Universidade Estadual de Campinas - UNICAMP
Faculdade de Engenharia Elétrica e de Computação - FEEC
Departamento de Telemática - DT

Codificação de Geodésicas Fechadas Simples em Superfícies Hiperbólicas

Marinaldo Felipe da Silva

Orientador : ***Prof. Dr. Reginaldo Palazzo Jr.***

Banca Examinadora:

Prof. Dr. Yuzo Iano-FEEC/UNICAMP

Prof. Dr. Marcelo Firer-IMECC/UNICAMP

Prof. Dr. Nirzi G. de Andrade-IM/UFRJ

Prof. Dr. Antonio de Andrade e Silva-DM/UFPB

Prof. Dr. Henrique Lazari-DM/UNESP/RIO CLARO

Tese de Doutourado apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, como parte dos requisitos exigidos para a obtenção do título de **Doutor em Engenharia Elétrica**.

Campinas - SP
Março de 2002

Aos meus pais

Antonio Felipe (in memoriam) e Maria das Neves.

A minha esposa Rai

e aos meus filhos Renata, Felipe e Vanessa.

Agradecimentos

Ao Prof. Dr. Reginaldo Palazzo Junior; pela sua dedicação, paciência e compreensão somadas a incomensurável capacidade de orientar.

Aos professores da banca examinadora: Prof. Dr. Yuzo Iano (FEEC/UNICAMP), Prof. Dr. Marcelo Firer (IMECC/UNICAMP), Prof. Dr. Nirzi Gonçalves de Andrade (IME/UFRJ), Prof. Dr. José Carmelo Interlando (IBILCE/UNESP), Prof. Dr. Antonio Andrade e Silva (DM/UFPB) e Prof. Dr. Henrique Lazari(UNESP/RC).

Ao Prof. Dr. Henrique Lazari pela colaboração irrestrita;

Aos Professores: Dr. Trajano Pires da Nóbrega, Dr. Antonio de Andrade e Silva, Dr. José Carmelo Interlando, Luis Carlos Fernandes, Dr. Edson Moschim, Dr. Basilio Milani e Dr. Paulo Roberto de Paula e Silva pelo incentivo e amizade no decorrer desta jornada;

Aos funcionários da FEEC: Washington, Warley, Noêmia e Ana Flávia pela colaboração e convivência amigável;

Aos amigos que direta ou indiretamente contribuíram para a realização deste trabalho: Henrique Lazari, Adauto, Adriano, Alessandro, Andréa, Donizete, Eliane, João de Deus e esposa, Luzinete, Magno Rocha, Márcia, Martinho, Samísia, Rodrigo, Romis, Rúbia, Patricia e Tatiane;

Ao meu filho Felipe pela acessoria computacional e plotagem de várias figuras;

À Universidade Federal de Rondônia-UNIR;

À CAPES, pelo apoio financeiro.

Resumo

Geodésicas fechadas sobre superfícies modulares associadas a classes de conjugação de matrizes hiperbólicas em $SL(2,\mathbb{Z})$ podem ser codificadas de duas formas distintas: os códigos aritméticos e os códigos geométricos. Neste trabalho é feita uma descrição completa das geodésicas fechadas para as quais tais códigos coincidem. Para tal o plano hiperbólico foi tesselado por triângulos de duas formas distintas, diferenciando portanto as classes de geodésicas codificáveis.

De modo à fornecer suporte a extensão desta pesquisa, foram construídos: o grupo de classes de funções $\mathcal{MCG}(\Sigma_1)$ do toro 1-vez puncionado, associado à tesselação do plano hiperbólico agora por quadrados; uma máquina de estados finitos, que mostra de que forma o grupo de classes de funções desta particular superfície age sobre o conjunto $\mathcal{F}$ das células maximais; e a máquina de palavras-diferença que estabelece sua estrutura automática.

Abstract

Closed geodesics on modular surfaces associated to classes of hyperbolic matrices in $SL(2,\mathbb{Z})$ are bound to coding in two distinct ways: arithmetic and geometric coding. In this work, a complete description of closed geodesics for which the two types of codes are equivalents is established.

To achieve this goal, the hyperbolic plane was tesselated by triangles under two different restrictions giving rise to two classes of coded geodesics.

In order to extend the results of this research we considered the mapping class group $\mathcal{MCG}(\Sigma_1)$ of the once punctured torus, associated to the fundamental region based on squares which tesselate the hyperbolic plane; a finite-state machine, which shows how the mapping class group acts on the set of maximal cells; and the word-difference machine which establishes the automatic structure of the mapping class group.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

$\equiv (\mathrm{mod}\, m)$:	congruência módulo m
$\mathbb{Z}_m$:	conjuntos dos números inteiros mod m
$\sim$:	relação de equivalência
$(G, *)$:	grupo
S_n :	grupo simétrico de ordem n
$G \backsimeq G'$:	G é isomorfo a G'
$\mathbb{R}^{2\times 2}$:	grupo aditivo das matrizes reais 2×2
$GL(2, \mathbb{R})$:	grupo linear geral
$SL(2, \mathbb{R})$:	grupo especial linear
$PSL(2, \mathbb{R})$:	grupo especial linear projetivo
$\chi(\Sigma)$:	característica de Eüler-Poicaré de Σ
$d_M(x, y)$:	distância de x à y no espaço métrico M
$\mathbb{H}^2$:	modelo do semiplano superior do plano hiperbólico
X° :	interior do conjunto X
$\overline{X}$:	fecho do conjunto X
∂X :	fronteira do conjunto X
$H(S)$:	entropia de S
$\overline{L}$:	comprimento médio
η :	eficiência
LZ :	método de Lempel-Ziv
$SL(2, \mathbb{Z})$:	grupo modular
$PSL(2, \mathbb{Z})$:	grupo modular projetivo
γ :	geodésica
$\mathcal{D}$:	região de Dirichlet

F :	região fundamental
(A) :	código aritmético associado a matriz A
$[[x]]$:	parte inteira do número real x
$\mathrm{Tr}\, A$:	traço da matriz A
A_j :	matriz do tipo $\begin{pmatrix} j & -1 \\ 1 & 0 \end{pmatrix}$
u_A, w_A :	pontos fixos da transformação A
D, E :	segmentos geodésicos rotulados por direita e esquerda
M :	superfície modular
$\mathbb{F}$:	tesselação de Farey
Δ :	triângulo ideal
Σ :	superfície possivelmente puncionada
Σ_p :	toro p-vezes puncionado
$\mathcal{MCG}(\Sigma)$:	grupo de classes de funções de Σ
$\mathcal{T}(\Sigma)$:	espaço de Teichmüller
$\mathcal{PML}(\Sigma)$:	espaço das laminações projetivas
$\Lambda(\Gamma)$:	conjunto limite de Γ
$\mathcal{A}$:	alfabeto
$\mathcal{L}$:	linguagem sobre um alfabeto
$\mathcal{A} \cup \{\$\}$:	alfabeto estendido
I_j :	células maximais
$\mathbf{e}_j$:	*train track* elementares
ι_j :	simetrias do quadrado
$\delta_j^{\pm 1}$:	torções de Dehn
I_j :	células maximais para $\mathcal{ML}(\Sigma_1)$

Capítulo 1

Introdução

O presente trabalho tem como objetivo a introdução de técnicas de codificação de fontes, caracterizadas como sistemas dinâmicos discretos. Estas técnicas levam em consideração a codificação de geodésicas no plano hiperbólico.

As técnicas de codificação de fontes (revisadas no Capítulo 3) são na realidade casos particulares de sistemas dinâmicos discretos no tempo. Os sistemas dinâmicos considerados neste trabalho são sistemas dinâmicos discretos no tempo: o tempo é discreto porque uma órbita $\{\phi^n(x)\}_{n\in\mathbb{Z}}$ é vista como a evolução de um ponto x nos instantes de tempo $1, 2, 3, ...$

Entretanto, sistemas dinâmicos contínuos no tempo, onde as órbitas são parametrizadas pelo conjunto dos números reais, são também de grande importância. Tais sistemas surgem naturalmente de equações diferenciais. A teoria dos sistemas dinâmicos cresceu devido a esforços para entender sistemas dinâmicos contínuos no tempo, tais como: fluxos geodésicos sobre superfícies com curvatura constante e negativa.

Atualmente, estuda-se com mais freqüência, os sistemas dinâmicos sobre espaços compactos, muito embora, exista também muito interesse em sistemas dinâmicos sobre espaços não-compactos. Por exemplo, fluxos geodésicos sobre espaços não-compactos de curvatura negativa e shifts sobre alfabetos infinito-enumeráveis.

A dinâmica simbólica, pela sua relevância, ocupa boa parte da teoria dos sistemas dinâmicos. Desta forma, faremos uma breve introdução desta teoria, e de como a dinâmica simbólica se ajusta no presente trabalho. Algumas noções básicas de dinâmica

são: continuidade, compacidade e conjugação topológica.

Os sistemas físicos podem ser descritos através dos valores de um número finito de medidas. Como existe o envolvimento temporal, estes valores mudam. Seja M (um espaço métrico) o conjunto destes possíveis valores para um sistema. Imaginemos um filme (exame) do sistema, com um quadro a cada segundo. Cada quadro depende do anterior, a saber, como o sistema se comportou durante o intervalo de um segundo. Assumindo que as leis que governam tal sistema são invariantes com o tempo, a dependência deste quadro com o quadro anterior fornece uma função $\phi : M \to M$, normalmente contínua. Assim, se x descreve o sistema no instante de tempo $t = 0\,\text{seg}$, então $\phi(x)$ descreve o sistema no instante de tempo $t = 1\,\text{seg}$, $\phi^2(x)$ em $t = 2\,\text{seg}$, e assim por diante. Assim, para estudar como o sistema se comporta no tempo, basta estudar como $x, \phi(x), \phi^2(x), \ldots$ se comporta. Estas considerações motivam a seguinte noção.

Um **sistema dinâmico** (M, ϕ) consiste de um espaço métrico compacto M juntamente com uma função contínua $\phi : M \to M$. Se ϕ é um homeomorfismo, então (M, ϕ) é dito ser **sistema dinâmico invertível**.

Frequentemente abreviamos (M, ϕ) por ϕ, para enfatizar **a dinâmica**.

