ON THE ASYMPTOTIC DISTRIBUTION OF CLOSED GEODESICS ON COMPACT RIEMANN SURFACES

BY

BURTON RANDOL

ABSTRACT. The set of lengths of closed geodesics on a compact Riemann surface is related to the Selberg zeta function in a manner which is evocative of the relationship between the rational primes and the Riemann zeta function. In this paper, this connection is developed to derive results about the asymptotic distribution of these lengths.

Suppose S is a compact Riemann surface, regarded as a quotient of the upper half-plane H^+ by a discontinuous group Γ. We assume that H^+ is endowed with the metric $y^{-2}((dx)^2 + (dy)^2)$, and we denote the volume of S by A. As is well known, the closed geodesics on S are in one-to-one correspondence with the conjugacy classes of Γ. In particular, if γ is an element of Γ other than the identity, it can, by a change of coordinates, be put into the normal form $z \to N_\gamma z$, with $N_\gamma > 1$. The number N_γ, which is evidently the same within a conjugacy class, is called the norm of the element. The quantity $\log N_\gamma$ is then the length of the closed geodesic corresponding to the conjugacy class of γ (see, e.g., [4]). Huber [3] has investigated the asymptotic distribution of the closed geodesics, and found that their statistical behavior is heavily influenced by the "small" eigenvalues of the Laplace operator on S. That is, by the eigenvalues, if any, in $(0, \frac{1}{4})$ for the problem $\Delta f + \lambda f = 0$ on S. Now questions of this sort can be very easily treated using the standard theory of the Selberg zeta function, and from this point of view, the analysis is very evocative of the relationship between the distribution of rational primes and the Riemann zeta function. In this paper, we will outline such a treatment.

We begin with a few definitions. From now on, the symbol γ will represent a closed geodesic on S. The symbol γ^n will mean the geodesic obtained from γ by n-fold iteration. A closed geodesic will be called primitive if it is not a positive integral power of any geodesic other than itself. By analogy with the theory of rational primes, we will define $\Lambda(\gamma)$ to be $\log N_{\gamma_0}$, where $\gamma = \gamma_0^n$, with γ_0 primitive.

Now the Selberg zeta function $Z(s)$ for the group Γ is an entire function of order 2, having a sequence of zeros at $1, 0, -1, -2, \ldots$, with the zero at $s = 1$ simple, and having additional zeros in the critical strip $0 < \text{Res} < 1$. The zeros in the critical strip are located at points which are solutions of the equations $s(1 - s) = \lambda_n$, where λ_n ranges through the sequence of eigenvalues, omitting $\lambda_0 = 0$, for the problem $\Delta f + \lambda f = 0$ on $\mathfrak{S}$. The multiplicity of such a zero is the same as the multiplicity of the corresponding eigenvalue ([1], [4], [8]). The expression $s(1 - s)$ is symmetric about the point $s = \frac{1}{2}$, so the zeros of $Z(s)$ within the critical strip occur in symmetric pairs about $s = \frac{1}{2}$. These zeros are all located on the union of the real segment $0 < s < 1$ with the critical line $\text{Res} = \frac{1}{2}$. It is possible to exhibit surfaces for which zeros occur on $\frac{1}{2} < s < 1$ [5], so the counterpart of the Riemann hypothesis for $Z(s)$ is not necessarily true. Such zeros correspond to eigenvalues in $(0, \frac{1}{4})$.

For $\text{Res} > 1$, the logarithmic derivative of $Z(s)$ is given by the formula

$$\frac{Z'(s)}{Z(s)} = \sum_{\gamma} \Lambda(\gamma)\left(1 - N_{\gamma}^{-1}\right)^{-1} N_{\gamma}^{-s}$$

[4, p. 240], [8]. From this we immediately deduce the following lemma:

LEMMA 1. *For* $\text{Res} > 1$, $Z'(s)/Z(s) = \sum_{\gamma}\Lambda(\gamma)N_{\gamma}^{-s} + Z'(s + 1)/Z(s + 1)$.

