이학석사 학위논문

Explicit Examples of Coding Geodesics on Surfaces of Constant Negative Curvature

(음곡률을 가지는 곡면의 측지선 코딩)

2014년 2월

서울대학교 대학원

수리과학부

김 수 진

Explicit Examples of Coding Geodesics on Surfaces of Constant Negative Curvature

(음곡률을 가지는 곡면의 측지선 코딩)

지도교수 임 선 희

이 논문을 이학석사 학위논문으로 제출함

2013년 10월

서울대학교 대학원

수리과학부

김 수 진

김 수 진의 이학석사 학위논문을 인준함

2013년 12월

위 원 장 ______________________ (인)

부 위 원 장 ______________________ (인)

위 원 ______________________ (인)

Explicit Examples of Coding Geodesics on Surfaces of Constant Negative Curvature

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Master of Science
to the faculty of the Graduate School of
Seoul National University

by

Sujin Kim

Dissertation Director : Professor Seonhee Lim

Department of Mathematical Sciences
Seoul National University

February 2014

Abstract

It is well-known that there are various ways of representing geodesics on a surface M of constant negative curvature. There are two different methods on the bottom line: one is geometric coding and the other is arithmetic coding. The former is the so-called Morse method which is coding a geodesic by the cutting sequence as it passes a fixed set of curves on M. The latter, Artin method, is the construction using concatenating two sequences, obtained by using a suitable boundary expansion, of two endpoints of a lift of the geodesic (a geodesic in the unit disc $\mathcal{D}$). In this thesis, we investigate the more mysterious Artin method for specific examples of surfaces and show that we obtain a sofic system by using Artin method.

Key words: geodesic coding, symbolic dynamics
Student Number: 2011-20262

Contents

List of Figures

Chapter 1

Introduction

In 1898, Hadamard [10] proposed to use symbolic dynamics techniques in order to study the geodesic flow on a surface $\mathcal{S}$ of constant negative curvature. He constructed a surface of constant negative curvature whose geodesic can be represented by a sequence through certain "coding procedure".

This idea was inherited to Morse, Artin, Koebe, Nielsen, and Hedlund in the 1920's and 1930's. Morse and Artin developed the theory of coding geodesics independently around the same time.

Morse adopted the method known as the cutting sequence method. He associated a bi-infinite sequence to each geodesic on any open surface of variable negative curvature in 1921 ([16], [17]). He observed that each side of a given fundamental domain R is naturally associated with a unique generator of the fundamental group Γ. After labelling the same generator on the images of each side under the Γ-action, each geodesic γ determines a sequence of generators which label the successive sides of the images of the fundamental domain R cut by γ. This assignment is called the *Morse coding* and the sequence is called the *cutting sequence* of γ.

The Morse coding is canonical in the sense that the cutting sequence of a geodesic is uniquely determined once a tessellation is chosen. However, the sequence set obtained by this method may have a complicated structure because if the fundamental region has a vertex in the disc, any geodesic passing through a vertex of its tessellation has multiple coding sequences. In order to establish a one-to-one correspondence between geodesics and sequences with certain well-defined admissibility rules, one should take a small deformation of any geodesic which approaches to a vertex of the net N too closely. Here, the net N means the images of all sides of the fundamental domain under Γ.

In this thesis, we shall concentrate on *generic* geodesics, that is, geodesics not passing to any vertex of the net including vertices on $\mathbb{S}^1$ too closely.

Another approach, by Artin, is the method which codes the endpoints at infinity of some suitable lift of a given geodesic to $\mathcal{H}$ ([1]). Lifts of any geodesic on $\mathcal{S}$ are also geodesics in the upper half plane $\mathcal{H}$ and have two endpoints on $\mathbb{R} \cup \{\infty\}$. Artin made a symbolic representation of a geodesic by juxtaposing the continued fraction expansions of its endpoints.

This idea was not entirely new. A similar method was found by Nielsen previously. For a surface whose fundamental domain R is a symmetric $4g$-sided polygon, he noticed that points on the unit circle $\mathbb{S}^1$ can be expressed as a one-sided infinite sequence whose letters are elements of the fundamental group Γ as in the case of the continued fraction expansion of a real number ([19]). Hedlund used this idea to represent a geodesic by juxtaposing the Nielsen expansions of its endpoints and proved the ergodicity of the geodesic flow on $\mathcal{D}/\Gamma$ using the fact that conjugate geodesics have shift equivalent sequences and vice versa ([12]).

Now, in this spirit, the so-called *Artin coding* of a given geodesic γ is defined by simply concatenating two one-sided infinite sequences corresponding to two endpoints of γ using a suitable boundary map. The Artin coding is called *arithmetic coding* since this method is of arithmetic nature. This method gives different sequences for a fixed geodesic depending on the choice of the boundary map even with the same tessellation of the plane.

For a given Fuchsian group or a given surface, we want to know how to relate two sets obtained by the Morse method and the Artin method. In the case of a torus with punctures (Section 4.1), a geodesic has the same sequence both in the Morse coding and in the Artin coding. Moreover, the shift space can be defined by a single rule, namely, any sequence does not contain a word of the form gg^{-1} as its finite block where g is a generator of the fundamental group. This is a *1-step Markov chain*.

It is known that the set of Morse coding sequences for geodesics not passing too closely to a vertex of the tessellation is a topological Markov chain if and only if the fundamental domain R does not have any vertex in $\mathcal{D}$, or equivalently, all vertices of R lie on the boundary $\mathbb{S}^1$. In this case, we will say that R is a Dirichlet region consisting of an ideal polygon.

The case of a surface with genus two has some difficulty because its fundamental domain has interior vertices, in other words, it has vertices in the disc $\mathcal{D}$. We may choose a boundary map having a nice property, called the *Markov property*, with Markov partition on the boundary (see the details in Section 3.2). The set consisting of sequences defined by this map has a special property, namely, it is a subshift of finite type. We will see that this process show that the set of all sequences obtained by Artin coding of $\mathbb{S}^1$ except for countable points is a sofic system.

In this paper, we are going to summarize several aspects about the main two methods describing geodesics on a surface of constant negative curvature. In Chapter 2, we introduce the notion of the shift spaces. The sets obtained by the various coding methods are examples of shift spaces. we shall introduce some terminologies and theorems related to the geometry of the surface in Section 2.1 and explain the basic notion of symbolic dynamics in Section 2.2. The geometric coding and the arithmetic coding will be introduced in Section 3.1 and Section 3.2, respectively. We will then give some descriptions of Artin method on specific examples, namely, a punctured torus and a close surface of genus two in Section 4.1 and 4.2.

Chapter 2

Preliminaries

Throughout this paper, we write $\mathcal{H}$ for the upper half plane, $\mathcal{D}$ for the open unit disc (the Poincaré disc) in the complex plane equipped with the hyperbolic metric. The boundary of the open unit disc is denoted by $\partial\mathcal{D}$, often by $\mathbb{S}^1$. Let $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{C}$ denote the set of natural numbers, integers, real numbers, and complex numbers, respectively.

We use both the unit disc model and the upper half plane model of the hyperbolic geometry without distinction. Define a map $\Delta : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ to be $\Delta(z) = \dfrac{z-i}{z+i}$. Then this map and the inverse $\Delta^{-1} : w \mapsto \Delta^{-1}(w) = i\dfrac{1+w}{1-w}$ are Möbius transformations (and hence conformal maps). The map Δ sends the real axis and the point at infinity to the unit circle and leaves the upper half plane invariant. Through this map, we have analogous facts in the unit disc model as in the upper half plane model.

2.1 Fuchsian groups and Dirichlet regions

Let $\hat{\mathbb{C}}$ denote the extended complex plane, namely, the union of the complex plane and the point at infinity, $\mathbb{C} \cup \{\infty\}$. The Möbius group of all orientation-preserving Möbius transformations of $\hat{\mathbb{C}}$ onto itself is isomorphic to the projective general linear group $PGL(2, \mathbb{C})$ over $\mathbb{C}$ (that is, the quotient group $GL(2,\mathbb{C})/\{\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix} : \lambda \neq 0\})$ since all matrices of the form λA ($\lambda \neq 0$) give the same Möbius transformation as A ([2] §4).

The projective special linear group $PSL(2,\mathbb{R})$ over $\mathbb{R}$ acts on the upper half plane $\mathcal{H}$ by Möbius transformations: to each matrix $\begin{pmatrix}a & b\\ c & d\end{pmatrix}$ with $a,b,c,d \in \mathbb{R}$, $ad-bc=1$, we assign the Möbius transformation $g(z) = \dfrac{az+b}{cz+d}$ $(z \in \mathcal{H})$. Notice that $PSL(2,\mathbb{R})$ contains all transformations of the form $z \mapsto \dfrac{az+b}{cz+d}$ with $ad-bc>0$.

The circle $\{z \in \mathcal{D} : |cz+d| = 1\}$ is called the *isometric circle* of $g \in \Gamma$, where $g(z) = (az+b)/(cz+d)$ since $|g'(z)| > 1$ inside this circle and $|g'(z)| < 1$ outside.

Let $\mathrm{Isom}(\mathcal{H})$ denote the group of all transformations of the upper half plane $\mathcal{H}$ onto itself preserving the hyperbolic distance in $\mathcal{H}$, namely, isometries of $\mathcal{H}$ and $\mathrm{Isom}^+(\mathcal{H})$ be the subgroup of $\mathrm{Isom}(\mathcal{H})$ consisting of all orientation-preserving isometries of $\mathcal{H}$.

Theorem 2.1.1. ([14], Theorem 4.1) *The group Isom($\mathcal{H}$) is generated by the Möbius transformations from $PSL(2,\mathbb{R})$ together with the transformation $z \mapsto -\bar{z}$. The group $PSL(2,\mathbb{R})$ is a subgroup of Isom($\mathcal{H}$) of index two.*

This theorem gives the characterization of all the isometries of the upper half plane $\mathcal{H}$. We will refer to transformations of the forms $\dfrac{az+b}{cz+d}$ $(ad-bc=1)$ as *orientation-preserving* and to transformations of the form $\dfrac{a\bar{z}+b}{c\bar{z}+d}$ $(ad-bc=-1)$ as *orientation-reversing* isometries. Thus the group $PSL(2,\mathbb{R})$ consists exactly of all orientation-preserving isometries in the upper half plane, $\mathrm{Isom}^+(\mathcal{H})$ ([14] pp.8-9, 11). The action of the group $PSL(2,\mathbb{R})$ extends from $\mathcal{H}$ to its Euclidean boundary $\mathbb{R} \cup \{\infty\}$.

The group $GL(2,\mathbb{C})$ is a topological group with respect to the metric induced from $\mathbb{R}^4$; for $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the norm $\|A\|$ is given explicitly by

$$\|A\| = (|a|^2 + |b|^2 + |c|^2 + |d|^2)^{1/2}.$$

A subgroup G of $GL(2,\mathbb{C})$ is said to be *discrete* if the subspace topology on G is the discrete topology, that is, G is a discrete set in the topological space $\mathrm{Isom}(\mathcal{H})$. The following are equivalent conditions of the discreteness of a subgroup G of $GL(2,\mathbb{C})$:

1. For $X, A_1, A_2, \ldots \in G$, if A_n converges to X, then $A_n = X$ for all sufficiently large n. Note that X may not be in G but in $GL(2,\mathbb{C})$.

2. For $A_n \in G$, if A_n converges to the identity matrix I, then $A_n = I$ for almost all n.

3. In case of $SL(2,\mathbb{C})$, for each positive k, the set $\{A \in G : \|A\| \leq k\}$ is finite, that is, G cannot have any limit points (this criterion shows that a discrete subgroup G of $SL(2,\mathbb{C})$ is *countable*).

Any discrete subgroup of $\mathrm{Isom}(\mathcal{H})$ contains a special subgroup which consists of orientation-preserving isometries. We are concerned with a *discrete* subgroup of orientation-preserving isometries in the upper half plane.

Definition 2.1.2. A *Fuchsian group* is a discrete group consisting of orientation-preserving isometries in the upper half plane $\mathcal{H}$, or equivalently, is a discrete subgroup of $PSL(2,\mathbb{R})$.

The set of all orientation-preserving transformations in any discrete subgroup of $\mathrm{Isom}(\mathcal{H})$ is a Fuchsian group. Therefore, the study of Fuchsican groups is of importance when we study discrete subgroups of $\mathrm{Isom}(\mathcal{H})$. In general, discrete subgroups of isometries satisfy a slightly weaker discontinuity condition.

Definition 2.1.3. Let X be a locally compact metric space, and let G be a group of isometries of X. We say that G acts *properly discontinuously* on X if the G-orbit of any point $x \in X$ is locally finite, in other words, for any compact subset $K \subset X$, $\{gx \in K : g \in G\}$ is a finite set.

This definition is equivalent to the fact that each orbit has no accumulation point in X and the order of the stabilizer of each point is finite. In fact, the discreteness of all orbits implies the discreteness of the group.

