Clay Mathematics Proceedings
Volume 7, 2007

Reciprocal Geodesics

Peter Sarnak

ABSTRACT. The closed geodesics on the modular surface which are equivalent to themselves when their orientation is reversed have recently arisen in a number of different contexts.We examine their relation to Gauss' ambiguous binary quadratic forms and to elements of order four in his composition groups.We give a parametrization of these geodesics and use this to count them asymptotically and to investigate their distribution.

This note is concerned with parametrizing, counting and equidistribution of conjugacy classes of infinite maximal dihedral subgroups of $\Gamma = PSL(2, \mathbb{Z})$ and their connection to Gauss' ambiguous quadratic forms. These subgroups feature in the recent work of Connolly and Davis on invariants for the connect sum problem for manifolds [**CD**]. They also come up in [**PR04**] (also see the references therein) in connection with the stability of kicked dynamics of torus automorphisms as well as in the theory of quasimorphisms of Γ. In [**GS80**] they arise when classifying codimension one foliations of torus bundles over the circle. Apparently they are of quite wide interest. As pointed out to me by Peter Doyle, these conjugacy classes and the corresponding reciprocal geodesics, are already discussed in a couple of places in the volumes of Fricke and Klein ([**FK**], Vol. I, page 269, Vol II, page 165). The discussion below essentially reproduces a (long) letter that I wrote to Jim Davis (June, 2005).

Denote by $\{\gamma\}_\Gamma$ the conjugacy class in Γ of an element $\gamma \in \Gamma$. The elliptic and parabolic classes (i.e., those with $t(\gamma) \leq 2$ where $t(\gamma) = |\text{trace } \gamma|$) are well-known through examining the standard fundamental domain for Γ as it acts on $\mathbb{H}$. We restrict our attention to hyperbolic γ's and we call such a γ primitive (or prime) if it is not a proper power of another element of Γ. Denote by P the set of such elements and by Π the corresponding set of conjugacy classes. The primitive elements generate the maximal hyperbolic cyclic subgroups of Γ. We call a $p \in P$ reciprocal if $p^{-1} = S^{-1}pS$ for some $S \in \Gamma$. In this case, $S^2 = 1$ (proofs of this and further claims are given below) and S is unique up to multiplication on the left by $\gamma \in \langle p \rangle$. Let R denote the set of such reciprocal elements. For $r \in R$ the group $D_r = \langle r, S\rangle$, depends only on r and it is a maximal infinite dihedral subgroup of

2000 *Mathematics Subject Classification.* Primary 11F06, Secondary 11M36.
Key words and phrases. Number theory, binary quadratic forms, modular surface.
Supported in part by the NSF Grant No. DMS0500191 and a Veblen Grant from the IAS.

Γ. Moreover, all of the latter arise in this way. Thus, the determination of the conjugacy classes of these dihedral subgroups is the same as determining ρ, the subset of Π consisting of conjugacy classes of reciprocal elements. Geometrically, each $p \in P$ gives rise to an oriented primitive closed geodesic on $\Gamma\backslash\mathbb{H}$, whose length is $\log N(p)$ where $N(p) = \left[\left(t(p) + \sqrt{t(p)^2-4}\right)/2\right]^2$. Conjugate elements give rise to the same oriented closed geodesic. A closed geodesic is equivalent to itself with its orientation reversed iff it corresponds to an $\{r\} \in \rho$.

The question as to whether a given γ is conjugate to γ^{-1} in Γ is reflected in part in the corresponding local question. If $p \equiv 3 \pmod 4$, then $c = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ is not conjugate to c^{-1} in $SL(2,\mathbb{F}_p)$, on the other hand, if $p \equiv 1 \pmod 4$ then every $c \in SL(2,\mathbb{F}_p)$ is conjugate to c^{-1}. This difficulty of being conjugate in $G(\bar{F})$ but not in $G(F)$ does not arise if $G = GL_n$ (F a field) and it is the source of a basic general difficulty associated with conjugacy classes in G and the (adelic) trace formula and its stabilization [**Lan79**]. For the case at hand when working over $\mathbb{Z}$, there is the added issue associated with the lack of a local to global principle and in particular the class group enters. In fact, certain elements of order dividing four in Gauss' composition group play a critical role in the analysis of the reciprocal classes.

In order to study ρ it is convenient to introduce some other set theoretic involutions of Π. Let ϕ_R be the involution of Γ given by $\phi_R(\gamma) = \gamma^{-1}$. Let $\phi_w(\gamma) = w^{-1}\gamma w$ where $w = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \in PGL(2,\mathbb{Z})$ (modulo inner automorphism ϕ_w generates the outer automorphisms of Γ coming from $PGL(2,\mathbb{Z})$). ϕ_R and ϕ_w commute and set $\phi_A = \phi_R \circ \phi_w = \phi_w \circ \phi_R$. These three involutions generate the Klein group G of order 4. The action of G on Γ preserves P and Π. For H a subgroup of G, let $\Pi_H = \{\{p\} \in \Pi : \phi(\{p\}) = \{p\} \text{ for } \phi \in H\}$. Thus $\Pi_{\{e\}} = \Pi$ and $\Pi_{\langle\phi_R\rangle} = \rho$. We call the elements in $\Pi_{\langle\phi_A\rangle}$ ambiguous classes (we will see that they are related to Gauss' ambiguous classes of quadratic forms) and of $\Pi_{\langle\phi_w\rangle}$, inert classes. Note that the involution $\gamma \to \gamma^t$ is, up to conjugacy in Γ, the same as ϕ_R, since the contragredient satisfies ${}^t g^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} g \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Thus $p \in P$ is reciprocal iff p is conjugate to p^t.

To give an explicit parametrization of ρ let

$$C = \left\{(a,b) \in \mathbb{Z}^2 : (a,b) = 1,\, a > 0,\, d = 4a^2 + b^2 \text{ is not a square}\right\}. \tag{1}$$

To each $(a,b) \in C$ let (t_0, u_0) be the least solution with $t_0 > 0$ and $u_0 > 0$ of the Pell equation

$$t^2 - du^2 = 4. \tag{2}$$

Define $\psi : C \longrightarrow \rho$ by

$$(a,b) \longrightarrow \left\{\begin{bmatrix} \dfrac{t_0 - bu_0}{2} & au_0 \\ au_0 & \dfrac{t_0 + bu_0}{2} \end{bmatrix}\right\}_\Gamma, \tag{3}$$

It is clear that $\psi((a,b))$ is reciprocal since an $A \in \Gamma$ is symmetric iff $S_0^{-1}AS_0 = A^{-1}$ where $S_0 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Our central assertion concerning parametrizing ρ is;

PROPOSITION 1. *$\psi : C \longrightarrow \rho$ is two-to-one and onto.* *

There is a further stratification to the correspondence (3). Let

$$\mathcal{D} = \{m \,|\, m > 0\,,\, m \equiv 0,1 \pmod 4\,, m \text{ not a square}\}\,. \tag{4}$$

Then

$$C = \bigcup_{d \in \mathcal{D}} C_d$$

where

$$C_d = \big\{(a,b) \in C \,|\, 4a^2 + b^2 = d\big\}\,. \tag{5}$$

Elementary considerations concerning proper representations of integers as a sum of two squares shows that C_d is empty unless d has only prime divisors p with $p \equiv 1 \pmod 4$ or the prime 2 which can occur to exponent $\alpha = 0, 2$ or 3. Denote this subset of $\mathcal{D}$ by $\mathcal{D}_R$. Moreover for $d \in \mathcal{D}_R$,

$$|C_d| = 2\,\nu(d) \tag{6}$$

where for any $d \in \mathcal{D}, \nu(d)$ is the number of genera of binary quadratic forms of discriminant d ((6) is not a coincidence as will be explained below). Explicitly $\nu(d)$ is given as follows: If $d = 2^\alpha D$ with D odd and if λ is the number of distinct prime divisors of D then

$$\nu(d) = \begin{cases} 2^{\lambda-1} & \text{if} \quad \alpha = 0 \\ 2^{\lambda-1} & \text{if} \quad \alpha = 2 \quad \text{and} \quad D \equiv 1 \pmod 4 \\ 2^{\lambda} & \text{if} \quad \alpha = 2 \quad \text{and} \quad D \equiv 3 \pmod 4 \\ 2^{\lambda} & \text{if} \quad \alpha = 3 \quad \text{or} \quad 4 \\ 2^{\lambda+1} & \text{if} \quad \alpha \geq 5\,. \end{cases} \tag{6'}$$

Corresponding to (5) we have

$$\rho = \bigsqcup_{d \in \mathcal{D}_R} \rho_d\,, \tag{7}$$

with $\rho_d = \psi(C_d)$. In particular, $\psi : C_d \longrightarrow \rho_d$ is two-to-one and onto and hence

$$|\rho_d| = \nu(d) \text{ for } d \in \mathcal{D}_R\,. \tag{8}$$

Local considerations show that for $d \in \mathcal{D}$ the Pell equation

$$t^2 - du^2 = -4\,, \tag{9}$$

can only have a solution if $d \in \mathcal{D}_R$. When $d \in \mathcal{D}_R$ it may or may not have a solution. Let $\mathcal{D}_R^-$ be those d's for which (9) has a solution and $\mathcal{D}_R^+$ the set of $d \in \mathcal{D}_R$ for which (9) has no integer solution. Then

(i) For $d \in \mathcal{D}_R^+$ none of the $\{r\} \in \rho_d$, are ambiguous.
(ii) For $d \in \mathcal{D}_R^-$, every $\{r\} \in \rho_d$ is ambiguous.

