

THE PICARD GROUP, CLOSED GEODESICS, AND ZETA FUNCTIONS

MARK POLLICOTT

ABSTRACT. In this article we consider the Picard group $SL(2, \mathbb{Z}[i])$, viewed as a discrete subgroup of the isometries of hyperbolic space. We fix a canonical choice of generators and then construct a Markov partition for the action of the group on the sphere at infinity.

Our main application is to the study of the zeta function associated to the associated three-dimensional hyperbolic manifold.

INTRODUCTION

It is well known that there is a close connection between the geodesic flow for the Modular surface and the continued fraction transformation on the unit interval. The relationship between the two is a consequence of the fact that the Fuchsian group associated to the Modular surface is the Modular group $SL(2, \mathbb{Z})$ and the action of this group on the real line (as the “boundary” of the Poincaré upper half plane) induces the continued fraction transformation on $[0, 1]$ in a very natural way. Modern accounts of this theory can be found in the articles of Adler-Flatto [AF] and Series [S].

In this note we shall describe an analogous correspondence in which we replace the Modular Fuchsian group by the Picard Kleinian group $SL(2, \mathbb{Z}[i])$ acting on the three-dimensional upper half plane model for hyperbolic space \mathbb{H}^3 . In particular, we shall describe a transformation on the rectangle $B = \{z \in \mathbb{C} : |\operatorname{Re}(z)| \leq 1/2, 0 \leq \operatorname{Im}(z) \leq 1/2\}$ which plays the role of the continued fraction transformation.

As our main application of these ideas, we shall give an alternative approach to the meromorphic extension of the associated Zeta function, following the derivation for the case of the Modular surface described in the work of Mayer [M].

The construction we describe should have applications to the construction of special flows and also to the problem of the Markov spectrum of the Picard group (cf. Schmidt [Sc]). We hope to return to these problems later.

1. THE PICARD GROUP

We define the Picard group $\Gamma = SL(2, \mathbb{Z}[i])$ to be the group of 2×2 matrices whose entries are the Gaussian integers $u + iv$ ($u, v \in \mathbb{Z}$) and such that every matrix has determinant 1.

Received by the editors September 30, 1993.

1991 *Mathematics Subject Classification*. Primary 30F40; Secondary 58F20.

© 1994 American Mathematical Society
0002-9947/94 \$1.00 + \$.25 per page

The Picard group Γ has a natural action on the extended complex plane $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ defined in terms of linear fractional transformations. More specifically, given any matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2, \mathbb{Z}[i])$ we define the action on $\widehat{\mathbb{C}}$ by

$$A: z \mapsto \frac{az + b}{cz + d} \quad \text{for } z \in \widehat{\mathbb{C}}.$$

We shall follow the convention (described in Magnus [Ma]) of adding the conjugating map $z \mapsto \bar{z}$ to this group action and denoting the larger group of actions, generated by both Γ and $*$, by $\Gamma^* = \langle \Gamma, * \rangle$.

Bearing in mind the prototypal example of the action of the Modular group $\mathrm{SL}(2, \mathbb{Z})$ on the real line and the continued fraction transformation $f: (0, 1) \rightarrow (0, 1)$ defined by $f(x) = \frac{1}{x} - [\frac{1}{x}]$, we want to replace the action $\Gamma^* \times \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ by a single transformation (acting on a smaller space).

We begin with the observation that a choice of generators for the action of Γ is given by the linear fractional transformations:

$$\begin{aligned} A: z &\mapsto z + 1 && \text{corresponding to the matrix } \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\ B: z &\mapsto z + i && \text{corresponding to the matrix } \begin{bmatrix} 1 & i \\ 0 & 1 \end{bmatrix}, \\ T: z &\mapsto -\frac{1}{z} = -\frac{\bar{z}}{|z|^2} && \text{corresponding to the matrix } \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \end{aligned}$$

and the group Γ^* is generated by the same elements plus the operation $*$. We next need to introduce some notation to specify certain rectangles in the complex plane. We denote

$$\begin{aligned} \mathcal{B}^+ &= \{z \in \mathbb{C}: 0 \leq \mathrm{Re}(z), \mathrm{Im}(z) \leq 1/2\}, \\ \mathcal{B}^- &= \{z \in \mathbb{C}: -1/2 \leq \mathrm{Re}(z) \leq 0 \leq \mathrm{Im}(z) \leq 1/2\}, \\ \mathcal{B} &= \mathcal{B}^+ \cup \mathcal{B}^- = \{z \in \mathbb{C}: |\mathrm{Re}(z)| \leq 1/2, 0 \leq \mathrm{Im}(z) \leq 1/2\}. \end{aligned}$$

Given $w \in \mathbb{C}$ we let $C(w) = \{z \in \mathbb{C}: |w - z| = 1\}$ denote the circle in \mathbb{C} which is centered at w and has unit radius. The following lemma is immediate.

Lemma 1. (i) *The image $T(\mathcal{B}^-)$ is the region in the quadrant $\mathrm{Re}(z) \geq 0$, $\mathrm{Im}(z) \geq 0$ exterior to both of the circles $C(1)$ and $C(i)$.*

(ii) *The image $T(\mathcal{B}^+)$ is the region in the quadrant $\mathrm{Re}(z) \leq 0$, $\mathrm{Im}(z) \geq 0$ exterior to both of the circles $C(-1)$ and $C(i)$.*

(iii) *The image $T(\mathcal{B})$ is the region in the half plane $\mathrm{Im}(z) \geq 0$ exterior to each of the circles $C(-1)$, $C(1)$, and $C(i)$. (See Figure 1.)*

It is convenient to subdivide the image $\mathcal{H} = T(\mathcal{B})$ into two regions, which we shall define to be:

$$\begin{aligned} \mathcal{H}_1 &= \{z: |\mathrm{Re}(z)| \geq 2\} \cup \{z: |\mathrm{Re}(z)| \geq 0, \mathrm{Im}(z) \geq 2\}, \\ \mathcal{H}_2 &= \{z: 0 \leq |\mathrm{Re}(z)| \leq 2, 0 \leq \mathrm{Im}(z) \leq 2, \\ &\quad \text{and } z \text{ is exterior to } C(-1), C(1), \text{ and } C(i)\}. \end{aligned}$$

In the case of \mathcal{H}_1 , we observe that this region can be covered by squares (with sides of length $1/2$) of the form

$$(1.1) \quad \begin{cases} \mathcal{B}^+(r, s) = \{z \in \mathbb{C}: \frac{r}{2} \leq \mathrm{Re}(z) \leq \frac{r+1}{2} \text{ and } \frac{s}{2} \leq \mathrm{Im}(z) \leq \frac{s+1}{2}\}, \\ \mathcal{B}^-(r, s) = \{z \in \mathbb{C}: \frac{r}{2} \leq -\mathrm{Re}(z) \leq \frac{r+1}{2} \text{ and } \frac{s}{2} \leq \mathrm{Im}(z) \leq \frac{s+1}{2}\} \end{cases}$$

