $\mathbb{H}/\Gamma(2)$ iff it normalises the covering group i.e. $\Gamma(2)$. It follows that, since $\Gamma(2)$ is a normal subgroup of $\mathrm{SL}(2,\mathbb{Z})$, the quotient group

$$\mathrm{SL}(2,\mathbb{Z})/\Gamma(2) \simeq \mathrm{SL}(2,\mathbb{F}_2) \simeq \mathfrak{S}_3$$

acts as a group of (orientation preserving) automorphisms of the surface $\mathbb{H}/\Gamma(2)$.

In our proof of Theorem 1.1 it was necessary to study the action of a Klein 4 group of automorphisms which contained orientation reversing automorphisms. For the proof of

Theorem 1.2 the analysis is simpler as what is important is the action of a cyclic group of order 3 generated by

$$\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

This automorphism:

(1) permutes the cusps;
(2) leaves each of the properly embedded ideal triangles invariant;
(3) has exactly 2 fixed points and these coincide with the barycenters of each of the embedded ideal triangles.

4. Counting arcs and ideal triangles

Let $p > 2$ be a prime. To prove the relation (2) we have to

- count the number of arcs of λ-length p incident to the cusp $1/0 = \infty$ on $\mathbb{H}/\Gamma(2)$.
- show that all the equilateral triangles with sides of λ-length p are properly immersed.

Both points follow from simple calculations using lifts of arcs to $\mathbb{H}$.

Lemma 4.1. *On $\mathbb{H}/\Gamma(2)$:*

(1) *There are $2(p-1)$ arcs of λ-length p incident to the cusp labelled ∞;*
(2) *These geodesics form $2(p-2)$ properly immersed equilateral ideal triangles incident at ∞.*

Proof. Suppose that α is a arc ending at the cusp labelled ∞. It has a lift to $\mathbb{H}$ which is a vertical line and after applying a tranformation in $\Gamma(2)$ of the form $z \mapsto z + 2n$ we may assume this lift has finite endpoint $\alpha^- \in [0, 2]$. By hypothesis α has λ-length p so that $\alpha^- = k/p$ for some k co prime with p and we have that

$$k \in \{1, 2, \dots p-1, p+1, \dots 2p-1\}$$

. This proves the first point.

For the second point, begin by noting that consecutive elements of this set of lifts that is those ending at $k/p, (k+1)/p$ respectively, are sides of an equilateral ideal triangle since

$$\det \begin{pmatrix} k+1 & k \\ p & p \end{pmatrix} = p$$

. It is easy to see that this is the only way to construct such an equilateral ideal triangle incident at ∞. Thus there are $2(p-2)$ such triangles on $\mathbb{H}/\Gamma(2)$

□

Lemma 4.2. *The projection of each of the triangles in the preceding lemma is properly immersed.*

Proof. This follows from considerations of parity of numerator, denominator pairs. By convention ∞ is $1/0$ and by hypothesis p is odd so the denominators have different parities. So for k co prime with p, ∞ and k/p are in different $\Gamma(2)$ orbits. Likewise the parity of k and $k+1$ are always different so that $k/p, (k+1)/p$ are in different $\Gamma(2)$ orbits.

□

5. Concluding remarks

We have proved Theorem 1.2 by counting properly immersed equilateral ideal triangles on the three punctured sphere. Our proof is a simple application of the theory of hyperbolic surfaces and the action of the modular group on them. It is interesting to note that whilst our proof has similarities with the proof of Theorem 1.1 it is also quite different from Heath-Brown's proof of Fermat's two squares theorem [5] as it uses a different group of automorphisms and a different trick not based on parity.

In further work in preparation we will investigate the relation between values of binary quadratic forms and λ-lengths of arcs on other hyperbolic surfaces.

References

[1] Aigner M., Ziegler G.M. *Representing numbers as sums of two squares.* In: Proofs from THE BOOK. Springer, Berlin, Heidelberg. (2010)
[2] Dolan, S., *A very simple proof of the two-squares theorem.* The Mathematical Gazette, 106(564), 511-511. (2021) doi:10.1017/mag.2021.120
[3] Elsholtz C.A *Combinatorial Approach to Sums of Two Squares and Related Problems.* In: Chudnovsky D., Chudnovsky G. (eds) Additive Number Theory. Springer, New York, NY. (2010)
[4] Lester R Ford, *Automorphic Functions*
[5] Heath-Brown, Roger. *Fermat's two squares theorem.* Invariant (1984)
[6] Greg McShane, Vlad Sergiescu, *Geometry of Fermat's sum of squares* `https://macbuse.github.io/squares.pdf`
[7] R. C. Penner, *The decorated Teichmueller space of punctured surfaces*, Communications in Mathematical Physics 113 (1987), 299–339.
[8] J-P. Serre, *A Course in Arithmetic*, Graduate Texts in Mathematics, Springer-Verlag New York 1973
[9] B. Springborn. The hyperbolic geometry of Markov's theorem on Diophantine approximation and quadratic forms. Enseign. Math., 63(3-4):333–373, 2017.
[10] Boris Springborn, *The worst approximable rational numbers* `https://arxiv.org/abs/2209.15542`
[11] D. Zagier, *A one-sentence proof that every prime p = 1 (mod 4) is a sum of two squares*, American Mathematical Monthly, 97 (2): 144
[12] Github Copilot `https://copilot.github.com/`
[13] copilot.vim `https://github.com/github/copilot.vim`

Institut Fourier 100 rue des maths, BP 74, 38402 St Martin d'Hères cedex, France
Email address: `mcshane at univ-grenoble-alpes.fr`