Theorem 2.1.4. ([14], Theorem 8.6) *Let Γ be a subgroup of $PSL(2, \mathbb{R})$. Then Γ is a Fuchsian group if and only if Γ acts properly discontinuously on $\mathcal{H}$.*

Corollary 2.1.5. ([14], Corollary 8.7) *Let Γ be a subgroup of $PSL(2, \mathbb{R})$. Then Γ acts properly discontinuously on $\mathcal{H}$ if and only if for all $z \in \mathcal{H}$, the Γ-orbit Γz of z is a discrete subset of $\mathbb{H}$.*

Therefore, if $z \in \mathcal{H}$ and $\{g_n\}$ is a sequence of distinct elements in a Fuchsian group Γ such that $\{g_n(z)\}$ has a limit point $\alpha \in \hat{\mathbb{C}}$, then $\alpha \in \mathbb{R} \cup \{\infty\}$. For any Fuchsian group Γ, the limit set $\Lambda(\Gamma)$ lies in $\mathbb{R} \cup \{\infty\}$; or, in the unit disc model, $\Lambda(\Gamma) \subset \mathbb{S}^1$. In the upper half plane model, the centers of all isometric circles belong to the real axis $\mathbb{R}$. Let C_Γ be the set of the centers of the isometric circles of all elements in Γ. It is known that the limit set $\Lambda(\Gamma)$ is equal to the set of all limit points of C_Γ. We shall classify Fuchsian groups in the unit disc model as follows([13]):

(a) Fuchsian groups is said to be of the *first kind* if every point of the unit circle is a limit point.

(b) Fuchsian groups is said to be of the *second kind* if its limit points are nowhere dense on the unit circle, namely, an empty set, a set containing one or two points, or a perfect (and therefore infinite) nowhere dense set.

Suppose that Γ is a group of isometries acting properly discontinuously on $\mathcal{D}$. There is a geometric realization for the set of representatives of orbits, called a *fundamental domain*, which is a subset of $\mathcal{D}$ containing exactly one point from each of these orbits. The following is a precise definition:

Definition 2.1.6. A closed regoin F of $\mathcal{D}$ is defined to be a *fundamental domain* (or *fundamental region*) for Γ if it satisfies the following :

(i) $\bigcup_{g\in\Gamma} g(F) = \mathcal{D}$,

(ii) $\mathring{F} \cap g(\mathring{F}) = \emptyset$ for all $g \in \Gamma - \{id\}$.

A closed region means a closure of a non-empty open set $\mathring{F}$, called the *interior* of F. The set $\partial F = F \backslash \mathring{F}$ is called the *boundary* of F. The collection $\{g(F) : g \in \Gamma\}$ is called the *tessellation* or *tiling* of $\mathcal{D}$.

Suppose now that Γ is a Fuchsian group acting discontinuously on $\mathcal{D}$. Fundamental regions of Fuchsian groups are our main concern. Any Fuchsian group possesses a nice fundamental region. Therefore, we are going to consider only the following special kind of fundamental domains in this article:

Definition 2.1.7. Let Γ be an arbitrary Fuchsian group and let $p \in \mathcal{D}$ be not fixed by any element of $\Gamma - \{id\}$. Let $D_p(\Gamma) = \{z \in \mathcal{D} : \rho(z,p) \leq \rho(z, g(p))$ for all $g \in \Gamma\}$ where ρ is the hyperbolic metric in $\mathcal{D}$. Then we call this set $D_p(\Gamma)$ the *Dirichlet region* for Γ centered at p. Equivalently, for each fixed $g \in PSL(2,\mathbb{R})$, if we define the hyperbolic half-plane $H_p = \{z : \rho(z,p) \leq \rho(z, g(p))\}$, then *Dirichlet region* for Γ centered at p is the intersection of such hyperbolic half-planes

$$D_p(\Gamma) = \bigcap_{g\in\Gamma, g\neq id} H_p(g),$$

and thus it is a *hyperbolically convex* region.

The set obtained from this definition is indeed a connected and convex fundamental region.

Theorem 2.1.8. ([14], Theorem 10.3) *If p is not fixed by any element of $\Gamma - \{id\}$, then $D_p(\Gamma)$ is a connected fundamental region for Γ.*

The tessellation $\{g(F) : g \in \Gamma\}$ of $\mathcal{D}$ by a Dirichlet region F is referred to as a *Dirichlet tessellation* and all its images under Γ are called *faces*. The intersection between two bordering faces is said to be an *edge*, and a *vertex* is the point of intersection of three or more bordering faces. We will sometimes call the collection of all edges in the tessellation the *net*. This Dirichlet tessellation has nice local properties, namely, local finiteness.

Proposition 2.1.9. ([14], Proposition 11.3) *The vertices of a Dirichlet region F are isolated, in other words, every vertex of F has a neighborhood containing no other vertices of F.*

Corollary 2.1.10. ([14], Corollary 11.4) *A compact Dirichlet region has a finite number of vertices.*

The following theorem suggests a geometric interpretation of the Γ-action in the fundamental domain:

Theorem 2.1.11. ([14], Theorem 11.8) *For some fixed Dirichlet region F, let $\{g_i\}$ be the subset of Γ consisting of those elements which pair the sides of F. Then $\{g_i\}$ is a set of generators for Γ.*

Theorem 2.1.12. ([14], Theorem 11.7 *Let F be a Dirichlet region for Γ. Let $\theta_1, \theta_2, \cdots, \theta_t$ be the internal angles at all congruent vertices of F. Let m be the order of the stabilizer in Γ of one of these vertices. Then $\theta_1 + \theta_2 + \cdots + \theta_t = 2\pi/m$.*

Remark. Therefore, if a vertex is not a fixed point, then we have $m = 1$ and $\theta_1 + \theta_2 + \cdots + \theta_t = 2\pi$. The following is known: as F is locally finite, there are only finitely many vertices in a congruent cycle. As the stabilizers of two points in a congruent set are conjugate subgroups of Γ, they have the same order.

We shall deal with Fuchsian group of finite covolume and hence the following theorem is worth recording.

Theorem 2.1.13 (Seigel's Theorem)**.** ([14], Theorem 13.1) *If the Fuchsian group* Γ *is such that* $\mu(\Gamma\backslash\mathcal{H}) < \infty$*, then* Γ *is* geometrically finite*, in other words, any Dirichlet region* $F = D_p(\Gamma)$ *has finitely many sides.*

There are three types of elements in $PSL(2,\mathbb{R}) = \{z \mapsto g(z) = \dfrac{az+b}{cz+d} : ad - bc = 1\}$ distinguished by the the absolute value of its trace $Tr(g) = |a+d|$. The transformation g is called *elliptic* if $Tr(g) < 2$; *parabolic* if $Tr(g) = 2$; *hyperbolic* if $Tr(g) > 2$. By solving the equation $z = \dfrac{az+b}{cz+d}$, we see that a hyperbolic transformation has two fixed points in $\mathbb{R} \cup \{\infty\}$ (one repulsive and one attractive), a parabolic transformation has one fixed point in $\mathbb{R} \cup \{\infty\}$ and an elliptic transformation has a pair of complex conjugate fixed points, and hence, one fixed point in $\mathcal{H}$ ([13] §2.1).

Definition 2.1.14. A Fuchsian group is said to be *cocompact* if the quotient space $\Gamma\backslash\mathcal{H}$ is compact, where $\mathcal{H}$ is the hyperbolic upper half plane $\{z \in \mathbb{C} : im(z) > 0\}$.

The theorem and corollary below explain the relationship between the compactness of the quotient space and the property of a Fuchsian group Γ:

Theorem 2.1.15. ([14], Theorem 14.2) *If a Fuchsian group* Γ *is cocompact, that is,* Γ *has a compact Dirichlet region, then* Γ *contains no parabolic elements.*

Since the compactness of a Dirichlet region for a Fuchsian group Γ is equivalent to the cocompactness of Γ ([14], Corollary 14.4), we have the following corollary:

Corollary 2.1.16. ([14], Corollary 14.8) *A Fuchsian group* Γ *is cocompact if and only if* $\mu(\Gamma\backslash\mathcal{H}) < \infty$ *and* Γ *has no parabolic elements.*

2.2 Subshifts of finite type and Sofic systems

By a *discrete-time dynamical system*, we mean a pair (X, f) of a non-empty set X and a map $f : X \to X$. From now on, we shall introduce the notion of shift spaces, which are well-known examples of discrete-time dynamical systems.

For any natural number $m > 1$, we call a set $\mathcal{A}_m = \{1, 2, \ldots, m\}$ an *alphabet* and its elements *symbols*. we refer a *word* as a finite sequence consisting of symbols.

Let (X, f) and (Y, g) be discrete-time dynamical systems. If there exists a surjective map $\pi : Y \to X$ such that $\pi \circ g = f \circ \pi$, then we say (X, f) is a *factor* of (Y, g) and (Y, g) is an *extension* of (X, f). In this case, the map π is called a *semiconjugacy*. The map π is also called a *factor map* or a *projection*.

$$\begin{tikzcd} Y \arrow[r, "g"] \arrow[d, "\pi"'] & Y \arrow[d, "\pi"] \\ X \arrow[r, "f"] & X \end{tikzcd}$$

Moreover, if the map π is invertible, π is called a *conjugacy*. In this case, X and Y are said to be *conjugate*.

Let Σ_m $(= A_m^{\mathbb{Z}})$ be the space of all two-sided infinite sequences of symbols in A_m, and Σ_m^+ $(= A_m^{\mathbb{N}})$ be the space of all one-sided infinite sequences of symbols in A_m. Define the left shift map σ on Σ_m or Σ_m^+ so that $\sigma(x)_i = x_{i+1}$ for all i. The pair (Σ_m, σ) is the *full two-sided shift*; (Σ_m^+, σ) the *full one-sided shift*. These are typical examples of symbolic dynamical systems.

From now on, we are concerned with two-sided shift in the remaining section.
The set Σ_m is a topological space (often called the *Cantor space*) with respect to the distance d defined by $d(x, y) = 2^{-n}$ with $n = \min\{|k| \mid x_k \neq y_k,\ k \in \mathbb{Z}\}$ for $x, y \in \Sigma_m$.
We are interested in a subset which can be a dynamical system with the

left shift map. We say that a bi-infinite word $z \in \Sigma_m$ *avoids* a set of words $X \subset \Sigma_m$ if z does not contain any word in X (we will call each word in X a *forbidden word*). A *symbolic dynamical system* S of Σ_m is defined by such a subset X.

Proposition 2.2.1. ([15], Proposition 1.5.1) *A subset S of Σ_m is a symbolic dynamical system if and only if it is closed for the topology and invariant under the left shift map.*

This proposition gives the following definition:

Definition 2.2.2. A *subshift* is a closed subset $X \subset \Sigma_m$, which remains invariant under the shift map σ and its inverse σ^{-1}.

Let $X_i \subset \Sigma_m (i = 1, 2)$ be two subshifts. By a *code* from X_1 to X_2, we mean a continuous map from X_1 to X_2 commuting with the shift map σ, that is, $c : X_1 \to X_2$ such that $c \circ \sigma = \sigma \circ c$. Note that a surjective code is a factor map.

There are two things that can describe a *natural* class of subshifts: *adjacency matrices* and their associated *directed graphs*. An adjacency matrix is an $m \times m$ matrix whose entries are zero or one. To given an adjacency matrix $A = (a_{ij})$, we associate a directed graph Λ_A having m vertices and the number a_{ij} of edges from vertex i to vertex j for all i, j. Conversely, a finite directed graph Λ with n vertices and no multiple edges determines an $n \times n$ adjacency matrix.

Assume that we have an $m \times m$ adjacency matrix $A = (a_{ij})$. A word or an infinite sequence x in the alphabet $\mathcal{A}_m$ is said to be *allowed* (*accessible* in some other text) provided that $a_{x_i x_{i+1}} > 0$ for every i, that is, if there exists an edge from x_i to x_{i+1} in the associated directed graph Λ_A for each i, where x_i is the index of the vertex visited at time i. Therefore, we consider an allowed word or sequence as a walk along directed edges in the graph Λ_A. A word or an infinite sequence is *forbidden* if it is not allowed.

For a given adjacency matrix A, let $\Sigma_A \subset \Sigma_m$ be the set of all allowed two-sided sequences as explained in the above. Then Σ_A is a closed shift-invariant subset, namely, a subshift. In this case, the pair (Σ_A, σ) is called the two-sided *vertex shift* determined by A.

Definition 2.2.3. A subshift $X \subset \Sigma_m$ is called a *subshift of finite type* (SFT) if there is a finite collection of finite forbidden words which can determine X, that is, there are finitely many words such that X consists of precisely the sequences in Σ_m that do not contain any of these words. X is called a *k-step SFT* if the length of any forbidden word is at most $k+1$. In particular, when $k = 1$, X is called a *topological Markov chain.*

A vertex shift is an example of a subshift of finite type, more precisely, a topological Markov chain. Moreover, we have the converse:

Proposition 2.2.4. ([8], Proposition 3.2.1) *Every subshift of finite type is isomorphic to a vertex shift.*

Proof. Let X be a k-step SFT. We want to find an adjacency matrix A and hence a conjugacy c from X to some vertex shift Σ_A^v, the set of all allowed sequences defined by A. Let $W_k(X)$ be the set of all words of length k appearing in X, and let n be the cardinality of $W_k(X)$. Construct a graph Λ as follows:

(1) The vertex set $V(\Lambda)$ consists of elements of $W_k(X)$.