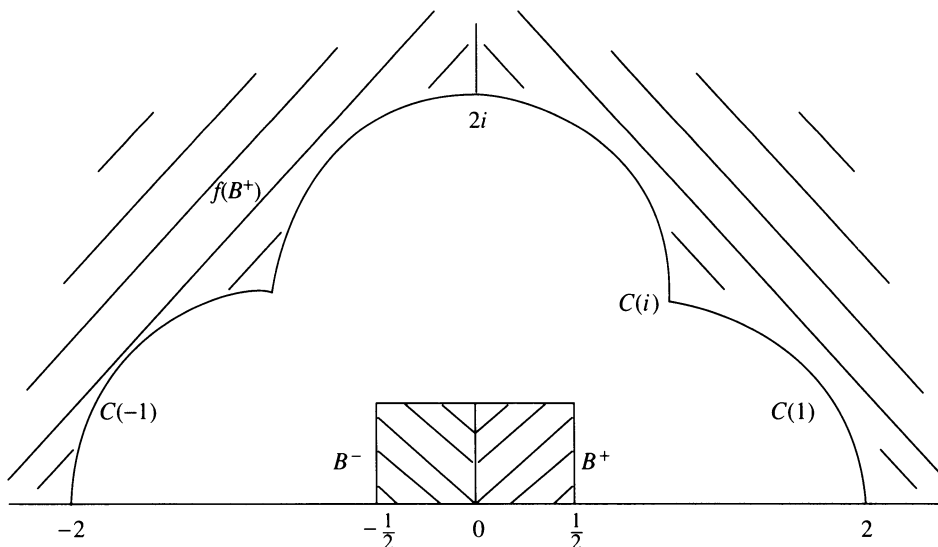


FIGURE 1

where $r, s = 4, 5, 6, \dots$. We denote the elements in this subdivision of \mathcal{H}_1 by \mathcal{P}_1 . It is easily verified that by a suitable combination of actions by A and B ("translation") and $*$ ("reflection in the real axis") each of the squares defined in (1.1) can be moved to either \mathcal{B}^+ or \mathcal{B}^- . The specific rules are summarized in the next lemma.

Lemma 2. (i) $A^{-r/2}B^{-s/2}\mathcal{B}^\pm(r, s) = \mathcal{B}^\pm$ if r, s are both even.
 (ii) $A^{-(r+1)/2}B^{-s/2}\mathcal{B}^\pm(r, s) = \mathcal{B}^\mp$ if r is odd and s is even.
 (iii) $*A^{-r/2}B^{-(s+1)/2}\mathcal{B}^\pm(r, s) = \mathcal{B}^\pm$ if r is even s is odd.
 (iv) $*A^{-r/2}B^{-s/2}\mathcal{B}^\pm(r, s) = \mathcal{B}^\mp$ if r, s are both odd.

For each square $\mathcal{B}^\pm(r, s)$ we denote the transformation onto either \mathcal{B}^+ or \mathcal{B}^- described by the left-hand side of the expressions in the above lemma by $D_{\mathcal{B}^\pm(r, s)}: \mathcal{B}^\pm(r, s) \rightarrow \mathcal{B} = \mathcal{B}^+ \cup \mathcal{B}^-$. (See Figure 2.)

In the case of \mathcal{H}_2 , we observe that this region can be divided in an obvious way into a finite number of pieces whose boundaries are contained in the curves:

- (a) $\{z \in \mathbb{C}: \operatorname{Im}(z) = \frac{r}{2}\}$ for $r = 0, \pm 1, \pm 2, \pm 3, \pm 4$;
- (b) $\{z \in \mathbb{C}: \operatorname{Re}(z) = \frac{s}{2}\}$ for $s = 0, 1, 2, 3, 4$;
- (c) $C(w)$ for $w = 0, \pm 1, i, i \pm 1$. (See Figure 3.)

We denote the elements in this subdivision of \mathcal{H}_2 by \mathcal{P}_2 . Since Lemma 2 still applies and each element $P \in \mathcal{P}_2$ lies in a square $\mathcal{B}^\pm(r, s) \subset \mathcal{H}_2$, we denote by $D_P = D_{\mathcal{B}^\pm(r, s)}|_P: P \rightarrow \mathcal{B} = \mathcal{B}^+ \cup \mathcal{B}^-$ the restriction of the map associated to that square.

We denote by \mathcal{P} the division of $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ into pieces formed by amalgamating the two subdivisions of \mathcal{H}_1 (into squares) and \mathcal{H}_2 (into a finite number of pieces). We observe that the elements of \mathcal{P} are not pairwise disjoint but that their interiors are. In particular, although \mathcal{P} does not form a partition of \mathcal{H} in the strictest sense, it does indeed form a partition in the sense of a measurable partition (with respect to Lebesgue measure, say).

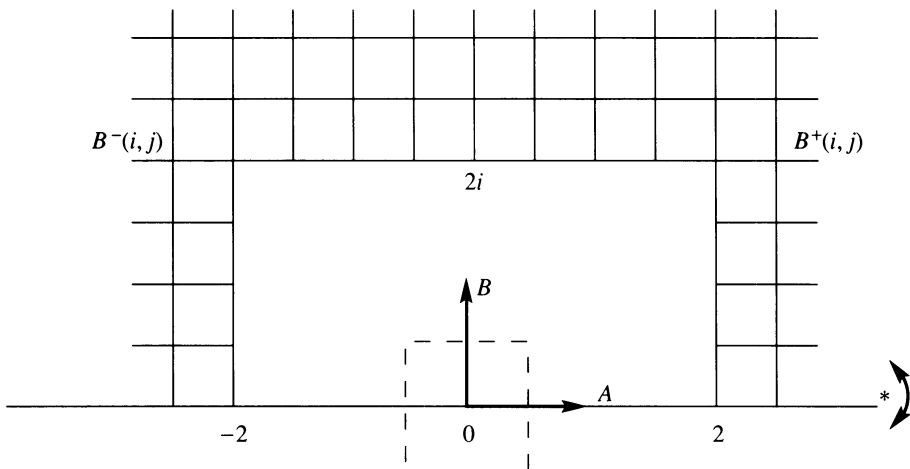


FIGURE 2

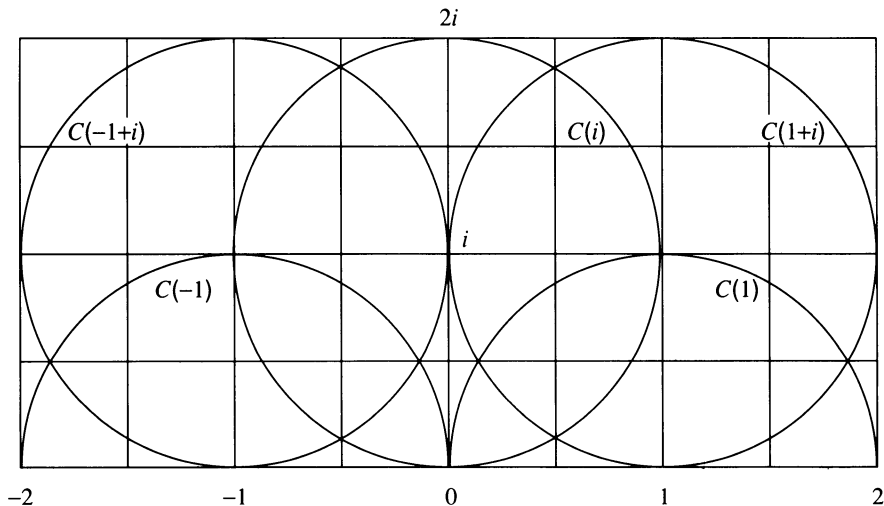


FIGURE 3

2. THE MARKOV CONDITION

For each $P \in \mathcal{P}$ we can define a map $f: P \rightarrow \mathcal{H}$ by $f(x) = T \circ D_P(x)$. We can extend this definition to the disjoint union of the sets \mathcal{P} , which by a slight abuse of notation we identify with \mathcal{H} and write $f: \mathcal{H} \rightarrow \mathcal{H}$. (This is justified on the grounds that any ambiguity only arises on a set of Lebesgue measure zero and is thus negligible for our purpose. Exactly the same argument applies as for the same complication for the continued fraction transformation.)