*Part of this Proposition is noted in ([**FK**], Vol. I, pages 267-269).

In this last case (ii) we can choose an explicit section of the two-to-one map (3). For $d \in \mathcal{D}_R^-$ let $C_d^- = \{(a,b) : b < 0\}$, then $\psi : C_d^- \longrightarrow \rho_d$ is a bijection.[†]

Using these parameterizations as well as some standard techniques from the spectral theory of $\Gamma\backslash\mathbb{H}$ one can count the number of primitive reciprocal classes. We order the primes $\{p\} \in \Pi$ by their trace $t(p)$ (this is equivalent to ordering the corresponding prime geodesics by their lengths). For H a subgroup of G and $x > 2$ let

$$\Pi_H(x) := \sum_{\substack{\{p\} \in \Pi_H \\ t(p) \leq x}} 1 \,. \tag{10}$$

THEOREM 2. *As $x \longrightarrow \infty$ we have the following asymptotics:*

$$\Pi_{\{1\}}(x) \sim \frac{x^2}{2\log x}\,, \tag{11}$$

$$\Pi_{\langle \phi_A \rangle}(x) \sim \frac{97}{8\pi^2}\, x(\log x)^2\,, \tag{12}$$

$$\Pi_{\langle \phi_R \rangle}(x) \sim \frac{3}{8}\, x\,, \tag{13}$$

$$\Pi_{\langle \phi_w \rangle}(x) \sim \frac{x}{2\log x} \tag{14}$$

and

$$\Pi_G(x) \sim \frac{21}{8\pi}\, x^{1/2}\, \log x\,. \tag{15}$$

(All of these are established with an exponent saving for the remainder).

In particular, roughly the square root of all the primitive classes are reciprocal while the fourth root of them are simultaneously reciprocal ambiguous and inert.

We turn to the proofs of the above statements as well as a further discussion connecting ρ with elements of order dividing four in Gauss' composition groups.

We begin with the implication $S^{-1}pS = p^{-1} \Longrightarrow S^2 = 1$. This is true already in $PSL(2,\mathbb{R})$. Indeed, in this group p is conjugate to $\pm\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1}\end{pmatrix}$ with $\lambda > 1$. Hence $Sp^{-1} = pS$ with $S = \begin{bmatrix} a & b \\ c & d\end{bmatrix} \Longrightarrow a = d = 0$, i.e., $S = \pm\begin{bmatrix} 0 & \beta \\ -\beta^{-1} & 0\end{bmatrix}$ and so $S^2 = 1$. If S and S_1 satisfy $x^{-1}px = p^{-1}$ then $SS_1^{-1} \in \Gamma_p$ the centralizer of p in Γ. But $\Gamma_p = \langle p \rangle$ and hence $S = \beta S_1$ with $\beta \in \langle p \rangle$. Now every element $S \in \Gamma$ whose order is two (i.e., an elliptic element of order 2) is conjugate in Γ to $S_0 = \pm\begin{bmatrix} 0 & 1 \\ -1 & 0\end{bmatrix}$. Hence any $r \in R$ is conjugate to an element $\gamma \in \Gamma$ for which $S_0^{-1}\gamma S_0 = \gamma^{-1}$. The last is equivalent to γ being symmetric. Thus each $r \in R$ is conjugate to a $\gamma \in R$ with $\gamma = \gamma^t$. (15′)

We can be more precise:

LEMMA 3. *Every $r \in R$ is conjugate to exactly four γ's which are symmetric.*

[†]For a general $d \in \mathcal{D}_R^+$ it appears to be difficult to determine explicitly a one-to-one section of ψ.

To see this associate to each S satisfying

(16) $$S^{-1}rS = r^{-1}$$

the two solutions γ_S and γ'_S (here $\gamma'_S = S\gamma_S$) of

(17) $$\gamma^{-1}S\gamma = S_0 .$$

Then

(18) $$\gamma_S^{-1}\, r\gamma_S = ((\gamma'_S)^{-1}\, r\gamma'_S)^{-1} \text{ and both of these are symmetric.}$$

Thus each S satisfying (17) affords a conjugation of r to a pair of inverse symmetric matrices. Conversely every such conjugation of r to a symmetric matrix is induced as above from a γ_S. Indeed if $\beta^{-1}r\beta$ is symmetric then $S_0^{-1}\,\beta^{-1}\,r\beta S_0 = \beta^{-1}\,r^{-1}\beta$ and so $\beta S_0^{-1}\,\beta^{-1} = S$ for an S satisfying (17). Thus to establish (16) it remains to count the number of distinct images $\gamma_S^{-1}\,r\gamma_S$ and its inverse that we get as we vary over all S satisfying (17). Suppose then that

(19) $$\gamma_S^{-1}\,r\gamma_S = \gamma_{S'}^{-1}\,r\gamma_{S'} .$$

Then

(20) $$\gamma_{S'}\,\gamma_S^{-1} = b \in \Gamma_r = \langle r\rangle .$$

Also from (18)

(21) $$\gamma_S^{-1}\,S\gamma_S = \gamma_{S'}^{-1}\,S'\,\gamma_{S'}$$

or

(22) $$\gamma_{S'}\,\gamma_S^{-1}\,S\,\gamma_S\,\gamma_{S'}^{-1} = S' .$$

Using (21) in (23) yields

(23) $$b^{-1}\,Sb = S' .$$

But bS satisfies (17), hence $bSbS = 1$. Putting this relation in (24) yields

(24) $$S' = b^{-2}S .$$

These steps after (22) may all be reversed and we find that (20) holds iff $S = b^2S'$ for some $b \in \Gamma_r$. Since the solutions of (17) are parametrized by bS with $b \in \Gamma_r$(and S a fixed solution) it follows that as S runs over solutions of (17), $\gamma_S^{-1}\,r\gamma_S$ and $(\gamma'_S)^{-1}r(\gamma'_S)$ run over exactly four elements. This completes the proof of (16). This argument should be compared with the one in ([**Cas82**], p. 342) for counting the number of ambiguous classes of forms. Peter Doyle notes that the four primitive symmetric elements which are related by conjugacy can be described as follows: If A is positive, one can write A as $\gamma'\gamma$ with $\gamma \in \Gamma$ (the map $\gamma \longrightarrow \gamma'\gamma$ is onto such); then A, A^{-1}, B, B^{-1}, with $B = \gamma\gamma'$, are the four such elements.

To continue we make use of the explicit correspondence between Π and classes of binary quadratic forms (see [**Sar**] and also ([**Hej83**], pp. 514-518). ‡ An integral binary quadratic form $f = [a, b, c]$ (i.e. $ax^2 + bxy + xy^2$) is primitive if $(a, b, c) = 1$. Let F denote the set of such forms whose discriminant $d = b^2 - 4ac$ is in $\mathcal{D}$. Thus

(25) $$F = \bigsqcup_{d\in\mathcal{D}} F_d .$$

with F_d consisting of the forms of discriminant d. The symmetric square representation of PGL_2 gives an action $\sigma(\gamma)$ on F for each $\gamma \in \Gamma$. It is given by $\sigma(\gamma)f = f'$

‡This seems to have been first observed in ([**FK**], Vol., page 268)

where $f'(x,y) = f((x,y)\gamma)$. Following Gauss we decompose F into equivalence classes under this action $\sigma(\Gamma)$. The class of f is denoted by $\bar{f}$ or Φ and the set of classes by $\mathcal{F}$. Equivalent forms have a common discriminant and so

$$\mathcal{F} = \bigsqcup_{d \in \mathcal{D}} \mathcal{F}_d \,. \tag{26}$$

Each $\mathcal{F}_d$ is finite and its cardinality is denoted by $h(d)$ - the class number. Define a map n from P to F by

$$p = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow{n} f(p) = \frac{1}{\delta} \,\mathsf{sgn}\,(a+d)\,[b, d-a, -c]\,. \tag{27}$$

where $\delta = \gcd(a, d-a, c) \geq 1$ and n satisfies the following

(i) n is a bijection from Π to F.
(ii) $n(\gamma p \gamma^{-1}) = (\det\gamma)\,\sigma(\gamma)\,n(p)$ for $\gamma \in PGL(2,\mathbb{Z})$.
(iii) $n(p^{-1}) = -n(p)$
(iv) $n(w^{-1}pw) = n(p)^*$
(v) $n(w^{-1}p^{-1}w) = n(p)'$

where

$$[a,b,c]^* = [-a,b,-c] \tag{28}$$

and

$$[a,b,c]' = [a,-b,c]\,. \tag{29}$$

The proof is a straight-forward verification except for n being onto, which relies on the theory of Pell's equation (2). If $f = [a,b,c] \in F$ and has discriminant d and if (t_d, u_d) is the fundamental positive solution to (2) (we also let $\epsilon_d := \frac{t_d + \sqrt{d}\,u_d}{2}$) and if

$$p = \begin{bmatrix} \frac{t_d - u_d b}{2} & a u_d \\ -c u_d & \frac{t_d + u_d b}{2} \end{bmatrix} \tag{30}$$

then $p \in P$ and $n(p) = f$. That p is primitive follows from the well-known fact (see [**Cas82**], p. 291) that the group of automorphs of f, $\mathsf{Aut}_\Gamma(f)$ satisfies

$$\mathsf{Aut}_\Gamma(f) := \{\gamma \in \Gamma : \sigma(\gamma) f = f\} = \left\{ \begin{pmatrix} \frac{t-bu}{2} & au \\ -cu & \frac{t+bu}{2} \end{pmatrix} : t^2 - du^2 = 4 \right\} \Big/ \pm 1 \tag{31}$$

