BULLETIN DE LA S. M. F.

MARK POLLICOTT

Distribution of closed geodesics on the modular surface and quadratic irrationals

Bulletin de la S. M. F., tome 114 (1986), p. 431-446

<http://www.numdam.org/item?id=BSMF_1986__114__431_0>

Bull. Soc. math. France,
114, 1986, p. 431-446.

DISTRIBUTION OF CLOSED GEODESICS ON THE MODULAR SURFACE AND QUADRATIC IRRATIONALS

BY

MARK POLLICOTT (*)

ABSTRACT. — In this note we show that closed geodesics, when ordered by length, are equidistributed with respect to the Poincaré measure. Our main application of this result is to distribution of quadratic irrationals in the unit interval with respect to an appropriate ordering.

0. Introduction

In a recent paper SARNAK proved some striking asymptotic results for averages of class numbers of quadratic forms [17]. A basic ingredient in his proofs was the relationship between quadratic forms and closed geodesics on the modular surface (and other surfaces corresponding to principal congruence subgroups of $SL(2, \mathbb{Z})$). The asymptotic averages for class numbers are then a consequence of certain "Prime Orbit Theorems" i. e. asymptotic estimates for numbers of closed geodesics in terms of an upper bound on their lengths.

In this paper we shall begin by proving a prime orbit theorem for the modular surface (claimed by SARNAK and WOO *cf.* [17]) and extend to the

(*) Texte reçu le 22 juillet 1985.

Mark POLLICOTT, Institut des Hautes Études Scientifiques, 35, route de Chartres, 91440 Bures-sur-Yvette, France.

BULLETIN DE LA SOCIÉTÉ MATHÉMATIQUE DE FRANCE — 0037-9484/1986/04 431 16/$3.60

modular surface a result of Bowen on the equidistribution of closed geodesics in terms of the Riemann measures for compact manifolds of constant negative curvature [4]. (We should recall that the modular surface is the classic example of a non-compact surface.)

As a first application we will be able to derive various results about the distribution of geodesics which also reflect geometric features of the surface. As a second application we shall invoke the correspondence studied by Sarnak to prove some curious results about the distribution of quadratic irrationals in the unit interval (with respect to a natural ordering).

The actual extension of Bowen's equidistribution result to the modular surface will be modelled on Parry's elegant proof in analogy with Dirichle-t's theorem (for the case of Axiom A flows on compact manifolds) [11]. The most important requirement for this analysis is a result on the domain of an appropriate zeta-function. This poses somewhat different problems to those encountered in the compact case ([11], [14]). Our approach is to translate the problem of the zeta-function into the setting of continued fraction transformations. We then prove the required results in this context by extending a theorem of Mayer [10].

In the first section we relate the modular surface to the continued fraction transformation. In the second section we define the necessary zeta function and deduce the results we need on its domain. In section 3 we prove the asymptotic and equidistribution results for closed geodesics on the modular surface. In the penultimate section we explain the relationship between quadratic irrationals, quadratic forms and closed geodesics. In the final section we use the preceeding results to study the distribution of quadratic irrationals.

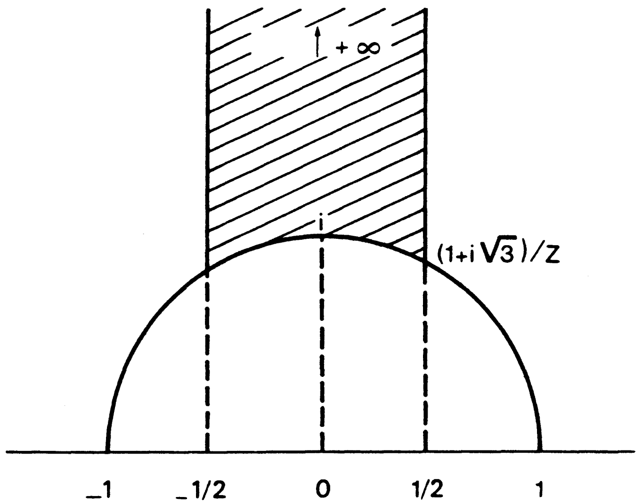
I would like to acknowledge the hospitality and support of I.H.E.S. whilest this paper was being written.

I am grateful to David Ruelle for some interesting remarks.

1. The modular surface

In this section we recall the close connection between the modular surface and the continued fraction transformation ([18], [19], [1]).