Seja (M, ϕ) um sistema dinâmico. A **órbita** de um ponto $x \in M$ é o conjunto $\{\phi^n(x)\}_{n\in\mathbb{Z}}$ quando ϕ é invertível e, $\{\phi^n(x)\}_{n\geq 0}$ caso contrário. Um **ponto periódico** é um ponto $x \in M$ tal que $\phi^n(x) = x$ para algum $n > 0$. Uma órbita é chamada uma **órbita periódica** se x é um ponto periódico.

Um modelo mais realístico de um sistema físico usa continuidade no tempo para modelar a evolução. O modelo matemático para isto é chamado um **fluxo contínuo** consistindo de uma família $\{\phi_t\}_{t\in\mathbb{R}} : X \to X$ de homeomorfismos ϕ_t de um espaço métrico compacto sobre si mesmo, tal que

1. $\phi_t(x)$ é conjuntamente contínua em t e x;

2. $\phi_s \circ \phi_t = \phi_{s+t}$, para todo $s, t \in \mathbb{R}$.

A **órbita** de um ponto $x \in X$ sobre um fluxo contínuo $\{\phi_t\}$ é o conjunto $\{\phi_t(x) : t \in \mathbb{R}\}$. As soluções de um sistema de equações diferenciais podem ser vistas como um fluxo

contínuo. Uma classe de fluxos contínuos de grande interesse na dinâmica simbólica é a classe dos fluxos geodésicos.

Dada uma superfície S, existem as noções clássicas de vetor tangente, curvatura e geodésica. **Geodésicas** são curvas que minimizam (localmente) distância. Sobre uma superfície compacta S de curvatura negativa, para cada ponto x sobre S e para cada vetor u tangente a S em x, existe uma única geodésica $\gamma_{u,x}$ passando por x na direção de u. As geodésicas são normalmente parametrizadas por $\mathbb{R}$, o parâmetro é visto como sendo o tempo. Denotamos por $\gamma_{u,x}(t)$ a posição da geodésica $\gamma_{u,x}$ no instante de tempo t.

O **fluxo geodésico** é o fluxo contínuo $\{\phi_t\}_{t\in\mathbb{R}} : X \to X$ definido por $\phi_t(x,u) = (y,v)$, onde $y = \gamma_{u,x}(t)$ e v é o vetor unitário tangente a $\gamma_{u,x}$ em y, como mostrado na Figura 1.1.

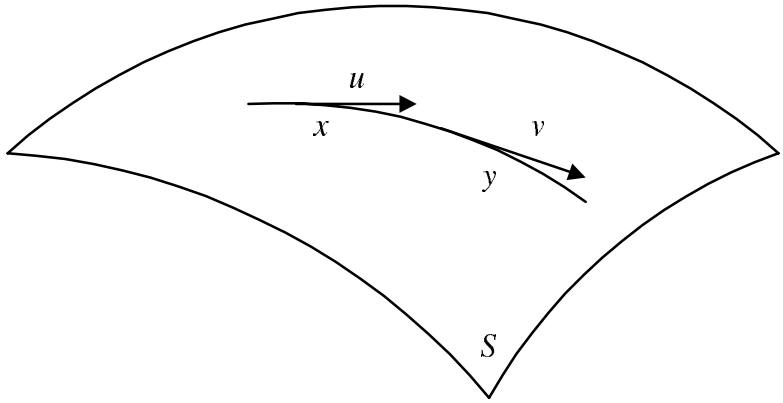

Figura 1.1: Fluxo geodésico sobre uma superfície.

No caso dos sistemas dinâmicos discretos no tempo, a idéia da dinâmica simbólica é associar seqüências de símbolos com órbitas de fluxos contínuos e, deduzir as propriedades das órbitas através das propriedades das seqüências.

Dois fluxos geodésicos $\{\phi_t\}_{t\in\mathbb{R}} : X \to X$ e $\{\psi_t\}_{t\in\mathbb{R}} : Y \to Y$ são **equivalentes** se existe um homeomorfismo (que preserva orientação) $\pi : X \to Y$ que mapeia órbitas de $\{\phi_t\}$ em órbitas de $\{\psi_t\}$.

Daremos agora uma pequena mostra de como este trabalho se insere no contexto da dinâmica simbólica.

No Capítulo 3, a técnica de codificação de Elias [24], pode ser vista como um fluxo geodésico que intersecta uma superficie S. A seqüência $,011010111010\cdots = \delta$ a ser codificada está associada ao fluxo geodésico que intersecta tal superfície.

No que segue, usaremos a notação da parte fracionária de um número real, isto é, cada $r \in \mathbb{R}$, pode ser escrito de maneira única como $r = n + s$, onde $n \in \mathbb{Z}$ e $s \in [0, 1)$. Neste caso escrevemos $r \equiv s \pmod 1$ e chamamos s a **parte fracionária** do número real r.

Com esta notação a função $\phi_r : \mathbb{R} \to \mathbb{R}$ definida por $\phi_r(x) = x + r \pmod 1$ é um homeomorfismo. Dessa forma $(\mathbb{R}, \phi_r)$ é um sistema dinâmico invertível.

Os automorfismos do toro introduzidos no Capítulo 5, são sistemas dinâmicos discretos no tempo que compartilham muitas propriedades dinâmicas com fluxos geodésicos.

Como exemplo, seja $M = \mathbb{T}^2$ o toro bi-dimensional, descrito como o quadrado unitário $[0, 1] \times [0, 1]$ com lados opostos rotulados conforme mostra a Figura 5.1, (*pág.*83). Assim, o ponto $(0, t)$ é colado no ponto $(1, t)$ para $0 \leq t \leq 1$ e $(s, 0)$ é colado em $(s, 1)$ para $0 \leq s \leq 1$. A função $\phi : \mathbb{T}^2 \to \mathbb{T}^2$ definida por $\phi(s, t) = (s + t, t) \pmod 1$ tem inversa contínua dada por $\phi^{-1}(s, t) = (t, s - t) \pmod 1$. Consequentemente, $(\mathbb{T}^2, \phi)$ é um sistema dinâmico invertível.

Chamamos a atenção para o fato de que a dinâmica $\phi(s, t)$ é objeto da Proposição 5.2.2.

As técnicas de codificação de geodésicas no plano hiperbólico ora propostas são: a codificação geométrica e a codificação aritmética.

O teorema central deste trabalho (4.1.21) estabelece condições de existência para a equivalência entre tais códigos.

Tendo em vista o papel que tais códigos poderão vir a desempenhar e pensando em fornecer suporte à continuidade da pesquisa para casos mais gerais, é construído o grupo de classes de funções para uma particular superfície. A codificação feita usando-se os elementos do grupo de classes de funções de uma superfície com curvatura constante e negativa tem a seguinte racionalidade: estaremos codificando não uma curva simplesmente, mas uma classe de curvas cuja curva em questão é a representante da

classe.

Embora não tenhamos conhecimento de trabalhos anteriores relacionados com esta temática na área de Engenharia, existem alguns artigos no contexto da matemática envolvendo geometria hiperbólica que foram relevantes para o desenvolvimento do presente trabalho.

Por exemplo, em [23], Morse propôs a codificação de geodésicas em relação a uma dada região de Dirichlet para um grupo Fuchsiano Γ.

Em [30],[29] e [16], foi proposto o uso de frações contínuas para a codificação de geodésicas em superfícies modulares. Outros trabalhos relacionados aparecem em [2], [27], [3] e [12].

Em [15], foi proposto um algoritmo para a construção de códigos aritméticos tendo como base o método de redução de Gauss.

Em [25], foi proposto um procedimento para a contrução do grupo de classes de funções para o toro 2-vezes puncionado, baseado no conceito de *train tracks* proposto em [4].

Uma das dificuldades desta área de pesquisa, reside no fato de que tais técnicas envolvem conceitos matemáticos de análise, álgebra, geometria e topologia de nível não elementar.

Portanto, para o estudo e proposta de novas técnicas é imperativo que métodos desenvolvidos sejam apropriadamente caracterizados para servirem de referencial no estudo em questão, bem como sejam minoradas as dificuldades através da consideração de casos particulares de fácil entendimento.

Na tentativa de atender esses objetivos, inserimos, no Capítulo 2, os conceitos de estruturas matemáticas (grupos e espaços métricos), números inteiros e algumas de suas propriedades que, no nosso entendimento são essenciais para o desenvolvimento da teoria de codificação proposta nos Capítulos 4 e 5.

No Capítulo 3, são abordados os sistemas de comunicações típico e simplificado, dos quais é descrito cada um dos blocos componentes. Sendo nosso objeto de estudo a codificação de fontes, inserimos os conceitos de informação, incerteza e entropia. Como não poderia deixar de ser, enunciamos o teorema de codificação de fontes, [24].

Ainda neste capítulo, fazemos uma revisão de algumas técnicas de codificação de fontes discretas. Na Seção 3.2, revisamos os códigos livres de prefixo, a saber, os códigos de Huffman, [6], [24]. Na seqüência descrevemos os códigos de Elias [24], por meio de um exemplo de compactação de uma fonte binária. Finalizando esta seção, apresentamos os métodos de $LZ77$ e $LZ78$ devidos a Abraham Lempel e Jacob Ziv [20], [21].

No Capítulo 4, desenvolvemos uma teoria de codificação de geodésicas. Para tal, o plano hiperbólico foi tesselado por triângulos. A Seção 4.1, reporta sobre a codificação de geodésicas no plano hiperbólico. É de fundamental importância observar que a divisão desta seção não é meramente formal, mas retrata realmente a seqüência lógica do problema mencionado em cada um de seus respectivos títulos. Tal codificação, está associada a uma matriz

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

pertencente ao grupo $SL(2,\mathbb{Z})$, que por sua vez está associada à transformação de Möbius

$$T : z \to \frac{az+b}{cz+d}.$$

De posse de tal transformação, encontramos seus pontos fixos, para os quais determinamos a geodésica no plano hiperbólico que os une. A partir dessa geodésica é feita a construção dos códigos geométricos e aritméticos, estudados nas Subseções 4.1.1 e 4.1.2, respectivamente. **As condições necessárias e suficientes para que um código geométrico coincida com um código aritmético é um dos principais resultados deste trabalho, e, é objeto do Teorema 4.1.21.**

Na Seção 4.2, consideramos a codificação de geodésicas no plano hiperbólico. Entretanto, existem duas diferenças em relação a codificação apresentada na Seção 4.1. A primeira, é que o semi-plano superior $\mathbb{H}^2$ é tesselado por meio de triângulos ideais, isto é, triângulos cujos vértices estão na fronteira do disco de Poincaré. Enquanto que na Seção 4.1, o plano foi tesselado por triângulos com um vértice no infinito e os outros dois em $\mathbb{H}^2$. A segunda grande diferença está na seqüência de codificação. Enquanto na Seção 4.1 os elementos da seqüência são potências de transformações de Möbius específicas, na presente seção, os elementos da seqüência são segmentos geodésicos devidamente rotulados.