More generally

$$\begin{aligned} Z(f) &:= \{\gamma \in PGL(2,\mathbb{Z}) \,|\, \sigma(\gamma) f = (\det\gamma) f\} \\ &= \left\{ \begin{pmatrix} \frac{t-bu}{2} & au \\ -cu & \frac{t+bu}{2} \end{pmatrix} : t^2 - du^2 = \pm 4 \right\} \Big/ \pm 1\,. \end{aligned} \tag{32}$$

$Z(f)$ is cyclic with a generator η_f corresponding to the fundamental solution $\eta_d = (t_1 + \sqrt{d}\,u_1)/2$, $t_1 > 0$, $u_1 > 0$ of

$$t^2 - du^2 = \pm 4\,. \tag{33}$$

If (9) has a solution, i.e. $d \in \mathcal{D}_R^-$ then η_d corresponds to a solution of (9) and $\epsilon_d = \eta_d^2$. If (9) doesn't have a solution then $\eta_d = \epsilon_d$. Note that $Z(f)$ has elements with $\det\gamma = -1$ iff $d_f \in \mathcal{D}_R^-$. (35)

From (ii) of the properties of the correspondence n we see that $Z(f)$ is the centralizer of p in $PGL(2,\mathbb{Z})$, where $n(p) = f$. (36)

Also from (ii) it follows that n preserves classes and gives a bijection between Π and $\mathcal{F}$. Moreover, from (iii), (iv) and (v) we see that the action of $G = \{1, \phi_w, \phi_A, \phi_R)\}$ corresponds to that of $\tilde{G} = \{1, *, \prime, -\}$ on $\mathcal{F}, \tilde{G}$ preserves the decomposition (27) and we therefore examine the fixed points of $g \in \tilde{G}$ on $\mathcal{F}_d$.

Gauss [**Gau**] determined the number of fixed points of $\prime$ in $\mathcal{F}_d$. He discovered that $\mathcal{F}_d$ forms an abelian group under his law of composition. In terms of the group law, $\Phi' = \Phi^{-1}$ for $\Phi \in \mathcal{F}_d$. Hence the number of fixed points of $\prime$ (which he calls ambiguous forms) in $\mathcal{F}_d$ is the number of elements of order (dividing) 2. Furthermore $\mathcal{F}_d/\mathcal{F}_d^2$ is isomorphic to the group of genera (the genera are classes of forms with equivalence being local integral equivalence at all places). Thus the number of fixed points of $\prime$ in $\mathcal{F}_d$ is equal to the number of genera, which in turn he showed is equal to the number $\nu(d)$ defined earlier. For an excellent modern treatment of all of this see [**Cas82**].

Consider next the involution $*$ on $\mathcal{F}_d$. If $b \in \mathbb{Z}$ and $b \equiv d \pmod 2$ then the forms $[-1, b, \frac{d-b^2}{4}]$ are all equivalent and this defines a class $J \in \mathcal{F}_d$. Using composition one sees immediately that $J^2 = 1$, that is J is ambiguous. Also, applying composition one finds that

$$J\,\overline{[a,b,c]} \,=\, \overline{[-a,b,-c]} \,=\, \overline{[a,b,c]}^{\,*}\,. \tag{37}$$

That is, the action of $*$ on $\mathcal{F}_d$ is given by translation in the composition group; $\Phi \to \Phi J$. Thus $*$ has a fixed point in $\mathcal{F}_d$ iff $J = 1$, in which case all of $\mathcal{F}_d$ is fixed by $*$. To analyze when $J = 1$ we first determine when J and 1 are in the same genus (i.e. the principal genus). Since $[1, b, \frac{b^2-d}{4}]$ and $[1, -b, \frac{b^2-d}{4}]$ are in the same genus (they are even equivalent) it follows that J and 1 are in the same genus iff $f = [1, b, \frac{b^2-d}{4}]$ and $-f$ are in the same genus. An examination of the local genera (see [**Cas82**], p. 33) shows that there is an f of discriminant d which is in the same genus as $-f$ iff $d \in \mathcal{D}_R$. Thus J is in the principal genus iff $d \in \mathcal{D}_R$. (38)

To complete the analysis of when $J = 1$, note that this happens iff $[1, b, \frac{b^2-d}{4}] \sim [-1, b\,\frac{d-b^2}{4}]$. That is, $[1, b, \frac{b^2-d}{4}] \sim (\det w)\,\sigma(w)[1, b, \frac{b^2-d}{4}]$. Alternatively, $J = 1$ iff $f = (\det\gamma)\,\sigma(\gamma)f$ with $f = [1, b, \frac{b^2-d}{4}]$ and $\det\gamma = -1$. According to (35) this is equivalent to $d \in \mathcal{D}_R^-$. Thus $*$ fixes $\mathcal{F}_d$ iff $J = 1$ iff $d \in \mathcal{D}_R^-$ and otherwise $*$ has no fixed points in $\mathcal{F}_d$. (39)

We turn to the case of interest, that is, the fixed points of $-$ on $\mathcal{F}_d$. Since $-$ is the (mapping) composite of $*$ and $\prime$ we see from the discussion above that the action $\Phi \longrightarrow -\Phi$ on $\mathcal{F}_d$ when expressed in terms of (Gauss) composition on $\mathcal{F}_d$ is given by

$$\Phi \,\longrightarrow\, J\,\Phi^{-1}\,. \tag{40}$$

Thus the reciprocal forms in $\mathcal{F}_d$ are those Φ's satisfying

$$\Phi^2 \,=\, J\,. \tag{41}$$

Since $J^2 = 1$, these Φ's have order dividing 4. Clearly, the number of solutions to (41) is either 0 or $\#\{B | B^2 = 1\}$, that is, it is either 0 or the number of ambiguous classes, which we know is $\nu(d)$. According to (38) if $d \notin \mathcal{D}_R$ then J is not in the principal genus and since Φ^2 is in the principal genus for every $\Phi \in \mathcal{F}_d$, it follows that if $d \notin \mathcal{D}_R$ then (41) has no solutions. On the other hand, if $d \in \mathcal{D}_R$ then we remarked earlier that $d = 4a^2 + b^2$ with $(a, b) = 1$. In fact there are $2\nu(d)$ such representations with $a > 0$. Each of these yields a form $f = [a, b, -a]$ in $\mathcal{F}_d$ and each of these is reciprocal by S_0. Hence for each such $f, \Phi = \bar{f}$ satisfies (41), which of course can also be checked by a direct calculation with composition. Thus for $d \in \mathcal{D}_R$, (41) has exactly $\nu(d)$ solutions. In fact, the $2\nu(d)$ forms $f = [a, b, -a]$ above project onto the $\nu(d)$ solutions in a two-to-one manner. To see this, recall (15′), which via the correspondence n, asserts that every reciprocal g is equivalent to an $f = [a, b, c]$ with $a = c$. Moreover, since $[a, b, -a]$ is equivalent to $[-a, -b, a]$ it follows that every reciprocal class has a representative form $f = [a, b, -a]$ with $(a, b) \in C_d$. That is $(a, b) \longrightarrow \overline{[a, b, -a]}$ from C_d to $\mathcal{F}_d$ maps onto the $\nu(d)$ reciprocal forms. That this map is two-to-one follows immediately from (16) and the correspondence n. This completes our proof of (3) and (8). In fact (15′) and (16) give a direct counting argument proof of (3) and (8) which does not appeal to the composition group or Gauss' determination of the number of ambiguous classes. The statements (i) and (ii) follow from (41) and (39). If $d \in \mathcal{D}_R^-$ then $J = 1$ and from (41) the reciprocal and ambiguous classes coincide. If $d \in \mathcal{D}_R^+$ then $J \neq 1$ and according to (14) the reciprocal classes constitute a fixed (non-identity) coset of the group A of ambiguous classes in $\mathcal{F}_d$.

To summarize we have the following: The primitive hyperbolic conjugacy classes are in 1-1 correspondence with classes of forms of discriminants $d \in \mathcal{D}$. To each such d, there are $h(d) = |\mathcal{F}_d|$ such classes all of which have a common trace t_d and norm ϵ_d^2. The number of ambiguous classes for any $d \in \mathcal{D}$ is $\nu(d)$. Unless $d \in \mathcal{D}_R$ there are no reciprocal classes in $\mathcal{F}_d$ while if $d \in \mathcal{D}_R$ then there are $\nu(d)$ such classes and they are parametrized by C_d in a two-to-one manner. If $d \notin \mathcal{D}_R^-$, there are no inert classes. If $d \in \mathcal{D}_R^-$ every class is inert and every ambiguous class is reciprocal and vice-versa. For $d \in \mathcal{D}_R^-$, C_d^- parametrizes the G fixed classes.

