COMPOSITIO MATHEMATICA

THOMAS A. SCHMIDT
MARK SHEINGORN

On the infinite volume Hecke surfaces

Compositio Mathematica, tome 95, nº 3 (1995), p. 247-262
<http://www.numdam.org/item?id=CM_1995__95_3_247_0>

Compositio Mathematica **95**: 247–262, 1995.

On the infinite volume Hecke surfaces

THOMAS A. SCHMIDT* & MARK SHEINGORN†

**Oregon State University, Corvallis, OR 97331*; †*Baruch College, Cuny, New York, NY 10010*

Received 28 September 1993; accepted in final form 6 January 1994

Abstract. An infinite volume Hecke surface, $G_\lambda \backslash \mathscr{H}$, is the Riemann surface associated with the Hecke triangle group of translation length $\lambda > 2$. This paper: (i) gives an algorithm producing the length spectrum for each of these surfaces employing an unramified double cover (by way of illustration, we tabulate the shortest 25 geodesics for the case $\lambda = 4$. We know of no other infinite volume surface for which this data exists). (ii) Establishes the existence of a Hall ray for *open* geodesics on $G_\lambda \backslash \mathscr{H}$. Our proof requires that $\lambda > \sqrt{8}$. The existence of a Hall ray means (after Haas) that the set consisting of the highest penetration of each geodesic into the fundamental horocycle is dense in some real half-line. It is necessary to use open geodesics—there is no Hall ray otherwise. This is in marked contrast to the finite volume case, as we prove.

1. Introduction

The *length spectrum* of a Riemann surface, the sequence of lengths of closed geodesics, has been the object of intense study since at least the advent of the Selberg Trace Formula, which involves this spectrum in a crucial manner. For more on these matters, we refer the reader to Peter Buser's recent book [Bu].

There has also been recent activity in the study of the length spectrum (especially as moduli) in more general differential geometric settings, for example see Croke [C] and Lalley [L]. Chiefly, this paper offers the explicit computation of the length spectrum of a one real parameter family of infinite volume surfaces with ramification—the Hecke triangle surfaces of the second kind. Most previous explicit computation has taken place on a single finite volume surface, although our previous paper [S-S] discusses computations for all finite volume Hecke triangle groups. We know of no other infinite volume surface for which such computations exist in the literature. (However, the referee has kindly pointed out that there are algorithms and (unpublished) computations for surfaces with a positive lower bound for the injectivity radius together with a free fundamental group. In this paper said lower bound is zero due to the cusp, and the surface is ramified.) Fittingly perhaps, it turns out that for these examples the shortest geodesic is the boundary geodesic.

Our second set of results concerns the existence of a Hall ray (defined below) for various classes of geodesics. In the case of the modular group, the existence of the Hall ray for closed geodesics is equivalent to the existence of the Hall ray for the Markoff spectrum (this last depends on Hall's theorem to the effect that every real number in [0, 1] may be represented as the sum of two regular continued fractions (CFs) with partial denominators less than 4; thus the name Hall ray). This argument goes over, *mutatis mutandis*, to the finite volume Hecke triangle groups. Not unexpectedly, the role of the CFs is played by the (Rosen-like [R]) next λ-CFs already used in [S-S].

But, as we prove, Hecke groups of the second kind do not have a Hall's ray if we are restricted to the use of (the very sparse set of) closed geodesics. We must also employ open geodesics—indeed geodesics with ends in the funnel (almost all geodesics have this property). Here the technique of establishing the existence of the Hall ray is geometric, not continued fraction based.

Haas-Series [H-S] referred to the sequence of closed geodesic penetrations as the *Cohn spectrum*. As above, the Cohn spectrum and the Markoff spectrum coincide for finite volume Hecke groups. The remarks of the previous paragraph show that these notions bifurcate for infinite volume (second kind) Hecke groups.

We thank A. Haas for pointing out an error in the original version of this work and the referee for help in improving the presentation.

2. Geometry and length spectrum of $H_\lambda \backslash \mathscr{H}$

The Hecke group G_λ is the Fuchsian group which, when considered as acting upon the Poincaré upper half-plane, is generated by $S: z \mapsto z + \lambda$ and $T: z \mapsto -1/z$. When $\lambda > 2$, G_λ is of the second kind; that is, it is a group of infinite volume. We consider the subgroup H_λ of index 2 in G_λ defined as follows:

$$H_\lambda =: \left\langle \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} \right\rangle$$

These generators are obviously parabolic; we denote them P_1 and P_2, respectively. The signature of $H_\lambda \backslash \mathscr{H}$ is $(0; \infty, \infty; 1)$; i.e., a sphere with two punctures and one hole. P_1 and P_2^{-1} are conjugated into one another by T, which means that this is a non-trivial rotational isometry of the surface that interchanges the cusps. This rotation fixes (the projection of) i and interchanges those of