In this section we want to verify that \mathcal{P} is a Markov partition for $f: \mathcal{H} \rightarrow \mathcal{H}$ in the following sense.

Theorem 1. *For each $P \in \mathcal{P}$, the image $f(P) \subset \mathcal{H}$ is a union of elements from \mathcal{P} .*

Proof. The proof depends on checking five different cases.

Case I. If P is any of the rectangles $\mathcal{B}^\pm(r, s) \in \mathcal{P}_1$ occurring in the subdivision of \mathcal{H}_1 or the special cases of $P = \mathcal{B}^\pm(2, 2) \in \mathcal{P}_2$, then either $DP = \mathcal{B}^+$ or $DP = \mathcal{B}^-$ (by Lemma 2). By Lemma 1, we see that in the first instance the image $f(P) = T(DP) = T(\mathcal{B}^+)$ is the region in the quadrant $\operatorname{Re}(z) \leq 0$, $\operatorname{Im}(z) \geq 0$ exterior to both of the circles $C(-1)$ and $C(i)$. In the second instance, we see that the image $f(P) = T(DP) = T(\mathcal{B}^-)$ is the region in the quadrant $\operatorname{Re}(z) \geq 0$, $\operatorname{Im}(z) \geq 0$ exterior to both of the circles $C(1)$ and $C(i)$. In either eventuality, the image $f(P)$ is verified to be a union of elements from \mathcal{P} by reference to Figures 2 and 3.

Case II. Consider the region $P \subseteq \mathcal{B}^+(3, 3)$ which is exterior to the circle $C(1+i)$ and the region $P' \subset \mathcal{B}^-(3, 3)$ which is exterior to the circle $C(-1+i)$. The corresponding images $DP \subset \mathcal{B}^-$ and $DP' \subset \mathcal{B}^+$ are the subset of \mathcal{B}^- exterior to $C(-1+i)$ and the subset of \mathcal{B}^+ exterior to $C(1+i)$, respectively. (See Figure 4.)

The image $f(P) = T(DP)$ is the region in the quadrant $\operatorname{Re}(z) \geq 0$, $\operatorname{Im}(z) \geq 0$ exterior to the three circles $C(1)$, $C(1+i)$, and $C(i)$. The image $f(P') = T(DP')$ is the region in the quadrant $\operatorname{Re}(z) \leq 0$, $\operatorname{Im}(z) \geq 0$ exterior to both of the circles $C(-1)$, $C(-1+i)$, and $C(i)$. (See Figure 5.)

In either eventuality, the image $f(P)$ is verified to be a union of elements from \mathcal{P} by reference to Figures 2 and 3.

Case III. Consider the four similar subcases:

- (i) the region $P^1 \subset \mathcal{B}^+(3, 2)$ interior to $C(1+i)$,
- (ii) the region $P^2 \subset \mathcal{B}^-(3, 2)$ interior to $C(-1+i)$,
- (iii) the region $P^3 \subset \mathcal{B}^+(2, 3)$ interior to $C(1+i)$,
- (iv) the region $P^4 \subset \mathcal{B}^-(2, 3)$ interior to $C(-1+i)$.

The images of these regions under D can be seen to take the form:

- (i) the region $DP^1 \subset \mathcal{B}^-$ interior to $C(-1)$,
- (ii) the region $DP^2 \subset \mathcal{B}^+$ interior to $C(1)$,
- (iii) the region $DP^3 \subset \mathcal{B}^-$ interior to $C(i)$,
- (iv) the region $DP^4 \subset \mathcal{B}^+$ interior to $C(i)$.

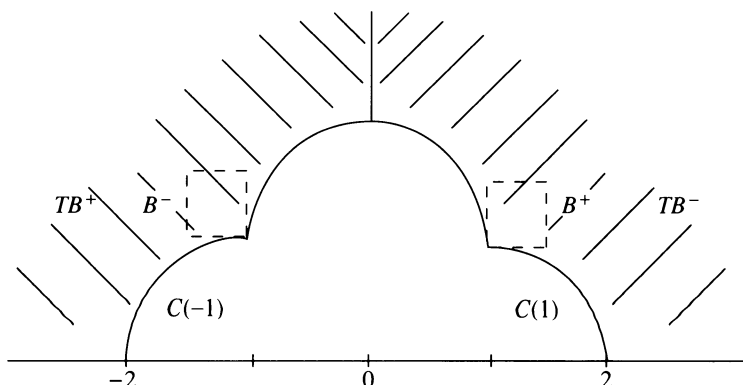


FIGURE 4

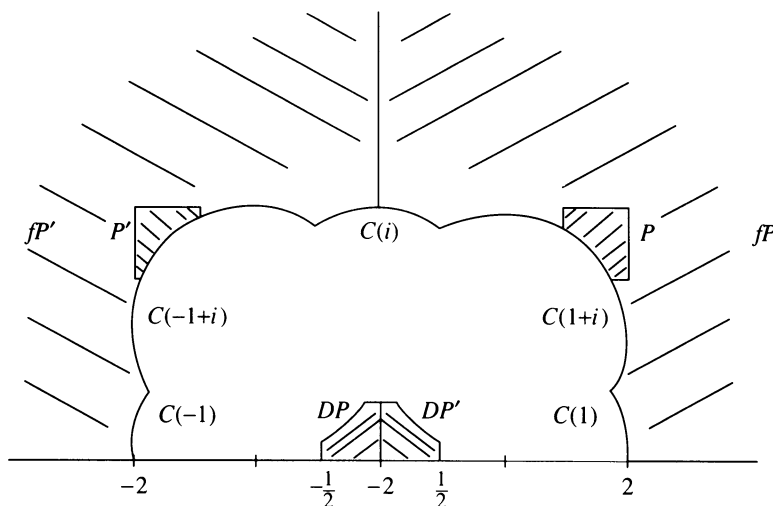


FIGURE 5

The subsequent images under T take the form:

- (i) $f(P^1) = T(DR^1)$ is the region in $\operatorname{Re}(z) \geq 1/2$, $\operatorname{Im}(z) \geq 0$ exterior to both $C(i)$ and $C(1)$,
- (ii) $f(P^2) = T(DR^2)$ is the region in $\operatorname{Re}(z) \leq -1/2$, $\operatorname{Im}(z) \geq 0$ exterior to both $C(i)$ and $C(-1)$,
- (iii) $f(P^3) = T(DR^3)$ is the region in $\operatorname{Re}(z) \leq 0$, $\operatorname{Im}(z) \geq 1/2$ exterior to both $C(-1)$ and $C(i)$,
- (iv) $f(P^4) = T(DR^4)$ is the region in $\operatorname{Re}(z) \geq 0$, $\operatorname{Im}(z) \geq 1/2$ exterior to both $C(i)$ and $C(1)$.

In each of these four subcases we see by comparison with Figures 2 and 3 that $f(P^i)$ ($i = 1, 2, 3, 4$) is a union of elements from \mathcal{P} . (See Figure 6.)

Case IV. Consider the following four similar subcases:

- (i) the region $P^1 \subset \mathcal{B}^+(1, 3)$ exterior to $C(i)$ but interior to $C(1+i)$,
- (ii) the region $P^2 \subset \mathcal{B}^-(1, 3)$ exterior to $C(i)$ but interior to $C(-1+i)$,
- (iii) the region $P^3 \subset \mathcal{B}^+(3, 1)$ exterior to $C(1)$ but interior to $C(1+i)$,
- (iv) the region $P^4 \subset \mathcal{B}^-(3, 1)$ exterior to $C(-1)$ but interior to $C(-1+i)$.