Here are some examples:

(i) If $d \in \mathcal{D}_R$ and $\mathcal{F}_d$ has no elements of order four, then $d \in \mathcal{D}_R^-$ (this fact seems to be first noted in [Re1]). For if $d \in \mathcal{D}_R^+$ then $J \neq 1$ and hence any one of our $\nu(d)$ reciprocal classes is of order four. In particular, if $d = p \equiv 1 \pmod 4$, then $h(d)$ is odd (from the definition of ambiguous forms it is clear that $h(d) \equiv \nu(d) \pmod 2$) and hence $d \in \mathcal{D}_R^-$. That is, $t^2 - pu^2 = -4$ has a solution (this is a well-known result of Legendre).

(ii) $d = 85 = 17 \times 5$. $\eta_{85} = \frac{9+\sqrt{85}}{2}$, $\epsilon_{85} = \frac{83+9\sqrt{85}}{2}$, $85 \in \mathcal{D}_R^-$ and $\nu(85) = h(85) = 2$. The distinct classes are $\overline{[1, 9, -1]}$ and $\overline{[3, 7, -3]}$. Both are ambiguous reciprocal and inert. The corresponding classes in ρ are

$$\left\{ \begin{bmatrix} 1 & 9 \\ 9 & 82 \end{bmatrix} \right\}_\Gamma \quad \text{and} \quad \left\{ \begin{bmatrix} 10 & 27 \\ 27 & 73 \end{bmatrix} \right\}_\Gamma .$$

(iii) $d = 221 = 13 \times 17$. $\eta_{221} = \epsilon_{221} = \frac{15+\sqrt{221}}{2}$ so that $221 \in \mathcal{D}_R^+$. $\nu(221) = 2$ while $h(221) = 4$. The distinct classes are $\overline{[1,13,-13]}$, $\overline{[-1,13,13]}$, $\overline{[5,11,-5]} = \overline{[7,5,-7]}$, $\overline{[-5,11,5]} = \overline{[-7,5,7]}$. The first two classes 1 and J are the ambiguous ones while the last two are the reciprocal ones. There are no inert classes. The composition group is cyclic of order four with generator either of the reciprocal classes. The two genera consist of the ambiguous classes in one genus and the reciprocal classes in the other. The corresponding classes in ρ_{221} are

$$\left\{\begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix}\right\}_\Gamma \quad \text{and} \quad \left\{\begin{bmatrix} 13 & 5 \\ 5 & 2 \end{bmatrix}\right\}_\Gamma .$$

The two-to-one correspondence from C_{221} to ρ_{221} has $(5,11)$ and $(7,5)$ going to the first class and $(5,11)$ and $(7,-5)$ going to the second class.

(iv)[§] $d = 1885 = 5 \times 13 \times 29$. $\eta_{1885} = \epsilon_{1885} = (1042 + 24\sqrt{1885}/2)$ so that $1885 \in \mathcal{D}_R^+$. $\nu(1885) = 4$ and $h(1885) = 8$. The 8 distinct classes are

$$\mathbf{1} = \overline{[1,43,-9]},\ \overline{[-1,43,9]} = J,\ \overline{[7,31,-33]},\ \overline{[-7,31,33]},$$
$$\overline{[21,11,-21]} = \overline{[-19,21,19]},\ \overline{[-21,11,21]} = \overline{[19,21,-19]},$$
$$\overline{[3,43,-3]} = \overline{[17,27,-17]},\ \overline{[-3,43,3]} = [-17,27,17].$$

The first four are ambiguous and the last four reciprocal. The composition group $\mathcal{F}_{1885} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4$ and the group of genera is equal to $\mathcal{F}_{1885}/\{\mathbf{1}, J\}$. The corresponding classes in ρ_{1885} are

$$\left\{\begin{bmatrix} 389 & 504 \\ 504 & 653 \end{bmatrix}\right\}_\Gamma , \quad \left\{\begin{bmatrix} 653 & 504 \\ 504 & 389 \end{bmatrix}\right\}_\Gamma ,$$

$$\left\{\begin{bmatrix} 5 & 72 \\ 72 & 1037 \end{bmatrix}\right\}_\Gamma , \quad \left\{\begin{bmatrix} 1037 & 72 \\ 72 & 5 \end{bmatrix}\right\}_\Gamma .$$

The two-to-one correspondence from C_{1885} to ρ_{1885} has the pairs $(21,11)$ and $(19,-21)$, $(21,-11)$ and $(19,21)$, $(3,43)$ and $(17,27)$, $(3,-43)$ and $(17,-27)$ going to each of the reciprocal classes.

[§]The classes of forms of this discriminant as well as all others for $d < 2000$ were computed using Gauss reduced forms, in Kwon [**Kwo**].

(v) Markov discovered an infinite set of elements of Π all of which project entirely into the set $\mathcal{G}_{3/2}$, where for $a > 1\, \mathcal{G}_a = \{z \in \mathcal{G};\, y < a\}$ and $\mathcal{G}$ is the standard fundamental domain for Γ. These primitive geodesics are parametrized by positive integral solutions $m = (m_0, m_1, m_2)$ of

$$m_0^2 + m_1^2 + m_2^2 = 3\, m_0 m_1 m_2\,. \tag{41$'$}$$

All such solutions can be gotten from the solution $(1,1,1)$ by repeated application of the transformation $(m_0, m_1, m_2) \to (3m_1m_2 - m_0, m_1, m_2)$ and permutations of the coordinates. The set of solutions to (41′) is very sparse [**Zag82**]. For a solution m of (41′) with $m_0 \geq m_1 \geq m_2$ let u_0 be the (unique) integer in $(0, m_0/2]$ which is congruent to $\epsilon \bar{m}_1 m_2 \pmod{m_0}$ where $\epsilon = \pm 1$ and $\bar{m}_1 m_1 \equiv 1 \pmod{m_0}$. Let v_0 be defined by $u_0^2 + 1 = m_0 v_0$, it is an integer since $(\bar{m}_1 m_2)^2 \equiv -1 \mod m_0$, from (41′). Set f_m to be $[m_0, 3m_0 - 2u_0, v_0 - 3u_0]$ if m_0 is odd and $\frac{1}{2}[m_0, 3m_0 - 2u_0, v_0 - 3u_0]$ if m_0 is even. Then $f_m \in F$ and let $\Phi_m = \bar{f}_m \in \mathcal{F}$. Its discriminant d_m is $9m_0^2 - 4$ if m_0 is odd and $(9m_0^2 - 4)/4$ if m_0 is even. The fundamental unit is given by $\epsilon_{d_m} = (3m + \sqrt{d_m})/2$ and the corresponding class in Π is $\{p_m\}_\Gamma$ with

$$p_m = \begin{bmatrix} u_0 & m_0 \\ 3u_0 - v_0 & 3m_0 - u_0 \end{bmatrix}. \tag{41$''$}$$

The basic fact about these geodesics is that they are the only complete geodesics which project entirely into $\mathcal{G}_{3/2}$ and what is of interest to us here, these $\{p_m\}_\Gamma$ are all reciprocal (see [**CF89**] p. 20 for proofs).

$m = (1,1,1)$ gives $\Phi_{(1,1,1)} = \overline{[1,1,-1]}$, $d_{(1,1,1)} = 5$, $\epsilon_5 = (3+\sqrt{5})/2$ while $\eta_5 = (1+\sqrt{5})/2$. Hence $d_5 \in \mathcal{D}_R^-$ and $\Phi_{(1,1,1)}$ is ambiguous and reciprocal. The same is true for $m = (2,1,1)$ and $\Phi_{(2,1,1)} = \overline{[1,2,-1]}$.

$m = (5,2,1)$ gives $\Phi_{(5,2,1)} = [5, 11, -5]$ and $d_{(5,2,1)} = 221$. This is the case considered in (iv) above. $\Phi_{(5,2,1)}$ is one of the two reciprocal classes of discriminant 221. It is not ambiguous.

For $m \neq (1,1,1)$ or $(2,1,1)$, $\eta_{d_m} = \epsilon_{d_m}$ and since Φ_m is reciprocal we have that $d_m \in \mathcal{D}_R^+$ and since Φ_m is not ambiguous, it has order 4 in $\mathcal{F}_{d_m}$.

We turn to counting the primes $\{p\} \in \Pi_H$, for the subgroups H of G. The cases $H = \{e\}$ and $\langle \phi_w \rangle$ are similar in that they are connected with the prime geodesic theorems for $\Gamma = PSL(2,\mathbb{Z})$ and $PGL(2,\mathbb{Z})$ [**Hej83**].