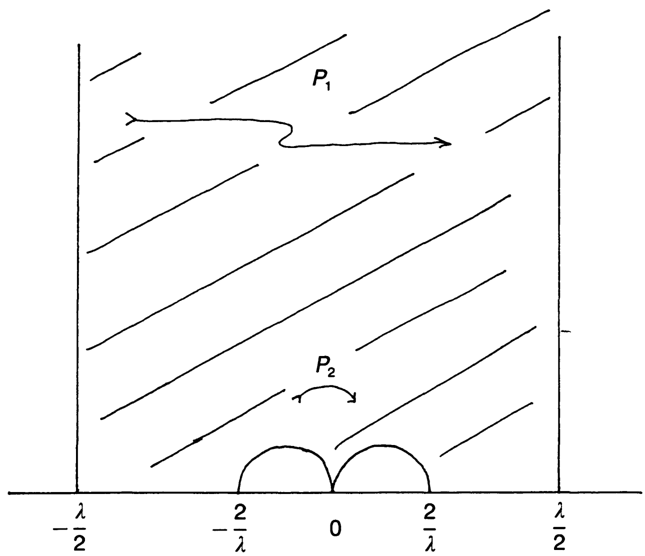

Fig. 1. Fundamental region for H_λ.

$\pm 2/\lambda + i\sqrt{1 - 4/\lambda^2}$, which are "antipodal" on the boundary geodesic. These last are, of course, identical when projected to $G_\lambda \backslash \mathscr{H}$.

The virtues of working this surface are that H_λ is freely generated by P_1 and P_2—thus, H_λ is a manifold, not merely an orbifold. This feature makes the manner in which short geodesics are constructed quite "visual". A fundamental region for this group is given just below. It should be noted that, the boundary geodesic, the hole, and the projection of i (which is an ordinary point on $H_\lambda \backslash \mathscr{H}$, of course) are symmetrically placed with respect to the cusps. (More precisely each is fixed, but only i point-wise, by T, which is an isometry of $H_\lambda \backslash \mathscr{H}$.)

We are going to give bounds for the trace of cyclically reduced words in P_1 and P_2. This is done by estimating the length contributed to a closed geodesic by: (i) the presence of a term P_1^n in its word ($|n| > 1$). (The contribution of a term in P_2^n will be the same since the conjugation by T, an isometry, interchanges these.) (ii) The presence of consecutive terms of the form $P_1^s \cdot P_2^{-t}$; $s, t = \pm 1$. Note that a closed geodesic must remain inside the Nielsen convex region $\mathscr{N}$ for G_λ, else it is captured by the funnel and cannot be closed.

Consider the array of fundamental regions for H_λ given by the one above and all translates by $S^{\pm m}$ for integral m. The array has a sequence of pairs of sides terminating at (*anchored at*) $\pm m\lambda$. The presence of a term P_1^n in a word is equivalent to the presence of a segment of the corresponding geodesic connecting two of the above pairs anchored at points $n\lambda$ apart. Such a segment has length at most $d(1/\lambda + i/\lambda, n\lambda - 1/\lambda + i/\lambda) = 2\log n + 4\log\lambda + O(1)$, by

Beardon ([Be], Eq. 7.2.1(i)) and also the height of the boundary geodesic (given below). Similarly, it has length at least $d(i, n\lambda+i) = 2\log n + 2\log\lambda + O(1)$.

Note that the sign of n merely determines the direction in which the segment is traversed, not its length and that this estimate works even when $n = 1$.

We have shown:

LEMMA 1. *The length of the geodesic given by a cyclically reduced hyperbolic word*

$$A = \prod_{j=1}^{k} P_1^{a_j} P_2^{b_j}$$

is less than

$$C \log \lambda \sum_j \min(2, |a_j| + |b_j|) + D \sum_j \log(|a_j| + |b_j|).$$

It is greater than $E\,\Sigma_j \log(|a_j| + |b_j|)$. *For each run of n consecutive j with $|a_j| = |b_j| = 1$, $Fn\log\lambda$ is added to the length. (Here B, C, D, E and F are independent of λ.)*

Of course this is equivalent to a corresponding bound on the trace of A. In creating our algorithm for the length spectrum of G_λ, we need only resolve any ambiguities arising from the above statement being with respect to a double cover of our basic groups. That is, we must locate those words ($A \in H_\lambda$, listed out as above) which have square roots in G_λ. This is easily done via the same technique we employed in [S-S]: i.e., obtaining a CF expansion of a fixed point of A, and checking for a shorter period.