PROOF.

$$\frac{Z'(s)}{Z(s)} = \sum_{\gamma} \Lambda(\gamma)\left(1 - N_{\gamma}^{-1}\right)^{-1} N_{\gamma}^{-s}$$

$$= \sum_{\gamma} \Lambda(\gamma)N_{\gamma}^{-s} + \sum_{\gamma} \Lambda(\gamma)(N_{\gamma} - 1)^{-1}N_{\gamma}^{-s}$$

$$= \sum_{\gamma} \Lambda(\gamma)N_{\gamma}^{-s} + \sum_{\gamma} \Lambda(\gamma)\left(1 - N_{\gamma}^{-1}\right)^{-1}N_{\gamma}^{-(s+1)}. \quad \text{Q.E.D.}$$

LEMMA 2. *Let* H *be a half-plane of the form* $\text{Res} < -\varepsilon < 0$, *minus the union of a sequence of congruent disks about the negative integers. Then there exists a constant* $C > 0$, *such that for* $s \in H$, $|Z'(s)/Z(s)| < C|s|$.

REMARK. Such an estimate is valid in more general regions, but the present form of the lemma will serve here.

PROOF. $Z(s)$ satisfies the functional equation

$$Z(1 - s) = Z(s)\exp\left(-A \int_0^{s-1/2} w\tan \pi w \, dw\right)$$

[8] which implies that $-Z'(1 - s)/Z(1 - s) = Z'(s)/Z(s) - A(s - \frac{1}{2})\tan \pi(s - \frac{1}{2})$. Since $Z'(s)/Z(s)$ is bounded in any half-plane of the form $\text{Res} > 1 + \varepsilon$, and $\tan \pi(s - \frac{1}{2})$ is bounded in the complement of a

union of congruent disks about the integers, this implies the desired result.

LEMMA 3. *Given real numbers a and b, there exists a sequence $T_n \uparrow \infty$ and a constant $C > 0$, such that $|Z'(x + iT_n)/Z(x + iT_n)| \leqslant CT_n^2$, for $a < x < b$.*

PROOF. $Z(s)$ is an entire function of order 2, and the reciprocals of the roots on the critical line are not square summable. Accordingly, $Z(s)$ has a canonical product expansion

$$(1) \qquad Z(s) = e^{as^2 + bs + c} s^n \prod_\rho (1 - s/\rho) e^{s/\rho + (1/2)(s/\rho)^2},$$

where ρ runs over the nonzero roots of $Z(s)$, and a, b, c, and n are constants whose exact value is not important for our purpose. One deduces easily from (1) that

$$(2) \qquad \frac{Z'(s)}{Z(s)} = 2as + b + \frac{n}{s} + \sum_\rho s^2 \rho^{-2} (s - \rho)^{-1}.$$

Supposing now $t_0 > 2$ fixed, consider the segment of the critical line $\frac{1}{2} + it$, with $t_0 - 1 < t < t_0 + 1$. Define $N(t)$ to be the number of roots on the critical line on the interval $\frac{1}{2} + ix$, with $0 < x < t$. Now it is known [6] that $N(t) = (A/4\pi)t^2 + O(t)$, so we easily conclude that the number of roots on the segment is $O(t_0)$. By the Dirichlet principle, there exists a $\frac{1}{2} + iT$ in the segment whose distance from both endpoints is greater than $\frac{1}{4}$, and whose distance from any root is greater than C/T, for some fixed $C > 0$. We conclude that the portion of the sum $\sum_\rho s^2 \rho^{-2} (s - \rho)^{-1}$ corresponding to the roots in the segment for the choice $s_x = x + iT$ is $O(T^2)$, since $|s_x^2 \rho^{-2}| = O(1)$ for these ρ, when $a < x < b$. To deal with the segments $\frac{1}{2} + it$ ($0 < t < t_0 - 1$), and $\frac{1}{2} + it$ ($t_0 + 1 < t < \infty$), we proceed as follows. The portions of the sum $\sum_\rho s_x^2 \rho^{-2} (s - \rho)^{-1}$ corresponding to the first and second segments, respectively, can be written

$$s_x^2 \int_0^{t_0 - 1} \left(\tfrac{1}{2} + it\right)^{-2} \left(s_x - \tfrac{1}{2} - it\right)^{-1} dN(t)$$

and

$$s_x^2 \int_{t_0 + 1}^\infty \left(\tfrac{1}{2} + it\right)^{-2} \left(s_x - \tfrac{1}{2} - it\right)^{-1} dN(t).$$

Recalling that $N(t) \sim (A/4\pi)t^2$, we easily conclude that both of these expressions are $O(T^2)$. Q.E.D.