Essa forma de tesselar o plano hiperbólico (por triângulos) tem a seguinte explicação: qualquer geodésica no plano hiperbólico intersectando a região fundamental corta lados adjacentes dessa região. Na tesselação feita por triângulos cujos vértices estão todos na fronteira do disco de Poincaré (conhecida como **tesselação de Farey**) qualquer geodésica intersecta a região fundamental (triângulo), isto é, é codificável. Tal região é identificada pelo grupo quociente $\mathbb{H}^2/\Gamma$ (região fundamental) sendo topologicamente caracterizada como a superfície esférica com 3 puncionamentos.

Pergunta: É possível realizar a codificação de geodésicas tendo como grupo quociente $\mathbb{H}^2/\Gamma$ (região fundamental) um polígono de n lados, $n \geq 4$, com apenas um ponto excepcional no ∞, onde as geodésicas cortam lados não adjacentes do polígono?

A resposta é sim!, e, para ver isto, inserimos no Capítulo 5, a codificação de geodésicas no plano hiperbólico, onde o plano é tesselado por quadrados conforme mostra a Figura 5.1. Neste modelo, a codificação é feita através de homeomorfismos (torções de Dehn) que no contexto dos sistemas dinâmicos, desempenham o papel da dinâmica.

Neste capítulo, construímos o grupo de classes de funções e a máquina de estados finitos para o toro 1-vez puncionado Σ_1. Na Seção 5.1, de uma forma geral, fornecemos os conceitos e definições necessários para o desenvolvimento do presente capítulo. Esta seção está desmembrada em quatro subseções. Na subseção 5.1.1 definimos a função de Markov e os grupos automáticos. Esta subseção contém uma ampla discussão sobre máquinas de palavras-diferença. Na Subseção 5.1.2, definimos as malhas simples e consideramos o conceito dos *train tracks* [4]. Este conceito envolve a escolha de uma estrutura hiperbólica sobre a superfície considerada. Na Subseção 5.1.3, contruímos o espaço $W(\tau)$ das células maximais geradas pelas combinações lineares das curvas irredutíveis γ sobre Σ e, finalmente, na Subseção 5.1.4 definimos as torções de Dehn. Tais torções são homeomorfismos sobre Σ agindo sobre as células maximais I_0, I_1, I_2 e I_3.

Na Seção 5.2 é feito, com detalhes, o estudo proposto na Seção 5.1 para uma parti- cular superfície Σ_1, a saber: **o toro 1-vez puncionado**. Ainda neste capítulo, construímos a máquina de estados finitos. Tal máquina fornece a ação do grupo de classes de funções $\mathcal{MCG}(\Sigma_1)$ sobre as células maximais formadoras da partição de Markov. Para

finalizar o capítulo construímos a máquina de palavras-diferença que fornece a estrutura automática para o grupo em questão.

No Capítulo 6, apresentamos as conclusões e propostas para futuros trabalhos.

Capítulo 2

Conceitos Fundamentais

Neste trabalho iremos codificar fontes contínuas. Esta codificação é feita de duas maneiras distintas, o código geométrico, em relação a uma região fundamental estabelecida a priori, está associado à geometria de um elemento geodésico pertencente a um grupo fuchsiano Γ, enquanto que o código aritmético está associado à expansão em frações contínuas do ponto fixo atrator de tal elemento.

Salientamos que embora os conceitos envolvidos sejam não muito elementares, tornar-se-ão bastantes simples quando nos restringirmos a uma particular superfície, para a qual também construímos o grupo de classes de funções. Baseado nesses parâmetros, "apontamos" o caminho para o estudo de superfícies mais gerais.

O presente capítulo proporciona uma introdução a alguns conceitos matemáticos indispensáveis para o desenvolvimento de tal teoria, para tal faremos uma resumida revisão dos conceitos abaixo relacionados.

Números inteiros: a congruência módulo m e os inteiros módulo m.

Relações de equivalência: classes de equivalência e representante de uma classe.

Grupos: definição, grupos de permutação, grupos cíclicos, grupos de simetrias, classe lateral, conjugação, automorfismo de grupos e alguns grupos especiais.

Espaços métricos: definição e exemplos, isometria, homeomorfismo, seqüências e resultados relacionados.

O plano hiperbólico: definição, modelos, métrica hiperbólica, geodésicas e conjunto limite.

As referências básicas para esse capítulo são: [1], [22], [8], [9], [10], [13], [19], [17] e [24].

2.1 Os Inteiros Módulo m

Seja m um inteiro fixo. Diz-se que dois números inteiros a e b são **côngruos** módulo m se $a-b$ é divisível por m. Neste caso escreve-se $a \equiv b \pmod{m}$. A **congruência** módulo m satisfaz as seguintes propriedades:

(i) $a \equiv a \pmod{m}$; (propriedade reflexiva).

(ii) Se $a \equiv b \pmod{m}$, então $b \equiv a \pmod{m}$; (propriedade simétrica).

(iii) Se $a \equiv b \pmod{m}$ e $b \equiv c \pmod{m}$, então $a \equiv c \pmod{m}$; (propriedade transitiva).

Nas questões de congruências é conveniente supormos que m é positivo, já que $a \equiv b \pmod{m}$ se, e somente se, $a \equiv b \pmod{(-m)}$. Suponhamos que $m \geq 2$. Do ponto de vista de congruências módulo m não há diferença entre 1 e $m+1$ ou entre 0 e m. Na verdade todos os múltiplos de m são identificados via congruências e pode-se pensar no conjunto $C_0 = \{y \in \mathbb{Z};\ y \equiv 0 \pmod{m}\} = \{0, \pm m, \pm 2m, ..., \pm km, ...\}_{k \geq 0}$ como um novo objeto. Trata-se, assim, de m novos objetos: $C_1 = \{1 \pm km\}_{k \geq 0}$, $C_2 = \{2 \pm km\}_{k \geq 0}, ...,$ $C_{m-1} = \{m-1 \pm km\}_{k \geq 0}$. Cada um desses objetos é chamado **um inteiro módulo** m. O **conjunto dos inteiros módulo** m é indicado como $\mathbb{Z}_m$.

Observação 2.1.1 *Desejamos ressaltar que os novos objetos, a saber, os inteiros módulo m, são representados na forma $C_x = \{y \in \mathbb{Z};\ y \equiv x \pmod{m}\}$ e o* ***representante*** *x se vincula a esta representação. Assim, temos que usar um dos seus representantes ao falarmos num inteiro módulo m. Sempre que definimos uma função ou uma operação entre os inteiros módulo m, somos obrigados a verificar que a definição independe dos representantes dos inteiros módulo m em questão. Quando isso ocorre dizemos que a função ou operação é* ***bem definida.***

2.2 Relações de Equivalência

Seja X um conjunto. Uma relação $\backsim$ é chamada de uma **relação de equivalência** se são válidas as seguintes propriedades:

(i) Para todo $x \in X$, $x \backsim x$. (reflexiva)

(ii) Se $x \backsim y$ então $y \backsim x$, $\forall\, x, y \in X$ (simétrica)

(iii) Se $x \backsim y$ e $y \backsim z$ então $x \backsim z$, $\forall\, x, y, z \in X$ (transitiva)

Note que a congruência é uma relação de equivalência. Como no caso das congruências, uma relação de equivalência $\backsim$ possibilita a definição de um novo objeto. Com efeito, se $x \in X$, seja $\overline{x} = \{y \in X;\, y \backsim x\}$; $\overline{x}$ é chamado a **classe de equivalência** módulo $\backsim$ determinada pelo elemento x de X. O elemento x por sua vez é chamado **representante** da classe de equivalência $\overline{x}$.

Observação 2.2.1 *Todo elemento de X define uma classe de equivalência, e uma determinada classe de equivalência pode admitir vários representantes. É importante salientar que uma classe de equivalência é, por sua natureza, definida a partir de um representante.*

2.3 Grupos

Seja G um conjunto não vazio munido de uma operação $* : G \times G \to G$ entre os pares de G definida por: $(x, y) \rightsquigarrow x * y$.

Diz-se que o par $(G, *)$ é um **grupo** se satisfaz as três seguintes propriedades:

1. $\forall\, a, b, c \in G$, $a * (b * c) = (a * b) * c$ (associatividade)

2. $\exists\, e \in G$ tal que $a * e = e * a = a$ $\forall\, a \in G$ (elemento neutro ou identidade)

3. $\forall\, a \in G$, $\exists\, b \in G$ tal que $a * b = b * a = e$ (elemento inverso).