Let $\mathbb{H}^+$ denote the Lobetechsky upperhalf plane $\{s \in \mathbb{C} \mid I(s) > 0\}$ with the Poincaré metric $ds^2 = (dx^2 + dy^2)/y^2$. With respect to this metric geodesics on $\mathbb{H}^+$ are either semi-circles centred on the real line or vertical lines [2]. The group $PSL(2, \mathbb{Z})$ acts on $\mathbb{H}^+$ as linear fractional transformations, $z \to (az+b)/(cz+d)$; $a, b, c, d \in \mathbb{Z}$, with $ad-bc=1$. Furthermore, all of these transformations are isometries with respect to the Poincaré metric [2]. The modular surface is the quotient space $\mathcal{M}=\mathbb{H}^+/PSL(2, \mathbb{Z})$, with the induced metric. The modular surface has three singular points: a cusp and two ramification points of orders 2 and 3 at i and $(1+\sqrt{3}i)/2$ respectively [8].



The curvature of $\mathcal{M}$ with respect to the Poincaré metric is $\kappa=-1$. Our interest in $\mathcal{M}$ stems from it being non-compact, although $\mathcal{M}$ has finite area with respect to the corresponding Riemann measure (The area of $\mathcal{M}$ is $2\pi^2/3$).

Consider now continued fractions in the unit interval [3]. If $0<x<1$ has the following expansion

$$x=\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\dots}}}; \qquad a_1, a_2, a_3, \dots \in \mathbb{Z}^+$$

then we denote this by $x=[a_1, a_2, a_3, \dots]$. The expansion terminates after finitely many terms if and only if x is rational.

Let $T : (0, 1) \to [0, 1)$ be the continued fraction transformation given by $Tx = 1/x - [1/x]$ i.e. the fractional part of $1/x$. If $x = [a_1, a_2, a_3, \ldots]$ is irrational then it is easy to see that $\{x\} = \bigcap_{n=1}^{\infty} T^{-(n-1)}[1/(a_n+1), 1/a_n]$. In general we shall not be interested in rational points.

The periodic points for T are precisely those numbers with periodic continued fraction transformations i.e. there exists $p > 0$ such that $a_{p+n} = a_n$, for all $n \geqslant 0$. We then denote $x = [\overline{a_1, a_2, \ldots, a_p}]$.

There is a unique T-invariant probability measure on (0, 1) equivalent to Lebesgue measure called the *Gauss measure*. The Gauss measure μ can be defined by the Radon-Nikodym derivative

$$\frac{d\mu}{dl}(x) = \frac{1}{\log 2} \cdot \frac{1}{(1+x)} \quad [3].$$

Let $T_1 \mathcal{M}$ be the unit tangent bundle of $\mathcal{M}$. Define the geodesic flow $\varphi_t : T_1 \mathcal{M} \to T_1 \mathcal{M}$ as follows: Given $(x, v) \in T_1 \mathcal{M}$. Let $\gamma : \mathbb{R} \to \mathcal{M}$ be the unique (unit speed) geodesic with $(\gamma(0), \dot{\gamma}(0)) = (x, v)$ then set $\varphi_t(x, v) = (\gamma(t), \dot{\gamma}(t))$. Thus closed φ-orbits correspond exactly to closed geodesics of the same length. In addition, there is a natural φ-invariant measure $\tilde{m}$ on $T_1 \mathcal{M}$ induced by the Riemann measure m on $\mathcal{M}$ and Lebesgue measure in each fibre ($\tilde{m}$ is called the Liouville measure).

The natural extension of the continued fraction transformation is given by

$$\hat{T}: \quad (0, 1) \times (0, 1) \to (0, 1) \times (0, 1);$$

where

$$\hat{T}: \quad [a_0, a_1, \ldots] \times [a_{-1}, a_{-2}, \ldots] \to [a_1, a_2, \ldots] \times [a_0, a_1, \ldots].$$

The product measure $\mu \times \mu$ is invariant under $\hat{T}$ and there is a natural correspondence between $\hat{T}$-closed orbits and T-closed orbits. (This definition of $\hat{T}$ is not adequate for the case of rational co-ordinates. However, since these form a set of $\mu \times \mu$-measure zero and do not contain the $\hat{T}$-periodic points this will not lead to any difficulties).

Let $r : (0, 1) \times (0, 1) \times \mathbb{Z}_2 \to \mathbb{R}^+$ be a function which is finite and continuous a. e. $(\mu \times \mu)$ and let

$$X = \{ (x, y, e; t) \in (0, 1) \times (0, 1) \times \mathbb{Z}_2 \times \mathbb{R}^+ \,|\, 0 \leqslant t \leqslant r(x, y, e),$$

where $(x, y, e; r(x, y, e))$ is identified with $(\hat{T}(x, y), e+1 \pmod 2; 0)$. We can define a suspended flow $\psi_t : X \to X$ by $\psi_t(x, y, e; q) = (x, y, e; q+t)$, with appropriate identifications.