(2) The elements of edge set $E(\Lambda)$ are defined as follows: there is a directed edge from a vertex $x_1x_2\cdots x_k$ to a vertex $x'_1x'_2\cdots x'_k$ if the words $x_1x_2\cdots x_kx'_k = x_1x'_1x'_2\cdots x'_k$ belong to the set of all words in X of length $k+1$, $W_{k+1}(X)$.

Then, this edge condition gives an $n \times n$ adjacency matrix A, and hence determines a vertex shift Σ_A^v. Clearly, Σ_A^v is a subset of X. If we define the map $c : X \to \Sigma_A^v$ by $c(x)_i = x_ix_{i+1}\cdots x_{i+k-1} \in W_k(X)$, c is an invertible surjective map, a conjugacy from X to Σ_A^v. $\square$

More generally, if we consider an $n \times n$ matrix with *nonnegative* integer entries, we can construct a finite directed graph with *multiple* directed edges between some pairs of vertices. To be precise, for an $n \times n$ adjacency matrix B with each entry nonnegative integer, there is a directed graph Λ with n vertices and with B_{ij} labelled edges from the vertex i to the vertex j. Let Σ_B^e be the set of all infinite directed paths in Λ. Then Σ_B^e is closed and shift-invariant and is called the *edge shift* determined by B. Note that any edge shift is isomorphic to a vertex shift in that we can construct a new graph by taking an edge of the original as a vertex. Hence it is also a subshift of finite type by the above proposition.

Corollary 2.2.5. ([8], Corollary 3.2.2) *Every subshift of finite type is isomorphic to an edge shift.*

We sometimes seek the weaker notion than subshifts of finite type; a generalization of subshifts of finite type, called *sofic shift*.

Definition 2.2.6. A subshift $X \in \Sigma_m$ is called *sofic* if it is a factor of a subshift of finite type, that is, there is an adjacency matrix A and a surjective code $c : \Sigma_A^e \to X$ such that $c \circ \sigma = \sigma \circ c$. Here, Σ_A^e means the edge shift determined by A.

Let Λ be a finite directed labelled graph, that is, the edges of Λ are labelled by a symbol of a fixed alphabet set. Note that, in this case, the graph Λ has *the same label on some different edges*, in other words, we allow that different edges of Λ have the same label (this is the only difference with an edge shift). The subset $X_\Lambda \subset \Sigma_m$ consisting of all infinite directed paths in Λ is closed and shift invariant. If a subshift (X, σ) is isomorphic to (X_Λ, σ) for some directed labelled graph Λ, then we say that Λ is a *presentation* of X or Λ *presents* X. Now we are ready to introduce the most important proposition in this section:

Proposition 2.2.7. ([8], Proposition 3.7.1) *A subshift $X \subset \Sigma_m$ is sofic if and only if it admits a presentation by a finite directed labelled graph.*

Remark. ([6]) There is "almost" the same notion as a sofic system in the coding theory. A *constrained system* or *constraint system* is the set X of all words (finite sequences) obtained from reading the labels of edge paths in a labelled directed graph Λ. In this case, we say that Λ represents X. A constrained system is equivalent to a regular language in automata theory which is recognized by an automaton, namely, the states of which are all accepting. A constrained system is called a sofic system in symbolic dynamics except that a sofic system usually refers to the *bi-infinite* symbol sequences generated by a labelled directed graph. A constrained system should not be confused with any particular labelled graph because a given constrained system can be presented by many different labelled directed graphs.

Chapter 3

Symbolic Coding of Geodesics

Recall that a discrete group of linear fractional transformations of the form $z \mapsto (az+b)/(cz+d)$ with $ad-bc=1$ is a *Fuchsian group* and when there are points of the unit circle $\mathbb{S}^1$ with dense orbits, that is, its limit set is the whole unit circle, Γ is said to be *of the first kind*. Let Γ be a finitely generated Fuchsian group of the first kind acting in the unit disc $\mathcal{D}$. The corresponding surface $\mathcal{S} = \mathcal{D}/\Gamma$ is a Riemann surface of constant negative curvature of finite area.

Recall that the circle $\{z \in \mathcal{D} : |cz+d| = 1\}$ is called the *isometric circle* of $g \in \Gamma$, where $g(z) = (az+b)/(cz+d)$ since $|g'(z)| > 1$ inside this circle and $|g'(z)| < 1$ outside. Such isometric circles are always circles orthogonal to $\mathbb{S}^1 = \partial\mathcal{D}$.

Throughout this article, we deal with the only case that the fundamental region R has *even corners*. This means that the net $N = \Gamma(\partial R)$ consists of complete geodesics in $\mathcal{D}$. Series ([5]) redefined this notion as follows:

Definition 3.0.8. we will say that R satisfies *property* $(*)$ if, for each side s of R,

(i) $C(s)$ is the isometric circle of g_s,

(ii) $C(s)$ lies completely in the net N, where $C(s)$ is the circle containing s orthogonal to $\mathbb{S}^1$ and the other side of R is identified with s by an element $g_s \in \Gamma$.

According to the paper [5], we can construct a fundamental domain for Γ which satisfies property $(*)$ whenever the quotient space $\mathcal{D}/\Gamma$ has positive area. Therefore, *from now on we shall suppose that R satisfy the property* $(*)$ *and moreover that R is not a triangle and dose not have elliptic vertices of order 2.*

In this setting, we also assume that the origin O is not fixed by any element of $\Gamma - \{id\}$. Then Γ has a fundamental region R (in fact, a Dirichlet region) R in $\mathcal{D}$ which can be chosen to be a polygon consisting of a finite number of circular arcs and containing the origin O of $\mathcal{D}$ inside (Theorem 2.1.8 and Theorem 2.1.13). A Dirichlet region R of Γ is a geodesically convex polygon. Note that *geodesic convexity* of R means that the geodesic arc joining any two points in R lies in R. Therefore, each extension of a side of R in the Poincaré disc $\mathcal{D}$ is a circle orthogonal to the boundary $\mathbb{S}^1$ of the disc, namely, R consists of a finite number of geodesic segments.

By property $(*)$, R has an even number of sides, say $\{s_1, s_2, \ldots, s_{2n}\}$. Moreover, each side s_i of R is identified with another side s_j by the isometry $g_i \in \Gamma$. Then the set $\{g_1, g_2, \ldots, g_{2n}\}$ is the symmetric set of generators for Γ and its inverses, denoted by Γ_R (Theorem 2.1.11).

Obviously, geodesics on $\mathcal{S}$ are the projections of circular arcs in $\mathcal{D}$ orthogonal to $\mathbb{S}^1$. Note that $\mathcal{S}$ is compact if and only if R has no cusps, vertices lying on $\mathbb{S}^1$ (Theorem 2.1.15 and Corollary 2.1.16).

We shall describe the Morse method and the Artin method of coding geodesics on the torus with n punctures and the closed surface with genus two.

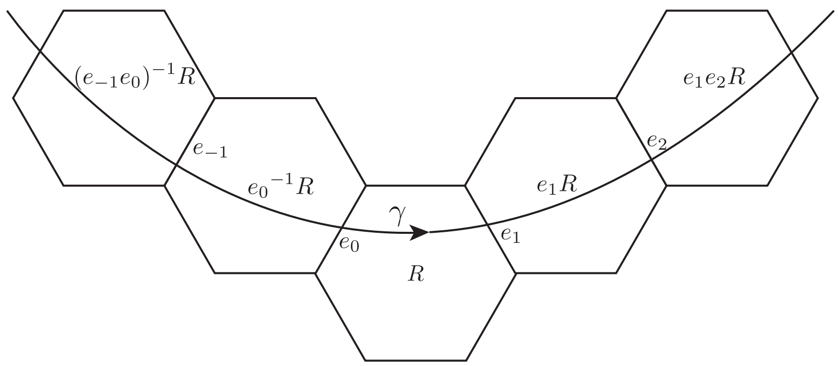

Figure 3.1: The Morse coding method

3.1 Geometric coding

Let $\mathcal{G}$ denote the set of all oriented geodesics in $\mathcal{D}$, $\mathcal{R} = \{\gamma \in \mathcal{G} : \gamma \cap R \neq \emptyset\}$, and let Σ be the set of all reduced bi-infinite sequences in Γ_R. Here, *reduced* means that a generator $g \in \Gamma_R$ dose not follow or precede its inverse g^{-1}.

For a given surface $\mathcal{S}$, after choosing a suitable fundamental domain R, we label each side s of R on the inside by the element g of Γ_R if the side s is identified with $g(s)$ in R. This is equivalent to labelling the side s on the outside by the element g^{-1}, and from now on we call the latter *exterior labelling*. Similarly, label all images of side s under Γ with the same generator g on the inside to obtain a labelling of the net $N = \Gamma(\partial R)$, the images of sides of R. With this convention, if gR and hR are adjacent along a side s, then the side of s which is interior to hR is labelled $g^{-1}h$, and that which is interior to gR is labelled $h^{-1}g$.

To an oriented geodesic $\gamma \in \mathcal{R}$, we associate a sequence by recording the labellings of the edges of the net N crossed by γ one by one in the direction of its orientation. This process is called *the Morse coding method*. More precisely, define the Morse coding sequence $x = x(n)$ of γ as follows: when the geodesic γ leaves the fundamental domain R through a side s, $x(1) = e_1$

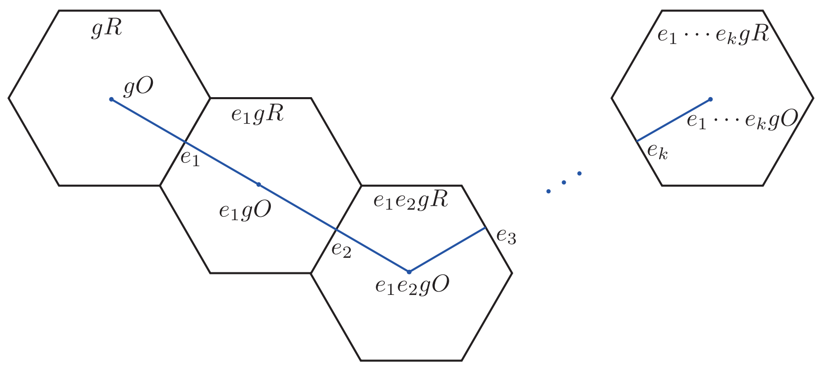

Figure 3.2: An edge path in Definition 3.1.1

is the exterior label of the side s. Similarly, $x(2) = e_2$ is given the exterior label of the side of e_1R intersecting with γ. $x(0) = e_0$ is the interior label of the side of R where γ is entering into R through the side and $x(-1) = e_{-1}$ is the interior label of the side of $e_0^{-1}R$ crossed by γ, so on (see Figure 3.1). The sequence for x is $\cdots e_{-1}e_oe_1e_2\cdots, e_i \in \Gamma_R$. This process provides us a bi-infinite sequence consisting of elements of Γ_R corresponding to a geodesic. We call it the *cutting sequence* or *geometric Morse code* of the given geodesic. Clearly, this sequence is always reduced since the geodesic does not go back-tracking and the occurrence of consecutive letters ee^{-1} for some $e \in \Gamma_R$ in the sequence implies that the geodesic cut the side (corresponding the generator e) twice in a row coming from opposite direction ([21]).

As mentioned, we have that $O \in R$ and that O is not a fixed point of any element of $\Gamma - \{id\}$. Let N^* be the dual net to N obtained by joining a geodesic segment from gO to hO whenever $g^{-1}h \in \Gamma_R$. Note that the dual net N^* may be regarded as the Cayley graph given by Γ since O is not a fixed point of any element of Γ.

Definition 3.1.1. Any path in N^* will be called an *edge path* (see Figure 3.2). We may associate such a path with the *polygonal path* which consists of the adjacent regions: with initial region gR ($g \in R$), for any word $w = e_1 \cdots e_k (e_i \in \Gamma_R)$, we obtain the associated polygonal path being connected

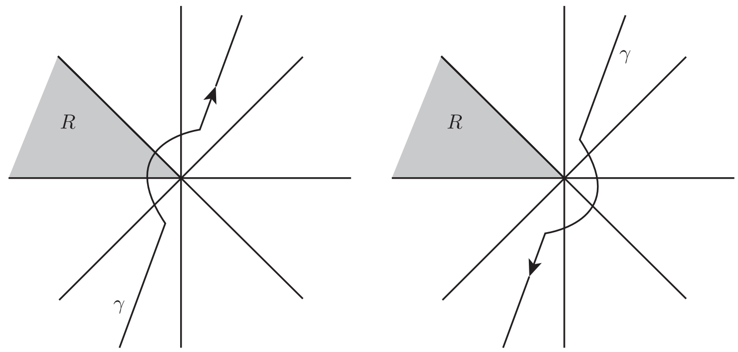

Figure 3.3: Deformation to the left around the vertex

by the regions $gR, e_1gR, \dots, e_1 \cdots e_kgR$. In this case, the corresponding edge path consisting of the geodesic segments joining $gO, e_1gO, \dots, e_1 \cdots e_kgO$ and the cutting sequence of this path is $e_1 \cdots e_k$. In particular, a path defined by a geodesic arc in this way is called a *geodesic edge path* and the corresponding cutting sequence is called a *geodesic word*([4]).