The four corresponding images under the map D are:

- (i) the region $DP^1 \subset \mathcal{B}^-$ exterior to $C(-1+i)$, but interior to $C(i)$,
- (ii) the region $DP^2 \subset \mathcal{B}^+$ exterior to $C(1+i)$, but interior to $C(i)$,
- (iii) the region $DP^3 \subset \mathcal{B}^-$ exterior to $C(-1+i)$, but interior to $C(-1)$,
- (iv) the region $DP^4 \subset \mathcal{B}^+$ exterior to $C(1+i)$, but interior to $C(1)$.

The final images under f are then of the form:

- (i) $f(P^1) = T(DR^1)$ is the region in $\operatorname{Re}(z) \geq 0$, $\operatorname{Im}(z) \geq 1/2$ exterior to both $C(i)$ and $C(1+i)$,
- (ii) $f(P^2) = T(DR^2)$ is the region in $\operatorname{Re}(z) \leq 0$, $\operatorname{Im}(z) \geq 1/2$ exterior to both $C(i)$ and $C(-1+i)$,
- (iii) $f(P^3) = T(DR^3)$ is the region in $\operatorname{Re}(z) \geq 1/2$, $\operatorname{Im}(z) \geq 0$ exterior to both $C(1+i)$ and $C(1)$,

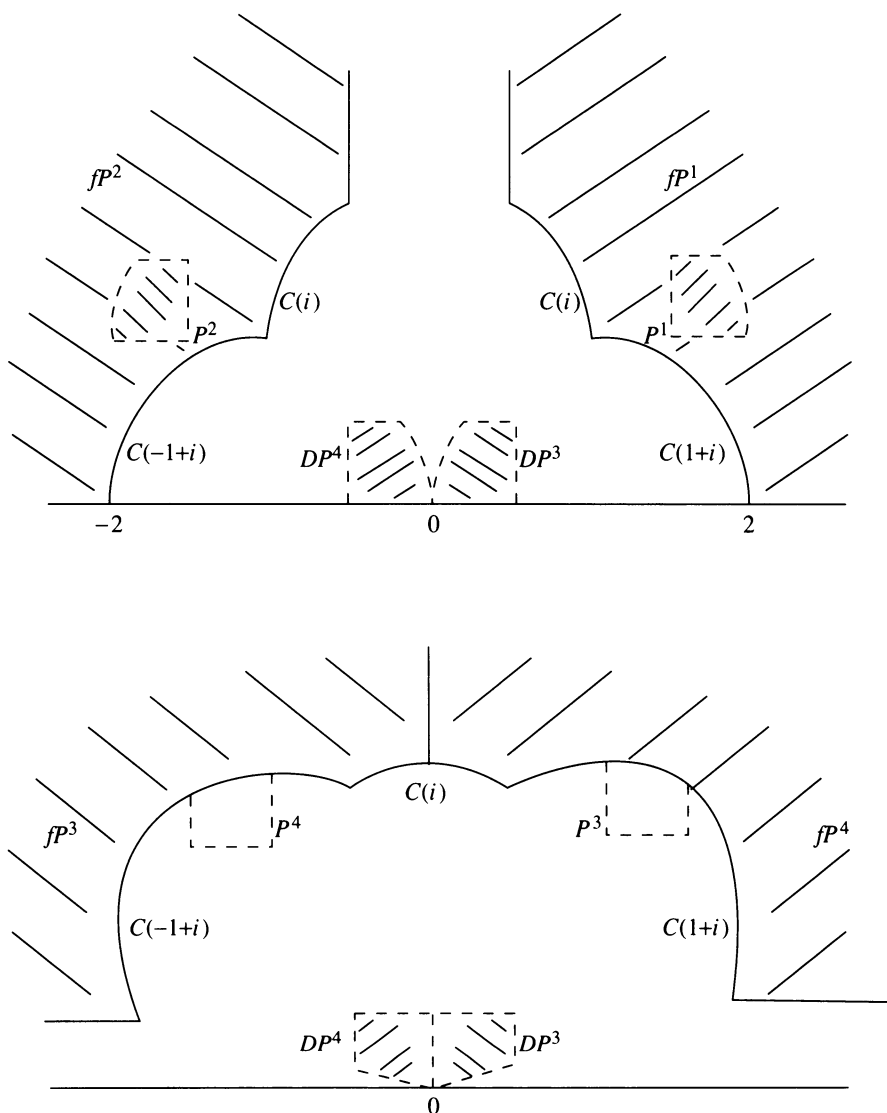


FIGURE 6

- (iv) $f(P^4) = T(DR^4)$ is the region in $\operatorname{Re}(z) \leq -1/2$, $\operatorname{Im}(z) \geq 0$ exterior to both $C(-1+i)$ and $C(-1)$.

In each instance, a comparison with Figures 2 and 3 shows that the images $f(P^i)$ ($i = 1, 2, 3, 4$) are each a union of elements in \mathcal{P} . (See Figure 7.)

Case V. Finally, we have to consider the remaining 12 regions which are of the form:

- (i) the region $P^1 \subset \mathcal{B}^+(3, 0)$ exterior to $C(1)$,
- (ii) the region $P^2 \subset \mathcal{B}^+(3, 1)$ exterior to $C(1+i)$,
- (iii) the region $P^3 \subset \mathcal{B}^+(3, 2)$ exterior to $C(1+i)$,
- (iv) the region $P^4 \subset \mathcal{B}^+(0, 3)$ exterior to $C(i)$,