The following proposition summarises results from [18], [19], [1].

PROPOSITION 1. — *There exists a suspension function r and $p : X \to T_1 \mathcal{M}$ such that*

(i) $p\psi_t = \varphi_t p$.

(ii) *A closed ψ-orbit corresponding to a closed*

$$T\text{-orbit}\,\{[a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]\}_{i=1}^{2n}$$

has least period $-2\log\prod_{i=0}^{2n}[a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]$.

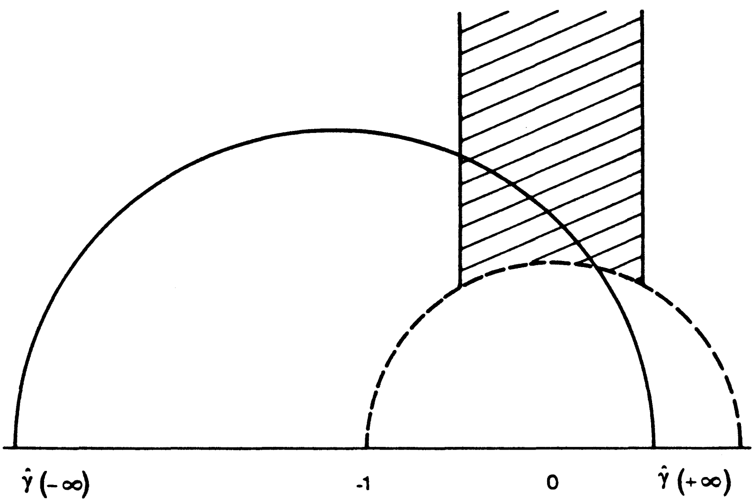
(iii) *The map p gives a one-one (period-preserving) correspondence between closed φ_t-orbits and closed ψ-orbits.*

(iv) *For a closed T-orbit $\{[a_i, \ldots, a_{2n}, a_1, \ldots, a_i]\}_{i=1}^{2n}$ the two corresponding closed geodesics (for $e=0$ or 1) have lifts to $\mathbb{H}$ with endpoints*

$$\hat{\gamma}(+\infty) = [a_1, \ldots, a_{2n}],\ \hat{\gamma}(-\infty) = -[a_1, \ldots, a_{2n}]^{-1}$$

or

$$\hat{\gamma}(+\infty) = [a_2, \ldots, a_{2n}, a_1],\ \hat{\gamma}(-\infty) = -[a_2, \ldots, a_{2n}, a_1]^{-1}.$$



Remark. — An alternative way of generating the continued fraction expansion for a geodesic was presented in [18]. Consider a geodesic γ and call the segment between successive crossings of the fixed geodesic passing through $+\infty$ and i a *rotation* (about either the cusp or the ramification point of order 3). Then a_i, $i=1, \ldots, 2n$ counts the number of successive clockwise (or anti-clockwise) rotations. A sequence which terminates (corresponding to a rational co-ordinate) is interpreted as a geodesic which "escapes" up the cusp to infinity.

2. The η-function

In order to pursue Parry's proof of equidistribution in the next section we need to establish certain properties of an appropriate complex function (*cf.* [11]).

Let $F: \mathcal{M} \to \mathbb{R}$ be a continuous function then we can weight a closed geodesic $\gamma_0 n \mathcal{M}$ by $l_F(\gamma)=\int_0^{l(\gamma)} F(\varphi_t x)\, dt$, where x is any point on γ i.e. $l_F(\gamma)$ is the integral of F along the length of the geodesic γ. Then the function we will need to study is given by

$$\eta(s, F)=\sum_\gamma l_F(\gamma) \exp - sl(\gamma), \qquad s \in \mathbb{C}.$$

Here the sum is over all closed geodesics, and is well-defined for $\mathcal{R}(s)>1$.

In practice, we shall only require the family of $F's$ to be L^1-dense. Set $f_e=f(x, y, e)=\int_0^{r(x, y, e)} Fp(x, y, e;\, t)\, dt$ then if γ_0, γ_1 correspond to $\{[a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]\}_{i=1}^{2n}$ by Proposition 1 then:

$$\exp - sl(\gamma_e)=\prod_{i=1}^{2n} [a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]^{2s}, \qquad e=0, 1$$

and

$$l_F(\gamma_0)+l_F(\gamma_1)=\sum_{i=0}^{2n} (f_0+f_1)([a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]).$$

By suitable approximation we need only consider the function

$$\hat{\eta}(s, f)=\sum_{n=1}^{\infty} \sum_{a_1, \ldots, a_{2n}} \{\sum_{i=0}^{2n} f([a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}])\} \times \prod_{i=1}^{2n} [a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]^{2s}$$

where f is a bounded holomorphic function.