This Morse method is natural and well-understood, however, it has a disadvantage that we have at least two different sequences corresponding to a geodesic when that geodesic meets a vertex of the net N.

In this article, for a given geodesic meeting a vertex, we replace it by one passing the left side of the vertex instead of the vertex, that is, we deform it to the left around the vertex. Imagine that we walk along the geodesic and detour to the left of the vertex (see Figure 3.3). This phenomenon happens when the fundamental region has an *interior vertex*, that is, a vertex in $\mathcal{D}$.

If v is a vertex of the net N, we find a cutting sequence corresponding to a relation in Γ from a small circle around v. By assumption, the sequence has even length and hence this sequence is represented by $e_1 \cdots e_{2n(v)}$ for $e_i \in \Gamma_R$. By a *cycle*, we mean any finite sequence of generators whose letters occur in the order of one of these relations: we call such a sequence with length $n(v)$ a *half cycle*, and any cycle with greater length a *long cycle*.

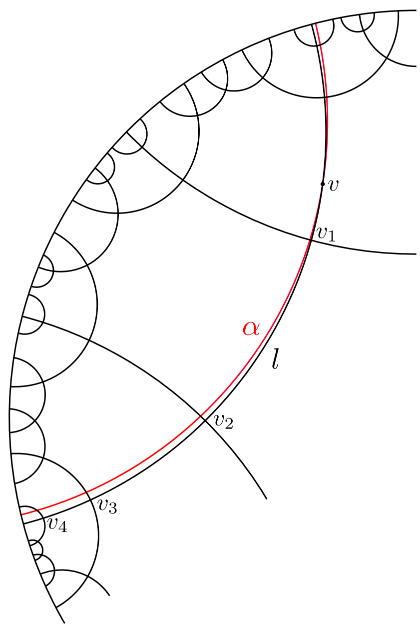

Figure 3.4: Consecutive cycles

Suppose that $v_1, \ldots, v_p$ are some successive vertices of the net N on some geodesic l in N. If α is a curve running almost parallel to and close to l on one side possibly cutting l before v_1 and after v_p (see Figure 3.4: in this case, two geodesics meet at a point v). The cutting sequence of α consists of cycles at $v_1, \ldots, v_p$ and the length of the cycle at each intermediary vertex v_i $(1 < i < p)$ is $n(v_i) - 1$. We call such cycles *consecutive*, and the sequence of consecutive cycles we call a *chain*. By a *polygonal chain* in N (or N^*), we mean a sequence of polygons in N (or N^*) each meeting the next in a common edge.

Definition 3.1.2. A chain is said to be *long* if it consists of cycles of lengths $n(v_1), n(v_2) - 1, \ldots, n(v_{p-1}) - 1, n(v_p)$, where $v_1, v_2, \cdots, v_p$ are the vertices of the net N dual to the corresponding polygonal chain. In other words, a long chain is a chain with half cycles at the both ends.

The main theorem (Theorem 3.1.8) needs some restriction on the fundamental region R. In our situation, we consider the only "good" cases that

$\mathcal{S}$ is a surface having no branch point and hence it satisfies the both of the below conditions.

Hypothesis. ([4], Hypothesis 2.1) *Let R be a fundamental region for a group Γ.*

(i) If R has three sides, then not all vertices of R lie in Int $\mathcal{D}$.

(ii) If R has four sides, and if all vertices of R lie in Int $\mathcal{D}$, then at least three geodesics in the net N cross at each vertex of R.

Remark. These hypotheses are necessary conditions so that our fundamental domain R has no branch point, that is, Γ has no elliptic elements. If R is a triangle, then any vertex in Int $\mathcal{D}$ will be a branch point. Therefore, we expect the good situation where at least one vertex lies on $\partial\mathcal{D}$. If R has four sides with all vertices in Int $\mathcal{D}$, and if there is a vertex where exactly two geodesics meet, the angle between the geodesics is equal to $\pi/2$ and all the interior angles are exactly equal to $\pi/2$; this means the fundamental domain is a rectangle which has the Euclidean plane as its universal cover, in other words, it is not covered by the hyperbolic disc $\mathcal{D}$. To exclude this case, we put the condition (ii).

Lemma 3.1.3. ([5], Lemma 2.2) *A geodesic lying in the net N cuts a geodesic edge path in N^* at most once.*

Proof. Let $E(\gamma)$ be a geodesic edge path in N^* connecting $g_1O, g_2O, \ldots, g_kO$. Suppose that a geodesic C in the net N intersects $E(\gamma)$ between g_1O and g_2O and again between $g_{k-1}O$ and g_kO. Since C is in the net N, the images of ∂R under Γ, C contains the geodesic segments of $g_1R \cap g_2R$ and $g_{k-1}R \cap g_kR$. This means that γ crosses from g_1R into g_2R and from $g_{k-1}R$ into g_kR, and hence the geodesics γ and C meet at two points, which is a contradiction. □

Lemma 3.1.4. ([5], Lemma 2.2) *Let s and s' be nonadjacent sides of R. Then the geodesics which extend s and s' do not intersect.*

Proof. Let $C(s), C(s')$ be the geodesics in the disc $\mathcal{D}$ containing the side s, s', respectively. Assume that $C(s)$ and $C(s')$ intersect at a point P so that it leads to be a contradiction. Let $s_1, \ldots, s_k$ be the sides of the fundamental region R between s and s' connecting in the order $s, s_1, \ldots, s_k, s'$. Then $k \geq 1$ since s and s' be nonadjacent sides of R.

Let A be the intersection point (the vertex of R) between s and s_1, B the intersection point (the vertex of R) between s' and s_k, and let γ be the geodesic segment between A and B. Since R is geodesically convex, γ lies in R, in other words, $s_1, \ldots, s_k$ lie in the triangle $\triangle ABP$. Since $C(s_k)$ meets with γ at the point B, $C(s_k)$ must cross $C(s)$, otherwise $C(s_k)$ intersects with γ once more or $C(s_k)$ should be the geodesic $C(s')$. Let P_k be the intersection point of $C(s)$ and $C(s_k)$. Then $C(s)$ and $C(s_k)$ intersect at P_k and there are $(k-1)$ sides $s_1, \ldots, s_{k-1}$ between s and s_k. Inductively arguing as above, it reduces to the case $k = 1$ and hence we may assume that s and s' connect through the only one side s_1.

Consider $g_1 R$, the image of R adjacent to the side s_1, that is, the exterior label of s is g. Let t, t' be the sides of $g_1 R$ which pass the point A, B, respectively. Then t and t' must intersect at a point Q within the triangle $\triangle ABP$ because $g_1 R$ is not a triangle(this means Q is not equal to P) and $t$and t' meets with $C(s)$ or $C(s')$ at most once. Therefore, one can find a copy of R within the triangle $\triangle ABP$. Similarly, after continuing this argument we obtain an infinite set $\{g_1 R, g_2 R, \ldots\}$ of (almost) disjoint copies of R contained in $\triangle ABP$. This contradicts to the fact that $\triangle ABP$ has finite area. $\square$

To each cycle or chain is associated a complementary cycle or chain. By a *complementary* cycle or chain, we mean the opposite boundary of the corresponding polygonal chain, and it also represents the same element of Γ. Clearly, the complement of a long cycle or chain is a path of strictly shorter length.

Lemma 3.1.5. ([4], Lemma 2.5) *The vertex angle along any path E which is either geodesic or which contains no long cycles are at most π^+. If the angle at v is π^+ then we may replace the cycle at v by its complementary cycle and obtain another path of the same length.*

Lemma 3.1.6. ([4], Lemma 2.6) *Let E be an edge path which is either geodesic or which contains no long chains. Then the occurrence of a sequence of vertex angles $\pi^+, \pi, \ldots, \pi, \pi^+$ along E is impossible.*

If E is an edge path in N^*, we denote the corresponding polygonal chain in N^* by $P(E)$.

Lemma 3.1.7. ([4], Lemma 2.7 or [21], Lemma 5.6) *Let E_1 and E_2 be edge paths containing no long chains and with coincident initial and final points. Then there are no copies gR of R lying strictly inside the region bounded by the polygonal chains $P(E_1), P(E_2)$ defined by E_1 and E_2.*

We say that an edge path is *shortest* provided that the corresponding word is a shortest possible representation of the element in Γ defined by the word. By the above proposition, the regions formed by two shortest polygonal paths with the same initial and final endpoints (possibly at infinity) should either coincide or be adjacent. Here, a shortest polygonal path means that the corresponding cutting sequence of the polygonal path is shortest. The paths differ only by taking complementary cycles round vertices of their common boundary. When two polygonal paths have common initial and final regions, the two paths have the same length (Lemma 3.1.5).

We state the main theorem in this section that explains which sequence can be a cutting sequence of a geodesic:

Theorem 3.1.8. ([21], Theorem 3.1) *Suppose that R is not a triangle and that R has even corners. Assume that R satisfies the conditions of Hypothesis(ii). Then,*

(i) an edge path is shortest if and only if it is reduced and contains no long cycles or long chains, and

(ii) the cutting sequences of geodesic arcs are shortest.

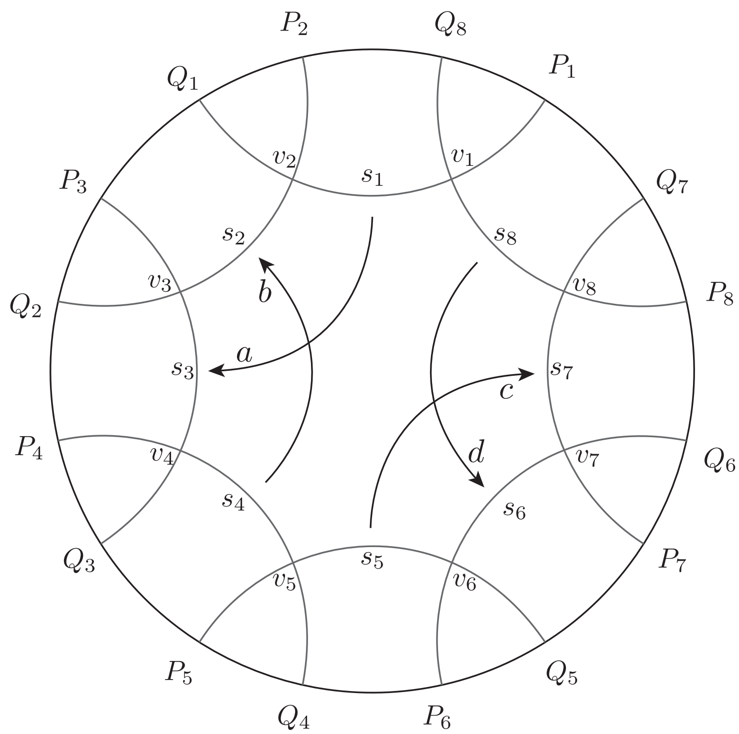

Figure 3.5: An example: labels on a surface with genus 2

3.2 Arithmetic coding

We can describe a real number as an infinite sequence using continued fraction expansion. Artin(1965) obtained a representation of geodesics in the Poincaré upper half plane $\mathcal{H}$ as bi-infinite sequences of positive integers by juxtaposing the continued fraction expansions of their endpoints. Arithmetic coding is the method which represents a geodesic as a bi-infinite sequence by concatenating two sequences of its end points analogous to the continued fraction expansion.

We find a one-sided sequence for a point on $\partial\mathcal{D}$ using a particular map, called a *boundary map* or a *boundary expansion*. First, we shall set some labels on the fundamental region R before introducing the map.

Assume that the sides $s_1, s_2, \ldots, s_{2n}$ of the fundamental domain R surround R in the counterclockwise order so that the region is on the left when we trace the boundary of R. The vertex v_i is the intersection of s_{i-1} and s_i (with $s_0 = s_{2n}$). Let P_i, Q_{i+1} be the end points on $\mathbb{S}^1$ of the complete geodesic $C(s_i)$ containing the side s_i, so that the order of points along $C(s_i)$ is $P_i, v_i, v_{i+1}, Q_{i+1}$ (see Figure 3.5).

For each i, let g_i be the label of s_i on the inside and $A(s_i)$ the arc $[P_iQ_i]$ on $\partial\mathcal{D}$. The given Fuchsian group Γ is of the first kind, and hence $\bigcup_{g\in\Gamma_R} A(g) = \partial\mathcal{D} = \mathbb{S}^1$.

Define a map $f_\Gamma : \partial\mathcal{D} \to \partial\mathcal{D}$ by $f|_{[P_iP_{i+1})}(\xi) = g_i(\xi)$. Here we assume that $P_{2n+1} = P_1$ so that f_Γ can be defined on $[P_{2n}P_1)$.

To be more specific, according to the same paper [5], this map $f = f_\Gamma$ satisfies the following properties:

Definition 3.2.1. (*Markov Property*)

(a) Except for a finite number of pairs of $x, y \in \mathbb{S}^1$: $x = gy$, $x, y \in \mathbb{S}^1$, $g \in \Gamma$ if and only if there exist nonnegative numbers $n, m \geq 0$ such tat $f^n(x) = f^m(y)$.