4. Se em um grupo $(G,*)$ verifica-se a propriedade: $a*b=b*a \quad \forall\, a,b\in G$

dizemos que o grupo $(G,*)$ é um **grupo abeliano** ou **comutativo.**

Exemplo 2.3.1 $(\mathbb{Z},+),(\mathbb{Q},+),(\mathbb{R},+),(\mathbb{C},+)$, *são grupos aditivos infinitos.*

Exemplo 2.3.2 $(\mathbb{Z}_m,\oplus)$, *é um grupo finito com exatamente* m *elementos, onde* $\oplus$ *é a operação soma módulo* m*, isto é,* $a\oplus b$ *é o resto da divisão de* $a+b$ *por* m*.*

Exemplo 2.3.3 *O grupo de Klein,* $K=\{e,a,b,c\}$ *é um grupo finito com* 4 *elementos, cuja operação* $*:K\times K\longrightarrow K$ *é definida conforme a tabela abaixo.*

$*$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Exemplo 2.3.4 *Seja* S *um conjunto não vazio e* $G=\{f:S\to S,\ \text{bijetivas}\}$*. Se* $*$ *é a operação composição de funções, isto é,* $*:G\times G\to G$ *que associa* $(g,f)\rightsquigarrow g*f=g\circ f$*, então* $(G,*)$ *é um grupo tendo* $I_s:S\to S$ *como elemento identidade. Esse grupo é chamado de* ***grupo de permutações do conjunto*** S*. Se* $S=\{1,2,...,n\}$ *denotaremos esse grupo por* S_n*, e temos que o número de elementos de* S_n *é exatamente* $n!$*.*

Em particular, S_3 é um exemplo de grupo, não abeliano, com exatamente 6 elementos. É usual denotar um elemento de S_n por:

$$f=\begin{pmatrix} 1 & 2 & 3 & ... & n \\ f(1) & f(2) & f(3) & ... & f(n)\end{pmatrix}.$$

Assim, o grupo S_3 é composto dos seguintes 6 elementos:

$$\begin{aligned} e &= \begin{pmatrix}1&2&3\\1&2&3\end{pmatrix} \quad ; f_1=\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}\\ f_2 &= \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix} \quad ; f_3=\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}\\ f_4 &= \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix} \quad ; f_5=\begin{pmatrix}1&2&3\\3&1&2\end{pmatrix}. \end{aligned}$$

Sejam G um grupo e H um subconjunto não vazio de G. Dizemos que H é um **subgrupo** de G, se H for ele próprio um grupo com a mesma operação de G. Para denotar que $H \subset G$ é um subgrupo de G, escrevemos $H \leq G$. Com esta notação, se denotarmos $\langle x \rangle = \{x^m : m \in \mathbb{Z}\} \subset G$, com $x^n = \begin{cases} x^{n-1} \cdot x, \ se \ n > 0 \\ e, \ se \ n = 0 \\ (x^{-1})^{-1}, \quad se \ n < 0 \end{cases}$, então o conjunto $\langle x \rangle$ é dito ser um **grupo cíclico** gerado pelo elemento $x \in G$, e o elemento x um **gerador** do grupo G.

Seja $\{1, 2, 3, ..., n\}$, $n \geq 3$, o conjunto dos vértices de um polígono regular de n lados. Considerando S_n o grupo de todas as permutações do conjunto de vértices $\{1, 2, 3, ..., n\}$, vamos agora ver um exemplo de um subgrupo de S_n, não abeliano, contendo exatamente $2n$ elementos.

Seja $\theta \in S_n$ a permutação determinada pelo efeito da **rotação** de um ângulo de $\frac{2\pi}{n} rd$ no sentido trigonométrico , isto é,

$$\theta = \begin{pmatrix} 1 & 2 & 3 & ... & n-1 & n \\ 2 & 3 & 4 & ... & n & 1 \end{pmatrix}.$$

Para fixar idéias, inserimos os polígonos regulares com 5 e 6 lados, como pode ser observado na Figura 2.1.

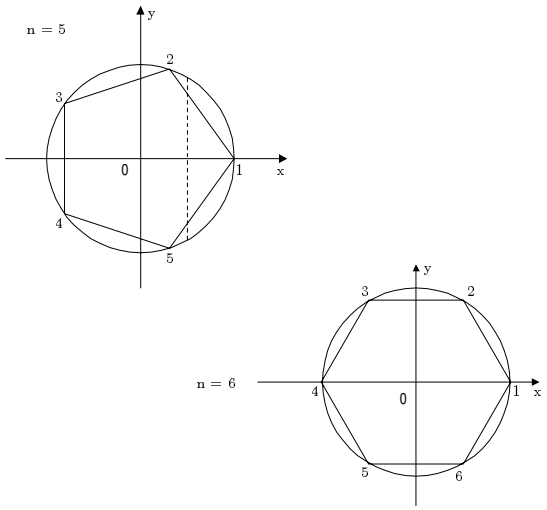

Figura 2.1: Polígonos de 5 e 6 lados.

Consideremos $r \in S_n$ a permutação determinada pelo efeito da reflexão da figura em torno do eixo $\overline{ox}$, isto é, se n é par, então r fixa os vértices 1 e $\dfrac{n+2}{2}$ e, é representado por

$$r = \begin{pmatrix} 1 & 2 & 3 & ... & \frac{n+2}{2} & ... & n-1 & n \\ 1 & n & n-1 & ... & \frac{n+2}{2} & ... & 3 & 2 \end{pmatrix}$$

se n é impar, então r fixa apenas o vértice 1 e, é representada por

$$r = \begin{pmatrix} 1 & 2 & 3 & ... & n-1 & n \\ 1 & n & n-1 & ... & 3 & 2 \end{pmatrix}.$$

Seja $D_n = \langle r, \theta \rangle$ o menor subgrupo de S_n contendo r e θ. Então $D_n = \{e, r, \theta, \theta^2, ..., \theta^{n-1}, r\theta, r\theta^2, ..., r\theta^{n-1}\}$ é um grupo não abeliano contendo exatamente $2n$ elementos. Tal grupo é chamado **grupo dihedral de ordem 2n** ou **grupo de simetrias do polígono regular de n lados**.

Observação 2.3.5 *O grupo $D_4 = \{e, r, \theta, \theta^2, \theta^3, r\theta, r\theta^2, r\theta^3\}$ das simetrias do quadrado é um exemplo de um grupo não abeliano contendo exatamente 8 elementos.*

Consideremos $X = \{1, 2, 3, 4, 5\}$. Um elemento de S_5, isto é, uma permutação, é dado por:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 1 & 5 & 2 \end{pmatrix}.$$

Podemos usar a notação de ciclos para representar uma dada permutação σ, através da seguinte regra: abra um parêntesis e escreva o menor inteiro que é trocado pela permutação. Liste a imagem deste inteiro sob a permutação σ seguido de sua imagem e assim por diante. Feche o parêntesis quando o ciclo se fechar, isto é, quando o primeiro inteiro da lista for imagem do último. Abra agora um novo parêntesis e escreva o menor inteiro que ainda não foi listado e continue o processo de forma análoga. No exemplo acima considerado

$$\sigma = (13)(245).$$

Uma permutação da forma $(a_1, a_2, ..., a_k)$ é chamada um **ciclo**, tendo em vista que a mesma permuta ciclicamente os elementos $a_1, a_2, ..., a_k$ de X e fixa os demais.

Proposição 2.3.6 *[1], Sejam G um grupo e $H \leq G$. Se $x, y \in G$, então a relação $x \equiv y \pmod{H} \Longleftrightarrow xy^{-1} \in H$ define uma relação de equivalência no conjunto G.*

Consideremos agora a classe de equivalência $\overline{x} = \{y \in G\,;\, y \equiv x \pmod{H}\}$.

Assim,

$y \in \overline{x} \Leftrightarrow y \equiv x \pmod{H} \Leftrightarrow yx^{-1} = h$, para algum $h \in H \Leftrightarrow y = hx$, para algum $h \in H$. Se denotarmos $Hx = \{hx\,; h \in H\}$, então $\overline{x} = Hx$, é chamada **uma classe lateral (à direita)** de H em G. Chamamos de **conjunto quociente** e representamos por G/H, ao conjunto $\{\overline{x} : x \in H\}$, isto é, $G/H = \{Hx : x \in G\}$ é o conjunto de todas as classes laterais (**à direita**) de H em G.

Suponhamos que G/H possui exatamente n classes laterais, assim sendo $G/H = \{Hx_1, Hx_2, ..., Hx_n\}$ onde, $x_1, x_2, ..., x_n \in G$. Como o conjunto $\{Hx_1, Hx_2, ..., Hx_n\}$ é uma partição de G temos que: $Hx_i \cap Hx_j = \varnothing$ se $i \neq j$ e além disso, $G = Hx_1 \overset{\circ}{\cup} Hx_2 \overset{\circ}{\cup} ... \overset{\circ}{\cup} Hx_n$, onde $\overset{\circ}{\cup}$ significa união disjunta.

Se X é um conjunto finito, representamos por $|X|$ o **número de elementos** de X. Se G é um grupo finito definimos a **ordem de G** como sendo o número $|G|$ de elementos de G.

Proposição 2.3.7 *[1], Sejam G um grupo e $x, y \in G$. A relação $x \underset{G}{\sim} y \Leftrightarrow \exists\, g \in G\,; y = gxg^{-1}$ define uma relação de equivalência em G.*

Se $x \underset{G}{\sim} y$ dizemos que x e y são **elementos conjugados** em G. A classe $\overline{x} = \{y \in G : x \underset{G}{\sim} y\}$ é chamada **classe de conjugação (em G) determinada pelo elemento $x \in G$.**

Vamos agora definir a noção de homomorfismo de grupos . Sejam $(G, *)$ e $(G', *')$ grupos e $\psi : G \to G'$ uma função de G em G'. Dizemos que ψ é um **homomorfismo** se $\psi(x * y) = \psi(x) *' \psi(y)$. Se o homomorfismo $\psi : G \to G'$ for bijetivo dizemos que ψ é um **isomorfismo** e neste caso dizemos que G é **isomorfo** a G' e denotamos por $G \backsimeq G'$. Um isomorfismo $\psi : G \to G$ diz-se um automorfismo de G.

O exemplo a seguir é de vital importâcia para os nossos propósitos, o mesmo será uma de nossas principais ferramentas de trabalho nos Capítulos 4 e 5.

Exemplo 2.3.8 *Considere o conjunto* $\mathbb{R}^{2\times 2}$ *das matrizes* 2×2 *com entradas reais, isto é,*

$$\mathbb{R}^{2\times 2}=\left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d\in\mathbb{R}\right\},$$

então o par $(\mathbb{R}^{2\times 2},+)$ *é um grupo, onde* $+$ *é a operação soma de matrizes. Se considerarmos o conjunto* $GL(2,\mathbb{R})\subset\mathbb{R}^{2\times 2}$*;*

$$GL(2,\mathbb{R})=\left\{A=\begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d\in\mathbb{R}\,;ad-bc\neq 0\right\}$$

então $(GL(2,\mathbb{R}),\cdot)$ *é um grupo, onde* $(\,\cdot\,)$ *denota a operação produto de matrizes. Tal grupo é denominado* ***grupo linear geral das matrizes*** 2×2 *sobre* $\mathbb{R}$*. Consideremos em* $GL(2,\mathbb{R})$ *o seguinte subconjunto:*

$$SL^{*}(2,\mathbb{R})=\{A\in GL(2,\mathbb{R}) : \det A=\pm 1\}.$$

Com a mesma operação de $GL(2,\mathbb{R})$, $SL^{*}(2,\mathbb{R})$ é um grupo, que tem como subgrupo natural o seguinte grupo:

$$SL(2,\mathbb{R})=\{A\in GL(2,\mathbb{R}) : \det A=1\}$$

chamado **grupo especial**.