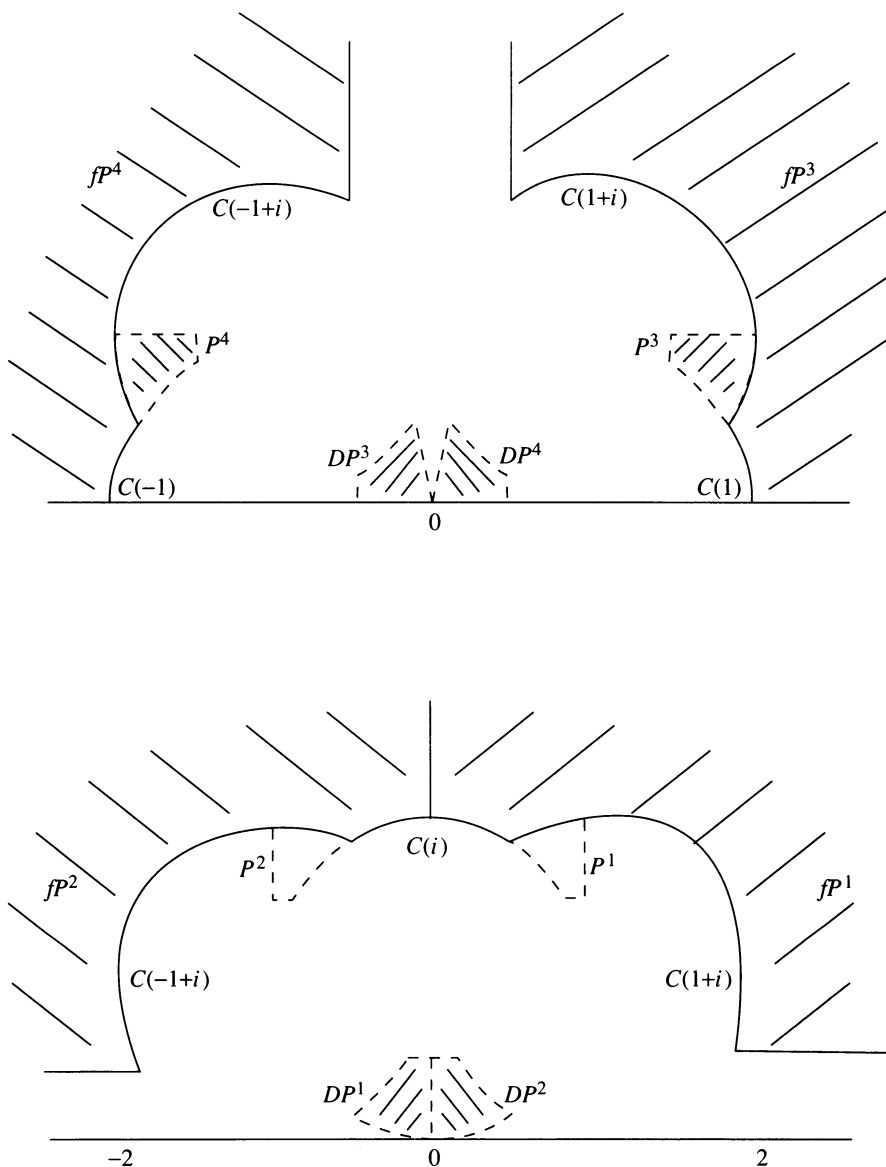


FIGURE 7

- (v) the region $P^5 \subset \mathcal{B}^+(1, 3)$ exterior to $C(1+i)$,
 - (vi) the region $P^6 \subset \mathcal{B}^+(2, 3)$ exterior to $C(1+i)$,
- in the quadrant $\operatorname{Re}(z) \geq 0$, $\operatorname{Im}(z) \geq 0$ and of the form:
- (vii) the region $P^7 \subset \mathcal{B}^-(3, 0)$ exterior to $C(-1)$,
 - (viii) the region $P^8 \subset \mathcal{B}^-(3, 1)$ exterior to $C(-1+i)$,
 - (ix) the region $P^9 \subset \mathcal{B}^-(3, 2)$ exterior to $C(-1+i)$,
 - (x) the region $P^{10} \subset \mathcal{B}^-(0, 3)$ exterior to $C(i)$,
 - (xi) the region $P^{11} \subset \mathcal{B}^-(1, 3)$ exterior to $C(-1+i)$

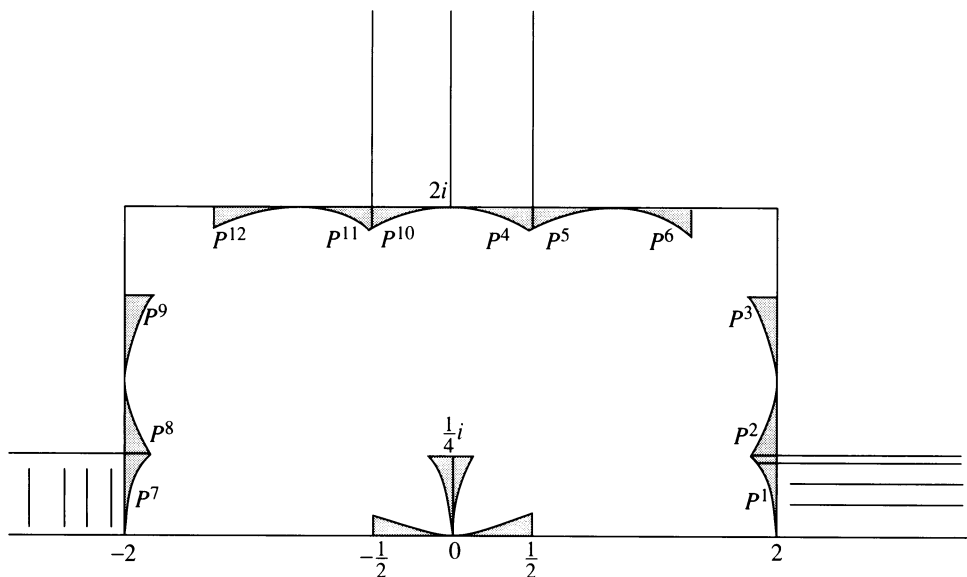


FIGURE 8

(xii) the region $P^{12} \subset \mathcal{B}^-(2, 3)$ exterior to $C(-1 + i)$

in the quadrant $\operatorname{Re}(z) \leq 0$, $\operatorname{Im}(z) \geq 0$.

The images under the map D of these various regions are:

- (a) for $i = 1, 2, 3$, $DR_i \subset \mathcal{B}^-$ is the region exterior to $C(-1)$;
- (b) for $i = 10, 11, 12$, $DR_i \subset \mathcal{B}^+$ is the region exterior to $C(1)$;
- (c) for $i = 4, 6, 8$, $DR_i \subset \mathcal{B}^+$ is the region exterior to $C(i)$;
- (d) for $i = 5, 7, 9$, $DR_i \subset \mathcal{B}^-$ is the region exterior to $C(i)$.

The final images under f are:

- (a) for $i = 1, 2, 3$, $f(P^i) = T(DP^i)$ is the region in $0 \leq \operatorname{Re}(z) \leq 1/2$, $\operatorname{Im}(z) \geq 0$ exterior to $C(i)$;
- (b) for $i = 10, 11, 12$, $f(P^i) = T(DP^i)$ is the region in $-1/2 \leq \operatorname{Re}(z) \leq 0$, $\operatorname{Im}(z) \geq 0$ exterior to $C(i)$;
- (c) for $i = 4, 6, 8$, $f(P^i) = T(DP^i)$ is the region in $\operatorname{Re}(z) \leq 0$, $0 \leq \operatorname{Im}(z) \leq 1/2$ exterior to $C(-1)$;
- (d) For $i = 5, 7, 9$, $f(P^i) = T(DP^i)$ is the region in $\operatorname{Re}(z) \geq 0$, $0 \leq \operatorname{Im}(z) \leq 1/2$ exterior to $C(1)$.

In these various cases we observe that $f(P^i) = T(DR_i)$ is a union of elements from \mathcal{P} (Figure 8). Since these five cases deal with all of the possibilities, the proof of the lemma is complete.

We can define a complementary map $\bar{f}: \mathcal{B} \rightarrow \mathcal{B}$ by conjugating with T , i.e., $\bar{f} = T \circ f \circ T^{-1}$ (with reference to Lemma 1). This is the analogue of the continued fraction map promised in the introduction. In practice, it is easier to define first the map $f: \mathcal{H} \rightarrow \mathcal{H}$, since the partition \mathcal{P} is notationally easier to describe. If we define the corresponding cover $\bar{\mathcal{P}} = \{\bar{P} = T(P): P \in \mathcal{P}\}$ for \mathcal{B} , then we have the following corollary:

Theorem 2. For each $\bar{P} \in \bar{\mathcal{P}}$, the image $\bar{f}(\bar{P}) \subset \mathcal{B}$ is a union of elements from $\bar{\mathcal{P}}$.