Since $t(p) \sim (N(p))^{1/2}$ as $t(p) \longrightarrow \infty$,

$$\Pi_{\{e\}}(x) = \sum_{\substack{t(p) \leq x \\ \{p\} \in \Pi}} 1 \sim \sum_{\substack{N(p) \leq x^2 \\ \{p\} \in \Pi}} 1 \,. \tag{42}$$

According to our parametrization we have

$$\sum_{\substack{N(p) \leq x^2 \\ \{p\} \in \Pi}} 1 = \sum_{\substack{d \in \mathcal{D} \\ \epsilon_d \leq x}} h(d) \,. \tag{43}$$

The prime geodesic theorem for a general lattice in $PSL(2,\mathbb{R})$ is proved using the trace formula, however for $\Gamma = PSL(2,\mathbb{Z})$ the derivation of sharpest known remainder makes use of the Petersson-Kuznetzov formula and is established in [**LS95**]. It reads

$$\sum_{\substack{N(p) \leq x \\ \{p\} \in \Pi}} 1 = Li(x) + O(x^{7/10}) \,. \tag{44}$$

Hence

$$\Pi_{\{e\}}(x) \sim \sum_{\substack{d \in \mathcal{D} \\ \epsilon_d \leq x}} h(d) \sim \frac{x^2}{2\log x} \text{ , as } x \longrightarrow \infty \,. \tag{45}$$

We examine $H = \langle \phi_w \rangle$ next. As $x \longrightarrow \infty$,

$$\Pi_{\langle \phi_w \rangle}(x) = \sum_{\substack{t(p) \leq x \\ \{p\} \in \Pi_{\langle \phi_w \rangle}}} 1 \sim \sum_{\substack{N(p) \leq x^2 \\ \{p\} \in \Pi_{\langle \phi_w \rangle}}} 1 \,. \tag{46}$$

Again according to our parametrization,

$$\sum_{\substack{N(p) \leq x^2 \\ \{p\} \in \Pi_{\langle \phi_w \rangle}}} 1 = \sum_{\substack{d \in \mathcal{D}_R^- \\ \epsilon_d \leq x}} h(d) \,. \tag{47}$$

Note that if $p \in P$ and $\phi_w(\{p\}) = \{p\}$ then $w^{-1} p\, w = \delta^{-1} p \delta$ for some $\delta \in \Gamma$. Hence $w\, \delta^{-1}$ is in the centralizer of p in $PGL(2,\mathbb{Z})$ and $\det(w\delta^{-1}) = -1$. From (36) it follows that there is a unique primitive $h \in PGL(2,\mathbb{Z})$, $\det h = -1$, such that $h^2 = p$. Moreover, every primitive h with $\det h = -1$ arises this way and if p_1 is conjugate to p_2 in Γ then h_1 is Γ conjugate to h_2. That is,

$$\sum_{\substack{N(p) \leq x^2 \\ \{p\} \in \Pi_{\langle \phi_w \rangle}}} 1 = \sum_{\substack{N(h) \leq x \\ \{h\}_\Gamma \\ \det h = -1}} 1 \,, \tag{48}$$

where the last sum is over all primitive hyperbolic elements in $PGL(2,\mathbb{Z})$ with $\det h = -1$, $\{h\}_\Gamma$ denotes Γ conjugacy and $N(h) = \sqrt{N(h^2)}$. The right hand side of (48) can be studied via the trace formula for the even and odd part of the spectrum of $\Gamma\backslash\mathbb{H}$ ([**Ven82**], pp. 138-143). Specifically, it follows from ([**Efr93**], p. 210) and an analysis of the zeros and poles of the corresponding Selberg zeta functions $Z_+(s)$ and $Z_-(s)$ that

$$B(s) := \prod_{\substack{\{h\}_\Gamma,\, \det h = -1 \\ h \text{ primitive}}} \left(\frac{1 - N(h)^{-s}}{1 + N(h)^{-s}} \right) \tag{49}$$

has a simple zero at $s = 1$ and is homomorphic and otherwise non-vanishing in $\Re(s) > 1/2$.

Using this and standard techniques it follows that

$$\sum_{\substack{N(h) \leq x \\ \det h = -1 \\ \{\gamma\}_\Gamma}} 1 \sim \frac{1}{2} \frac{x}{\log x} \text{ as } x \longrightarrow \infty . \tag{50}$$

Thus

$$\Pi_{\langle \phi_w \rangle}(x) \sim \sum_{\substack{d \in \mathcal{D}_R^- \\ \epsilon_d \leq x}} h(d) \sim \frac{x}{2 \log x} \text{ as } x \longrightarrow \infty . \tag{51}$$

The asymptotics for $\Pi_{\langle \phi_R \rangle}$, $\Pi_{\langle \phi_A \rangle}$ and Π_G all reduce to counting integer points lying on a quadric and inside a large region. These problems can be handled for quite general homogeneous varieties ([**DRS93**], [**EM93**]), though two of the three cases at hand are singular so we deal with the counting directly.

$$\Pi_{\langle \phi_R \rangle}(x) = \sum_{\substack{\{\gamma\} \in \Pi_{\langle \phi_R \rangle} \\ t(\gamma) \leq x}} 1 = \sum_{\substack{t_d \leq x \\ d \in \mathcal{D}_R}} \nu(d) . \tag{52}$$

According to (16) every $\gamma \in R$ is conjugate to exactly 4 primitive symmetric $\gamma \in \Gamma$. So

$$\begin{aligned} \Pi_{\langle \phi_R \rangle}(x) &= \tfrac{1}{4} \sum_{\substack{t(\gamma) \leq x \\ \gamma \in P \\ \gamma = \gamma^t}} 1 \\ &\sim \tfrac{1}{4} \sum_{\substack{N(\gamma) \leq x^2 \\ \gamma \in P \\ \gamma = \gamma^t}} 1 . \end{aligned} \tag{53}$$

Now if $\gamma \in P$ and $\gamma = \gamma^t$, then for $k \geq 1$, $\gamma^k = (\gamma^k)^t$ and conversely if $\beta \in \Gamma$ with $\beta = \beta^t$, β hyperbolic and $\beta = \gamma_1^k$ with $\gamma_1 \in P$ and $k \geq 1$, then $\gamma_1 = \gamma_1^t$. Thus we have the disjoint union

$$\bigsqcup_{k=1}^{\infty}\{\gamma^k : \gamma\in P, \gamma=\gamma^t\}$$

$$= \left\{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in\Gamma : t(\gamma) > 2 , \gamma=\gamma^t\right\}$$

$$= \left\{\begin{pmatrix} a & b \\ b & d \end{pmatrix} : ad-b^2 = 1 , 2 < a+d, a,b,d \in\mathbb{Z}\right\}. \tag{54}$$

Hence as $y \longrightarrow \infty$ we have,

$$\psi(y) := \#\left\{\gamma = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \in\Gamma : 2 < t(\gamma) \le y\right\}$$

$$\sim \#\left\{\gamma = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \in\Gamma : 1 < N(\gamma) \le y^2\right\}$$

$$= \sum_{k=1}^{\infty}\#\left\{\gamma\in P : \gamma = \gamma^t , N(\gamma) \le y^{2/k}\right\}$$

$$= \#\{\gamma \in P : \gamma = \gamma^t , N(\gamma) \le y^2\} + O(\psi(y)\log y)\,. \tag{55}$$

Now $\gamma \longrightarrow \gamma^t\gamma$ maps Γ onto the set of $\begin{bmatrix} a & b \\ b & d \end{bmatrix}$, $ad-b^2=1$ and $a+d \ge 2$, in a two-to-one manner. Hence

$$\psi(y) = \frac{1}{2}\left|\left\{\gamma \in \Gamma : \text{ trace } (\gamma^t\gamma) \le y\right\}\right| - 1\,. \tag{56}$$

This last is just the hyperbolic lattice point counting problem (for Γ and $z_0 = i$) see ([**Iwa95**], p. 192) from which we conclude that as $y \longrightarrow \infty$,

$$\psi(y) = \frac{3}{2}\,y + O(y^{2/3})\,. \tag{57}$$

Combining this with (55) and (53) we get that as $x \longrightarrow \infty$

$$\Pi_{\langle\phi_R\rangle}(x) \sim \sum_{\substack{d\in\mathcal{D}_R\\ \epsilon_d \le x}} \nu(d) \sim \frac{3}{8}\,x\,. \tag{58}$$

The case $H = \langle\phi_A\rangle$ is similar but singular. Firstly one shows as in (16) (this is done in ([**Cas82**], p. 341) where he determines the number of ambiguous forms and classes) that every $p\in P$ which is ambiguous is conjugate to precisely 4 primitive p's which are either of the form

$$w^{-1}p\,w = p^{-1} \tag{59}$$

or

$$w_1^{-1}\, p\, w_1 \;=\; p^{-1} \quad \text{with} \quad w_1 \;=\; \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}, \tag{60}$$

called of the first and second kind respectively.

Correspondingly we have

$$\sum_{\substack{d\,\in\,\mathcal{D} \\ \epsilon_d\,\leq\, x}} \nu(d) \;\sim\; \Pi_{\langle\phi_A\rangle}(x) \;=\; \Pi^{(1)}_{\langle\phi_A\rangle}(x) \;+\; \Pi^{(2)}_{\langle\phi_A\rangle}(x)\,. \tag{61}$$

An analysis as above leads to

$$\Pi^{(1)}_{\langle\phi_A\rangle}(x) \;\sim\; \frac{1}{4}\,\#\left\{a^2 - bc \;=\; 1\,;\, 1 \;<\; a \;<\; \frac{x}{2}\right\} \;=\; \frac{1}{2}\sum_{1<a<\,\frac{x}{2}} \tau(a^2-1) \tag{62}$$

where $\tau(m) \;=\; \#$ of divisors of m.