(b) f is *Markov* in the following sense: there is a finite or countable partition on $\mathbb{S}^1$ into intervals $\{I_i\}_{i=1}^{\infty}$ such that

- (Mi) f is strictly monotonic on each $\bar{I}_i$ and extends to a C^2-function $\bar{f}_i$ on $\bar{I}_i$,
- (Mii) $f(I_k) \cap I_j \neq \emptyset \Rightarrow f(I_k) \supseteq I_j$, for any j, k,
- (Miii) (*transitivity condition*) $\bigcup_{r=0}^{\infty} f^r(I_k) \supseteq I_j$ for any j, k,
- (Miv) (*finiteness condition*) If $\bar{I}_i = [a_i, b_i]$ then $\{\bar{f}_i(a_i), \bar{f}_i(b_i)\}_{r=0}^{\infty}$ is finite.

Moreover, the partition $\{I_i\}_{i=1}^{\infty}$ is finite if and only if if $\mathcal{D}/\Gamma$ is compact, or equivalently, if R has no cusps.

(c) (*expansion condition*)

(Ei) If there are no cusps, then there exists a positive number N such that $\inf\limits_{x\in\mathbb{S}^1} |(f^N)'(x)| > \lambda > 1$,

(Eii) A cusp of R is a periodic point for f with derivative one. There is a subset $K \subset \mathbb{S}^1$, consisting of a union of intervals I_i, so that if $f_K(x) = f^{n(x)}(x)$, $n(x) = \min\{n > 0 : f^n(x) \in K\}, x \in K$, is the first return map induced on K, then there is a number N such that $\inf\limits_{x\in\mathbb{S}^1} |(f_K^N)'(x)| > \lambda > 1$.

With this Markov map f, by f-*expansion* of $\xi \in \mathbb{S}^1$, we mean the one-sided infinite sequence $\xi_f = i_0i_1i_2\cdots$ obtained by iterating the map f where $f^n(x) \in \bar{I}_{i_n}, n = 0, 1, 2, \cdots$. Then, by the condition (Mii), the word i_ri_s appears if and only if $f(\bar{I}_r) \supseteq \bar{I}_s$ and hence the set Σ_f of all sequences $\xi_f = i_0i_1i_2\cdots$ for any $\xi \in \mathbb{S}^1$ is a subshift of finite type. More precisely, Σ_f is a 1-step subshift of finite type, or called a *Markov chain* since the forbidden word is of length at most 2, the partition $\{\bar{I}_i\}$ is called a *Markov partition.*

The following theorem in the paper [5] guarantees the existence of such a Markov map under certain conditions:

Theorem 3.2.2. ([5], Theorem 2.1) *Let Γ be a finitely generated Fuchsian group of the first kind, with a fundamental region R satisfying the property $(*)$. Then there is a Markov map $f_\Gamma : \mathbb{S}^1 \to \mathbb{S}^1$ which is orbit equivalent to Γ on $\mathbb{S}^1$. Moreover,*

(1) if R has no parabolic cusps, the Markov partition is finite and f_Γ satisfies properties (Miii), (Miv), (Ei), (Eii) of Definition 3.2.1.

(2) If R has parabolic cusps, the Markov partition is countable. There is a subset $K \subseteq \mathbb{S}^1$, consisting of a finite union of sets in the partitions, minus the countable set of points which eventually map onto one of the cusps, such that the first return map induced by f_Γ on K has properties (Miii), (Miv), (Ei), and (Eii).

Consider the set of all endpoints of intervals forming Markov partition. Then this Markov map leaves such partitioning points invariant:

Lemma 3.2.3. ([5], Lemma 2.3) *There is a finite or countable set $W \subset \mathbb{S}^1$ with $f(W) \subseteq W$ which partitions $\mathbb{S}^1$ into intervals; W is finite if and only if R has no parabolic vertices.*

In order to compare with the geometric coding, we shall replace the partition $\{\bar{I}_i\}$ by $\{[P_iP_{i+1}) = \langle g_i^{-1} \rangle\}$, and then now we may represent a sequence consisting of the generators Γ_R of Γ. For $\xi \in \mathbb{S}^1$, $\xi_{f_\Gamma} = g_{i1}g_{i2}\ldots$ if ${f_\Gamma}^n(\xi) \in \langle g_{in} \rangle$, $n \in \mathbb{N}$. This is called the *Artin method* or *Artin-type coding.*

In this situation, however, the rules deciding which sequences are"valid" are no longer of finite type.

Definition 3.2.4. A finite sequence $e_1e_2\cdots e_n (\in \Gamma_R^n)$ is said to be *admissible* if $\displaystyle\bigcap_{r=1}^{n} f^{-r}([e_r^{-1}]) \neq \emptyset$.

Let Σ^+ be the set of all admissible sequences consisting of elements of Γ_R, that is, $\Sigma^+ = \{e_1e_2\cdots \in \Gamma_R^{\mathbb{N}} : e_ke_{k+1}\cdots e_{k+l}$ is admissible for any $k, n \in \mathbb{N}\}$. Then this subgroup $\Sigma^+ \subset \displaystyle\prod_{i=0}^{\infty} \Gamma_R$ has the special property:

Theorem 3.2.5. ([21], Lemma 4.1) *The subshift Σ^+ is a sofic system. More precisely, there is an alphabet $\mathcal{B}$, and a finite-to-one map $\beta : B \to \Gamma_R$, and a subshift of finite type $\Sigma_B \subset \displaystyle\prod_{i=0}^{\infty} B$, so that the induced map $\bar{\beta} : \Sigma_B \to \Sigma^+$ is surjective and injective except at a countable set of points where it is two-to-one.*

There is a criterion whether a given sequence is a member of the set Σ^+:

Theorem 3.2.6. ([21], Theorem 4.2) *A word (finite sequence) occurs in Σ^+ if and only if it is shortest and contains no counterclockwise half-cycles.*

Corollary 3.2.7. *An infinite sequence occurs in Σ^+ if and only if it is shortest not containing counterclockwise half-cycles and not ending in an infinite chain of counterclockwise cycles.*

To be specific, we are going to go after the paper [20].

With proceeding along one side to the vertex, we obtain a cycle of congruent vertices and corresponding generators in two directions. First, if you started at a vertex v_i with the side s_i, you would obtain a cycle of congruent vertices $v_i = w_1, w_2, \cdots, w_p$ and the corresponding generators $g_i = h_1, h_2, \cdots, h_p$. More specific process is the following: let v_i be a vertex of R and s_i an adjacent side. Initiate $w_1 = v_i$, $t_1 = s_i$, and $h_1 = g_i$ where g_i is the label of s_i on inside of R. Then $w_2 = h_1(w_1)$ is another vertex and $t_2 = h_1(t_1)$ is an adjacent side. Let t'_2 be the other side of R adjacent to t_2. Let $w_3 = h_2(w_2), t_3 = h_2(t'_2)$ where h_2 is an element of Γ such that t'_2 is identified with another side of R by h_2. We will return to v_i in a finite number of steps, after continuing this process, $(w_p, t'_p) = (w_1, t_1)$, $h_p \cdots h_1$ fixes $w_1 = v_i$ at last. Schematically, we have a part of tessellation around the vertex v_i shown as Figure 3.6.

There are no vertices of R congruent to v_i other than those just found, whence $w_1, w_2, \cdots, w_p$ constitute a cycle.

Definition 3.2.8. ([9], Sec.26) A complete set of congruent vertices of a fundamental region is called an *ordinary cycle*.

We say the anticlockwise sequence $h_1^{-1} h_2^{-1} \cdots h_p^{-1}$ is *in left-hand (L) cyclic order*. Similarly, if you started at a vertex v_{i+1} with the side s_i, you would obtain a cycle $v_{i+1} = z_1, z_2, \cdots, z_q$ and generators $g_i = j_1, j_2, \cdots, j_q$. The clockwise sequence $j_1^{-1} j_2^{-1} \cdots j_q^{-1}$ is said to be *in right-hand (R) cyclic order*.

To connect a vertex cycle with a relation, we are going to introduce two theorems without showing proofs:

Theorem 3.2.9. ([9], Theorem 14) *The sum of the angles at the vertices of an ordinary cycle is $2\pi/k$, where k is an integer. If $k > 1$, each vertex of the cycle is a fixed point of an elliptic transformation of period k.*

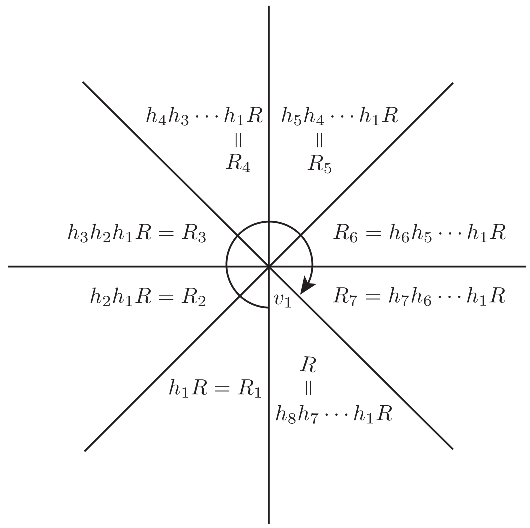

Figure 3.6: An example of a vertex cycle in left-hand

Theorem 3.2.10. ([9], Theorem 15) *Each ordinary cycle determines a relation of the form* $(h_p h_{p-1} \cdots h_2 h_1)^k = 1$ *satisfied by the transformations connecting congruent sides of R.*

By the above theorems, there are integers μ, ν such that $(h_1^{-1} h_2^{-1} \cdots h_p^{-1})^\mu = (j_1^{-1} j_2^{-1} \cdots j_q^{-1})^\nu = 1$. If $v_i \in \mathcal{D}$, $h_p \cdots h_1$ is elliptic and has order $\mu \in \mathbb{N}$, otherwise (i.e., $v_i \in \mathbb{S}^1$) it is parabolic. For all elliptic vertices v_i, the relations $(h_p h_{p-1} \cdots h_1)^\mu = (h_1^{-1} h_2^{-1} \cdots h_p^{-1})^\mu$ form a complete set of relations for Γ.

Under our circumstance, Γ satisfies property $(*)$ (Definition 3.0.8) and hence the numbers $p\mu, q\nu$ of sides of the net N meeting at v_i, v_{i+1} are even numbers; $p\mu = 2l, q\nu = 2k$.

Definition 3.2.11. We call L cycles of lengths $l-1, l, l+1$, *D-, H-, S-* L cycles respectively and similarly, for R cycles of lengths $k-1, k, k+1$. Here, D means deficient, half for H and superfluous for S. By a *full cycle*, we mean a cycle of length $2l$ or $2k$.

Clearly, a full cycle is equal to the identity in Γ.

Definition 3.2.12. ([20]) If $h = g_i$, we write $h^+ = g_{i+1}$ and $h^- = g_{i-1}$. If $B = b_1 \cdots b_r$, $B^1 = b_1 \cdots b_{r+1}$, $C = c_1 \cdots c_s$ are L [respectively, R] cycles with $c_1^{-1} = (b_{r+1}^{-1})^+$ [respectively, $c_1^{-1} = (b_{r+1}^{-1})^-$], we say B and C are *adjacent* or *consecutive* L [respectively, R] cycles. A sequence $B_1, B_2, \ldots, B_r$ of consecutive L cycles, where B_1, B_r are H-cycles and $B_2, \ldots, B_{r-1}$ are D-cycles, will be called an L *H-chain*; such a sequence with B_1 an L D-cycle is an L *D-chain*, often denoted by $DD \cdots DH$.

Proposition 3.2.13. ([20], Proposition 1.1) *A sequence* $e_1 \cdots e_p \in (\Gamma_R)^p$ *is admissible if and only if*

(1) $gg^{-1}, g \in \Gamma_R$*, dose not occur,*

(2) No R H-cycles occur,

(3) No L S-cycles occur,

(4) No L H-chains occur.

Consider a map $\pi : \Sigma^+ \to \mathbb{S}^1$ defined by $\pi(e_1 e_2 \cdots) = \bigcap_{r=1}^{\infty} f^{-r}([e_r^{-1}])$. Then the image $\pi(e_1 e_2 \cdots)$ has exactly one point due to the finite intersection property and the expansion condition of f. However, the map π is not one-to-one: for $x \in \mathbb{S}^1$ with $f^n(s) \in \{P_1, P_2, \ldots, P_{2n}\}$ for some $n \geq 0$, x has two representation because P_i can be written either as infinite sequence of consecutive R D-cycles ($DD \cdots$), or as an infinite sequence of consecutive L cycles ($HDD \cdots$).

Also, for $x \in \Sigma^+$, if x does not end in an infinite string of R D-cycles, $(\pi \circ \sigma)(x) = (f \circ \pi)(x)$ where σ denotes the shift map. Therefore, it is convenient to take the representation of x as a sequence terminating with L cycles whenever $x \in \mathbb{S}^1$ has two symbolic expressions in Σ^+. Actually, this is why we adopt the definition of f with the arc intervals $\{[P_i P_{i+1})\}$ rather than $\{(P_i P_{i+1}]\}$.