Para os nossos propósitos é também de relevada importância o grupo especial linear projetivo

$$PSL(2,\mathbb{R})=SL(2,\mathbb{R})/\{\pm I\}.$$

Exemplo 2.3.9 *Considere* $G=\{z\to\dfrac{az+b}{cz+d} : a,b,c,d\in\mathbb{R}\ e\ ad-bc=1\}$*, e* $\varphi : SL(2,\mathbb{Z})\to G$ *definida por* $\varphi\begin{pmatrix} a & b \\ c & d \end{pmatrix}=\dfrac{az+b}{cz+d}$*. Então* φ *é um homomorfismo sobrejetor. Portanto, os grupos* $SL(2,\mathbb{Z})$ *e* G *são isomorfos, isto é,* $SL(2,\mathbb{Z})\cong G$*.*

Note que, $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\in\operatorname{Ker}\varphi\Leftrightarrow\dfrac{az+b}{cz+d}=z,\ \forall z\in\mathbb{C}$, daí, obtemos $cz^2+(d-a)\,z-b=0$, $\forall\, z\in\mathbb{C}$, $z\neq-\dfrac{c}{d}$. Note ainda que, $SL(2,\mathbb{Z})$ e G têm comportamentos

geométricos diferentes quando agem sobre $\mathbb{R}^2$ (identificando $\mathbb{R}^2$ com $\mathbb{C}$). Por exemplo, $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (-1, 1)$ e $-\frac{1}{z} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{-1}{1+i} = -\frac{1}{2} + \frac{1}{2}i$.

Vamos estender G a $\widetilde{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, definindo $T(-\frac{c}{d}) = \infty$ e $T(\infty) = T_1(0)$, onde $T_1(z) = T(\frac{1}{z}) = \frac{bz+a}{dz+c}$.

Consideremos o subgrupo $PSL(2, \mathbb{Z}) \subset PSL(2, \mathbb{R})$. Tal subgrupo é a imagem do subgrupo $SL(2, \mathbb{Z}) \subset SL(2, \mathbb{R})$ pela projeção contínua $\pi : SL(2, \mathbb{R}) \to PSL(2, \mathbb{R})$. Este grupo é chamado **grupo modular** e, é isomorfo ao grupo das tranformações lineares fracionárias $T : \widetilde{\mathbb{C}} \to \widetilde{\mathbb{C}}$ definidas por

$$z \to \frac{az+b}{cz+d} : a, b, c, d \in \mathbb{Z}$$

chamadas **transformações de Möbius.** Assim,

$$PSL(2, \mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{Z} \text{ e } ad - bc = 1 \right\}.$$

Sejam $X \neq \varnothing$ um conjunto e G um grupo. Uma **ação de G sobre X** é uma função $* : G \times X \to X$ definida por $(g, x) \to g * x$ satisfazendo:

1. $ex = x$,
2. $(g_1 g_2)(x) = g_1(g_2 x)$, $\forall\, x \in X$ e $\forall\, g_1, g_2 \in G$.

A **órbita** de um elemento $x \in A$ por um grupo G é o conjunto $G_x = \{gx; g \in G\}$.

Exemplo 2.3.10 *O grupo $G = PSL(2, \mathbb{Z})$ age sobre o conjunto $X = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ como transformações de Möbius.*

Um **grafo** é uma tripla ordenada (N, A, g), onde

N = um conjunto não-vazio de **vértices** (**nós** ou **nodos**)

A = um conjunto de **arestas** (**arcos**)

g = uma função que associa cada aresta a a um par não ordenado x-y de vértices chamados **extremos** de A.

2.4 Espaços Métricos

2.4.1 Conceitos básicos

Uma **métrica** num conjunto M é uma função $d : M \times M \to R$, que associa a cada par ordenado $(x, y) \in M \times M$ um número real $d(x, y)$, chamado a **distância** de x a y, satisfazendo as seguintes condições para quaisquer $x, y, z \in M$:

$d_1)$ $d(x, x) = 0$;

$d_2)$ Se $x \neq y$ então $d(x, y) > 0$;

$d_3)$ $d(x, y) = d(y, x)$;

$d_4)$ $d(x, z) \leq d(x, y) + d(y, z)$.

Um **espaço métrico** é um par (M, d), onde M é um conjunto e d é uma métrica em M.

Exemplo 2.4.1 *A reta, ou seja, o conjunto $\mathbb{R}$, dos números reais é o exemplo mais importante de espaço métrico. A distância entre dois pontos $x, y \in \mathbb{R}$, é dada por $d(x, y) = |x - y|$.*

Seja a um ponto num espaço métrico, e $r > 0$ um número real. A **bola aberta** de centro a e raio r é o conjunto $B(a; r)$ dos pontos de M cuja distância ao ponto a é menor do que r. Ou seja,

$$B(a; r) = \{x \in M : d(a, x) < r\}.$$

Um ponto $a \in M$ chama-se um **ponto isolado** quando ele é uma bola aberta em M, ou seja, quando existe $r > 0$ tal que $B(a; r) = \{a\}$. Um espaço métrico chama-se **discreto** quando todo ponto de M é isolado.

Sejam $(M, d_M), (N, d_N)$ espaços métricos. Uma bijeção $f : M \to N$ chama-se uma **isometria** se $d_N(f(x), f(y)) = d_M(x, y)$ para quaisquer $x, y \in M$. A composta de duas isometrias e a inversa de uma isometria são ainda isometrias. Uma aplicação $f : M \to N$ é dita **contínua no ponto** $a \in M$ quando, para todo $\varepsilon > 0$ dado, é possível encontrar $\delta > 0$ tal que $d(f(x), f(y) < \varepsilon$ sempre que $d(a, x) < \delta$. Um **homeomorfismo** de M

sobre N é uma bijeção contínua $f : M \to N$ cuja inversa $f^{-1} : N \to M$ também é contínua. Neste caso, dizemos que M e N são **homeomorfos.**

Seja X um subconjunto de um espaço mérico M. Um ponto $a \in X$ diz-se um **ponto interior** a X quando é centro de uma bola aberta contida em X, ou seja, quando existe $r > 0$ tal que $d(a, x) < r \Rightarrow x \in X$. Chama-se **o interior** de X em M ao conjunto X° formado pelos pontos interiores a X. A **fronteira** de X em M é o conjunto ∂X, formado pelos pontos $b \in M$ tais que toda bola aberta de centro b contém pelo menos um ponto de X e um ponto de $M - X$.

O interior do conjunto $[0, 1)$ em $\mathbb{R}$ é o intervalo aberto $(0, 1)$ e sua fronteira consta dos pontos 0 e 1 apenas.

Um subconjunto A de um espaço métrico M diz-se **aberto** em M quando todos os seus pontos são interiores, isto é, $A^\circ = A$. Um subconjunto $F \subset M$ é dito ser **fechado** quando $M - F$ é aberto.

Seja X um subconjunto de um espaço métrico M. Uma **cobertura** de X é uma família $\mathcal{C} = (\mathcal{C}_\lambda)_{\lambda \in L}$ de subconjuntos de M tal que $X \subset \left(\cup_{\lambda \in L} C_\lambda\right)$. Um espaço métrico chama-se **compacto** quando toda cobertura aberta possui uma subcobertura finita. Um subconjunto X de um espaço métrico M chama-se um **subconjunto compacto** se, e somente se, de cada cobetura de X por conjuntos abertos for possível extrair uma subcobertura finita.

Exemplo 2.4.2 *Seja $\mathbb{Q}$ o conjunto dos números racionais. O interior de $\mathbb{Q}$ em $\mathbb{R}$ é vazio pois nenhum intervalo aberto pode ser formado apenas por números racionais. Por outro lado, a fronteira de $\mathbb{Q}$ é toda reta $\mathbb{R}$ porque qualquer intervalo aberto contém números racionais e irracionais.*

Observação 2.4.3 *Uma propriedade de que goza um espaço métrico M chama-se uma **propriedade topológica** quando todo espaço métrico homeomorfo a M também goza daquela propriedade. As propriedades topológicas se destinguem das propriedades **métricas** de M, que são aquelas que são preservadas por isometrias.*

Uma **seqüência** num conjunto M é uma aplicação $x : \mathbb{N} \to M$, definida no conjunto $\mathbb{N} = \{1, 2, ..., n, ...\}$. O valor que a seqüência x assume no número $n \in \mathbb{N}$ será indicado

por x_n, em vez de $x(n)$, e chamar-se-á o n-**ésimo termo** da seqüência. Usaremos a notação $(x_n)_{n\in\mathbb{N}}$, ou (x_n) para representar a seqüência. Uma seqüência no espaço métrico M é **limitada** quando o conjunto de seus pontos é limitado. Diz-se que um ponto $a \in M$ é **limite de uma seqüência** (x_n) quando, para todo número $\epsilon > 0$ dado arbitrariamente, é possível obter um $n_0 \in \mathbb{N}$ tal que $n > n_0 \Rightarrow d(x_n, a) < \epsilon$. Escreve-se $a = \lim x_n$. Quando existe $a = \lim x_n \in M$, diz-se que a seqüência de pontos $x_n \in M$ é **convergente.** Uma seqüência de números reais é dita **limitada** se existe $A > 0$ tal que $|x_n| < A$, $\forall\ n \in \mathbb{N}$ e, toda seqüência limitada de números reais possui uma subseqüência convergente, $(cf.\ [22])$.

Um ponto a diz-se **aderente** a um subconjunto X de um espaço métrico M quando $d(a, X) = 0$, ou seja, para cada $\varepsilon > 0$, podemos encontrar $x \in X$ tal que $d(a, x) < \varepsilon$. O **fecho** de X num espaço métrico M é o conjunto $\overline{X}$ dos pontos de M que são aderentes a X.