Following the usual conventions, we can associate to this Markov partition \mathcal{P} an (infinite) matrix $A(P, P')_{P, P' \in \mathcal{P}}$ with zero-one entries defined by

$$A(P, P') = \begin{cases} 1 & \text{if } f(P) \supset P' \\ 0 & \text{otherwise} \end{cases} \quad \text{where } P, P' \in \mathcal{P}.$$

We then define a subshift of finite type (with an infinite alphabet) on the space

$$\Sigma_A = \left\{ (x_n) \in \prod_{n=0}^{+\infty} \mathcal{P} : A(x_n, x_{n+1}) = 1, n \geq 0 \right\}$$

by $\sigma_A: \Sigma_A \rightarrow \Sigma_A$, $\sigma_A(x_n) = (x_{n+1})$. Furthermore, the derivative of the map $f: \mathcal{H} \rightarrow \mathcal{H}$ takes the form:

$$f'(z) = (T \circ D)'(z) = T'(Dz) = -\frac{1}{(D, z)^2},$$

where $Dz \in B$. Thus, we can deduce that $|f'(z)| \geq \inf\{\frac{1}{|w|^2} : w \in B\} = 2$ whenever $z \in \mathcal{H}$. In particular, this estimate on the derivative implies that we have the bound

$$\text{diam}(\pi\{(x_n) : x_0 = P_0, x_1 = P_1, \dots, x_n = P_n\}) \leq \frac{1}{2^n}.$$

By letting $n \rightarrow \infty$ we see that the map $\pi: \Sigma_A \rightarrow \mathcal{H}$ given by $\pi(x_n) = \bigcap_{n=0}^{\infty} f^{-n}x_n$ is well defined.

Similarly, we can associate to the Markov partition $\bar{\mathcal{P}}$ an (infinite) matrix $\bar{A}(\bar{P}, \bar{P}')_{\bar{P}, \bar{P}' \in \bar{\mathcal{P}}}$ with zero-one entries defined by

$$\bar{A}(\bar{P}, \bar{P}') = \begin{cases} 1 & \text{if } \bar{f}(\bar{P}) \supset \bar{P}' \\ 0 & \text{otherwise} \end{cases} \quad \text{where } \bar{P}, \bar{P}' \in \bar{\mathcal{P}},$$

the subshift of finite type (with an infinite alphabet) $\sigma_{\bar{A}}: \Sigma_{\bar{A}} \rightarrow \Sigma_{\bar{A}}$, and the map $\bar{\pi}(y_n) = \bigcap_{n=0}^{\infty} \bar{f}^{-n}y_n$.

Proposition 1. (i) The maps $\pi: \Sigma \rightarrow \mathcal{H}$ and $\bar{\pi}: \Sigma \rightarrow B$ are surjective and one-to-one except on a set of zero Lebesgue measure.

(ii) $\pi \circ \sigma_A = f \circ \pi$ and $\bar{\pi} \circ \sigma_{\bar{A}} = \bar{f} \circ \bar{\pi}$.

(iii) \bar{A} is the transpose of A .

(The proof for expanding maps goes through with only minor modifications.)

3. COMPLEX CONTINUED FRACTION EXPANSIONS

It is interesting to compare the coding with respect to the partitions of orbits of a point $z \in \mathbb{C}$ and its complex continued fraction expansion. Given a number $z \in \mathbb{C}$, we can consider its expansion as a complex continued fraction of the following form

$$(3.1) \quad z = s_0 - \frac{1}{s_1 - \frac{1}{s_2 - \frac{1}{s_3 + \dots}}}$$

where $s_n = u_n + iv_n \in \mathbb{Z}[i]$ are Gaussian integers. This is clearly a natural generalization of the usual continued fraction of real numbers in which z would be a real number and s_n a natural number. Complex continued fraction expansions of the type (3.1) were originally studied by Hurwitz [H] and Ford [F].

We shall denote the expression on the left-hand side of (3.1) by $[s_0; s_1, s_2, \dots]$. It is not difficult to see that the expansion of an arbitrary point z is far from being unique.

Given a point $z_0 \in \mathbb{C}$ we shall describe a simple algorithm for generating this expansion. Assume that $z_0 \in B(a_0 + ib_0, 1/2) = \{u + iv \in \mathbb{C} : |a_0 - u|, |b_0 - v| < 1/2\}$; then we set $s_0 = a_0 + ib_0 \in \mathbb{Z}[i]$. It then follows that $E(z_0) := A^{-a_0}B^{-b_0}z_0 \in B(0, 1/2)$. We choose to exclude those values of z_0 such that $E(z_0) = 0$ (which form only a countable set). We denote the image $T \circ E(z_0)$ by z_1 and observe that by our last assumption we have $z_1 \in \mathbb{C}$ and more particularly that either $z_1 \in \mathcal{H}$ or $\overline{z_1} \in \mathcal{H}$.

Proceeding inductively, if we assume that $z_n = (T \circ E)^n z_0 \in B(a_n + ib_n, 1/2)$, then we set $s_n = a_n + ib_n \in \mathbb{Z}[i]$. If we exclude those values such that $E(z_n) = 0$, then the image $z_{n+1} := T \circ E(z_n)$ and by assumption $z_{n+1} \in \mathbb{C}$ (in fact, either $z_{n+1} \in \mathcal{H}$ or $\overline{z_{n+1}} \in \mathcal{H}$).

Thus, providing we exclude the countable set of points

$$\delta = \bigcup_{n=0}^{\infty} (E \circ (T \circ E)^n)^{-1}(0)$$

which are mapped onto zero at some stage of the algorithm, we have that for any $n \geq 0$

$$(3.2) \quad z \in B(s_0, 1/2) \cap (T \circ E)^{-1}B(s_1, 1/2) \cap \dots \cap (T \circ E)^{-n}B(s_n, 1/2).$$

Since the derivative $(T \circ E)'(z) = T'(Ez) = -1/(Ez)^2$, where $Ez \in B(0, 1/2)$, we can deduce that $|(T \circ E)'(z)| \geq \inf\{1/|w|^2 : w \in B(0, 1/2)\} = 2$.

In particular, this estimate on the derivative implies that we can bound the diameter of the intersection in (3.2) by $\text{diam}(B(s_0, 1/2) \cap (T \circ E)^{-1}B(s_1, 1/2) \cap \dots \cap (T \circ E)^{-n}B(s_n, 1/2)) \leq 1/2^n$ and letting n tend to infinity, we deduce that $z = [s_0; s_1, s_2, \dots]$. (Observe that there is an ambiguity in this algorithm in the case where either $\text{Re}(z_n)$ or $\text{Im}(z_n)$ takes a half integer value, for some $n \geq 0$. However, the set of such points is still only a set of zero Lebesgue measure.)

In order to relate this to the transformation $f: \mathcal{H} \rightarrow \mathcal{H}$ described in the previous section, observe that if we start with $z = z_0 \in \mathcal{H}$, then either $f(z) = z_1$ or $f(z) = \overline{z_1}$. The two possibilities arise depending on whether $E(z) \in B$ or $E(z) \in \overline{B}$ (i.e., whether $E(z)$ lies in either the upper or lower half planes). Similarly, if we start with $z_0 \in \overline{\mathcal{H}}$ and set $z = \overline{z_0}$, then either $f(z) = z_1$ or $f(z) = \overline{z_1}$.

Associate to each element $P \in \mathcal{P}$ the unique Gaussian integer $i(P) = u_P + iv_P \in \mathbb{Z}[i]$ with $P \subseteq B(u_P + iv_P, 1/2)$. If we have an itinerary $f^n(z) \in P_n$ (where each P_n is an element of \mathcal{P}) for $n \geq 0$, then we can associate to the sequence of elements in the Markov partition $\{P_n\} \in \Sigma^+$ a well-defined sequence of Gaussian integers $\{w_n\}$, where $w_n = i(P_n)$. (Notice that since $P_n \subset \mathcal{H}$, we must always have that $\text{Im}(w_n) \geq 0$.)

Proposition 2. Consider $z \in \mathcal{H}$ and assume that:

- (a) $z = [s_0; s_1, s_1, \dots]$ (using the above algorithm); and
- (b) $z = \pi(x_n)$, for $(x_n) \in \Sigma^+$.

Then, except on a set of zero measure, we have that either $i(x_n) = s_n$ or $i(x_n) = \overline{s_n}$ (whichever has nonnegative imaginary part).