The asymptotics on the r.h.s. of (62) may be derived elementarily as in Ingham [**Ing27**] (for a power saving in the remainder see [**DFI94**]) and one finds that

$$\Pi^{(1)}_{\langle\phi_A\rangle}(x) \;\sim\; \frac{3}{2\pi^2}\,x(\log x)^2 \;\text{ as }\; x \longrightarrow \infty\,. \tag{63}$$

$\Pi^{(2)}_{\langle\phi_A\rangle}(x)$ is a bit messier and reduces to counting

$$\frac{1}{4}\,\#\left\{(m,n,c)\,:\; m^2-4 \;=\; n(n-4c)\,,\; 2 < m \;\leq\; x\right\}\,. \tag{64}$$

This is handled in the same way though it is a bit tedious, yielding

$$\Pi^{(2)}_{\langle\phi_A\rangle}(x) \;\sim\; \frac{85}{8\pi^2}\,x(\log x)^2\,. \tag{65}$$

Putting these together gives

$$\sum_{\substack{d\,\in\mathcal{D} \\ \epsilon_d\,\leq\, x}} \nu(d) \;\sim\; \Pi_{\langle\phi_A\rangle}(x) \;\sim\; \frac{97}{8\pi^2}\,x(\log x)^2 \;\text{ as }\; x \longrightarrow \infty\,. \tag{66}$$

Finally we consider $H = G$. According to the parametrization we have

$$\Pi_G(x) \;=\; \sum_{\substack{\{p\}\,\in\Pi_G \\ t(p)\,\leq\, x}} 1 \;=\; \sum_{\substack{d\,\in\mathcal{D}^-_R \\ t_d\,\leq\, x}} \nu(d) \;\sim\; \sum_{\substack{d\,\in\,\mathcal{D}^-_R \\ \epsilon_d\,\leq\, x}} \nu(d)\,. \tag{67}$$

As in the analysis of $\Pi_{\langle\phi_R\rangle}$ and $\Pi_{\langle\phi_A\rangle}$ we conclude that

(68)

$$\Pi_G(x) \;\sim\; \frac{1}{4}\,\#\left\{\gamma \;=\; \begin{bmatrix} a & b \\ b & c \end{bmatrix} \in PGL(2,\mathbb{Z});\, \det\gamma \;=\; -1\,,\;\; 2 < a+c \;\leq\; \sqrt{x}\right\}.$$

Or, what is equivalent, after a change of variables:

$$\Pi_G(x) \sim \frac{1}{4}\sum_{m\,\leq\,\sqrt{x}} r_f(m^2+4) \tag{69}$$

where $r_f(t)$ is the number of representations of t by $f(x_1, x_2) \;=\; x_1^2 + 4x_2^2$. This asymptotics can be handled as before and gives

$$\sum_{\substack{d\,\in\mathcal{D}^-_R \\ \epsilon_d\,\leq\, x}} \nu(d) \;\sim\; \Pi_G(x) \;\sim\; \frac{21}{8\pi}\sqrt{x}\,\log x\,. \tag{70}$$

This completes the proof of Theorem 2.

Returning to our enumeration of geodesics, note that one could order the elements of Π according to the discriminant d in their parametrization and ask about the corresponding asymptotics. This is certainly a natural question and one that was raised in Gauss (see [**Gau**], §304).

For H a subgroup of G define the counting functions ψ_H corresponding to Π_H by

$$\psi_H(x) = \sum_{\substack{d \in \mathcal{D} \\ d \leq x}} \#\{\Phi \in \mathcal{F}_d : h(\Phi) = \Phi, h \in H\}. \tag{71}$$

Thus according to our analysis

$$\psi_{\{e\}}(x) = \sum_{\substack{d \in \mathcal{D} \\ d \leq x}} h(d) \tag{72}$$

$$\psi_{\langle \phi_A \rangle}(x) = \sum_{\substack{d \in \mathcal{D} \\ d \leq x}} \nu(d) \tag{73}$$

$$\psi_{\langle \phi_R \rangle}(x) = \sum_{\substack{d \in \mathcal{D}_R \\ d \leq x}} \nu(d) \tag{74}$$

$$\psi_{\langle \phi_w \rangle}(x) = \sum_{\substack{d \in \mathcal{D}_R^- \\ d \leq x}} h(d) \tag{75}$$

$$\psi_G(x) = \sum_{d \in \mathcal{D}_R^-, d \leq x} \nu(d). \tag{76}$$

The asymptotics here for the ambiguous classes was determined by Gauss ([**Gau**], §301), though note that he only deals with forms $[a, 2b, c]$ and so his count is smaller than (73). One finds that

$$\psi_{\langle \phi_A \rangle}(x) \sim \frac{3}{2\pi^2} x \log x, \text{ as } x \longrightarrow \infty. \tag{77}$$

As far as (74) goes, it is immediate from (1) that

$$\psi_{\langle \phi_R \rangle}(x) \sim \frac{3}{4\pi} x, \text{ as } x \longrightarrow \infty. \tag{78}$$

The asymptotics for (72) and (75) are notoriously difficult problems. They are connected with the phenomenon that the normal order of $h(d)$ in this ordering appears to be not much larger than $\nu(d)$. There are Diophantine heuristic arguments that explain why this is so [**Hoo84**], [**Sar85**]; however as far as I am aware, all that is known are the immediate bounds

$$(1 + o(1)) \frac{3}{2\pi^2} x \log x \leq \psi_{\{e\}}(x) \ll \frac{x^{3/2}}{\log x}. \tag{79}$$

The lower bound coming from (77) and the upper bound from the asymptotics in [**Sie44**],

$$\sum_{\substack{d \in \mathcal{D} \\ d \leq x}} h(d) \log \epsilon_d = \frac{\pi^2}{18\zeta(3)} x^{3/2} + O(x \log x).$$

In [Ho] a more precise conjecture is made:

$$\psi_{\{e\}}(x) \sim c_2\, x(\log x)^2 . \tag{80}$$

Kwon [**Kwo**] has recently investigated this numerically. To do so she makes an ansatz for the lower order terms in (80) in the form: $\psi_{\{e\}}(x) = x[c_2(\log x)^2 + c_1(\log x) + c_0] + O(x^\alpha)$ with $\alpha < 1$. The computations were carried out for $x < 10^7$ and she finds that for $x > 10^4$ the ansatz is accurate with $c_0 \simeq 0.06, c_1 \simeq -0.89$ and $c_2 \simeq 4.96$. It would be interesting to extend these computations and also to extend Hooley's heuristics to see if they lead to the ansatz.

The difficulty with (76) lies in the delicate issue of the relative density of $\mathcal{D}_R^-$ in $\mathcal{D}_R$. See the discussions in [**Lag80**] and [**Mor90**] concerning the solvability of (9). In [**R3́6**], the two-component of $\mathcal{F}_d$ is studied and used to get lower bounds of the form: Fix t a large integer, then

$$\sum_{\substack{d\,\in\mathcal{D}_R^+\\ d\,\le\,x}} 1 \ \text{ and } \sum_{\substack{d\,\in\,\mathcal{D}_R^-\\ d\,\le\,x}} 1 \gg_t \frac{x(\log\log x)^t}{\log x} . \tag{81}$$

On the other hand each of these is bounded above by $\sum_{\substack{d\,\in\mathcal{D}_R\\ d\,\le\,x}} 1$, which by Landau's thesis or the half-dimensional sieve is asymptotic to $c_3\, x\big/\sqrt{\log x}$. (81) leads to a corresponding lower bound for $\psi_G(x)$. The result [**R3́6**] leading to (81) suggests strongly that the proportion of $d \in \mathcal{D}_R$ which lie in $\mathcal{D}_R^-$ is in $\left(\frac{1}{2}, 1\right)$ (In [**Ste93**] a conjecture for the exact proportion is put forth together with some sound reasoning). It seems therefore quite likely that

$$\frac{\psi_G(x)}{\psi_{\langle\phi_R\rangle}(x)} \longrightarrow c_4 \quad \text{as} \quad x \longrightarrow \infty\,, \quad \text{with} \quad \frac{1}{2} < c_4 < 1\,. \tag{82}$$

It follows from (78) and (79) that it is still the case that zero percent of the classes in Π are reciprocal when ordered by discriminant, though this probability goes to zero much slower than when ordering by trace. On the other hand, according to (82) a positive proportion, even perhaps more than 1/2, of the reciprocal classes are ambiguous in this ordering, unlike when ordering by trace.

We end with some comments about the question of the equidistribution of closed geodesics as well as some comments about higher dimensions. To each primitive closed $p \in \Pi$ we associate the measure μ_p on $X = \Gamma\backslash\mathbb{H}$ (or better still, the corresponding measure on the unit tangent bundle $\Gamma\backslash SL(2,\mathbb{R})$) which is arc length supported on the closed geodesic. For a positive finite measure μ let $\bar\mu$ denote the corresponding normalized probability measure. For many p's (almost all of them in the sense of density, when ordered by length) $\bar\mu_p$ becomes equidistributed with respect to $\overline{dA} = \frac{3}{\pi}\,\frac{dxdy}{y^2}$ as $\ell(p) \to \infty$. However, there are at the same time many closed geodesics which don't equidistribute w.r.t. $\overline{dA}$ as their length goes to infinity. The Markov geodesics ($41''$) are supported in $\mathcal{G}_{3/2}$ and so cannot equidistribute with respect to $\overline{dA}$. Another example of singularly distributed closed geodesics is that of the principal class 1_d ($\in \Pi$), for $d \in \mathcal{D}$ of the form $m^2 - 4$, $m \in \mathbb{Z}$. In this case $\epsilon_d = (m + \sqrt{d})/2$ and it is easily seen that $\bar\mu_{1_d} \to 0$ as $d \to \infty$ (that is, all the mass of the measure corresponding to the principal class escapes in the cusp of X).