In order to actually prove the results we need on this $\hat{\eta}$-function we shall define $\chi(z)=\chi_{s,\omega}(z)=z^{2s}\exp\omega f(z)$ (where $\mathscr{R}(s)>1/2$, $\omega\in\mathbb{C}$) and consider the following zeta function studied by Mayer [10]:

$$\zeta(z\chi)=\exp\sum_{n=1}^{\infty}\frac{z^n}{n}\sum_{a_1\ldots a_n}\prod_{i=1}^{n}\chi([a_i,\ldots,a_n,a_1,\ldots,a_{i-1}])$$

We shall now briefly recall Mayer's results on the meromorphic domain of this zeta function, where we have been obliged to make modifications to suit our present needs. We let $D_r=\{z \mid |z-1|<r\}$ and define B to be the space of analytic functions $g: D_{3/2}\to\mathbb{C}$ which have a uniform extension to $\bar{D}$.

Define a Ruelle-Perron-Frobenius operator $L_1: B\to B$ by

$$(L_1 g)(z)=\sum_{n=1}^{\infty} g\left(\frac{1}{z+n}\right)\chi\left(\frac{1}{z+n}\right).$$

Similarly we define a second operator $L_2: B\to B$ by

$$(L_2 g)(z)=\sum_{n=1}^{\infty} g\left(\frac{1}{z+n}\right)\chi\left(\frac{1}{z+n}\right)\left(\frac{1}{z+n}\right)^2.$$

By mimicing the arguments of Mayer [10] (after Ruelle [15]) we can prove the following

(i) $L_1, L_2: B\to B$ are nuclear operators and we write their eigenvalues as $\lambda_1^1, \lambda_2^1, \lambda_3^1, \ldots$, and $\lambda_1^2, \lambda_2^2, \lambda_3^2, \ldots$ respectively.

(ii) The following are both entire functions of z:

$$d_1(z,\chi)=\exp\sum_{n=1}^{\infty}\frac{z^n}{n}\sum_{a_1\ldots a_n}\frac{\prod_{i=1}^{n}\chi([a_i,\ldots,a_n,a_i,\ldots,a_{i-1}])}{1-(-1)^n\prod_{i=1}^{n}[a_i,\ldots,a_n,a_1,\ldots,a_{i-1}]^2}$$

$$d_2(z,\chi)=\exp\sum_{n=1}^{\infty}\frac{(-z)^n}{n}$$

$$\times\sum_{a_1,\ldots,a_n}\frac{\prod_{i=1}^{n}\chi([a_i,\ldots,a_n,a_i,\ldots,a_{i-1}])[a_i,\ldots,a_n,a_1,\ldots,a_{i-1}]^2}{1-(-1)^n\prod_{i=1}^{n}[a_i,\ldots,a_n,a_1,\ldots,a_{i-1}]^2}$$

and thus $z\to\zeta(z\chi)=d_1(z,\chi)/d_2(z,\chi)$ is meromorphic in the entire complex plane [assuming $\mathscr{R}(s)>1/2$].

(iii) We can write $d_1(z, \chi)=\prod_{i=1}^{\infty}(1-z\lambda_i^1)$ and $d_2(z, \chi)=\prod_{i=1}^{\infty}(1-z\lambda_i^2)$.

We are particularly interested in the dependence of $\zeta(z\chi)$ on χ (and in particular on s and ω).

Both Ruelle [15] and Mayer [10] observed that $d_1(z, \chi)$ and $d_2(z, \chi)$ depend analytically on L_1 and L_2, respectively (on the Banach space of Fredholm Kernals on B). Furthermore, since $\chi \to L_1$, L_2 are linear and bounded (provided $\mathscr{R}(s)>1/2$) it follows that $(s, \omega) \to d_1(\pm 1, \chi_{s,\omega})$, $d_2(\pm 1, \chi_{s,\omega})$ are analytic, for $\mathscr{R}(s)>1/2$. We can therefore conclude that

$$\hat{\eta}(s, f)=\frac{d}{d\omega}\log(\zeta(\chi_{s,\omega})\zeta(-\chi_{s,\omega}))\big|_{\omega=0}$$

is meromorphic as a function of s in a neighborhood of $\mathscr{R}(s)=1$ (in fact for $\mathscr{R}(s)>1/2$).