Now suppose k is an integer $\geqslant 2$, and T, $c > 1$. Then

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Z'(s)}{Z(s)} s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s \, ds$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\sum_\gamma \Lambda(\gamma) N_\gamma^{-s} \right) s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s ds$$

$$(3) \qquad + \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Z'(s+1)}{Z(s+1)} s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s ds$$

$$= \frac{1}{k!} \sum_{N_\gamma < T} \Lambda(\gamma)(1 - N_\gamma/T)^k$$

$$+ \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Z'(s+1)}{Z(s+1)} s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s ds,$$

by [2, Theorem 40].

The integrals which occur at the beginning and end of (3) are of the form

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Z'(s+\delta)}{Z(s+\delta)} s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s ds, \quad \text{with } \delta = 0 \text{ or } 1.$$

The function $Z'(s+\delta)/Z(s+\delta)$ is meromorphic, with simple poles located at the set obtained by translating the zeros of $Z(s)$ by $-\delta$. By virtue of Lemmas 2 and 3, it is permissible, in evaluating this type of integral, to shift the line of integration into the half-plane Res $< -\delta$, provided the new line does not pass through a pole of the integrand, and provided we take account of the residues thereby picked up. We thus easily find, if the line of integration is shifted leftward to pass through a suitable point $c' < -1$, that

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Z'(s+\delta)}{Z(s+\delta)} s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s ds$$

$$= \sum_{\substack{\rho \in S_\delta^k \\ \text{Re}\rho > c'}} c_\rho^k(\delta) T^\rho$$

$$+ \frac{1}{2\pi i} \int_{c'-i\infty}^{c'+i\infty} \frac{Z'(s+\delta)}{Z(s+\delta)} s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s ds,$$

where S_δ^k is the set of poles of $(Z'(s+\delta)/Z(s+\delta))s^{-1}(s+1)^{-1} \cdots (s+k)^{-1}$, and $c_\rho^k(\delta)$ is the residue at $s = \rho$.

Now if $T > 1$, it follows from Lemma 2 that the last integral tends to zero as $c' \to -\infty$ through, say, points of the form $-(n + \frac{1}{2})$, so we conclude that for $T > 1$,

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Z'(s+\delta)}{Z(s+\delta)} s^{-1}(s+1)^{-1} \cdots (s+k)^{-1} T^s ds = \sum_{\rho \in S_\delta^k} c_\rho^k(\delta) T^\rho.$$

We thus obtain the following result from (3):

THEOREM 1. *Suppose $k > 2$, $T > 1$. Then*

$$\sum_{N_\gamma < T} \Lambda(\gamma)(T - N_\gamma)^k = k! \sum_{\rho \in S_0^k} c_\rho^k(0)T^{k+\rho} - k! \sum_{\rho \in S_1^k} c_\rho^k(1)T^{k+\rho}.$$

Now if we define $\Psi_0(T) = \sum_{N_\gamma < T}\Lambda(\gamma)$, and $\Psi_j(T) = \int_0^T \Psi_{j-1}(t)\,dt$ $(j = 1, 2, \ldots,)$, it is well known that $\Psi_j(T) = (j!)^{-1}\sum_{N_\gamma < T}\Lambda(\gamma)(T - N_\gamma)^j$. Theorem 1 then becomes

THEOREM 1'. *Suppose $k > 2$, $T > 1$. Then*

$$\Psi_k(T) = \sum_{\rho \in S_0^k} c_\rho^k(0)T^{k+\rho} - \sum_{\rho \in S_1^k} c_\rho^k(1)T^{k+\rho}.$$

In particular, bearing in mind the location of the poles of $Z'(s)/Z(s)$ and the fact that the residue at $s = 1$ is 1, we find that for $k > 2$, $\Psi_k(T) = ((k + 1)!)^{-1}T^{k+1} + O(T^{k+\rho_0})$, where ρ_0 is either $\frac{1}{2}$, if $Z(s)$ has no roots in $(\frac{1}{2}, 1)$, or the largest root in this interval, if it does.