On renormalizing one finds that for K and L compact geodesic balls in X,

$$\lim_{d\to\infty} \frac{\mu_{1_d}(L)}{\mu_{1_d}(K)} \to \frac{\mathsf{Length}(g \cap L)}{\mathsf{Length}(g \cap K)},$$

where g is the infinite geodesics from i to $i\infty$.

Equidistribution is often restored when one averages over naturally defined sets of geodesics. If S is a finite set of (primitive) closed geodesics, set

$$\bar{\mu}_S = \frac{1}{\ell(S)} \sum_{p\in S} \mu_p$$

where $\ell(S) = \sum_{p\in S} \ell(p)$.

We say that an infinite set S of closed geodesics is equidistributed with respect to μ when ordered by length (and similarly for ordering by discriminant) if $\bar{\mu}_{S_x} \to \mu$ as $x \to \infty$ where $S_x = \{p \in S : \ell(p) \leq x\}$. A fundamental theorem of Duke [**Duk88**] asserts that the measures $\mu_{\mathcal{F}_d}$ for $d \in \mathcal{D}$ become equidistributed with respect to $\overline{dA}$ as $d \to \infty$. From this, it follows that the measures

$$\sum_{\substack{t(p)=t\\ p\in\Pi}} \mu_p = \sum_{\substack{t_d=t\\ d\in\mathcal{D}}} \mu_{\mathcal{F}_d}$$

become equidistributed with respect to $\overline{dA}$ as $t \to \infty$. In particular the set Π of all primitive closed geodesics as well as the set of all inert closed geodesics become equidistributed as the length goes to infinity. However, the set of ambiguous geodesics as well as the G-fixed closed geodesics don't become equidistributed in $\Gamma\backslash PSL(2,\mathbb{R})$ as their length go to infinity. The extra logs in the asymptotics (63) and (70) are responsible for this singular behaviour. Specifically, in both cases a fixed positive proportion of their mass escapes in the cusp. One can see this in the ambiguous case by considering the closed geodesics corresponding to $[a, 0, -c]$ with $4ac = t^2 - 4$ and $t \leq T$. Fix $y_0 > 1$ then such a closed geodesic with $\sqrt{c/a} \geq y_0$ spends at least $\log\,(\sqrt{c/a}/y_0)$ if its length in $\mathcal{G}_{y_0} = \{z \in \mathcal{G}; \Im(z) > y_0\}$. An elementary count of the number of such geodesics with $t \leq T$, yields a mass of at least $c_0T(\log T)^3$ as $T \longrightarrow \infty$, with $c_0 > 0$ and independent of y_0. This is a positive proportion of the total mass $\sum_{\substack{t(\{\gamma\})\leq T\\ \gamma\in\pi_{\langle\phi_A\rangle}}} \ell(\{\gamma\})$, and, since it is independent of y_0, the claim follows. The argument for the case of G-fixed geodesics is similar.

We expect that the reciprocal geodesics are equidistributed with respect to dg in $\Gamma\backslash PSL(2,\mathbb{R})$, when ordered by length. One can show that there is $c_1 > 0$ such that for any compact set $\Omega \subset \Gamma\ PSL(2,\mathbb{R})$

$$\liminf_{x\longrightarrow\infty} \overline{\mu}_{\rho_x}(\Omega) \geq c_1 \mathrm{Vol}(\Omega)\,. \tag{83}$$

This establishes a substantial part of the expected equidistribution. To prove (83) consider the contribution from the reciprocal geodesics corresponding to $[a, b, -a]$ with $4a^2 + b^2 = t^2 - 4$, $t \leq T$. Each such geodesic has length $2\log((t+\sqrt{t^2-4})/2)$. The equidistribution in question may be rephrased in terms of the Γ action on the space of geodesics as follows. Let V be the one-sheeted hyperboloid $\{(\alpha, \beta, \gamma) : \beta^2 - 4\alpha\gamma = 1\}$. Then $\rho(PSL(2,\mathbb{R}))$ acts on the right on V by the symmetric square representation and it preserves a Haar measure dv on V. For $\xi \in V$ let Γ_ξ be the stabilizer in Γ of ξ. If the orbit $\{\xi\rho(\gamma) : \gamma \in \Gamma_\xi\backslash\Gamma\}$ is discrete in V then $\sum_{\gamma\in\Gamma_\xi\backslash\Gamma} \delta_{\xi\rho(\gamma)}$ defines a locally finite $\rho(\Gamma)$-invariant measure on V. The equidistribution question is that of showing that ν_T becomes equidistributed with respect to dv, locally in V, where

$$\nu_T := \sum_{4<t\le T} \ \sum_{4a^2+b^2=t^2-4} \ \sum_{\gamma\in\Gamma_{\xi(a,b)}\backslash\Gamma} \delta_{\xi(a,b)\,\rho(\gamma)} \tag{84}$$

and $\xi(a,b) = \left(\frac{a}{\sqrt{t^2-4}}, \frac{b}{\sqrt{t^2-4}}, \frac{-a}{\sqrt{t^2-4}}\right)$.

Let Ω be a nice compact subset of V (say a ball) and fix $\gamma\in\Gamma$, then using the spectral method [**DRS93**] for counting integral points in regions on the two-sheeted hyperboloid $4a^2+b^2-t^2=-4$ one can show that

$$\sum_{4<t\le T} \ \sum_{\substack{4a^2+b^2=t^2-4\\ \gamma\notin i\,\Gamma_{\xi(a,b)}}} \delta_{\xi(a,b)\,\rho(\gamma)}\,(\Omega) = c(\gamma,\Omega)T + 0\left(T^{1-\delta} \parallel \gamma \parallel^A\right) \tag{85}$$

where $\delta>0$ and $A<\infty$ are fixed, $c(\gamma,\Omega)\ge 0$ and $\parallel\gamma\parallel = \sqrt{\mathrm{tr}(\gamma'\gamma)}$. The c's satisfy

$$\sum_{\|\gamma\|\le\xi} c(\gamma,\Omega) \gg \mathrm{Vol}(\Omega)\log\xi \quad \text{as } \xi\longrightarrow\infty\,. \tag{86}$$

Hence, summing (85) over γ with $\parallel\gamma\parallel \le T^{\epsilon_0}$ for $\epsilon_0>0$ small enough but fixed, we get that

$$\nu_T(\Omega) \gg \mathrm{Vol}(\Omega)\,T\log T\,. \tag{87}$$

On the other hand for any compact $B\subset V$, $\nu_T(B) = O(T\log T)$ and hence (83) follows.

In this connection we mention the recent work [**ELMV**] in which they revisit Linnik's methods and give a proof along those lines of Duke's theorem mentioned on the previous page. They show further that for a subset of $\mathcal{F}_d$ of size d^{ϵ_0} with $\epsilon_0>0$ and fixed, any probability measure which is a weak-star limit of the measures associated with such closed geodesics has positive entropy.

The distribution of these sets of geodesics is somewhat different when we order them by discriminant. Indeed, at least conjecturally they should be equidistributed with respect to $d\bar{A}$. We assume the following normal order conjecture for $h(d)$ which is predicted by various heuristics [**Sar85**], [**Hoo84**]; For $\alpha>0$ there is $\epsilon>0$ such that

$$\#\{d\in\mathcal{D} : d\le x \text{ and } h(d)\ge d^\alpha\} = O\left(x^{1-\epsilon}\right)\,. \tag{88}$$

According to the recent results of [**Pop**] and [**HM**], if $h(d)\le d^{\alpha_0}$ with $\alpha_0 = 1/5297$ then *every* closed geodesic of discriminant d becomes equidistributed with respect to $d\bar{A}$ as $d\longrightarrow\infty$. From this and Conjecture (88) it follows that each of our sets of closed geodesics, including the set of principal ones, becomes equidistributed with respect to $d\bar{A}$, when ordered by discriminant.

An interesting question is whether the set of Markov geodesics is equidistributed with respect to some measure ν when ordered by length (or equivalently by discriminant). The support of such a ν would be one-dimensional (Hausdorff). One can also ask about arithmetic equidistribution (e.g. congruences) for Markov forms and triples.