Exemplo 2.4.4 *Seja $M = \mathbb{R}$ e $X = [0, 1)$, então 1 é aderente a X e, $\overline{X} = [0, 1]$.*

Uma **partição** de um conjunto $X \subset M$, é uma coleção finita $P = \{I_0, I_1, ..., I_{j-1}\}$ de conjuntos abertos e disjuntos cuja união de seus fechos cobre X, isto é, $X = \overline{I_0} \cup \overline{I_1} \cup ... \cup \overline{I_{j-1}}$.

Um subconjunto $X \subset M$ diz-se **denso em** M quando $\overline{X} = M$. Por exemplo, $\mathbb{Q}$ é denso em $\mathbb{R}$.

Sejam M, N espaços métricos. Uma aplicação $f : M \to N$ é **aberta** quando para cada $A \subset M$ aberto, sua imagem $f(A)$ é um subconjunto aberto em N. Ou seja, quando f transforma abertos em abertos.

2.4.2 O Plano hiperbólico

Seja $\mathbb{C}$ o plano complexo. Usaremos as notações usuais para as partes real e imaginária de $z = x + iy \in \mathbb{C}$, $\mathrm{Re}\, z = x$ e $\mathrm{Im}\, z = y$. Consideremos o conjunto

$$\mathbb{H}^2 = \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$$

munido da métrica

$$d_{\mathbb{H}^2}(z,w) = \log \frac{|z-\overline{w}| + |z-w|}{|z-\overline{w}| - |z-w|}.$$

Com esta estrutura, o semi-espaço $\mathbb{H}^2$ é chamado de **espaço hiperbólico bi-dimensional** ou **plano hiperbólico** e a métrica acima definida de **métrica hiperbólica.** Esse modelo é conhecido como **modelo de Lobachevski**.

Para os nossos propósitos (Capítulo 5) iremos considerar um outro modelo da geometria hiperbólica plana, tomando o conjunto $\mathbb{D} = \{z \in \mathbb{C} : |z| \leq 1\}$ munido da métrica

$$d_{\mathbb{D}}(z,w) = \log \frac{|1-z\overline{w}| + |z-w|}{|1-z\overline{w}| - |z-w|}$$

conhecido como **disco de Poincaré**.

Uma curva $\gamma : [a,b] \to \mathbb{H}^2$ é dita ser uma **geodésica**, se γ minimiza distâncias entre pontos de seu traçado e, uma transformação em $\mathbb{H}^2$ é dita uma **isometria** se preserva distância hiperbólica em $\mathbb{H}^2$.

Um **grupo fuchsiano** é um subgrupo discreto Γ de $PSL(2,\mathbb{R})$.

Enunciaremos alguns resultados fundamentais cujas demonstrações encontram-se nas referências [8], [19], [17].

1. As geodésicas em $\mathbb{H}^2$ são semicirculos e semiretas ortogonais ao eixo real $\mathbb{R}$.

2. Qualquer transformação $T \in PSL(2,\mathbb{R})$ leva geodésicas de $\mathbb{H}^2$ em geodésicas de $\mathbb{H}^2$.

3. O grupo $PSL(2,\mathbb{R})$ está contido no grupo das isometrias de $\mathbb{H}^2$.

4. A área de um triangulo hiperbólico Δ, com ângulos α, β e γ é dada por $\mu(\Delta) = \pi - \alpha - \beta - \gamma$.

5. A área hiperbólica é invariante por isometrias, que são as trasnformações de $PSL(2,\mathbb{R})$.

6. A fronteira de $\mathbb{H}^2$ é o conjunto $\partial_\infty \mathbb{H}^2 = \{z \in \mathbb{C};\ \mathrm{Im}\, z = 0\} \cup \{\infty\}$.

Exemplo 2.4.5 *Considere* $\mathbb{H}^2 = \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\} \subset \mathbb{C}$. *A fronteria de* $\mathbb{H}^2$ *em* $\mathbb{C}$ *é* $\mathbb{R} \cup \{\infty\}$, *pois dados* $a \in \mathbb{R} \cup \{\infty\}$ *e* $r > 0$, *qualquer bola aberta de centro* a *e raio* r *contém pontos de* $\mathbb{H}^2$ *e de* $\mathbb{C} - \mathbb{H}^2 = \{z \in \mathbb{C} : \mathrm{Im}\, z \leq 0\}$.

Já vimos que um grupo fuchsiano Γ é um subgrupo discreto de $PSL(2,\mathbb{R})$. Vamos agora introduzir o conceito de grupo fuchsiano do primeiro tipo.

Sejam Γ um grupo fuchsiano e $z \in \mathbb{H}^2$.

(i) Chamamos de conjunto limite de Γ determinado por um ponto z ao conjunto

$$\Lambda_z = \{\zeta \in \mathbb{H}^2; \zeta \text{ é ponto de acumulação de } \Gamma(z)\}.$$

(ii) Chamamos de conjunto limite de Γ ao conjunto

$$\begin{aligned} \Lambda(\Gamma) &= \{\zeta \in \mathbb{H}^2; \zeta \text{ é ponto de acumulação de } \Gamma(z) \text{ para algum } z \in \mathbb{H}^2\} \\ &= \cup_{z\in\mathbb{H}^2}\Lambda_z. \end{aligned}$$

Diremos que um grupo fuchsiano Γ é do primeiro tipo se $\Lambda(\Gamma) = \partial_\infty\mathbb{H}^2$. O grupo Γ será dito do segundo tipo se $\Lambda(\Gamma) \neq \partial_\infty\mathbb{H}^2$.

Exemplo 2.4.6 *Seja* $\Gamma = PSL(2,\mathbb{Z})$, *então* $\Lambda(\Gamma) = \mathbb{R} \cup \{\infty\} = \partial_\infty\mathbb{H}^2$, *isto é,* $PSL(2,\mathbb{Z})$ *é um grupo fuchsiano do primeiro tipo.*

Um grupo fuchsiano Γ é dito **co-compacto** se o espaço quociente $\mathbb{H}^2/\Gamma$ for compacto.

A **característica de Eüler** de uma superfície Σ é o numero $\chi(\Sigma) = \mathcal{V} - \mathcal{A} + \mathcal{F}$, onde $\mathcal{V}$, $\mathcal{A}$ e $\mathcal{F}$ são os números de vértices, arestas e faces de Σ respectivamente.

Na Seção 3.1, abordaremos os modelos de sistemas de comunicações, típico e simplificado, e faremos algumas definições em teoria da informação as quais serão úteis no desenvolvimento da Seção 3.2.

Capítulo 3

Revisão de Técnicas de Codificação de Fontes Discretas

3.1 Sistemas de Comunicações: Modelos e Definições

3.1.1 Modelo simplificado do sistema de comunicações

Neste trabalho iremos abordar alguns casos de codificação de fontes. O diagrama de blocos de um sistema de comunicações típico é mostrado na Figura 3.1. Como nosso interesse maior é o processo de codificação de fontes, podemos simplificar o modelo ilustrado na Figura 3.1, considerando os blocos codificador de canal, modulador, canal, demodulador e decodificador de canal, como um único bloco chamado **canal discreto sem ruído.** Tal procedimento, está suportado no fato de que tanto o código corretor de erros a ser utilizado apresenta uma capacidade de correção bastante grande como o modulador emprega uma modulação apropriada para a transmissão de dados pelo canal com ruído. Desta forma, o modelo do sistema de comunicações típico passa então a ser o modelo conhecido como **sistema de comunicações simplificado**, mostrado na Figura 3.2.

A seguir definimos cada um dos blocos componentes de um sistema de comunicações típico.

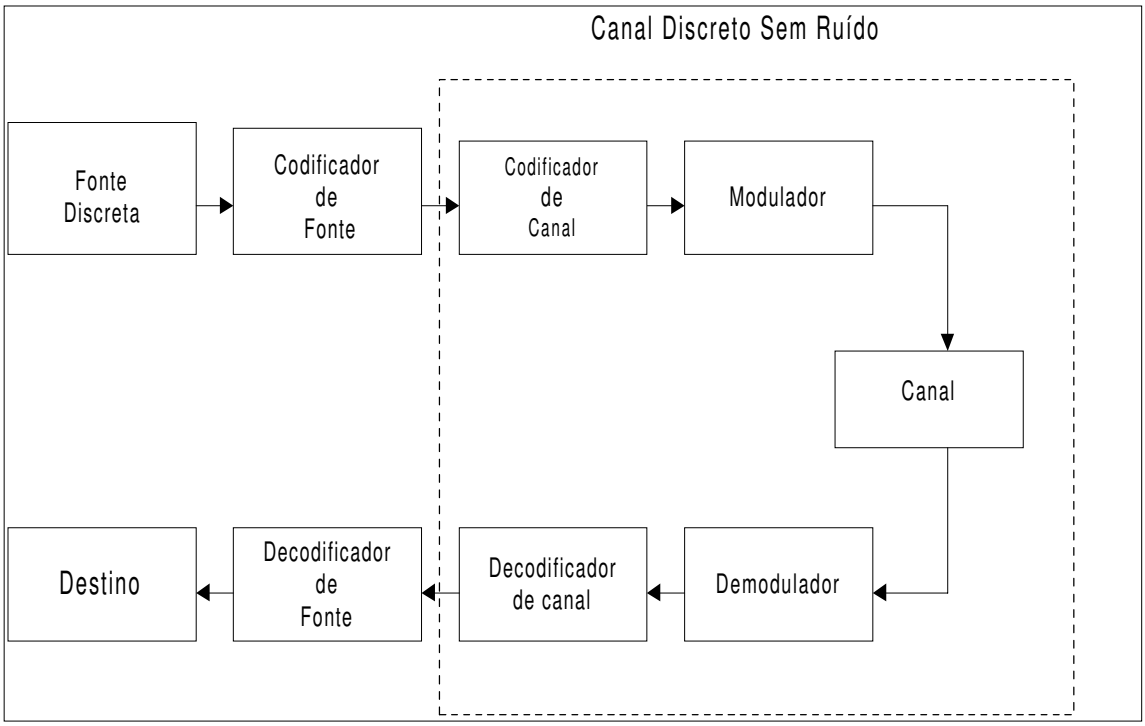

Figura 3.1: Diagrama de blocos de um sistema de comunicações.