Using Lemma 1, one can create a complete list of shortest geodesics on $H_\lambda\backslash\mathscr{H}$. Here is a way to visualize the process. Geodesics on $H_\lambda\backslash\mathscr{H}$ may be described as winding about a cusp, then traversing at least half of the boundary geodesic (with more winding about the hole a possibility), and then winding about the other cusp, and so on. Each of these processes adds an estimable amount of length to the geodesic. (Cusp loops with corresponding behavior at the cusps are isometric via T.) Here is the data precisely:

Boundary geodesic: This is easily shown to be fixed by the (primitive in H_λ) hyperbolic

$$(P_1 P_2^{-1}) = \begin{pmatrix} \lambda & -1 \\ 1 & 0 \end{pmatrix}^2,$$

the first power of the latter being in $G_\lambda - H_\lambda$. The length is thus $\approx 2\log\lambda$. Therefore winding about the hole m times adds $\approx 2m\log\lambda$ to the length. (Here m may be a half-integer.)

Loops about a cusp: Up to a constant independent of λ, looping about a cusp m times adds no more than $d(i/\lambda, m\lambda + i/\lambda)$ and no less than $d(i, m\lambda + i)$. (The former comes from the fact that the boundary geodesic runs between

$$\pm\lambda/2 + i\sqrt{\lambda^2/4 - 1} \quad \text{and} \quad \frac{\pm\lambda + i\sqrt{\lambda^2 - 4}}{\lambda^2 - 2}.$$

Geodesic loops must stay "above" this geodesic, else they are captured by the funnel.) From this it is easy to compute the added length as $\approx \log m + 2\log\lambda$.

As an example, $P_1^m \cdot P_2^n$ is just $P_1^{m-1} \cdot P_1 P_2^{-1} \cdot P_2^{n+1}$. This is $n - 1$ trips about ∞, half the boundary geodesic, and $m + 1$ trips about zero. It is easy to directly compute the trace of $P_1^m \cdot P_2^n$, and this has the same order of magnitude (with respect to m and n) as the estimates from Lemma 1.

3. The length spectrum of $G_\lambda \backslash \mathscr{H}$

With an obvious abuse of notation, let S and T be the usual matrix representatives of the above elements S and T. Let $U = ST$; we define $g_i = U^{i-1}S$ and $h_i = Tg_i^{-1}T$. In fact, h_i is the transpose of g_i. Let $\mathscr{S}_m = \{g_i, h_i\}_{i=1}^m$ and $\mathscr{S} = \bigcup_m \mathscr{S}_m$.

We say that a word W in S and T is *T-reduced* if T appears in W to at most the first power.

LEMMA 2. *Each conjugacy class of G_λ, other than those of T and $\{U^j\}_{j=1}^\infty$ has a representative which is a word in $\mathscr{S}$ with only positive exponents.*

Proof. The finite covolume Hecke groups are of the form G_{λ_q} ($=: G_q$, by abuse of notation), with $\lambda = \lambda_q = 2\cos\pi/q$ and $q \geqslant 3$ an integer. The relations in G_q are $T^2 = U^q = I$. These give rise to $g_{[q/2]+j} = h_j$ for $j = 1, \ldots, [q/2]$. With q and λ fixed, we have the homomorphism

$$\begin{aligned} \phi_q &: G_\lambda \to G_q \\ S &\mapsto S \\ T &\mapsto T \end{aligned}$$

There is an injective set-map (but not a homomorphism) which takes a T- and U-reduced W of G_q to the same word in the generators of G_λ.

$$\begin{aligned} i_q &: G_q \to G_\lambda \\ W &\mapsto W \end{aligned}$$

The ϕ_q are not injective, but are stable in the sense that for any reduced W in G_λ, there is an integer $Q(W)$ such that for all $q > Q(W)$, $W = i_q(\phi_q(W))$. Indeed, $Q(W)$ is simply the largest exponent of U or U^{-1} to be found in W; the meaning here is that $\phi_q(W)$ is reduced in G_q with respect to $T^2 = I$ and $U^q = I$. We note that $Q(i_q(A_q)) \leqslant q$ for any reduced A_q of G_q.

Given reduced W of G_λ, let $q > Q(W)$ be odd, then [S-S] shows that the G_q-conjugacy class of $\phi_q(W)$ has a member expressible in positive terms of $\{g_{q,i}, h_{q,i}\}_{i=1}^{[q/2]}$. Let $A_q\phi_q(W)A_q^{-1}$ be such a conjugate. Let $A = i_q(A_q)$. Since $g_j = i_q(g_{q,j})$ and similarly for h_j, we are done.

REMARK. Although the above could be proved in a more direct manner, the stratification of G_λ by q-stably reduced sets, the $i_q(G_q)$, is the avatar in the structure of the infinite volume G_λ of what we called the "q principle" in [S-S]. We are stressing that a T- and U-reduced word for G_q is also such a word for $G_{q'}$ for all $q' > q$. Indeed, for all G_λ with $\lambda > \lambda_q$. And since $U^{\pm 1}$ appears in a given T-reduced word W of G_λ to a maximal exponent, W may be thought of as being "in" G_q for all $q > Q(W)$. For the G_q, the manner in which the geometric aspects of these images of W depend on q is exactly what we called the q principle. We take the opportunity here to point out that the fact that the theta group, $G_{q=\infty}$ is the limit of the G_q has been given the name of "discrete deformation" and the fact that every finitely generated Fuchsian group with cusps admits such a limiting sequence of Fuchsian groups with a chosen cusp of the original group being replaced by elliptics of increasing order has been shown by J. Wolfart [W].