It is easy, using standard differencing techniques and the monotonicity of $\Psi_0(T)$, to pass from Theorem 1' to information about $\Psi_0(T)$ itself. In the process some information may be lost, however. The simplest technique yields the following result:

THEOREM 2 (CF. THE NACHTRAG TO [3]). *Suppose $Z(s)$ has roots $\alpha_1, \ldots, \alpha_n$ in $(\frac{3}{4}, 1)$. Then there exist constants $c_1, \ldots, c_n$, such that*

$$\Psi_0(T) = T + c_1 T^{\alpha_1} + \cdots + c_n T^{\alpha_n} + O(T^{3/4}).$$

If $Z(s)$ has no roots in $(\frac{3}{4}, 1)$, $\Psi_0(T) = T + O(T^{3/4})$.

REMARKS. 1. This result has a somewhat unsatisfactory appearance inasmuch as any roots which lie in $(\frac{1}{2}, \frac{3}{4}]$ do not figure explicitly. It would be quite interesting to determine the extent to which the statement can be sharpened.

2. One can, by imitating the standard procedure for going from $\Psi_0(T)$ to $\pi(T)$ in prime number theory, easily deduce from Theorem 2 that $\sum_{N_\gamma < T} 1 = liT$, $T + O(T^\beta/\log T)$, where $\beta = \max(\frac{3}{4}, \alpha_1, \ldots, \alpha_n)$.

PROOF OF THEOREM 2. For a given positive number d, define the second difference operator Δ, by setting $\Delta f(x) = f(x + 2d) - 2f(x + d) + f(x)$. Now it follows from Theorem 1', that

$$\Psi_2(T) = \frac{1}{6} T^3 + \underset{\mathrm{I}}{\sum} c_\rho^2(0)T^{2+\rho} + \underset{\mathrm{II}}{\sum} c_\rho^2(0)T^{2+\rho}$$
$$+ \underset{\mathrm{III}}{\sum} c_\rho^2(1)T^{2+\rho} + \underset{\mathrm{IV}}{\sum} c_\rho^2(1)T^{2+\rho} + O(T^{3/4}),$$

where the first sum ranges over the finite set of elements of S_0^2 which have real part greater than $-\frac{5}{4}$ and do not lie on the line Res $= \frac{1}{2}$, the second sum ranges over the elements of S_0^2 which lie on Res $= \frac{1}{2}$, the third sum ranges over the finite set of elements of S_1^2 which have real part greater than $-\frac{5}{4}$ and do not lie on the line Res $= -\frac{1}{2}$, and the fourth sum ranges over the elements of S_1^2 which lie on Res $= -\frac{1}{2}$.

Now by computation, $\Delta(\frac{1}{6}T^3) = Td^2 + d^3$. Also, since in general $\Delta f(x) = \int_x^{x+d} dy \int_y^{y+d} f''(t)\,dt$, we obtain the mean value theorem $\Delta T^\theta = d^2\theta(\theta - 1)\tilde{T}^{\theta-2}$, where $\tilde{T} \in [T,\, T+2d]$. Since $\Psi_0(T)$ is nondecreasing, it follows from the second equality that