We now need more information on the location of the poles for $\hat{\eta}(s, f)$. From (iii) we have that

$$\zeta(\pm\chi_{s,\omega})=\prod_{i=1}^{\infty}(1\mp\lambda_i^1)/\prod_{i=1}^{\infty}(1\mp\lambda_i^2)$$

and therefore

$$\hat{\eta}(s, f)=-\sum_{i=1}^{\infty}\left(\frac{d\lambda_i^1}{d\omega}\right)(s, 0)\,.\,\frac{2\lambda_i^1(s, 0)}{1-(\lambda_i^1(s, 0))^2} + \sum_{i=1}^{\infty}\left(\frac{d\lambda_i^2}{d\omega}\right)(s, 0)\,.\,\frac{2\lambda_i^2(s, 0)}{1-(\lambda_i^2(s, 0))^2}.$$

Using very simple arguments (*cf.* [12], [14]) one can see that if $s=1+it$ then

(a) 1 is never an eigenvalue for L_2;

(b) 1 is an eigenvalue for L_1 if and only if $t=0$.

Furthermore, since $\lambda_1^1=1$ is a simple eigenvalue of $L_1(s=1)$ it follows that $s=1$ is a simple pole for $s\to\hat{\eta}(s, f)$. Hence $s\to\hat{\eta}(s, f)$ has a single (simple) pole on $\mathscr{R}(s)=1$, at $s=1$, with residue $((d\lambda_1^1/d\omega)/(d\lambda_1^1/ds))(0, 1)$. (The existence of these derivatives is a consequence of the analyticity of the eigenvalues.)

Following Walters ([22], p. 134) we know that

$$\log\lambda_1^1(\sigma, \omega)=P(-\sigma\log z^2+\omega f(z))$$

$$=\sup\left\{\int_0^1 [-\sigma\log z^2+\omega f(z)]\,d\rho+h(\rho)\,\middle|\,\rho\ T\text{-invariant}\right\}$$

where $0<z<1$ and $\sigma,\ \omega\in\mathbb{R}$. If $\omega>0$ then

$$P(-\log z^2+\omega f(z))\geqslant h(\mu)-\int\log z^2\,d\mu+\omega\int f\,d\mu$$

$$=P(-\log z^2)+\omega\int f\,d\mu$$

where μ is the Gauss measure, since

$P(-\log z^2)=h(\mu)-\int\log z^2\,d\mu$ (*cf.* [22]). Thus

$$\lim_{\omega\to 0^+}\frac{P(-\log z^2+\omega f)-P(-\log z^2)}{\omega}\geqslant\int f\,d\mu.$$

A similar argument with $\omega<0$ gives that

$$\lim_{\omega\to 0^-}\frac{P(-\log z^2+\omega f)-P(-\log z^2)}{\omega}\geqslant\int f\,d\mu$$

and we conclude that

$$\frac{d}{d\omega}P(-\log z^2+\omega f)_{|\,\omega=0}=\int f\,d\mu.$$

The same sort of reasoning can be applied to prove that

$$\frac{d}{d\sigma}P(-\sigma\log z^2)_{|\,\sigma=1}=-\int_0^1\log z^2\,d\mu=\frac{\pi^2}{6\log}.$$

We summarize our conclusions in the following proposition.

PROPOSITION 2. — (i) *The function* $s\to\hat{\eta}(s,f)$ *is analytic in a neighborhood of* $\mathcal{R}(s)\geqslant 1$, *except for a simple pole at* $s=1$ *with residue*

$$\frac{6.\log 2}{\pi^2}\int_0^1 f\,dz.$$

(ii) *The function* $s \to \eta(s, F)$ *is analytic in a neighborhood of* $\mathcal{R}(s) \geqslant 1$, *except for a simple pole at* $s=1$, *with residue* $\int F\, dm$ *(For an* $L^1(m)$ *dense set of functions* F*).*

Remark. — In Mayer's original proof he defined L and B slightly differently. However, since the complex variable $s \in \mathbb{C}$ occur as an exponent we have to make changes.

Remark. — If we had tried to follow more closely the proof for axiom A flows then we encounter problems. Although analogous results to the compact case on the spectra of the appropriate Ruelle operator are still valid the difficulty comes in trying to apply there to extending the zeta function ([16], p. 93).

Remark. — Traditionally the analysis of closed geodesics on (compact) manifolds of constant negative curvature is based on the study of the Selberg trace formula. It is not known to the author if such an approach could be used to prove Proposition 2 (ii).