The importance of $\mathscr{S}$ is that all of its elements may be taken as matrices with *non-negative* entries. In particular, traces of products (other than those of the form g_1^m or h_1^m) increase. The ideas of B. Fine [F] can be applied, as in [S-S], to obtain trace-class representatives in order of size of trace and thereby to list out all closed geodesics in order of length.

The only closed geodesic which does not arise from an $\mathscr{S}$-word is the closed geodesic from

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

In [S-S] it was shown that $\operatorname{tr}(g_{q,j}) > \operatorname{tr}(g_{q,j-1})$ for all $2 \leqslant j \leqslant [q/2]$. As $\operatorname{tr}(h_j) = \operatorname{tr}(g_j)$, the same is true for the h_j.

Call the sum of the exponents in a reduced word W of $\mathscr{S}$ the block length of W. For G_q, the hyperbolic elements of block length 1 give the shortest geodesics. This is no longer the case for G_λ. Indeed, the boundary geodesic corresponding to U is the shortest geodesic on G_λ. But, there are also hyperbolics of higher block length which have smaller trace than some of the g_j.

LEMMA 3. *Let W be a reduced positive $\mathscr{S}$-word and $i < j$ be integers, then* $\operatorname{tr}(g_i W) < \operatorname{tr}(g_j W)$. *The hyperbolic element of smallest trace of block length l is $g_1^{l-1} g_2$.*

Proof. Let

$$g_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}.$$

We use induction, hence we prove the first statement with j replaced by $i + 1$. Since $a_i = d_i$ for all i, it is sufficient to prove that $a_{i+1} > a_i$, $b_{i+1} > b_i$ and $c_{i+1} > c_i$. However, $a_i > c_i$ is shown in [S-S], and since $a_{i+1} = b_i$ and $c_{i+1} = a_i$, we are done. The second statement follows directly from the first—note that the element $h_1^{l-1} h_2$ also has this trace.

EXAMPLE. The shortest geodesics when $\lambda = 4$.

The shortest geodesic is the boundary geodesic, as U is of trace λ. Now, $\operatorname{tr}(g_2) = 2\lambda$ and $\operatorname{tr}(g_3) = 2\lambda^2 - 2$. Consider $g_1^l g_2$, which is of trace $(l + 2)\lambda$. This trace is less than that of g_2 whenever $l < 2\lambda^2 - 2/\lambda$. We also consider the trace for g_3 and so on, as well as words in both g_i and h_j. In particular of these latter, one has $\operatorname{tr}(g_1^n h_1^m) = nm\lambda^2 + 2$.

Thus when $\lambda = 4$, using that the length of the geodesic corresponding to a hyperbolic of trace t is $2 \log(t + \sqrt{t^2 - 4})/2$ we obtain Table 1.

Of course, the first eight hyperbolics as labeled remain the first eight shortest geodesics for any $G_\lambda \backslash \mathscr{H}$ with $\lambda > 4$. That is, the initial length ordering of geodesics is eventually stable as expressions in λ under the continuous deformation of increasing λ. (See below for the significance of initial ordering switches.)

The final two entries of Table 1 give two distinct geodesics of the same length. We intend to discuss multiplicity (and intersection) phenomena to greater detail in forthcoming work. Here we will just remark that length spectrum multiplicity for our one-parameter family of surfaces arises in each of two ways. The first, where the trace polynomials of non-conjugate hyperbolics are identical, may be thought of as structural—this phenomenon exists even for the Γ_q. In addition, as λ increases, size order of two traces may reverse. As λ is a continuous variable, there will be an (algebraic) value for which the traces are identical. This sort of multiplicity may be thought of as *accidental.* (These two species of multiplicity exist on many families of surfaces; our point is that the Hecke surfaces are already complex enough to exhibit both.)

Table 1. The shortest geodesics, $\lambda = 4$

Hyperbolic	Trace	Length of geodesic
U	4	2.63392
g_2	8	4.12687
g_1g_2	12	4.95578
$g_1^2g_2$	16	5.53732
g_1h_1	18	5.77454
$g_1^3g_2$	20	5.98645
$g_1^4g_2$	24	6.35263
$g_1^5g_2$	28	6.66185
g_3	30	6.80017
$g_1^6g_2$	32	6.92952
$g_1^2h_1$	34	7.05099
$g_1^7g_2$	36	7.16549
$g_1^8g_2$	40	7.37651
$g_1^9g_2$	44	7.56735
g_1g_3	46	7.65634
$g_1^{10}g_2$	48	7.74153
$g_1^3h_1$	50	7.82325
$g_1^{11}g_2$	52	7.90175
$g_1^{12}g_2$	56	8.05007
$g_1^{13}g_2$	60	8.18813
$g_1^2g_3$	62	8.25375
$g_1^{14}g_2$	64	8.31728
$g_1^4h_1$	66	8.37885
$g_1^{15}g_2$	68	8.43858
g_1h_2	68	8.43858