(4) $\Psi_0(T) \leqslant d^{-2}\Delta\Psi_2(T) \leqslant \Psi_0(T + 2d)$

It also follows from the equalities for the difference operator that

$$d^{-2}\Delta\left(\tfrac{1}{6}T^3 + \sum_{\mathrm{I}} c_\rho^2(0) T^{2+\rho} + \sum_{\mathrm{III}} c_\rho^2(1) T^{2+\rho} \right)$$

$$= T + c_1 T^{\alpha_1} + \cdots + c_n T^{\alpha_n} + O(T^{3/4}) + O(d),$$

where the α_j's are as described in the statement of the theorem, and

$$c_j = (2 + \alpha_j)(1 + \alpha_j)c_{\alpha_j}^2(0).$$

In order to estimate $d^{-2}\Delta(\Sigma_{\mathrm{II}} c_\rho^2(0) T^{2+\rho})$, we need two estimates for $\Delta(c_\rho^2(0) T^{2+\rho})$. To begin with, it is evident that for such a ρ, $d^{-2}\Delta(c_\rho^2(0) T^{2+\rho}) = O(d^{-2}|\rho|^{-3}T^{5/2})$. On the other hand, it follows from the mean value theorem for the difference operator that we also have the estimate $d^{-2}\Delta(c_\rho^2(0) T^{2+\rho}) = O(|\rho|^{-1}T^{1/2})$. Now, using the previously introduced counting function $N(t)$ for the zeros on the critical line, it follows that for $M > 1$,

$$d^{-2}\Delta\left(\sum_{\mathrm{II}} c_\rho^2(0) T^{2+\rho} \right) = O\left(T^{1/2}\int_1^M t^{-1}dN(t) + d^{-2}T^{5/2}\int_M^\infty t^{-3}dN(t) \right)$$

$$= O(MT^{1/2} + M^{-1}d^{-2}T^{5/2}),$$

since $N(t) \sim (A/4\pi)t^2$. If we now set $M = T^{1/4}$, and $d = T^{3/4}$, we conclude that $\Delta(\Sigma_{\mathrm{II}} c_\rho^2(0) T^{2+\rho}) = O(T^{3/4})$. In almost exactly the same way, we find that $d^{-2}\Delta(\Sigma_{\mathrm{IV}} c_\rho^2(1) T^{2+\rho}) = O(T^{-1/4})$. Thus $d^{-2}\Delta\Psi_2(T) = T + c_1 T^{\alpha_1} + \cdots + c_n T^{\alpha_n} + O(T^{3/4})$. Returning to (4), we conclude from the left side of the inequality that $\Psi_0(T) \leqslant T + c_1 T^{\alpha_1} + \cdots + c_n T^{\alpha_n} + O(T^{3/4})$, and from the right that $T + c_1 T^{\alpha_1} + \cdots + c_n T^{\alpha_n} + O(T^{3/4}) \leqslant \Psi_0(T)$, which proves Theorem 2.

CONCLUDING REMARK. It would be interesting to determine the extent to which the methods of this paper extend to more general spaces (cf. [1] and [7]).

References

1. R. Gangolli, *Zeta functions of Selberg's type for compact space forms of symmetric spaces of rank* 1 (preprint).

2. G. H. Hardy and M. Riesz, *The general theory of Dirichlet's series*, Cambridge Univ. Press, Cambridge, 1952.

3. H. Huber, *Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen.* II, Math. Ann. **142** (1960/61), 385–398. MR **23** #A3845.

4. H. P. McKean, *Selberg's trace formula as applied to a compact Riemann surface*, Comm. Pure Appl. Math. **25** (1972), 225–246.

5. B. Randol, *Small eigenvalues of the Laplace operator on compact Riemann surfaces*, Bull. Amer. Math. Soc. **80** (1974), 996–1000.

6. _____, *The asymptotic behavior of the zeros of zeta functions of Selberg's type* (preprint).

7. D. Ruelle, *Zeta functions for expanding maps and Anosov flows* (preprint).

8. A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) **20** (1956), 47–87. MR **19**, 531.

INSTITUT DES HAUTES ÉTUDES SCIENTIFIQUES, BURES-SUR-YVETTE, FRANCE

DEPARTMENT OF MATHEMATICS, CITY UNIVERSITY OF NEW YORK, GRADUATE SCHOOL AND UNIVERSITY CENTER, NEW YORK, NEW YORK, 10036