To associate a two-sided infinite sequence with a geodesic, we juxtapose two f-expansions (one-sided infinite sequences) of the end points of the geodesic. However, we cannot be sure that the new sequence is still admissible. Define a new map $\bar{f} : \partial\mathcal{D} \to \partial\mathcal{D}$ by $\bar{f}|_{(Q_{i-1}Q_i]} = g_i$ using the another partition $\{(Q_{i-1}Q_i]\}_{i=1}^{2n}$ instead of $\{[P_{i-1}P_i)\}_{i=1}^{2n}$. Clearly, this map $\bar{f}$ has the same property (Definition 3.2.1) as the map f and the admissibility rules are still the same in Proposition 3.2.13 by interchanging 'R' and 'L'. Then we now check the admissibility of the sequence using the map $\bar{f}$:

Lemma 3.2.14. ([20], Lemma 2.1) *The sequence $e_1e_2\cdots$ is admissible for f if and only if the inverse sequence $\cdots e_2^{-1}e_1^{-1}$ is admissible for $\bar{f}$.*

Imagine that a directed geodesic γ has two end points ξ, η on the boundary $\partial\mathcal{D}$ and γ goes from η to ξ. In this case we call ξ the positive endpoint, η the negative endpoint of γ. If $e_1e_2\cdots$ is the f-expansion of ξ and $f_1f_2\cdots$ is the $\bar{f}$-expansion of η, the arithmetic coding sequence $\gamma(\xi,\eta) = \eta^{-1} * \xi$ will be defined as $\cdots f_2^{-1}f_1^{-1}e_1e_2\cdots$ whenever $\cdots f_2^{-1}f_1^{-1}e_1e_2\cdots$ is admissible.

Let Σ be the set of bi-infinite admissible sequences with left shift map σ. There are some propositions which explain the action of Γ on the space Σ as a symbolic dynamical system. First, two are something about the Γ-action on Σ^+:

Proposition 3.2.15. ([20], Proposition 2.2) *Let $x = e_1e_2\cdots \in \Sigma^+$, $g \in \Gamma_R$. Then*

(1) either $g(x) = ge_1e_2\cdots$ whenever $ge_1e_2\cdots \in \Sigma^+$,

(2) or $g(x) = e_2e_3\cdots$ if $g = e_1^{-1}$.

Proposition 3.2.16. ([20], Proposition 2.3) *Suppose $x \in \mathbb{S}^1$, and $g \in \Gamma$. Let $x = e_1e_2\cdots$, $g(x) = f_1f_2\cdots$ be the f-expansions of x, $g(x)$. Then there are $s,t > 0$ such that $ge_1e_2\cdots e_s = f_1f_2\cdots f_t$ in Γ and $e_{s+i} = f_{t+1}, i > 0$*

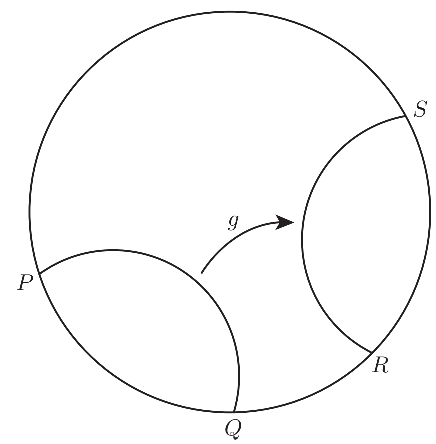

Figure 3.7: Conjugate geodesics in Proposition 3.2.17

The following proposition says that the map f is orbit equivalent to Γ on $\mathbb{S}^1$. To put it concretely, except for a finite number of pairs of points, if $(x, y) \in \mathbb{S}^1 \times \mathbb{S}^1$ with $x = gy$ for some $g \in \Gamma$ then there are nonnegative integers n, m such that $f^n(x) = f^m(y)$, and vice versa. Such a pair (x, y) is a set of endpoints of an admissible geodesic. Therefore the admissible geodesics are conjugate under the action of Γ if and only if the corresponding sequences are shift equivalent:

Proposition 3.2.17. ([20], Proposition 2.4) *Let $(P, Q), (R, S) \in \mathbb{S}^1 \times \mathbb{S}^1$ be such that $Q^{-1} * P, R^{-1} * S \in \Sigma$. Then there exists an element $g \in \Gamma$ with $gP = R, gQ = S$ if and only if there is a number n so that $\sigma^n(Q^{-1} * P) = R^{-1} * S$.*

Recall $\mathcal{R} = \{\gamma \in \mathcal{G} : \gamma \cap R \neq \emptyset\}$, and Σ denotes the set of all bi-infinite reduced sequences in Γ_R. Define $\mathcal{A} = \{\gamma = \gamma(\xi, \eta) \in \Sigma : \eta^{-1} * \xi \in \Sigma$, where ξ, η are the positive and negative endpoints of $\gamma\}$.

Lemma 3.2.18. ([21], Lemma 5.1) *If $\gamma \in \mathcal{R}$ and $\eta^{-1} * \xi$ is not shortest, then γ is a side of the net N.*

Lemma 3.2.19. ([21], Lemma 5.2) *Suppose that* $\gamma \in \mathcal{R} \triangle \mathcal{A}$ *and that* $\eta^{-1} * \xi$ *is shortest. Then* $\eta_0^{-1} * \xi_0$ *lies in a cycle or a chain and* γ *passes near* $v = v(\xi, \eta)$.

We say that a sequence contains a *pseudo half cycle* when the sequence begins or ends in an infinite chain of cycles of lengths $n(v_1), n(v_2), \cdots$ at vertices $v_1, v_2, \cdots$. Edge paths of geodesics passing a vertex too nearby always contain half cycles. More specifically, the following is known:

Lemma 3.2.20. ([21], Lemma 5.3) *Suppose* γ *passes near* v. *Then:*

(i) if $\gamma \in \mathcal{R}$ *and cuts off* v *on* R, *then* $E(\gamma)$ *contains a chain beginning or ending in a half cycle or pseudo half cycle and including the cycle at* v.

(ii) if $\eta^{-1} * \xi$ *is shortest and* $v = v(\xi, \eta)$, *then* $E(\eta^{-1} * \xi)$ *has the same property as in (i).*

We write down the proposition in order to relate the two sets $\mathcal{A}$ with $\mathcal{R}$:

Proposition 3.2.21. ([21], Proposition 5.6) *Suppose* $\eta_0^{-1} * \xi_0$ *lies in a cycle or chain and that* γ *passes near* $v(\xi, \eta)$. *Then*

(i) $\eta^{-1} * \xi \in \mathcal{A} \Rightarrow$ *(*γ *goes clockwise around* $v \Leftrightarrow \gamma \in \mathcal{R}$*),*

(ii) $\eta^{-1} * \xi \notin \mathcal{A} \Rightarrow$ *(*γ *goes counterclockwise around* $v \Leftrightarrow \gamma \in \mathcal{R}$*)*

Chapter 4

Artin Method for Some Examples of Surfaces

Recall $\mathcal{R} = \{\gamma \in \mathcal{G} : \gamma \cap R \neq \emptyset\}$, and Σ denotes the set of all bi-infinite reduced sequences in Γ_R. Define $\mathcal{A} = \{\gamma = \gamma(\xi, \eta) \in \Sigma : \eta^{-1} * \xi \in \Sigma$, where ξ, η are the positive and negative endpoints of $\gamma\}$.

Theorem 4.0.22. ([21], Theorems I and II) *There is a bijection $T : \mathcal{A} \to \mathcal{R}$ such that $T\sigma = \tau T$. $\mathcal{A}$ is partioned into a finite number of pieces with geodesic boundaries such that on each piece, T is some fixed element of Γ, and $T_{\mathcal{A}\cap\mathcal{R}}$ = id.*

4.1 Torus with punctures

These are the simplest examples demonstrating the Morse coding because there is no interior vertex of a fundamental domain R and hence there is no ambiguity when we choose generators of sides cut by geodesic. Moreover, the set Σ of all reduced sequences are realized as Morse coding sequences of elements of Γ_R of geodesics of $\mathcal{S}$ for a properly chosen fundamental domain.

The fundamental group of a torus M_n with n punctures is a free group with $n + 1$ generators. If we consider each puncture as an ideal vertex, we may choose its fundamental domain R as an ideal polygon on a Poincaré disc

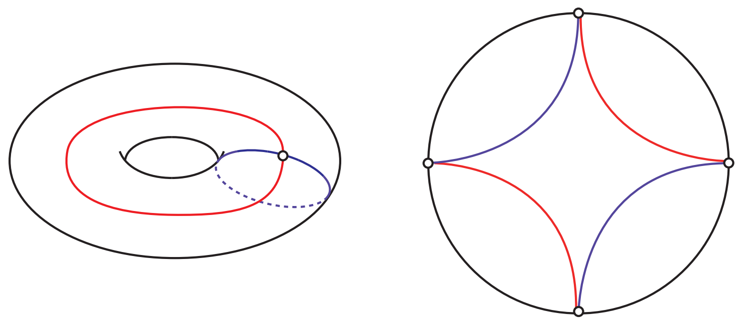

Figure 4.1: Torus with one puncture

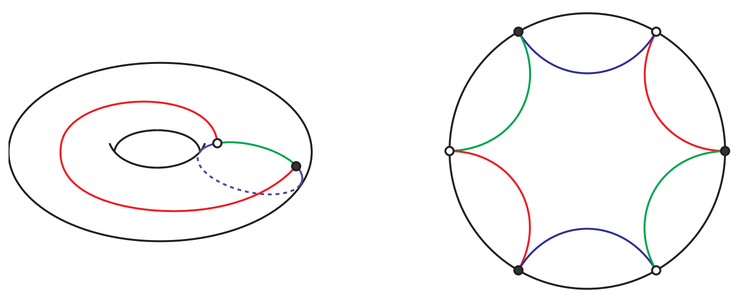

Figure 4.2: Torus with two punctures

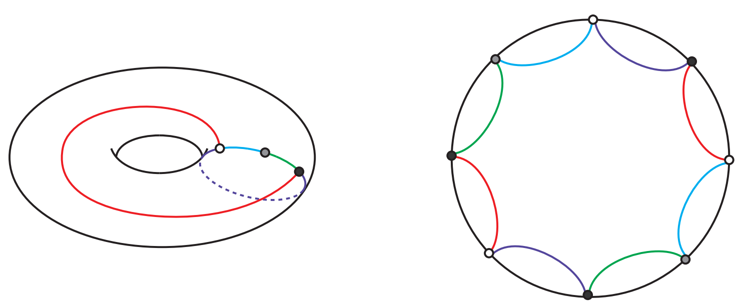

Figure 4.3: Torus with three punctures

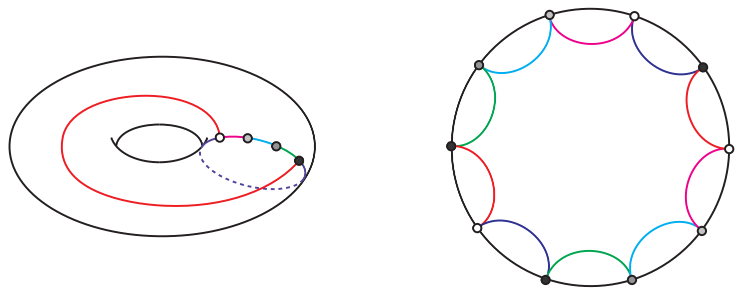

Figure 4.4: Torus with four punctures

(this means R has all its vertices on $\partial\mathcal{D} = \mathbb{S}^1$), which projects to M_n.

In this case the relevant shift space is the space of all bi-infinite reduced sequences whose letters are its generators. We shall mention that the geometric coding sequence of a geodesic except those having an endpoint as a cusp is the same as the sequence obtained by the boundary expansion coding.

Theorem 4.1.1. *If we can choose an ideal polygon as a fundamental domain,* $\mathcal{R} = \mathcal{A}$ *and* $\sigma = \tau$.

Moreover, the set of all admissible sequences in this case is a subshift of finite type. In particular, it is a topological Markov chain since the reduced word is the only admissibility rule and hence the forbidden words are of length 2.

The figures explain how to construct a fundamental domain in each case (see Figure 4.1-4.4).

4.2 A closed surface with genus two

We look into the arithmetic coding of a closed surface $\mathcal{S}$ with genus two.