The dihedral subgroups of $PSL(2, \mathbb{Z})$ are the maximal elementary noncyclic subgroups of this group (an elementary subgroup is one whose limit set in $\mathbb{R} \cup \{\infty\}$ consists of at most 2 points). In this form one can examine the problem more generally. Consider for example the case of the Bianchi groups $\Gamma_d = PSL(2, O_d)$ where O_d is the ring of integers in $\mathbb{Q}(\sqrt{d})$, $d < 0$. In this case, besides the issue of the conjugacy classes of maximal elementary subgroups, one can investigate the conjugacy classes of the maximal Fuchsian subgroups (that is, subgroups whose limit sets are circles or lines in $\mathbb{C} \cup \{\infty\}$ = boundary of hyperbolic 3-space $\mathbb{H}^3$). Such classes correspond precisely to the primitive totally geodesic hyperbolic surfaces of finite area immersed in $\Gamma_d\backslash\mathbb{H}^3$. As in the case of $PSL(2, \mathbb{Z})$, these are parametrized by orbits of integral orthogonal groups acting on corresponding quadrics (see Maclachlan and Reid [**MR91**]). In this case one is dealing with an indefinite integral quadratic form f in four variables and their arithmetic is much more regular than that of ternary forms. The parametrization is given by orbits of the orthogonal group $O_f(\mathbb{Z})$ acting on $V_t = \{x : f(x) = t\}$ where the sign of t is such that the stabilizer of an $x(\in V_t(\mathbb{R}))$ in $O_f(\mathbb{R})$ is not compact. As is shown in [**MR91**] using Siegel's mass formula (or using suitable local to global principles for spin groups in four variables (see [**JM96**]) the number of such orbits is bounded independently of t (for $d = -1$, there are 1,2 or 3 orbits depending on congruences satisfied by t). The mass formula also gives a simple formula in terms of t for the areas of the corresponding hyperbolic surface. Using this, it is straight-forward to give an asymptotic count for the number of such totally geodesic surfaces of area at most x, as $x \to \infty$ (i.e., a "prime geodesic surface theorem"). It takes the form of this number being asymptotic to $c.x$ with c positive constant depending on Γ_d. Among these, those surfaces which are noncompact are fewer in number, being asymptotic to $c_1 x/\sqrt{\log x}$.

Another regularizing feature which comes with more variables is that each such immersed geodesic surface becomes equidistributed in the hyperbolic manifold $X_d = \Gamma_d\backslash\mathbb{H}^3$ with respect to $d\tilde{\mathrm{V}}\mathrm{ol}$, as its area goes to infinity. There are two ways to see this. The first is to use Maass' theta correspondence together with bounds towards the Ramanujan Conjectures for Maass forms on the upper half plane, coupled with the fact that there is basically only one orbit of $O_f(\mathbb{Z})$ on $V_t(\mathbb{Z})$ for each t (see the paper of Cohen [**Coh05**] for an analysis of a similar problem). The second method is to use Ratner's Theorem about equidistribution of unipotent orbits and that these geodesic hyperbolic surfaces are orbits of an $SO_{\mathbb{R}}(2, 1)$ action in $\Gamma_d\backslash SL(2, \mathbb{C})$ (see the analysis in Eskin-Oh [**EO**]).

Acknowledgements

Thanks to Jim Davis for introducing me to these questions about reciprocal geodesics, to P. Doyle for pointing out some errors in my original letter and for the references to Fricke and Klein, to E. Ghys and Z. Rudnick for directing me to the references to reciprocal geodesics appearing in other contexts, to E. Lindenstrauss and A. Venkatesh for discussions about equidistribution of closed geodesics and especially the work of Linnik, and to W. Duke and Y. Tschinkel for suggesting that I prepare this material for this volume.

References

[Cas82] J. W. S. Cassels – "Rational quadratic forms", in *Proceedings of the International Mathematical Conference, Singapore 1981 (Singapore, 1981)* (Amsterdam), North-Holland Math. Stud., vol. 74, North-Holland, 1982, p. 9–26.

[CD] F. X. Connolly & J. F. Davis – "L-theory of $\mathrm{PSL}_2(\mathbb{Z})$ and connected sums of manifolds", in preparation.

[CF89] T. W. Cusick & M. E. Flahive – *The Markoff and Lagrange spectra*, Mathematical Surveys and Monographs, vol. 30, American Mathematical Society, Providence, RI, 1989.

[Coh05] P. B. Cohen – "Hyperbolic equidistribution problems on Siegel 3-folds and Hilbert modular varieties", *Duke Math. J.* **129** (2005), no. 1, p. 87–127.

[DFI94] W. Duke, J. B. Friedlander & H. Iwaniec – "Bounds for automorphic L-functions. II", *Invent. Math.* **115** (1994), no. 2, p. 219–239.

[DRS93] W. Duke, Z. Rudnick & P. Sarnak – "Density of integer points on affine homogeneous varieties", *Duke Math. J.* **71** (1993), no. 1, p. 143–179.

[Duk88] W. Duke – "Hyperbolic distribution problems and half-integral weight Maass forms", *Invent. Math.* **92** (1988), no. 1, p. 73–90.

[Efr93] I. Efrat – "Dynamics of the continued fraction map and the spectral theory of $\mathrm{SL}(2,\mathbb{Z})$", *Invent. Math.* **114** (1993), no. 1, p. 207–218.

[ELMV] M. Einsiedler, E. Lindenstrauss, P. Michel & A. Venkatesh – "Distribution properties of periodic torus orbits on homogeneous spaces", in preparation.

[EM93] A. Eskin & C. McMullen – "Mixing, counting, and equidistribution in Lie groups", *Duke Math. J.* **71** (1993), no. 1, p. 181–209.

[EO] A. Eskin & H. Oh – "Representations of integers by an invariant polynomial and unipotent flows", preprint.

[FK] R. Fricke & F. Klein – "Theorie der Elliptischen Modulfunktionen", Vol. I and II, *Leipzig*, 1890 and 1892.

[Gau] C. Gauss – "Disquisitiones Arithmeticae".

[GS80] E. Ghys & V. Sergiescu – "Stabilité et conjugaison différentiable pour certains feuilletages", *Topology* **19** (1980), no. 2, p. 179–197.

[Hej83] D. A. Hejhal – *The Selberg trace formula for* $\mathrm{PSL}(2,\mathbb{R})$*. Vol. 2*, Lecture Notes in Mathematics, vol. 1001, Springer-Verlag, Berlin, 1983.

[HM] G. Harcos & P. Michel – "The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points", preprint (2005).

[Hoo84] C. Hooley – "On the Pellian equation and the class number of indefinite binary quadratic forms", *J. Reine Angew. Math.* **353** (1984), p. 98–131.

[Ing27] A. Ingham – "Some asymptotic formulae in the theory of numbers", *J. London Math. Soc.* **2** (1927), p. 202–208.

[Iwa95] H. Iwaniec – *Introduction to the spectral theory of automorphic forms*, Biblioteca de la Revista Matemática Iberoamericana, Revista Matemática Iberoamericana, Madrid, 1995.

[JM96] D. G. James & C. Maclachlan – "Fuchsian subgroups of Bianchi groups", *Trans. Amer. Math. Soc.* **348** (1996), no. 5, p. 1989–2002.

[Kwo] S. Kwon – "On the average of class numbers", Princeton University undergraduate thesis, (2006).

[Lag80] J. C. Lagarias – "On the computational complexity of determining the solvability or unsolvability of the equation $X^2 - DY^2 = -1$", *Trans. Amer. Math. Soc.* **260** (1980), no. 2, p. 485–508.

[Lan79] R. P. Langlands – "Stable conjugacy: definitions and lemmas", *Canad. J. Math.* **31** (1979), no. 4, p. 700–725.

[LS95] W. Z. Luo & P. Sarnak – "Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbf{H}^2$", *Inst. Hautes Études Sci. Publ. Math.* (1995), no. 81, p. 207–237.

[Mor90] P. Morton – "On the nonexistence of abelian conditions governing solvability of the -1 Pell equation", *J. Reine Angew. Math.* **405** (1990), p. 147–155.

[MR91] C. Maclachlan & A. W. Reid – "Parametrizing Fuchsian subgroups of the Bianchi groups", *Canad. J. Math.* **43** (1991), no. 1, p. 158–181.

[Pop] A. Popa – "Central values of Rankin L-series over real quadratic fields", preprint (2005).

[PR04] L. Polterovich & Z. Rudnick – "Stable mixing for cat maps and quasi-morphisms of the modular group", *Ergodic Theory Dynam. Systems* **24** (2004), no. 2, p. 609–619.

[R3́6] L. Rédei – "Über einige Mittelwertfragen im quadratischen Zahlkörper", *Jnl. Reine Angew Math.* **174** (1936), p. 15–55.

[Sar] P. Sarnak – "Prime Geodesic Theorems", thesis, (1980), Stanford.

[Sar85] P. C. Sarnak – "Class numbers of indefinite binary quadratic forms. II", *J. Number Theory* **21** (1985), no. 3, p. 333–346.

[Sie44] C. L. Siegel – "On the theory of indefinite quadratic forms", *Ann. of Math. (2)* **45** (1944), p. 577–622.

[Ste93] P. Stevenhagen – "The number of real quadratic fields having units of negative norm", *Experiment. Math.* **2** (1993), no. 2, p. 121–136.

[Ven82] A. B. Venkov – "Spectral theory of automorphic functions", *Proc. Steklov Inst. Math.* (1982), no. 4(153), p. ix+163 pp. (1983), A translation of Trudy Mat. Inst. Steklov. **153** (1981).

[Zag82] D. Zagier – "On the number of Markoff numbers below a given bound", *Math. Comp.* **39** (1982), no. 160, p. 709–723.

Department of Mathematics, Princeton University, Princeton NJ 08544-1000
E-mail address: `sarnak@math.princeton.edu`