3. Distribution of closed geodesics on the modular surface

In this section we want to apply the results we obtained in the previous section to consider the distribution of closed geodesics on the modular surface. Here we shall employ Parry's number theory analogy [11].

We can write

$$\int_1^\infty t^{-s}\, d\Lambda_F(t) = \sum_\gamma l_F(\gamma)\, e^{-s l(\gamma)} = \frac{\int F\, dm}{s-1} + \varphi(s)$$

where $\varphi(s)$ is analytic in a neighborhood of $\mathcal{R}(s) \geqslant 1$ and $\Lambda_F(t) = \sum_{e^{l(\gamma)} \leqslant t} l_F(\gamma)$ is a partial summation over closed geodesics γ of lengths $l(\gamma)$ satisfying $e^{l(\gamma)} \leqslant t\ (t>1)$. It immediately follows from the Ikehara-Wiener Tauberian theorem ([23], p. 127) that $\Lambda_F(t)/t \to \int F\, dm$ as $t \to +\infty$. (We denote this as $\Lambda_F(t) \sim \int F\, dm\, t$.)

Let m_t be the probability measure formed by equidistributing Lebesgue measure around closed geodesics of length less than $t>0$ *i.e.*

$m_t(F) = \sum_{l(\gamma) \leqslant t} l_F(\gamma) / \sum_{l(\gamma) \leqslant t} l(\gamma)$. Then m_t converges to the Riemann measure on M (in the weak* topology) as t increases. This yields our main result.

THEOREM 1. — *The closed geodesics on M are equidistributed* (*by length*) *according to Riemann measure m.*

Remark. — Using some elementary manipulations of partial summations we can show that if

$$\pi_F(t) = \sum_{e^{l(\gamma)} \leqslant t} \frac{l_F(\gamma)}{l(\gamma)} \qquad \text{and} \qquad \pi(t) = \pi_1(t) = \text{Card}\{\gamma \mid \exp l(\gamma) \leqslant t\}$$

then $\pi_F(t)/\pi(t) \to \int F\,dm$. This is another way of expressing the above equidistribution result (*cf.* [11], § 7).

As an immediate corollary we have the following

THEOREM 2 (Sarnak and Woo). — $\pi(t) \sim t/\log t$.

(*We write* $f(t) \sim g(t)$ *if* $f(t)/g(t) \to 1$ *as* $t \to +\infty$.)

Because of the way in which the symbolic dynamics were constructed we can use these estimates to deduce results of a more geometric flavour. A closed geodesic γ corresponding to a periodic continued fraction $[a_1, \ldots, a_{2n}]$ first circles a_1 times in one direction and then a_2 times in the opposite direction, and so on. Thus $\omega(\gamma) = 2n$ is simply the number of times a geodesic "changes direction" during its length. Similarly

$$\alpha_k(\gamma) = \sum_{i=1}^{2n} \chi_{[1/k+1,\,1/k)}[a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}])$$

is the number of times a geodesic orbits the cusp or ramification point *exactly* k times any one direction.

By suitably approximating by analytic functions we can take $f = 1$ *and* $f = \chi_{[1/k+1,\,1/k]}$ successively to deduce the following

PROPOSITION 3. — (i)

$$\left(\sum_{e^{l(\gamma)} \leqslant t} \omega(\gamma)\right)/\pi(t) \sim \left[6 \frac{\log 2}{\pi^2}\right] \log t.$$

(ii)

$$\left(\sum_{e^{l(\gamma)} \leqslant t} \alpha_k(\gamma)\right)/\pi(t) \sim \left[\frac{6}{\pi^2} \log\left(\frac{(k+1)^2}{k(k+2)}\right)\right] \log t.$$

Thus whilest $\omega(\gamma)$ and $\alpha_k(\gamma)$ are erratically distributed under the ordering by $l(\gamma)$, "on the average" they grow as $l(\gamma)$. [In particular, $\omega(\gamma)$ and $\alpha_k(\gamma)$ are of order $l(\gamma)\log l(\gamma)$.]

Remark. — A natural generalization of our results would be to surfaces formed from co-finite normal subgroups of $SL(2, \mathbb{Z})$. Unfortunately, the approach we adopted towards extending the η-function in section 2 does not readily adapt to these cases.

4. Quadratic irrationals

In this section we shall recount the beautiful relationship between quadratic forms, quadratic irrationals and closed geodesics on the modular surface (as used by SARNAK [17]).