Let Γ be the fundamental group $\pi_1(\mathcal{S})$ of the surface $\mathcal{S}$. Then Γ is a finitely generated Fuchsian group of the first kind acting on the unit disc

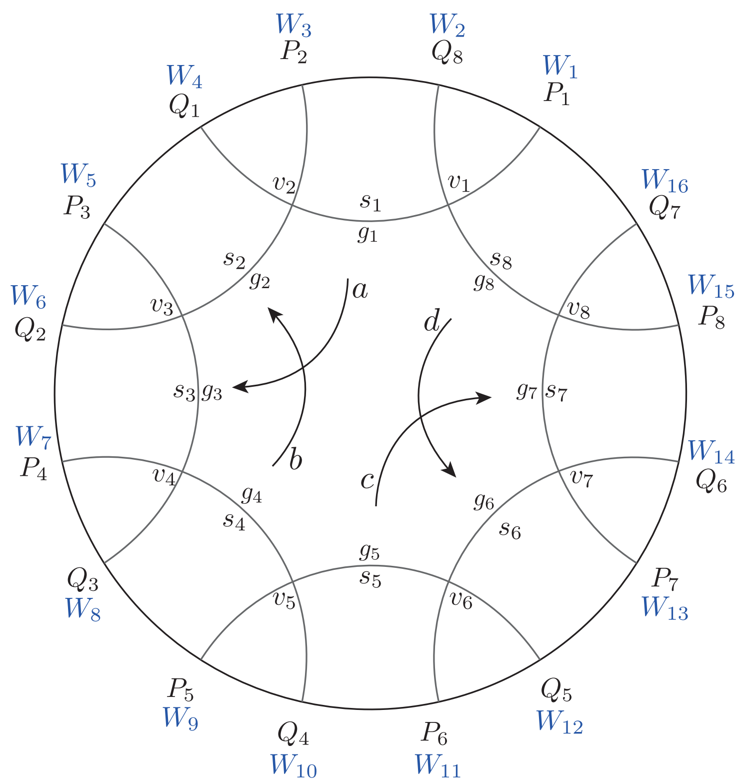

Figure 4.5: A fundamental region of a surface with genus 2

$\mathcal{D}$. Choose an octagon R in $\mathcal{D}$ as its Dirichlet region as the previous setting as Section 3.1 and 3.2. Then Γ is generated by side pairing maps by Theorem 2.1.11, say $\{g_1, g_2, \ldots, g_8\}$. Since the sides s_1 and s_3 are identified by g_1 and g_3, one element is the inverses of the other $g_1 = (g_3)^{-1}$. we will write $a = g_1 = (g_3)^{-1}$. Similarly, $b = g_4 = (g_3)^{-1}$.

The boundary map $f : \partial\mathcal{D} \to \partial\mathcal{D}$ is defined by $f|_{[P_iP_{i+1})}(\xi) = g_i(\xi)$, and let Σ^+ be the set of all admissible sequences consisting of elements of Γ_R. Then $\Sigma^+ = \{e_1e_2\cdots \in \Gamma_R^{\mathbb{N}} : e_ke_{k+1}\cdots e_{k+l}$ is admissible for any $k, n \in \mathbb{N}\}$ $= \{\xi_{f_\Gamma} = g_{i1}g_{i2}\ldots : \xi \in \mathbb{S}^1, {f_\Gamma}^n(\xi) \in \langle g_{in}\rangle$ for each $n \in \mathbb{N}\}$ according to Section 3.2.

Let $\mathcal{A}$ be the alphabet (the set of all symbols) of Σ^+, then $\mathcal{A}$ consists of the elements of Γ_R.

Let us state Theorem 3.2.5 again and prove it.

Theorem. ([21], Theorem 4.1) *The subshift Σ^+ is a sofic system. More precisely, there is an alphabet $\mathcal{B}$, and a finite-to-one map $\beta : \mathcal{B} \to \mathcal{A}$, and a subshift of finite type $\Sigma_{\mathcal{B}} \subset \prod_{i=0}^{\infty} \mathcal{B}$, so that the induced map $\bar{\beta} : \Sigma_{\mathcal{B}} \to \Sigma^+$ is surjective and injective except at a countable set of points where it is two-to-one.*

Series proved the above theorem using a Markov map and a Markov partition ([21], Lemma 4.1). We are going to follow two of her papers first and then give another proof of Theorem 3.2.5.

Proof by Series. Theorem 3.2.2 assures us that there exists a Markov map in our situation because all conditions are satisfied.

We will choose another alphabet $\mathcal{B}$ and *another admissibility rule* so that the new system $\Sigma_{\mathcal{B}}$ is a subshift of finite type. As mentioned in Section 3.2, it happens when the admissible sequences are defined with Markov partition.

Put another label W_i on $\{P_i, Q_i\}_{i=1}^{8}$ in the counterclockwise order: $W_1 = P_1$, $W_2 = Q_8$, $W_3 = P_2$, and so on (see Figure 4.5). Define the arc intervals $I = \{I_i = [W_i W_{i+1})\}_{i=1}^{16}$ with $W_{17} = W_1$. Let the f-expansion of a point ξ on the boundary of the disc $\mathbb{S}^1$ be the one-sided infinite sequence $\xi_f = i_0 i_1 i_2 \cdots$ obtained by iterating f where $f^n(x) \in \bar{I}_{i_n}, n = 0, 1, 2, \cdots$. Here, the map f is defined by $f(x) = g_k(x)$ where $x \in I_i \subset [P_k P_{k+1})$ for some k.

Recall that the word $i_r i_s$ appears if and only if $f(\bar{I}_r) \supseteq \bar{I}_s$ and hence the set $\Sigma_{\mathcal{B}}$ of all sequences $\xi_f = i_0 i_1 i_2 \cdots$ for any $\xi \in \mathbb{S}^1$ is a subshift of finite type. The alphabet $\mathcal{B}$ of $\Sigma_{\mathcal{B}}$ is the set $\{1, 2, \cdots, 16\}$.

Let us check this map f defined on the partition $\{I_i = [W_i W_{i+1})\}_{i=1}^{16}$ has the Markov property (Definition 3.2.1).

(a) By Proposition 3.2.16, the map f is orbit-equivalent to Γ on the unit circle $\mathbb{S}^1$.

(b) (i) f is equal to a fixed element of Γ_R on each I_i.

(ii) Since each element of Γ is an isometry (and hence a continuous map), f sends an arc interval to an arc interval. By Lemma 3.2.3, $f(I_k)$ is a union of adjacent arc intervals in I. Therefore, if $f(I_k) \cap I_j \neq \emptyset$, then $f(I_k) \supseteq I_j$ for any j, k.

Therefore, the set $\Sigma_{\mathcal{B}}$ is a topological Markov chain, namely, a one-step subshift of finite type.

If we set a map $\beta : \mathcal{B} \to \mathcal{A}$ by $2i-1, 2i \mapsto g_i^{-1}$ for each $i = 1, 2, \ldots, 8$, then β is a two-to-one continuous map because both intervals $[I_{2i-1} I_{2i})$ and $[I_{2i} I_{2i+1})$ are properly contained in $\langle g_i^{-1} \rangle = [P_i P_{i+1})$. Hence the set Σ^+ is a factor of the set $\Sigma_{\mathcal{B}}$, and show that Σ^+ is a sofic system. □

We are now going to construct a finite directed labelled graph to suggest another proof of the above theorem using the Automata theory.

$$\begin{tikzcd}
& (v_4,s_3) \arrow[d, dotted] & (v_3,s_2) \arrow[d, dotted] & (v_2,s_1) \arrow[d, dotted] & (v_5,v_4) \arrow[d, dotted] & (v_8,s_7) \arrow[d, dotted] & (v_7,s_6) \arrow[d, dotted] & (v_6,s_5) \arrow[d, dotted] & (v_1,v_8) \arrow[d, dotted] \\
(v_1,s_1) \arrow[ur, "a"] & (v_4,s_4) \arrow[ur, "b"] & (v_3,s_3) \arrow[ur, "a^{-1}"] & (v_2,s_2) \arrow[ur, "b^{-1}"] & (v_5,s_5) \arrow[ur, "c"] & (v_8,s_8) \arrow[ur, "d"] & (v_7,s_7) \arrow[ur, "c^{-1}"] & (v_6,s_6) \arrow[ur, "d^{-1}"] & (v_1,s_1)
\end{tikzcd}$$

Diagram 4.1: Vertex cycle in the left-hand cyclic order

Another proof by constructing a finite directed graph. First, we want to know relation sequences in the fundamental domain of the surface with genus 2 using vertex cycles. This is the process to find congruent vertices of R as explained in the Section 3.2.

Let v_1 be a vertex of R and s_1 an adjacent side; then the vertex and the side are identified with $v_4 = g_1(v_1)$, $s_3 = g_1(s_1)$ by $g_1 = a$. Take the other side of R adjacent to v_4, s_4. The generator of the side s_4 on inside is $g_4 = b$. After we identify the pair (v_4, s_4) by g_4, we arrive at (v_3, s_2). Then, by taking the other side, we pass to the pair (v_3, s_3) (see Figure 4.5). Repeat this process. One can observe the whole process in Diagram 4.1.

The generators in left-hand(L) cyclic order $a^{-1}, b^{-1}, a, b, c^{-1}, d^{-1}, c, d$ form a relation: the counterclockwise sequence satisfies $a^{-1}b^{-1}abc^{-1}d^{-1}cd = 1$. Similarly, we obtain the generators in right-hand(R) cyclic order a^{-1}, b, a, d^{-1}, c^{-1}, d, c, b^{-1} by starting with the pair (v_2, s_1): the clockwise sequence satisfies $a^{-1}bad^{-1}c^{-1}dcb^{-1} = 1$ ($\circledast$).

Under this circumstance, we investigate forbidden words using Proposition 3.2.13:

Proposition. ([20], Proposition 1.1) *A sequence $e_1 \cdots e_p \in (\Gamma_R)^p$ is admissible if and only if*

(1) $gg^{-1}, g \in \Gamma_R$, dose not occur,

(2) No R H-cycles occur,

(3) No L S-cycles occur,

	gg^{-1}	R H-cycle	L S-cycle	L H-cycle	L D-cycle
a	aa^{-1}	$ad^{-1}c^{-1}d$	$abc^{-1}d^{-1}c$	$abc^{-1}d^{-1}$	abc^{-1}
b	bb^{-1}	$bcd^{-1}c^{-1}$	$bc^{-1}d^{-1}cd$	$bc^{-1}d^{-1}c$	$bc^{-1}d^{-1}$
a^{-1}	$a^{-1}a$	$a^{-1}bad^{-1}$	$a^{-1}b^{-1}abc^{-1}$	$a^{-1}b^{-1}ab$	$a^{-1}b^{-1}a$
b^{-1}	$b^{-1}b$	$b^{-1}a^{-1}ba$	$b^{-1}abc^{-1}d^{-1}$	$b^{-1}abc^{-1}$	$b^{-1}ab$
c	cc^{-1}	$cb^{-1}a^{-1}b$	$cda^{-1}b^{-1}a$	$cda^{-1}b^{-1}$	cda^{-1}
d	dd^{-1}	$dcb^{-1}a^{-1}$	$da^{-1}b^{-1}ab$	$da^{-1}b^{-1}a$	$da^{-1}b^{-1}$
c^{-1}	$c^{-1}c$	$c^{-1}dcb^{-1}$	$c^{-1}d^{-1}cda^{-1}$	$c^{-1}d^{-1}cd$	$c^{-1}d^{-1}c$
d^{-1}	$d^{-1}d$	$d^{-1}c^{-1}dc$	$d^{-1}cda^{-1}b^{-1}$	$d^{-1}cda^{-1}$	$d^{-1}cd$

Table 4.1: forbidden words

(4) No L H-chains (of the form $HDD\cdots DH$) occur.

Then the set Σ^+ consists of all reduced one-sided sequences not containing R H-cycles, L S-cycles, and L H-chains. Now that it is possible for L H-chains of length $(3n+8)$ to exist for any positive integer n, the system Σ^+ can not have a finite list of forbidden words. Using the relation sequences ($\circledast$), we obtain the table of forbidden words (the left three columns in Table 3.1).

The set Σ^+ has forbidden words of length at most 5 but it may have an infinite D-cycle chain not bounded by some L Half cycles. Therefore, we should be careful when we deal with L D-cycles. Hence, we first construct a directed graph which gives sequences in Σ^+ containing no L D-cycle. After then, we add L H-cycles and L D-cycles without making any L H-chain.

We set the temporary admissibility rule in this case as follows:

(R1) $gg^{-1}, g \in \Gamma_R$, dose not occur,

(R2) No R H-cycles occur,

(R3) No L S-cycles occur,

(H4) No L H-cycles occur,

(D5) No L D-cycles occur.

Note that the condition (D5) implies the two conditions (R3), (H4). Thus, if we want to consider the whole rules (R1), (R2), (R3), (H4), and (D5), we need to check the only three rules (R1), (R2), (D5).

Let $W_4(\Sigma^+)$ be the set of all words (called *factors*) of length 4 appearing in some sequence of Σ^+ (that is, it is not containing any forbidden words: words of the form gg^{-1} and R H-cycles). We divide this set $W_4(\Sigma^+)$ into three subsets $W_F(\Sigma^+)$, $W_H(\Sigma^+)$, $W_D(\Sigma^+)$ which are pairwise disjoint.

1. Fisrt, we obtain $W_F(\Sigma^+)$ from $W_4(\Sigma^+)$ excluding all words containing any L D-cycle. Note that $W_F(\Sigma^+)$ has no words containing any L D-cycle or L H-cycle, in other words, $W_F(\Sigma^+)$ is the set containing all words of length 4 satisfying the conditions (R1), (R2), (D5).

2. The second set $W_H(\Sigma^+)$ is obtained from $W_4(\Sigma^+)$ - $W_F(\Sigma^+)$ choosing all words containing any L H-cycle. $W_H(\Sigma^+)$ consists of exactly 8 L H-cycles.