Table 2. Accidental multiplicity

Hyperbolic	Trace(λ)	Length ($\lambda = 4 + \varepsilon$)
$g_1^{15}g_2 = \begin{pmatrix} 16\lambda & \lambda^2 + 15\lambda - 1 \\ 1 & \lambda \end{pmatrix}$	17λ	$8.43858 + 0.3097\varepsilon + \cdots$
$g_1h_2 = \begin{pmatrix} \lambda^3 & \lambda^2 + 1 \\ \lambda^2 - 1 & \lambda \end{pmatrix}$	$\lambda^3 + \lambda$	$8.43858 + 1.4418\varepsilon + \cdots$

To illustrate this phenomenon, we offer Table 2, in which we show the trace of a primitive hyperbolics fixing lifts of each of two geodesics and the length of these geodesics as functions of λ as well as give an expansion to first order terms of the real deformation variable ε in a neighborhood of $\lambda = 4$ (the place where both traces are 68, of course).

As Buser notes [Bu, p. 273], often surprisingly small initial portions of the length spectrum serve as moduli for a surface. As the notation indicates, $G_\lambda \backslash \mathscr{H}$ is determined by one real parameter. It is not surprising then, that the length of the boundary geodesic of $G_\lambda \backslash \mathscr{H}$ also serves as a parameter determining the surface (up to Teichmüller class). This may be seen canonically using two facts: first, via $\mathrm{SL}(2, R)$ conjugation, every surface with signature $(0; 2, \infty; 1)$—sphere with one puncture, one elliptic fixed point of order 2, and one hole—may be represented by a Fuchsian group with an elliptic element fixing i and a parabolic element fixing ∞. It will thus have the usual fundamental region Hecke triangle group of infinite volume; i.e., bounded by the unit circle and two vertical lines through $\pm\lambda/2$. Second, the formula for the hyperbolic fixing the lift of the boundary geodesic to this fundamental region given in the next section.

4. The Hall ray

We shall show that, in general, the Hecke triangle groups possess a Hall ray. We take the geometric definition of a surface possessing a Hall ray: there exists a real H such that for any h greater than H there are geodesics with height arbitrarily close to h. The *height* of a geodesic is the largest euclidean radius of any lift of the geodesic to $\mathscr{H}$. (The Hall ray is the ray running from lim inf of such H and $+\infty$.) The known examples to date of surfaces with Hall rays are all related to the modular surface, see [H2].

Establishing the existence of a Hall ray turns out to be easiest to prove for groups of the second kind with $\lambda > \sqrt{8}$. Using λCF's, it is a routine matter to show that each of the Hecke groups of the first kind admits a Hall ray. That is, the original Hall argument applicable to the modular surface, can be extended to all finite volume Hecke groups. The geometric issues in the first kind and second kind cases are similar; we give the proof of the first kind case for completeness.

4.1 The Hall ray for G_λ with $\lambda = 2\cos\pi/q$

The naive height of a lift to the Poincaré upper half-plane of a geodesic on a Riemann surface with a single cusp is the euclidean diameter of this lift. The height of a geodesic is the lim sup of the naive heights of all lifts of this geodesic. The Cohn–Markoff spectrum is the set of all heights of geodesics of the surface. If the Cohn–Markoff spectrum of a surface contains a ray of values, the surface is said to have a Hall ray. In the literature to date, all surfaces known to have a Hall ray are (commensurate with) coverings of the modular surface. We show that for $q > 4$, G_q admits a Hall ray, with all values greater than or equal to $\mu = 6\lambda_q$ being included in the ray. Since A. Haas [H2] has shown that the case of $q = 4$ follows from the existence of the classical Hall ray for $q = 3$, our result implies that every G_q has a Hall ray.

Our proof is an adaptation of Hall's original, as presented in [C-F]. We thank A. Haas for pointing out an error in our original proof. The determination of the precise beginning value of the Hall ray for G_3 is notoriously difficult. Here we only identify a lower bound beyond which the ray extends.

One cannot use continued fraction expansions for the G_q which involve only positive signs, as in the ordinary continued fractions of the classical setting. Here we use the next λ fractions, as in [S-S]. The crucial advantage in the present setting over the use of the other option, the Rosen λ continued fractions, is the simple detection of the natural ordering of the reals.