4. CLOSED GEODESICS

We now come to a geometric interpretation of the Markov Partition \mathcal{P} for the map $f: \mathcal{H} \rightarrow \mathcal{H}$.

We denote by $\mathbb{H}^3 = \{(z, t) \in \mathbb{C} \times \mathbb{R} : t > 0\}$ the upper half plane model for hyperbolic space, to which we associate the Poincaré metric $ds^2 = \frac{dz^2 + dt^2}{t^2}$. We let G be the group of all isometries of \mathbb{H}^3 with respect to this metric. There is a simple correspondence between the linear fractional transformations of \mathbb{C} and the group G . Given any linear fractional transformation

$$A: z \mapsto \frac{az + b}{cz + d}, \quad \text{where } a, b, c, d \in \mathbb{C} \text{ satisfy } ad - bc = 1$$

we can associate an isometry $i(A) \in G$ of the form

$$i(A): (z, t) \mapsto (Az, t^*), \quad \text{where } t^* = t^*(A, z).$$

The linear map i is not quite injective, but its kernel consists only of the two elements $\{I, -I\}$. Nor is the map i quite surjective, but we only need to extend the definition of i by adding the conjugating map $*$: $\mathbb{C} \rightarrow \mathbb{C}$, which gives rise to the isometry $i(*)$: $(z, t) \mapsto (\bar{z}, t)$. For a fuller discussion of these matters we refer the reader to Beardon [B].

If we consider the Picard group Γ^* , then the image $i(\Gamma^*) \subset G$ is a discrete group of isometries on \mathbb{H}^3 (i.e., Kleinian groups). We denote the associated quotient manifold $M = \mathbb{H}^3/\Gamma^*$ and recall the following standard result:

Proposition 3. M is a noncompact finite-volume three manifold with constant negative sectional curvatures.

Remark. The finiteness of the volume should correspond to the existence of finite invariant measures on \mathcal{H} and \mathcal{B} (compare with [S]).

Any geodesic γ can be lifted to a geodesic $\tilde{\gamma}$ in the covering space \mathbb{H}^3 . Furthermore, we know that any geodesic $\tilde{\gamma}$ in \mathbb{H}^3 is a semicircular arc, which meets the complex plane \mathbb{C} perpendicularly at distinct points $x, y \in \mathbb{C}$, say, or is a vertical line which meets \mathbb{C} perpendicularly at a single point. For simplicity, we shall exclude the latter case (which we justify on the grounds that they form a set of zero measure in the space of geodesics). Thus, we shall always identify the geodesic $\tilde{\gamma}$ by the pair $(x, y) \in \mathbb{C} \times \mathbb{C}$ -diagonal.

The different choices of lift $\tilde{\gamma}$ of a given geodesic γ in M are related by the action of the group Γ . In particular, if (x, y) are the end points of a given lift, then necessary and sufficient conditions for a second pair to be end points of another lift is that they should be of the form (gx, gy) , for some $g \in \Gamma$. In particular, we see that by acting by a suitable element $A^n B^m (*)^\alpha$, if necessary, we can assume that $x \in \mathcal{H}$.

In the event that γ is a closed geodesic on M , it corresponds to a unique free homotopy class in $[g] \in \Gamma^* \simeq \pi_1(M)$ and $gx = x$, $gy = y$. We can write a representative element g in terms of the generators in the form

$$g = (TA^{n_1}B^{m_1})(TA^{n_2}B^{m_2}) \cdots (TA^{n_k}B^{m_k}).$$

The fixed point $gx = x$ has a (periodic) complex continued fraction expansion $x = [s_0; s_1, s_2, \dots]$, by §3, and by Proposition 2 it is an image $x = \pi(x_n)$, for $(x_n) \in \Sigma^+$. Finally, by Proposition 1, it is a periodic point $f^k x = x$ for $f: \mathcal{H} \rightarrow \mathcal{H}$.

Proposition 4. *There is a one-to-one correspondence between the closed geodesics for M and the periodic points $f^n x = x \in \mathcal{H}$. Furthermore, the closed geodesic corresponding to x has length*

$$l = \log |(f^n)'(x)| = \sum_{i=0}^{n-1} \log |f'(f^i x)|.$$

Proof. The correspondence was explained above. The length is given by the identity $|\operatorname{tr}(g)| = e^l + e^{-l}$, i.e.,

$$e^l = \frac{|\operatorname{tr} M| + (|\operatorname{tr} M|^2 - 4)^{1/2}}{2}.$$

If $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then the fixed point $gx = x$ satisfies $cx^2 + (d-a)x - b = 0$. Solving for x and substituting into the formula $g'(x) = \frac{1}{(cx+d)^2}$ completes the proof.

Remark. In [F'] Ford showed that for every irrational complex number $\xi \in \mathbb{C}$ there are infinitely many solutions $p, q \in \mathbb{Z}[i]$ to the approximation

$$(4.1) \quad \left| \xi - \frac{p}{q} \right| \leq \frac{1}{c|q|^2}$$

if $c = 3^{1/2}$. For a given $\xi \in \mathbb{C}$, we let $c(\xi) \geq 3^{1/2}$ denote the smallest value of c for which there are infinitely many solutions to (4.1). As observed in [F'] and [Sc], the value $c = c(\xi)$ is realized for

$$\xi_0 = \frac{(3^{1/2} - 2) + i}{2} \in \mathcal{B},$$

which is fixed by $g_0 = TB^2TA^{-2}TB^{-2}TA^2$ (cf. Case V in Theorem 1). (The value of $c(\xi)$ is the maximum height of the geodesic linking ξ_0 and $T\xi_0$, i.e., $c(\xi) = \frac{|\xi_0 - g\xi_0|}{2}$.) The partition \mathcal{P} makes it particularly easy to describe and compute other values. We give a few other examples:

- (a) $c(\xi_1) = 6^{1/2} = 2.44\dots$, for ξ_1 fixed by $g_1 = TA^{-1}B^{-1}TAB^{-1}$ (cf. Case I in Theorem 1);
- (b) $c(\xi_2) = 2$, for ξ_2 fixed by $g_2 = TB^2A^2TA^2B^{-2}$ (cf. Case II in Theorem 1);
- (c) $c(\xi_3) = 17^{1/4} = 2.03\dots$, for ξ_3 fixed by $g_3 = TA^2$ (cf. III in Theorem 1).

5. ZETA FUNCTIONS

In this section we want to consider our principal application of the Markov partition that we have constructed. This is to give an alternative method of constructing the analytic extension of the Selberg zeta function. We shall follow closely the related analysis for the Modular surface described in Mayer [M] and Pollicott [P].

For each element $P \in \mathcal{P}$ we can associate a unique square $\mathcal{B}(P) = \mathcal{B}^+(r, s)$ or $\mathcal{B}^-(r, s)$ containing P . We next choose Given $\delta > 0$ sufficiently small, we can define open neighbourhoods of each of these squares by

- (1) $U(P) = \{z \in \mathbb{C} : \inf_{w \in \mathcal{H}_2 \cap \{\operatorname{Re}(z) > 0\}} |z - w| < \delta\}$ if $P \in \mathcal{H}_2 \cap \{\operatorname{Re}(z) > 0\}$,
- (2) $U(P) = \{z \in \mathbb{C} : \inf_{w \in \mathcal{H}_2 \cap \{\operatorname{Re}(z) < 0\}} |z - w| < \delta\}$ if $P \in \mathcal{H}_2 \cap \{\operatorname{Re}(z) < 0\}$,
- (3) $U(P) = \{z \in \mathbb{C} : \inf_{w \in \mathcal{B}(P)} |z - w| < \delta\}$ if $P \in \mathcal{H}_1$.