Fonte: gerador da informação a ser transmitida (ser humano ou máquina). Pode gerar um sinal contínuo ou uma seqüência de símbolos discretos.

Codificador de Fonte: converte o sinal de saída da fonte em uma seqüência de dígitos binários. Se a fonte é contínua, uma conversão A/D (analógica para digital) é feita. É projetado de forma a minimizar o número de bits por unidade de tempo necessário para representar o sinal de saída da fonte e de forma que este sinal possa ser reconstituído sem ambiguidades.

Codificador de Canal: transforma a seqüência de saída do codificador de fonte em uma seqüência codificada (palavra-código) pela adição de redundância para combater os efeitos do ruído introduzido através do canal.

Modulador: converte a saída do codificador de canal em uma forma de onda adequada para transmissão através do canal.

Canal: meio físico que liga o transmissor ao receptor. O sinal modulado a ser conduzido pelo canal pode ou não sofrer ação de ruído.

Demodulador: a partir do sinal recebido do canal, estima o sinal transmitido e envia para o decodificador de canal a versão digital correspondente.

Decodificador de Canal: tenta corrigir possíveis erros (se houver) nos dígitos fornecidos pelo demodulador, produzindo uma estimativa dos dígitos da saída do codificador de fonte.

Decodificador de Fonte: transforma a seqüência estimada de saída do decodificador de canal em uma estimativa da saída da fonte. Se a fonte for contínua este processo pode envolver uma conversão D/A (digital para analógico).

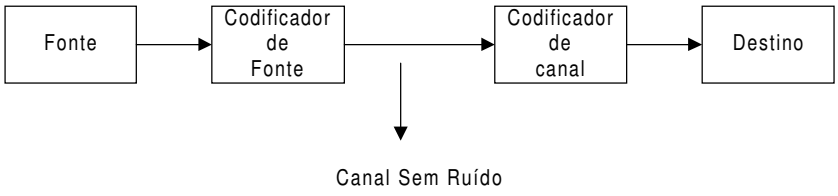

Figura 3.2: Diagrama de blocos de um sistema de comunicações simplificado.

De modo a alcançar os objetivos em decorrência da simplificação adotada, o processo de codificação de canal garantirá que a taxa de erros será adequadamente baixa. Esta hipótese está bem fundamentada através do Teorema de Codificação de canal (Teorema 3.1.4, pag. 31) proposto por Shannon (1948), isto é, através de uma codificação de canal adequada da informação, erros induzidos pelo canal podem ser reduzidos a qualquer nível desejado sem sacrificar a taxa de transmissão da informação.

Tipos de erros

Erros Aleatórios: os efeitos do ruído afetam cada símbolo transmitido de forma independente, isto é, se um bit[1] transmitido possui uma probabilidade p de ser transmitido incorretamente, então essa probabilidade independe dos demais bits transmitidos. Os canais que produzem este tipo de erro são chamados **canais sem memória**.

[1] bit = **bi**nary digi**t**

Erros em surto: o ruído não é independente de transmissão para transmissão. Os erros ocorrem de forma agrupada ou em surtos. Os canais que produzem este tipo de erro são chamados **canais com memória** ou **canais de erros em surtos.**

Observação 3.1.1 *Alguns canais produzem erros aleatórios e erros em surtos. Esses canais são chamados de* ***canais compostos****.*

3.1.2 Codificação de fontes: medidas características

O processo geral de codificação, seja de fonte ou de canal, tem como base um mapeamento um-a-um (injetivo), isto é, a conversão de uma dada seqüência ou conjunto de dígitos (ou símbolos) em uma outra seqüência ou conjunto de dígitos (ou símbolos). Este processo de conversão tem como objetivo realizar o tratamento dos símbolos ou seqüência de símbolos de modo a garantir ou alcançar a imunidade às interferências ou a confiabilidade desejada.

Um dos objetivos da teoria da informação é estabelecer um conjunto de medidas quantitativas da informação dada capacidade de determinados sistemas transmitirem, armazenarem e processarem a informação. Alguns dos problemas tratados aqui têm a ver com a proposta de métodos de codificação de fontes para serem utilizados em sistemas de comunicações de modo a torná-los mais eficientes segundo critérios de otimalidade convenientes aos propósitos do problema em consideração.

Nas próximas subseções apresentamos alguns conceitos da teoria da informação, tais como: informação, incerteza, entropia de uma fonte e o teorema da codificação da fonte, que servirão como referencial para os métodos de codificação de fontes. Todavia, essas medidas não são abordadas nos demais capítulos.

Conceitos de informação e incerteza

Quando se descreve um processo de seleção de um objeto entre vários outros aparecem naturalmente as noções de **informação** e **incerteza**.

Suponha que dispomos de um dispositivo (fonte) capaz de emitir três símbolos diferentes A, B e C. Sem perda de generalidades, suponha que o primeiro símbolo emitido pela fonte foi A. Enquanto se espera pelo aparecimento do próximo símbolo mantemos uma certa **incerteza** sobre qual símbolo aparecerá. Quando o símbolo aparece essa incerteza "desaparece" e podemos dizer que se "ganhou" uma certa **informação**.

Podemos deduzir que a informação corresponde a uma medida da incerteza. A questão que se coloca é: como que podemos medir essa incerteza?

A incerteza de uma fonte está assim associada à freqüência relativa de cada um dos símbolos diferentes por ela produzidos. Sendo M o número de símbolos diferentes produzidos pela fonte S, uma forma conveniente de medir a incerteza é supor que os M símbolos são igualmente prováveis, isto é, $p = \dfrac{1}{M}$ e com isso associar o valor $\log_2 p^{-1}$ a cada símbolo. Este valor é denominado a **quantidade de informação** atrelada ao correspondente símbolo. Com isso, a **incerteza** está relacionada com a freqüência relativa (probabilidade) enquanto que a **informação** está relacionada com o logaritmo na base 2 (medida em bits) do inverso da freqüência relativa.

Entropia de uma fonte

Considere uma fonte discreta com M símbolos onde cada símbolo ocorre com probabilidade p_i tal que:

$$\sum_{i=1}^{M} p_i = 1.$$

A **informação** associada ao aparecimento do i-ésimo símbolo será

$$u_i = -\log_2 p_i.$$

A incerteza da fonte pode ser definida como a **incerteza média** para o aparecimento de um dado símbolo. Suponha que uma seqüência emitida pela fonte consista de N símbolos, onde cada símbolo i ocorre N_i vezes. Toda vez que o símbolo i ocorre, a informação associada é u_i. Assim, a quantidade de informação associada às N_i ocorrências será $N_i u_i$. Com isso, a quantidade total de informação na seqüência com N

símbolos será $\sum_{i=1}^{M} N_i u_i$. Portanto, a **informação total média** será

$$\frac{1}{N}\sum_{i=1}^{M} N_i u_i = \sum_{i=1}^{M} \frac{N_i}{N} u_i$$

Para uma seqüência muito longa de símbolos, a razão $\frac{N_i}{N}$ (frequência relativa) converge para a probabilidade de ocorrência do símbolo i, p_i. Assim sendo, a informação média,ou **entropia** H, é dada por:

$$H = \sum_{i=1}^{M} p_i u_i,$$

ou seja,

$$H = -\sum_{i=1}^{M} p_i \log_2 p_i \; bits/símbolo.$$

Considerando agora que existem M símbolos todos equiprováveis, isto é, $p_i = \frac{1}{M}$, então a entropia é dada por

$$\begin{aligned} H &= -\sum_{i=1}^{M} \frac{1}{M} \log_2 \frac{1}{M} = -\left(\frac{1}{M} \log_2 \frac{1}{M}\right) M = \log_2 M \\ \therefore \quad & H = \log_2 M. \end{aligned}$$

Para M símbolos, o máximo de H é atingido quando os M símbolos são equiprováveis. Por outro lado, quando dos M símbolos apenas um deles aparece temos: $H = 0$.

A Figura 3.3, ilustra claramente a situação de uma fonte que emite o símbolo 0 com probabilidade p e o símbolo 1 com probabilidade $(1-p)$.

$$H(p) = \frac{-\{p \cdot \log(p) + (1-p) \cdot \log(1-p)\}}{\log 2}$$

A curva é simétrica, e atinge seu máximo no ponto $(p = 0.5)$. Aproxima-se de 0 quando a probabilidade de aparecimento de um dos pontos se aproxima de zero ou de 1.

Veremos a seguir alguns exemplos que serão utilizados a posteriori.

Exemplo 3.1.2 *A entropia da fonte que emite quatro símbolos* $\{A, B, C, D\}$ *com probabilidades* $\{0.5, 0.25, 0.125, 0.125\}$, *respectivamente, é:*

$$H = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - 2 \cdot \frac{1}{8} \cdot \log_2\left(\frac{1}{8}\right) = 1.75 \; bits.$$

Exemplo 3.1.3 *Analogamente, a entropia da fonte que emite os símbolos* $\{A, B, C\}$ *com probabilidades* $\{0.25, 0.25, 0.5\}$, *respectivamente, é* $H = 1.5$ *bits.*

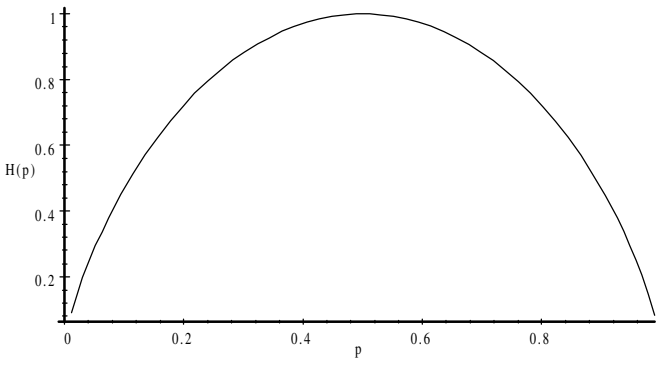

Figura 3.3: Função entropia binária

3.1.3 Teorema de codificação de fonte discreta sem memória

Considere o Exemplo 3.1.2 da seção anterior. Suponha que se pretende armazenar uma seqüência muito longa de M símbolos.