Consider a primitive indefinite quadratic form e. g.

$$Q(x, y)=ax^2+bxy+cy^2$$

where a, b, c are coprime integers [denoted $(a, b, c)=1$] and the discriminant $d=b^2-4ac$ satisfies $d>0$ and d is not a perfect square. We say two such forms Q and Q' are equivalent if we can transform from one to the other under substitutions of the type

$$x'=\alpha x+\beta y$$
$$y'=\gamma x+\delta y \qquad \text{where} \quad \alpha, \beta, \gamma, \delta\in\mathbb{Z}, \quad \text{and} \quad \alpha\delta-\beta\gamma=1.$$

This relation partitions forms with the same discriminant $d>0$ into equivalence classes. Gauss showed that there are only finitely many distinct classes for any particular $d>0$ ([5], § 6).

Those substitutions which preserve a particular form $Q=[a, b, c]$ are called the *automorphs* of Q. There is a very elegant way representing these automorphs in terms of the solution of Pell's equation. Let $(u, t)\in\mathbb{Z}\times\mathbb{Z}\diagdown(0, 0)$ be any solution of Pell's equation

$$t^2-du^2=4.$$

(There exist infinitely many such solutions.) Then this solution defines an automorph for $Q=[a, b, c]$ by chosing

$$\alpha=1/2(t-bu), \qquad \beta=-cu,$$

$$\gamma = au, \qquad \delta = 1/2\,(t + bu).$$

Let (t_0, u_0) be the solution for which $\varepsilon_d = (t_0 + u_0\sqrt{d})/2$ is least. Then all solutions (u, d) are generated by

$$1/2\,(t + u\sqrt{d}) = \pm [1/2\,(t_0 + u_0\sqrt{d})]^n, \qquad n \in \mathbb{Z}.$$

(For an interesting account of Pell's tenuous connection with the equation thet bears his name, *see* [7].)

The following proposition is due to SARNAK [17].

PROPOSITION 4 (Sarnak). — *There is a one-to-one correspondence between closed geodesics on the modular surface and equivalence classes of quadratic forms. Furthermore, under this bijection the geodesic* γ *corresponding to an equivalence class with discriminant* $d > 0$ *has length* $2 \log \varepsilon_d$.

There is associated with a given quadratic form $[a, b, c]$ two real quadratic irrationals (algebraic numbers of degree at most 2) $\theta_1 < \theta_2$ which are the roots of the polynomial $P(z) = az^2 + bz + c$. The points $\theta_1 < \theta_2$ on the real line define a unique geodesic $\hat{\gamma}$ in $\mathbb{H}^+$ which meets the real line at $\hat{\gamma}(-\infty) = \theta_1$ and $\hat{\gamma}(+\infty) = \theta_2$. The induced (closed) geodesic γ on $\mathcal{M}$ is precisely the geodesic corresponding to $[a, b, c]$ in the above proposition.

We recall the standard fact that (reduced) quadratic irrationals are precisely those numbers whose continued fraction expansion is periodic ([9], p. 144) and ([18], § 3).

In summary, there is a natural one-one correspondence between:

(i) Quadratic irrationals $\theta_i = [a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]$ with i odd (or i even).

(ii) Equivalence classes of quadratic forms (of discriminant $d > 0$); and

(iii) Closed geodesics on the modular surface of length $2 \log \varepsilon_d$.

Thus we can induce a partial ordering on the quadratic irrationals $0 < \theta < 1$ by $\theta \leqslant \theta'$ if and only if $\varepsilon_d \leqslant \varepsilon_{d'}$.

(We remark that $\sqrt{d} \leqslant \varepsilon_d \ll \exp\sqrt{d}$ [7].)

Remark. — Another abstract application of the modular surface (which is unrelated to the above) is to index isomorphism classes of elliptic curves (or equivalently, homothety classes of lattices of $\mathbb{C}$) [20].

5. Distribution of quadratic irrationals

The quadratic irrationals in the unit interval are precisely those numbers with a periodic continued fraction expansion. Thus not only is the set of all such numbers dense in (0, 1) but this suggests that there is a high degree of structure in their distribution. Davenport and Schmidt chose a particular ordering to prove the following approximation result [6]: There exists $C>0$ such that for any $0<\alpha<1$ there are arbitrarily large $T>0$ and quadratic irrationals $\theta=\theta(T)$, corresponding to a quadratic form $[a, b, c]$ with $|a|, |b|, |c| < T^{1/3}$, satisfying $|\alpha-\theta| < C/T$.

We shall now consider the question of how evenly the quadratic irrationals are distributed in the unit interval (with respect to the ordering by ε_d).