3. Then the set $W_D(\Sigma^+)$ consists of all remaining elements in $W_4(\Sigma^+)$, namely, $W_D(\Sigma^+) = W_4(\Sigma^+) - W_F(\Sigma^+) - W_H(\Sigma^+)$. This set $W_D(\Sigma^+)$ consists of words of $W_4(\Sigma^+)$ containing an L D-cycle *not an L H-cycle* and hence there are 96 elements in $W_D(\Sigma^+)$.

Clearly, $W_4(\Sigma^+) = W_F(\Sigma^+) \sqcup W_H(\Sigma^+) \sqcup W_D(\Sigma^+)$.

Then we now construct a directed graph Λ_F as follows: each element in $W_F(\Sigma^+)$ is a vertex of Λ_F, and there is a directed edge from $x_1x_2x_3x_4$ to $x_1'x_2'x_3'x_4'$ provided that the concatenated word $x_1x_2x_3x_4x_4' = x_1x_1'x_2'x_3'x_4'$ is an allowed word in Σ^+, in other words, it is not an L S-cycle. We will call this rule the *edge rule* from now on. In this graph Λ_F, we cannot obtain a sequence containing any of L D-cycles and hence L H-cycles.

Now, we construct two directed graphs Λ_D and Λ_H upon the above admissibility rules.

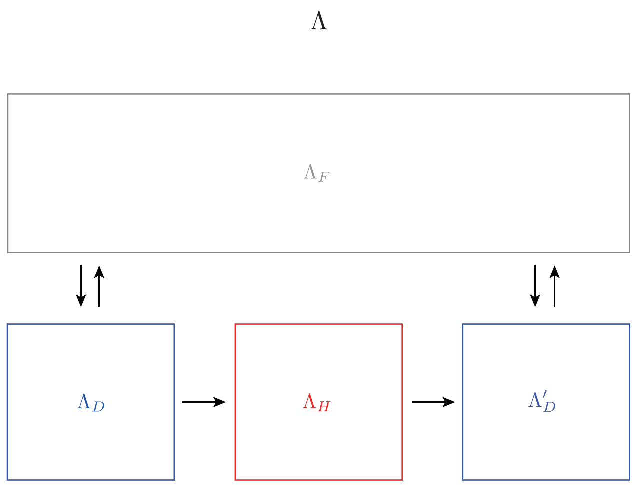

Figure 4.6: The schema of the finite directed graph Λ

Λ_D: The vertex set $V(\Lambda_D)$ of the graph Λ_D is equal to the set $W_D(\Sigma^+)$, and the edge set $E(\Lambda_D)$ is obtained by the edge rule.

Λ_H: The vertex set $V(\Lambda_H)$ of the graph Λ_H is equal to the set $W_H(\Sigma^+)$, and the edge set $E(\Lambda_H)$ is obtained by the edge rule. No two L H-cycles satisfy the above rule, the edge set $E(\Lambda_H)$ is an empty set.

Finally, we are going to connect the finite directed labelled graphs Λ_H and Λ_D to Λ_F and denote the whole graph by Λ. Then a sequence containing any of L D-cycles, L H-cycles can be expressed by a directed path in the graph Λ. In order not to produce any L H-chain through this graph Λ, we should regulate the "bridging" between three graphs Λ, Λ_H and Λ_D.

A sequence in Σ^+ may be the form $\cdots DH \cdots HD \cdots$ even though a L H-chain is not allowed. We use a copy Λ'_D of the graph Λ_D to make one-sided paths between Λ_H and Λ_D: the allowed path goes from Λ_D to Λ_H and Λ_H to Λ'_D. Put a directed edge from each vertex of Λ_D to a vertex of Λ_H and a directed edge from each vertex of Λ_H to a vertex of Λ'_D complying with the edge rule.

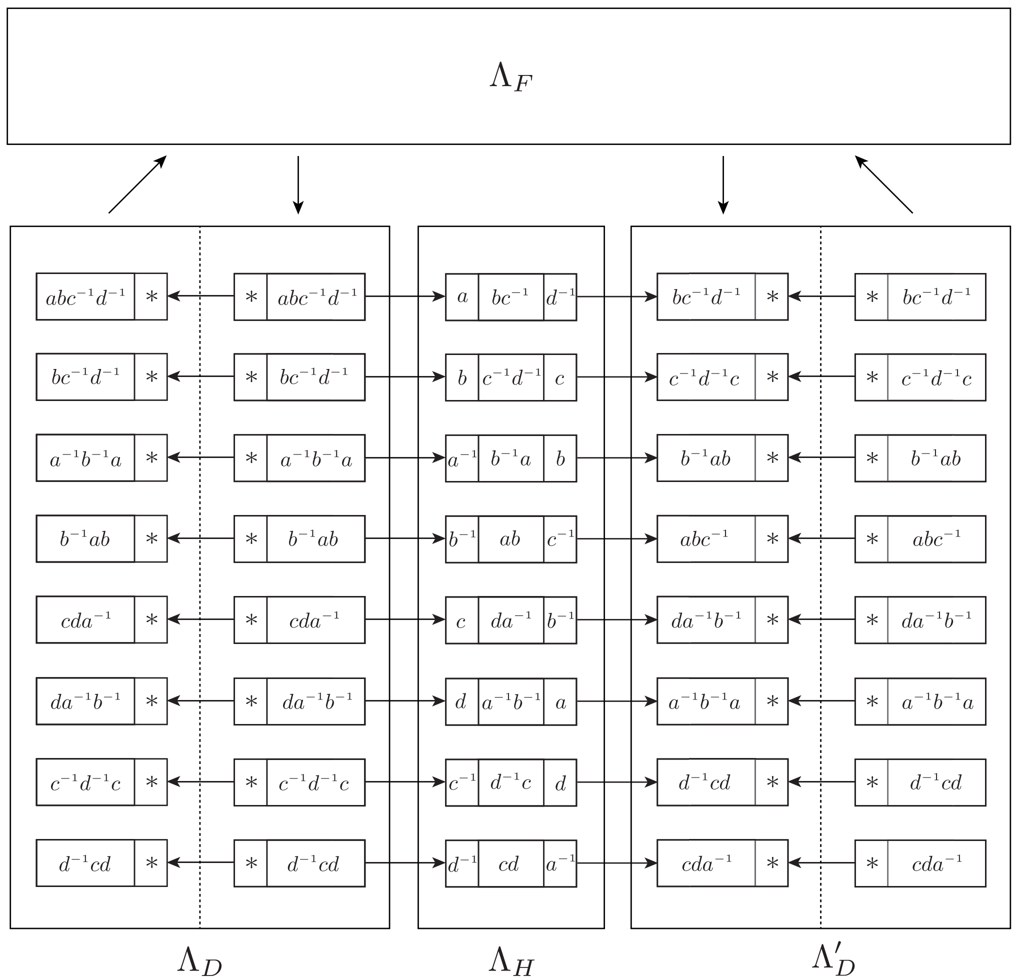

Figure 4.7: The directed edges between Λ_F, Λ_H and Λ_D

Then each vertex of Λ_F connects with all vertices of three graphs Λ_D, Λ_H and Λ'_D, respectively, of course satisfying the edge rule. Figure 4.6 shows the "bridging" between three graphs Λ, Λ_H and Λ_D (Λ'_D) schematically. Note that since there are no words containing an L D-cycle in the graph Λ_F, there is no edge between Λ_F and Λ_H by the edge rule. To be specific, we give a simple view of directed edges between Λ_D and Λ_H (see Figure 4.7: here, an asterisk means a letter which does not make the word an L H-cycle in each case. For example, in the case $*abc^{-1}$, $*$ can be a letter among six symbols $b, a^{-1}, b^{-1}, c, c^{-1}, d^{-1}$. Note that all edges are multiple).

This merged graph Λ does not produce any sequence containing L H-chain ($HDD \cdots DH$) because we obtain an L D-cycle or an L H-cycle only when we go through Λ_D (and Λ_H). All elements of the whole factor set $W_4(\Sigma^+)$ are included in Λ and we can travel any path sequence in Σ^+ through the edges in Λ defined by the rule (R1), (R2), (R3).

Each of the above graphs Λ_F, Λ_D, Λ_H, Λ'_D is defined by a finite set of forbidden words, and hence all of them give a vertex shift. Clearly, all vertex shifts defined by the graphs are shift invariant, and hence the vertex shift of the merged graph Λ is also vertex shift; we can walk along directed edges of the graph Λ visiting a vertex infinitely many times. That means each vertex in Λ has at least one incoming edge and at least one outgoing edge.

Now if, after we put the last letter of each vertex in the graph Λ on the outgoing edges from the vertex, we delete the labels of all vertices in Λ, then we will obtain a directed graph with labelled edges. In other words, we establish the edge shift (a finite directed graph with labelled edges) equal to Σ^+ from this finite directed graph Λ with multiple edges by interchanging the vertex set $V(\Lambda)$ and the edge set $E(\Lambda)$. Then Λ is a presentation of Σ^+ and hence Σ^+ is sofic by Proposition 2.2.7. $\square$

Remark. The directed graph Λ can give an infinite string of consecutive L cycles of the form $HDD\cdots$. This sequence corresponds with a point among $P_1, P_2, \cdots, P_8$. The point $x \in \mathbb{S}^1$ has two representations in Σ^+ whenever

$f^k(x) \in \{P_1, P_2, \cdots, P_8\}$ for some $k \geq 0$ since P_i can be written either as an infinite sequence of consecutive R D-cycles, or an infinite string of consecutive L cycles of the form $HDD\cdots$.

Bibliography

[1] E. Artin, *Ein Mechanisches System mit quasi ergodischen Bahnen*, Collected Papers. Addison Wesley(1965), 499-501.

[2] A. Beardon, *The Geometry of Discrete Groups*, Springer-Verlag, 1983.

[3] T. Bedford, M. Keane, and C. Series, *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, Oxford University Press, 1991.

[4] J. Birman and C. Series, *Dehn's algorithm revisited, with applications to simple curves on surfaces*, Proc. Conf. on combinatorial Group Theory, Utah, 1984, (Ed. Gersten and Stallings).

[5] R. Bowen and C. Series, *Markov maps for Fuchsian groups*, Inst. Hautes Études Sci. Publ. Math., 50(1979), 153-170.

[6] B. Marcus, R. Roth, and P. Siegel, *Modulation codes for digital data storage*, In Different aspects of coding theory (San Francisco, CA, 1995), pages 41-94, American Mathematical Society, Providence, RI, 1995.

[7] M. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Springer, 2009.

[8] M. Brin and G. Stuck, *Introduction to Dynamical Systems*, Cambridge University Press, 2002.

[9] L. R. Ford, *Automorphic Functions*, McGraw Hill, New York, 1929.

[10] J. Hadamard, *Les surfaces à courbures opposées et leurs lignes géodésiques*, J. Math. Pures et Appl. (5) 4(1898), 27-73.

[11] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[12] G. A. Hedlund, *On the metrical transitivity of geodesics on closed surfaces of constant negative curvature*, Ann. Math. 35(1934), 787-808

[13] S. Katok, *Fuchsian groups*, University of Chicago Press, 1992.

[14] S. Katok, *Fuchsian groups, geodesic flows on surfaces of constant negative curvature and symbolic coding of geodesic*, Clay Mathematics Proceedings. Volume 8, 2008.

[15] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, 2002.

[16] M. Morse, *A one-to-one representation of geodesics on a surface of negative curvature*, Amer. J. Math. XLIII(1921), 33-51.

[17] M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. XXII(1921), 84-100.

[18] M. Morse, *Symbolic dynamics*, Institute for advanced Study Notes, Princeton(1966) (unpublished), (First written 1938).

[19] J. Nielsen, *Untersuchungen zur Topologie der geschlossen zweiseitige Flächen*, Act. Math. 50(1927), 189-358.

[20] C. Series, *Symbolic dynamics for geodesic flows*, Acta Math, 146(1981), 103-128.

[21] C. Series, *Geometrical Markov coding of geodesics on surfaces of constant negative curvature*, Ergodic Theory Dynamical Systems 6(1986), 601-625.

국문초록

일정한 음곡률을 가지는 곡면에서의 측지선들을 다양한 방식으로 표현할 수 있다는 것이 알려져 있다. 그 중에서도 본질적으로 다른 두 가지 방법이 있는데, 하나는 곡면이 가지고 있는 기하적인 성질을 이용하는 것이고, 다른 하나는 실수를 연분수의 항들을 이용하여 수열로 나타낼 수 있는 것을 응용한 것이다. 전자는 Morse 코딩이라고 불리는데 곡면에 특정한 곡선들의 집합을 정해두고, 방향을 가지고 있는 측지선이 이 곡선들과 만날때마다 곡선에 정의된 라벨을 순서대로 기록하여 수열을 만드는 방식이다. 후자는 Artin 코딩이라고 한다. 이는 주어진 측지선의 올림의 두 끝점에 대해 푸앵카레 디스크의 경계에서 정의된 함수를 이용하여 수열을 얻고, 이 두 수열을 방향에 맞게 붙여서 해당 측지선에 대응시킨다. 특히, 이 논문에서 구체적으로 종수가 2인 닫힌 곡면의 경우에 Artin 코딩으로 얻어지는 이동공간이 유한 방향 그래프로 표현됨을 보임으로써 Sofic system임을 증명하였다.

주요어휘: 측지선 코딩, 기호 동역학
학번: 2011-20262