The next λ fractions are given by use of positive partial quotients, but only negative signs between. Thus, we use the symbol

$$[a_0;\, a_1,\, a_2 \ldots] := a_0\lambda_q - 1/(a_1\lambda_q - 1/(a_2\lambda_q \ldots),$$

where all $a_i, i > 0$ are positive integers; by mild abuse of notation, we will refer to the a_i as the partial quotients of such an expansion. In this form, λ_q itself has a periodic expansion:

$$\lambda_q = [\overline{2, 1_{q-3}}],$$

where we use an overline to indicate a period, and 1_n to represent a string of n consecutive 1's. The period of λ_q indicates the maximal number of consecutive 1's which appear in any (reduced) next λ fraction.

It is known, see say [S-S], that every real can be expanded in a next λ fraction. Thus given a lift of a geodesic on the surface $G_q \backslash \mathscr{H}$, the feet of any of its lifts can be expanded out in such continued fractions. Of course, any translation of this lift will have the same naive height. Thus we may normalize lifts so as to consider those with one foot between 0 and λ_q and the other greater than λ_q. The larger foot will have expansion $\beta = [a_0; a_1, \ldots]$, while the smaller will be $\alpha = -[0; a_{-1}, a_{-2}, \ldots]$. Since the expansion of the former corresponds to the image of infinity under $S^{a_0}TS^{a_1}T\ldots$, it is easily seen that TS^{-a_0} takes this to $[a_1; a_2, \ldots]$ while sending the other foot to $-[0; a_0, a_{-1}, \ldots]$. Since the diameter of the original lift is

$$[a_0;\, a_1, \ldots] + [0;\, a_{-1},\, a_{-2}, \ldots],$$

it is clear that the set of the lim sup of all such sums for allowable double sequences of the a_i is the same as the Cohn–Markoff spectrum of G_q.

THEOREM. *Let $q > 4$. Any real number greater than $6\lambda_q$ is contained in the Markoff spectrum of G_q.*

We use the following two lemmas.

LEMMA 4. *Let $q > 4$. Any real number can be written in the form*

$$a\lambda_q + [0; b_1, b_2 \ldots] + [0; c_1, c_2 \ldots],$$

where a is an integer and the b_i and c_i do not exceed 7.

This follows from

LEMMA 5. *Let $q > 4$. Any real number in the interval*

$$[-2\lambda_q, 2[0; \overline{7}]]$$

can be written in the form

$$a\lambda_q + [0; b_1, b_2 \ldots] + [0; c_1, c_2 \ldots],$$

where a is an integer and the b_i and c_i do not exceed 7.

Let $F(k)$ be the set of real numbers α such that the next λ fraction of α has no partial quotient larger than k and $-1/\lambda_q \leqslant \alpha \leqslant -1/[\overline{k}]$. Of course, Lemma 2 is equivalent to

$$F(7) + F(7) = [-2\lambda_q, 2[0; \overline{7}]].$$

To prove the above equation by [C-F; pp. 47–51], we need only show that $F(7)$ is obtained by a Cantor dissection process for which each interval removed is smaller than either of the subintervals created by its removal.

Our deletion process has a single rule: from an interval

$$[[0; a_1, \ldots, a_r, W, \overline{1_{q-3}, 2}], [0; a_1, \ldots, a_r, m, \overline{k}]],$$

where $2 \leqslant m \leqslant k$ and $W = w_1, w_2, \ldots, w_n$ is of minimal length we remove

$$[[0; a_1, \ldots, a_r, m-1, \overline{k}], [0; a_1, \ldots, a_r, m, \overline{1_{q-3}, 2}]].$$

It is easily verified that after all (the countably many) applications of this rule to

$$-\lambda_q \leqslant \alpha \leqslant [0; \bar{k}]$$

exactly $F(k)$ remains. In practice, k will be equal to 7; W will consist of consecutive partial quotients from $1_{q-3}, 2$; and W must not end with 1 as a partial quotient.

We now show that for $k \geqslant 7$ the lengths of each excised interval is shorter than that of either of the subintervals which this excision creates. We label:

$$M_1 = [[0; a_1, \ldots, a_r, W, \overline{1_{q-3}, 2}], [0; a_1, \ldots, a_r, m-1, \bar{k}]];$$

$$I = [[0; a_1, \ldots, a_r, m-1, \bar{k}], [0; a_1, \ldots, a_r, m, \overline{1_{q-3}, 2}]];$$

$$M_2 = [[0; a_1, \ldots, a_r, m, \overline{1_{q-3}, 2}], [0; a_1, \ldots, a_r, m, \bar{k}]].$$