(In practice, we require that $\delta < 2^{1/2} - 1$.) For each $P \in \mathcal{P}$ we can consider $TP = T(\mathcal{B}(P)) \subset \mathcal{B}^+$ or \mathcal{B}^- . We also see that whenever $P' \subset f(P)$, $DT(U(P)) \subset U(P')$ and $\bar{U}(P') \subset f(U(P))$.

Consider the space A_∞ of *real analytic* functions on the domain D which have a complex analytic extension (in two variables) to the neighbourhoods

$$U^C(P) = U(R) \times (-\varepsilon, \varepsilon)^2 \subset \mathbb{R}^2 \times i\mathbb{R}^2 = \mathbb{C}^2.$$

We let f_P^{-1} denote the local inverse to the map $f_P: P \rightarrow \mathcal{H}$, and we shall use the same notation to denote the analytic extension to $U^C(P)$ (providing that $\varepsilon > 0$ and $\delta > 0$ are sufficiently small for it to be well defined). Observe that whenever $(f_P)^{-1}P' \subset P$, $\operatorname{closure}((f_P)^{-1}U(P')) \subset U(P)$.

We denote by A_∞^0 the Banach space of functions which are real analytic on the disjoint union of the sets $U(R_i)$ and which have a uniformly bounded complex analytic extension to the sets $U^C(R_i)$. Compare with [R].

Consider the operators $L_{s,0}: A_\infty^0 \rightarrow A_\infty^0$ defined by:

$$(L_{s,0}h)(x_1, x_2) = \sum_{P: R_i \subset f(P)} |(f_P^{-1})'(x_1, x_2)|^s h(f_P^{-1}(x_1, x_2))$$

whenever $(x_1, x_2) \in U^C(R_i)$.

Remark. Since the sum has an infinite number of terms, we should require that $\operatorname{Re}(s) > 1$ for convergence, and we introduce the variables $x = (x_1, x_2)$ to emphasize that this is a function of *two* complex variables. This approach is needed because in the definition of the operator we use the function $|(f_P^{-1})'|$ which is *real analytic* (rather than complex analytic in the plane) in the variables $x = (x_1, x_2)$.

We can similarly define operators $L_{s,j}: A_\infty^j \rightarrow A_\infty^j$ acting on the space A_∞^j , for $j = 1, 2$, of real analytic differential j -forms on the disjoint union of the sets $U(R_i)$ and which have a uniformly bounded complex analytic extension to the sets $U^C(R_i)$; cf. [R].

Lemma 3. *For $\operatorname{Re}(s) > 1$ the operators $L_{s,j}$, $j = 0, 1, 2$, are nuclear of order zero and we have the identity*

$$(5.1) \quad \sum_{f^n x = x} e^{-s \log |(f^n)'(x)|} = \operatorname{trace}(L_{s,0}^n) - \operatorname{trace}(L_{s,1}^n) + \operatorname{trace}(L_{s,2}^n), \quad n \geq 1.$$

The proof of this result (in a more general context) is given in [R]. The identity corresponding to (5.1) for the Modular group is described in Mayer [M].

Consider the expressions

$$\zeta(z, s) = \exp \left(\sum_{n=0}^{\infty} \frac{z^n}{n} \sum_{f^n x = x} \exp(-s \log |(f^n)'(x)|) \right)$$

and

$$Z(s) = \exp \left(\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{n} \sum_{f^n x=x} \exp(-(s+k) \log |(f^n)'(x)|) x \right)$$

and observe that $Z(s) = \prod_{k=0}^{\infty} \zeta(1, s+k)$, for $\operatorname{Re}(s) > 1$.

By applying Grothendieck's Fredholm theory [M] and Proposition 4, we get the following:

Corollary 1. *The zeta function $\zeta(s, z)$ is analytic for*

$$\operatorname{Re}(s) > 1 \quad \text{and} \quad |z| < \limsup_{n \rightarrow +\infty} \left| \sum_{f^n x=x} \exp(-s \cdot |\log |(f^n)'(x)|)| \right|^{1/n}$$

and has a meromorphic extension to $\mathbb{C} \times \mathbb{C}$.

Corollary 2. *The zeta function $Z(s)$ is analytic for $\operatorname{Re}(s) > 1$ and has a meromorphic extension to \mathbb{C} .*

We now return to the geodesic flow for M . The manifold M has a countable number of closed geodesics whose lengths we denote by l_n , $n \geq 0$. The Selberg zeta function is defined to be the complex function given by the Euler product

$$(5.2) \quad Z_{\text{Selberg}}(s) = \prod_{k=0}^{\infty} \prod_{k=0}^{\infty} (1 - e^{-(s+k)l_n})$$

(which converges on the half plane $\operatorname{Re}(s) > 1$). Using our previous interpretation of lengths of closed geodesics in Proposition 4, we can rewrite the Selberg Zeta function as follows:

Lemma 4. $Z_{\text{Selberg}}(s) = Z(s)$.

In particular, comparing Lemma 3 and Lemma 4 we get a short proof of the following result.

Theorem 3. *The complex function $Z_{\text{Selberg}}(s)$ extends to a meromorphic function on \mathbb{C} ; cf. [Fr].*

Remark. We note that results of this type are of importance in applications to number theory; cf. [Sa]. It is expected that there should also be associated equi-distribution results (compare with [P]).

REFERENCES

- [AF] R. Adler and L. Flatto, *Cross section maps for geodesic flows I*, Ergodic Theory and Dynamical Systems II (A. Katok, ed.), Birkhäuser, Boston, MA, 1982.
- [B] A. Beardon, *The geometry of discrete groups*, Springer, Berlin, 1983.
- [F] L. Ford, *Rational approximations to irrational complex numbers*, Trans. Amer. Math. Soc. **25** (1918), 1–42.
- [F'] ———, *On the closeness of approach of complex rational fractions to a complex irrational number*, Trans. Amer. Math. Soc. **27** (1925), 146–154.
- [Fr] D. Fried, *The zeta functions of Ruelle and Selberg I*, Ann. Sci. École Norm. Sup. (4) **19** (1986), 491–517.
- [H] A. Hurewitz, *Über die Entwicklung Komplexer Grössen in Kettenbrüche*, Acta Math. **11** (1887), 187–200.

- [G] A. Grothendieck, *La theorie de Fredholm*, Bull. Soc. Math. France **84** (1956), 319–384.
- [Ma] J. Magnus, *Non-euclidean tessellations and their groups*, Academic Press, New York, 1974.
- [M] D. Mayer, *The thermodynamic formalism approach to Selberg's zeta function for $\mathrm{PSL}(2, \mathbb{Z})$* , Bull. Amer. Math. Soc. (N.S.) **25** (1991), 55–60.
- [P] M. Pollicott, *Distribution of closed geodesic on the modular surface and quadratic irrationals*, Bull. Soc. Math. France **114** (1986), 431–446.
- [R] D. Ruelle, *The zeta functions for expanding maps and Anosov flows*, Invent. Math. **34** (1978), 231–242.
- [Sa] P. Sarnak, *The arithmetic and geometry of some hyperbolic three manifolds*, Acta Math. **5** (1984), 253–295.
- [S] C. Series, *The modular surface and continued fractions*, J. London Math. Soc. (2) **31** (1985), 69–80.
- [Sc] A. Schmidt, *Farey triangles and Farey quadrangles in the complex plane*, **21** (1967), 241–295.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, ENGLAND
E-mail address: mp@maths.warwick.ac.uk