A quadratic irrational $\theta=[a_1, \ldots, a_{2n}]$ corresponds to a pair of closed geodesic γ of length $2\log\varepsilon_d$. From the symbolic dynamics outlined in section 1 these closed geodesics correspond to periodic orbits for the suspended flow (where the corresponding base points for the continued fraction transformation (the base transformation) are $\{[a_i, \ldots, a_{2n}, a_1, \ldots, a_{i-1}]\}_{i=1}^{2n}$). In this context the results of section 2 can be used, as in section 3, as follows. Let $\mu_t\,(t>0)$ be the purely atomic probability measure on (0, 1) by equidistributing measure over quadratic irrationals $0<\theta<1$, with $\varepsilon_d\leqslant t$ i. e.

$$\mu_t=(\textstyle\sum_{\varepsilon_d\leqslant t}\delta_\theta)/p(t) \qquad \text{where} \quad P(t)=\text{Card}\{\theta\,|\,\varepsilon_d\leqslant t\}.$$

Then μ_t converges to the Gauss measure μ (in the weak* topology) as $t\to+\infty$. In this sense we can say the following.

THEOREM 3. — *When the quadratic irrationals $0<\theta<1$ are ordered according to ε_d then they are distributed according to the Gauss measure.*

In particular, quadratic irrationals are not equidistributed (i. e. distributed according to Lebesgue measure) although the limiting measure is *equivalent* to Lebesgue measure.

Following the lines of section 3 we can deduce additional information about the distribution of digits in the continued fraction expansions of quadratic irrationals.

Let $\theta=[a_1, \ldots, a_p]$ and denote: $\omega(\theta)=p$, the period of the continued fraction expansion; and $\alpha_k(\theta)$ the number of times $a_i=k$, $i=1, \ldots, p$. Then we have the following asymptotic avarages.

PROPOSITION 5:

(i) $$\left(\sum_{\varepsilon_d \leq t} \alpha_k(\theta)\right)/P(t) \sim \frac{1}{2}\left[\frac{6}{\pi^2}\log\left(\frac{(k+1)^2}{k(k+2)}\right)\right]\log t;$$

(ii) $$\left(\sum_{\varepsilon_d \leq t} \omega(\theta)\right)/P(t) \underline{\infty} \frac{1}{2}\left[\frac{6\log 2}{\pi^2}\right]\log t.$$

This proposition shows that although the values of $\alpha_k(\theta)$ and $\omega(\theta)$ are erratically distributed with respect to ε_d, "on the average" $\omega(\theta)$ and $\alpha_k(\theta)$ grow as the logarithm of ε_d [and in particular $\alpha_k(\theta)$ and $\omega(\theta)$ are of order $\varepsilon_d \log \varepsilon_d$].

We recall Khintchine's classical result that for almost all $x=[a_1, a_2, a_3, \ldots]$

(3) $$\frac{1}{N}\sum_{i=1}^{N} \log a_i \rightarrow \log \prod_{k=1}^{\infty}\left[\frac{(k+1)^2}{k(k+2)}\right]^{\log k/\log 2}$$

If we take $\beta(\theta)=\sum_{i=1}^{P} \log a_i$ [and set $f([x_0, x_1, \ldots])=\log x_0$ in section 2] then we have the following proposition.

PROPOSITION 6:

$$\left(\sum_{\varepsilon_d \leq t} \beta(\theta)\right)/\left(\sum_{\varepsilon_d \leq t} \omega(\theta)\right) \rightarrow \log \prod_{k=1}^{\infty}\left[\frac{(k+1)^2}{k(k+2)}\right]^{\log k/\log 2}.$$

Of course, since the algebraic numbers are denumberable (and thus form a set of zero Lebesgue and Gauss measure) these two results are independent. However, Proposition 6 supports our choice of ordering of quadratic irrationals.

REFERENCES

[1] ADLER (R.) and FLATTO (L.). – Cross section maps for geodesic flows, I, in *Ergodic Theory and Dynamical Systems II*, KATOK (A.) Ed., Birkhäuser, Stuttgart, 1982.

[2] ARNOLD (V. I.) and AVEZ (A.). – *Ergodic problems of classical mechanics*, Benjamin, New York, 1968.

[3] BILLINGSLEY (P.). – *Ergodic theory and information*, Wiley, New York, 1965.

[4] BOWEN (R.). – The equidistribution of closed geodesics, *Amer. J. Math.*, Vol. 94, 1972, pp. 413-423.

[5] DAVENPORT (H.). – *Multiplicative number theory*, G.T.M., 74, Springer, New York, 1980.