Note that

$$[\ldots, m, \overline{1_{q-3}, 2}] = [\ldots, m-1].$$

Thus we let

$$p_{r+1}/q_{r+1} = [0; a_1, \ldots, a_r, m-1].$$

We now call $[\bar{k}]$ simply κ. Basic continued fraction formuli, analogous to [CF; (9)], give

$$|I| = \frac{1}{(\kappa q_{r+1} - q_r)q_{r+1}}.$$

If $\tilde{p}_{r+1}/\tilde{q}_{r+1} = [0; a_1, \ldots, a_r, m]$, then $\tilde{p}_{r+1} = p_{r+1} + \lambda_q p_r$ and similarly for the denominators. Thus

$$\begin{aligned} |M_2| &= \frac{\kappa \tilde{p}_{r+1} - p_r}{\kappa \tilde{q}_{r+1} - q_r} - \frac{p_{r+1}}{q_{r+1}} \\ &= \frac{\kappa\lambda_q - 1}{(\kappa q_{r+1} + (\kappa\lambda_q - 1)q_r)q_{r+1}}. \end{aligned}$$

We must show that

$$(\kappa\lambda_q - 1)(\kappa q_{r+1} - q_r) \geqslant \kappa q_{r+1} + (\kappa\lambda_q - 1)q_r.$$

This last equation is equivalent to

$$\frac{\kappa(\kappa\lambda_q - 2)}{2(\kappa\lambda_q + 1)} \geqslant \frac{q_r}{q_{r+1}}.$$

Now,

$$\frac{q_r}{q_{r+1}} = 1/[m-1, a_r, a_{r-1}, \ldots, a_1].$$

This last equation is bounded above by λ_q, because our excision rule dictates that $m - 1 \geqslant 1$ and that there can be at most $q - 3$ consecutive 1's in the sequence $m - 1$, a_r, $a_{r-1}, \ldots, a_1$. The right-hand side is at most λ_q and if $\kappa \geqslant 6\lambda_q$, we are done.

We now show that $|M_1| \geqslant |I|$. Let $\omega = [W, \overline{1_{q-3}, 2}]$. Thus

$$\frac{|M_1|\kappa - \omega}{(\kappa q_{r+1} - q_r)(\omega q_{r+1} - q_r)}.$$

It suffices to show that

$$q_{r+1}(\kappa - \omega) \geqslant \omega q_{r+1} - q_r.$$

But, $2\lambda_q > \omega$; hence if $\kappa > 4\lambda_q$, then

$$q_{r+1}(\kappa - \omega) > q_{r+1}(\kappa - 2\lambda_q) > 2\lambda_q q_{r+1} > \omega q_{r+1} > \omega q_{r+1} - q_r.$$

Thus $\kappa \geqslant 6\lambda_q$ suffices for this part as well. In conclusion, if $k = 7$ then any of the removed intervals is shorter than either of the subintervals which its removal causes.

We now prove our theorem. From the above,

$$F(7) + F(7) = [-2\lambda_q, 2[0; \bar{7}]].$$

This interval has length greater than λ_q, hence we may express any $\mu \geqslant 6\lambda_q$ as

$$a\lambda_q + [0; b_1, b_2, \ldots] + [0; c_1, c_2, \ldots],$$

where $a \geqslant 8$ and $b_i \leqslant 7$, $c_i \leqslant 7$. Let

$$\alpha = [b_{k_1}; \ldots, b_1, a, c_1, \ldots, c_{k_1}, b_{k_2}, \ldots, b_1, a, c_1, \ldots, c_{k_2}],$$

which is say $[a_0; a_1, a_2, \ldots]$, where k_i is a strictly increasing sequence of integers. From our definition of the spectrum for these continued fractions, it is clear that the spectral value for α is indeed μ.

Since the cover of $G_q\backslash\mathscr{H}$ corresponding to the commutator subgroup of G_q can be shown to be of genus $(q-1)/2$ and of a single cusp if q is odd (a similar formula holds for even q, but then one has two cusps), arguments as in [H2] (but without use of Millington's theorem) show that

COROLLARY. *For every genus g, there is a Riemann surface with a single cusp which admits a Hall ray.*

REMARKS. First, a most vexing question concerning the classical Markoff spectrum deals with the mysterious portion lying below the Hall ray and above 3, the limit point of the discrete part. We have no insight into whether or not an analogous portion exists in the case of the Hecke groups. But we can say that a discrete portion does exist. This follows from the construction of low height geodesics given in [S].

Second, the observant reader may think that a step or two of the above proof seems overly cautious. We point out here that the approximants p_i/q_i for a next λ fraction can display a mildly perturbing phenomenon. In fact, it can happen that $q_i > q_{i+1}$. Of course, this never happens for the ordinary continued fractions, and [R] showed that it never happens for the Rosen λ fractions. We give an example:

Let $q = 5$ and consider the approximants to $[1, m, 1, 1, \ldots]$, with $m \geqslant 2$. Here $q_2 = m\lambda_5 + m - 1$ which is clearly greater than $q_3 = (m-1)\lambda_5 + m$.

One might now be concerned about the convergence of the next λ fractions to the reals which they represent, but it is easily checked that the q_i are increasing unless a partial denominator is 1 and when a string of consecutive 1's ends, then the next q_{i+n} must in fact be larger. This allows convergence proofs to go through.

4.2 The Hall ray for G_λ with $\lambda > \sqrt{8}$

Let $\mathscr{N}$ be the Nielsen convex region of G_λ. This is given by cutting off the funnel in the standard fundamental region by a circle, $\mathscr{C}$, perpendicular to both the unit circle ($\mathscr{U}$) and the line $\Re(z) = \lambda/2$.

It is easy to check that $\mathscr{C}$ has center at $\lambda/2$ and diameter $\sqrt{\lambda^2-4}$ (which is larger than 2). The point of intersection of $\mathscr{C}$ with $\mathscr{U}$ is $(2/\lambda, \sqrt{1-4/\lambda^2})$; that with the line $\Re(z) = \lambda/2$ is $(\lambda/2, \sqrt{\lambda^2/4-1})$. (In the SFR, the entire geodesic is comprised of the arc between these two points and a companion which is the reflection of this arc by $z \to -\bar{z}$.)

The projection of $\mathscr{C}$ to $G_\lambda \backslash \mathscr{H}$ is (infinitely many copies of) the boundary geodesic. The diameter of $\mathscr{C}$ along $\mathfrak{R}$ is, apart from its endpoints which are limit points (hyperbolic fixed points since the boundary geodesic is closed—in fact, they are fixed by

$$U = \begin{pmatrix} \lambda & -1 \\ 1 & 0 \end{pmatrix} = S_\lambda T,$$

a matrix which is (crucially) elliptic if $\lambda < 2$!), comprised of a doubly infinite sequence of free sides. Likewise for the translates by S_λ of this diameter. These translates take up the entire real axis, apart from intervals of length less than 2. These intervals contain all limit points, of course. Take a height greater than 1. Draw an h-line of this naive height. This line may be slid along the real axis so that both its feet lie in the free side amalgamations. (The first set of intervals.)

It is clear that the naive height of this geodesic is its true height as the entire portion of the geodesic outside translates of the SFR lies beyond the boundary geodesic. This geodesic has two infinite ends—it is not closed in either direction.

What if we require the geodesics we consider to be closed? If we restrict our attention to closed geodesics, then there may be no Hall ray. This is easily illustrated as follows:

Assume λ is large, 100 say, then all closed geodesics have feet in

$$\mathscr{F} = \bigcup_{n \in \mathbb{Z}} [-1 + 50n, 1 + 50n]$$

Any geodesic of height h, for $2 + 50n < h < 48 + 50n$; $n \in \mathbb{Z}$ cannot have both feet in $\mathscr{F}$. Thus there is no Hall ray for closed geodesics. We see that the notions of *Markoff spectrum* (a closed geodesic—infinite λCF phenomenon) and *Cohn spectrum* (a geodesic height phenomenon) have bifurcated in this context.

Acknowledgements

T. A. Schmidt thanks Alf van der Poorten and Macquarie University, where this work was initiated; Widener University for generous granting of leaves of absence; and the NSF-RUI program. M. Sheingorn thanks the NSF and PSC-CUNY for supporting this research.

References

[Be] A. Beardon, *The Geometry of Discrete Groups.* Graduate Texts in Mathematics 91 (New York: Springer-Verlag) 1983.

[Bu] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces.* Progress in Mathematics 106 (Boston: Birkhäuser) 1992.

[C] C. Croke, Rigidity for surfaces of non-positive curvatures, *Comm. Math. Helv.* 65 (1990) 150–169.

[CF] T. Cusick and M. Flahive, *The Markoff and Lagrange Spectra*, Math. Surveys and Monographs 30 (Providence: AMS) 1989.

[F] B. Fine, Trace classes and quadratic forms in the modular group, *Can. J. Math.*, to appear.

[H] A. Haas, Diophantine approximation on hyperbolic Riemann surfaces, *Acta Math.* 156 (1986) 33–82.

[H2] A. Haas, Geometric Markoff theory and a theorem of Millington, in: A. Pollington and W. Moran (eds.) *Number theory with an emphasis on the Markoff Spectrum* (New York: Dekker) (1993) pp. 107–112.

[H-S] A. Haas and C. Series, Hurwitz constants and diophantine approximation on Hecke groups, *JLMS* (2) 34 (1986) 219–234.

[L] S. Lalley, Mostow rigidity and the Bishop-Steger dichotomy for surfaces of variable negative curvature, *Duke Math. J.* 68 (1992) 237–269.

[R] D. Rosen, A class of continued fractions associated with certain properly discontinuous groups, *Duke Math. J.* 21 (1954) 549–563.

[S-S] T. Schmidt and M. Sheingorn, Length spectra for Hecke triangle surfaces, *Math. Z.*, to appear.

[S] M. Sheingorn, Low height Hecke triangle group geodesics, *Cont. Math.* 143 (1993) 545–560.

[W] J. Wolfart, Diskrete deformation fuchsscher gruppen und ihrer automorphen formen, *J. f.d.r.u.a. Math.* 348 (1984) 203–220.