Pergunta: Como devemos codificar os símbolos A, B, C e D em bits de modo a minimizar a memória ocupada?

Se fizermos, $A = 00, B = 01, C = 10, D = 11$ iremos gastar 2 bits por símbolo.

Se fizermos, $A = 1, B = 01, C = 000, D = 001$ iremos gastar

$$1 \cdot 0.5 + 2 \cdot 0.25 + 2 \cdot 3 \cdot 0.125 = 1.75 \; bits/símbolo.$$

Este exemplo sugere que se use palavras-código curtas para representar os símbolos mais prováveis e palavras-código longas para representar símbolos menos prováveis. Um exemplo desta abordagem é o código de Morse[2].

A maior parte das fontes reais são redundantes. Para se conseguir uma transmissão eficiente a informação redundante pode ser removida dos símbolos a transmitir. Esta operação em que não se perde informação, chama-se **compactação de dados** ou **compressão de dados sem perdas.**

[2] Inventado por Samuel Morse em 1838, o Código Morse permitiu a transmissão de sinais elétricos que consistiam de pulsos curtos (pontos) ou longos (traços), como ilustrados na Tabela 3.1. A combinação desses pontos e traços formava caracteres, os quais, por sua vez formavam palavras que formavam frases. A aviação beneficiou-se desse meio de comunicação para a transmissão de mensagens relativas à navegação, controle e segurança de vôo.

A ._	J . _ _ _	S ...	2 .._ _
B _. . .	K _._	T _	3 ... _ _
C ._.	L . _ ..	U .._	4 .... _
D _..	M _ _	V ... _	5 .....
E .	N _.	W . _ _	6 _ ....
F .._.	O _ _ _	X _ .. _	7 _ _ ...
G _ _ .	P . _ _ .	Y _ . _ _	8 _ _ _ ..
H	Q _ _ ._	Z _ _ .. . ._ _._	9 _ _ _ _ .
I ..	R . _ .	1 ._ _ _ _	0 _ _ _ _ _

Tabela 3.1: Código Morse

O objetivo da mesma é representar de forma eficiente os dados gerados por uma fonte de informação. O dispositivo que realiza esse procedimento é chamado **codificador de fonte.**

Uma das formas de construir esse codificador é conhecer a estatística da fonte. Supondo que se usa um código de comprimento variável podemos:

- Associar palavras-código curtas aos símbolos mais frequentes;
  ou
- Associar palavras-códigos longas aos símbolos poucos frequentes.

Pretende-se que o **codificador** satisfaça às seguintes características:

- O alfabeto do código seja binário;
- O código produzido tenha **decodificação única** (isto é, dada uma seqüência codificada não existe ambiguidade com relação à seqüência decodificada.).

Considere o codificador de fonte, Figura 3.2, aplicado a uma fonte com K símbolos. Considere que o símbolo s_k corresponde à palavra-código binária b_k formada por l_k dígitos binários. O comprimento médio das palavras-código é dado por:

$$\overline{L} = \sum_{k=0}^{K-1} p_k l_k. \tag{3.1}$$

Sendo L_{min} o comprimento médio mínimo, a **eficiência** da codificação é dada por:

$$\eta = \frac{L_{min}}{\overline{L}}. \tag{3.2}$$

Note que a eficiência do codificador é máxima quando $L_{min} \to \overline{L}$.

Sendo S uma fonte discreta e sem memória, o **teorema da codificação da fonte** enunciado a seguir, diz que L_{min} é a entropia da fonte.

Teorema 3.1.4 (Shannon, 1948**)** *Dada uma variável aleatória S com incerteza $H(S)$, existe um código unicamente decodificável com alfabeto r-ário para a variável aleatória S cujo comprimento médio das palavras-código $\overline{L}$ satisfaz*

$$\frac{H(S)}{\log r} \leq \overline{L} \leq \frac{H(S)}{\log r} + 1.$$

Com base nesse teorema, podemos escrever

$$\eta = \frac{H(S)}{\overline{L}}.$$

Tendo como base o teorema da codificação da fonte, diversas técnicas de codificação de fontes discretas foram propostas. Algumas destas propostas tais como, o código de Huffman, de Elias e os métodos de Lempel-Ziv $LZ\,77$ e $LZ\,78$ serão descritas na próxima Seção.

3.2 Códigos de Huffman, Elias e Métodos de Lempel-Ziv

Na maioria das aplicações de codificação de fontes discretas, o codificador é alimentado por símbolos da saída da fonte, tais que os mesmos são caracterizados como sendo variáveis aleatórias identicamente distribuídas e estatisticamente independentes.

Neste tipo de aplicação os métodos de atribuição de palavras-código não levam em conside- ração a memória associada à fonte, isto é, a dependência estatística que existe entre os símbolos.

Técnicas de codificação de tais fontes são bem conhecidas. Dentre elas destacamos a técnica de **codificação de Huffman** e a técnica de **codificação Aritmética**.

No caso da codificação de Huffman, cada símbolo na entrada do codificador terá associado uma palavra-código. No caso da codificação aritmética cada símbolo implicará na divisão de um intervalo previamente fixado correspondente à representação da seqüência de símbolos.

Todavia, fontes discretas podem apresentar dependência entre os símbolos emitidos. Quando isso acontece, dizemos que tais fontes possuem memória. Para esta classe de fontes tanto os códigos de Huffman quanto o aritmético não fazem uso eficiente da memória ou da dependência entre os símbolos emitidos. Em geral, a modelagem utilizada, para o caso de fontes com memória, é a de cadeias de Markov ou as possíveis generalizações. Todavia, existem fontes com memória cuja estatística é desconhecida. Nestes casos, as técnicas de codificação universal são utilizadas com grande eficiência. Um estudo detalhado desta técnica pode ser encontrado em [18].

Um código é dito **universal** quando:

1. A codificação é realizada em blocos;

2. O desempenho é assintoticamente ótimo em algum sentido.

Segundo essa formulação os codificadores universais prestam-se muito bem para codificar fontes com características estatísticas desconhecidas de antemão e também mutáveis com o tempo. Isto porque a codificação realiza-se em blocos. Toda a informação relativa à caracterização estocástica da fonte tem que ser extraída do bloco que está sendo codificado, e não do conhecimento a priori ou do passado inteiro da fonte.

Os trabalhos [20] e [21] publicados por Jacob **Ziv** e Abraham **Lempel** no final da década de 70 revolucionaram a codificação universal. Hoje em dia quase todos os compressores de dados de uso genérico (compactadores de arquivos, compactadores de dados para armazenamento em disco, compressores de dados para comunicação via

Símbolo da Fonte	Probabilidade	Código I	Código II	Código III
s_0	0.5	0	0	0
s_1	0.25	1	10	01
s_2	0.125	00	110	011
s_3	0.125	11	111	0111

Tabela 3.2: Códigos livres de prefixo ou não.

modem) baseiam-se em um dos métodos propostos por Lempel e Ziv. Portanto, vale a pena conhecê-los para que aumentemos o rol de ferramentas úteis para novos algoritmos de comunicações.

3.2.1 Códigos livre de prefixo: códigos de Huffman

Considere uma fonte discreta sem memória composta pelos símbolos $\{s_1, s_2, ..., s_k\}$, onde cada símbolo ocorre com probabilidade $p_1, p_2, ..., p_k$, respectivamente. A cada símbolo s_k corresponde uma palavra-código C_k constituída de uma seqüência $(m_{k_1}, m_{k_2}, ..., m_{k_n})$, onde os m_{k_i} para $i = 1, ..., n$ pertencem ao alfabeto do código e, n é o **comprimento** da palavra-código.

Esta correspondência tem que satisfazer a propriedade de que o código seja unicamente decodificável. Uma classe importante de códigos satisfazendo tal propriedade é a classe dos **códigos livre de prefixo.**

Seja s_k o símbolo da fonte codificado como $(m_{k_1}, m_{k_2}, ..., m_{k_n})$. Qualquer seqüência $(m_{k_1}, m_{k_2}, ..., m_{k_i})$, com $i \leq n$, é dita ser um **prefixo** da palavra-código C_k. Um código diz-se **livre de prefixo** quando para quaisquer duas palavras-código C_i e C_j, C_i não é prefixo de C_j.

Observação 3.2.1 *Um código livre de prefixo é sempre unicamente decodificável.*

Exemplo 3.2.2 *Considere a Tabela* 3.2*:*

O Código I, não é um código livre de prefixo uma vez que a palavra-código 0 referente

ao símbolo da fonte s_0 é prefixo da palavra código 00, referente a codificação do símbolo s_2. O Código III também não é um código livre de prefixo mas o Código II é.

Uma classe de códigos livre de prefixo são os **códigos de Huffman.** O princípio geral em que eles se baseiam é atribuir a cada símbolo da fonte uma seqüência de símbolos do alfabeto do código cujo comprimento é o mais aproximado possível da quantidade de informação associada ao aparecimento deste símbolo.

O algoritmo de codificação proposto por Huffman é o seguinte:

1. S é o conjunto de símbolos da fonte;
2. Os símbolos da fonte são ordenados em ordem decrescente de probabilidade;
3. Aos dois símbolos menos prováveis é atribuído um 0 e um 1;
4. Esses dois símbolos são combinados de tal forma a dar origem ao conjunto S'. A probabilidade do novo símbolo é a soma das probabilidades dos símbolos combinados;
5. Se só resta um símbolo ir para 6. Caso contrário voltar para 2;
6. O código para cada um dos símbolos iniciais é obtido partindo do símbolo final e registrando a seqüência de bits encontrada até chegar a cada um dos símbolos iniciais.

Exemplo 3.2.3 *Considere uma fonte com símbolos* $\{s_0, s_1, s_2, s_3, s_4\}$ *caracterizados pelas probabilidades* $\{p_0, p_1, p_2, p_3, p_4\}$. *Suponha que se deseja codificar tais símbolos através de um alfabeto binário. A Figura 3.4 ilustra o procedimento usado na codificação de Huffman.*

Símbolo	s_0	s_1	s_2	s_3	s_4
Probabilidade	0.4	0.2	0.2	0.1	0.1

Obtendo-se o seguinte código: $C = \{00, 10, 11, 010, 011\}$.

O comprimento médio deste código é :

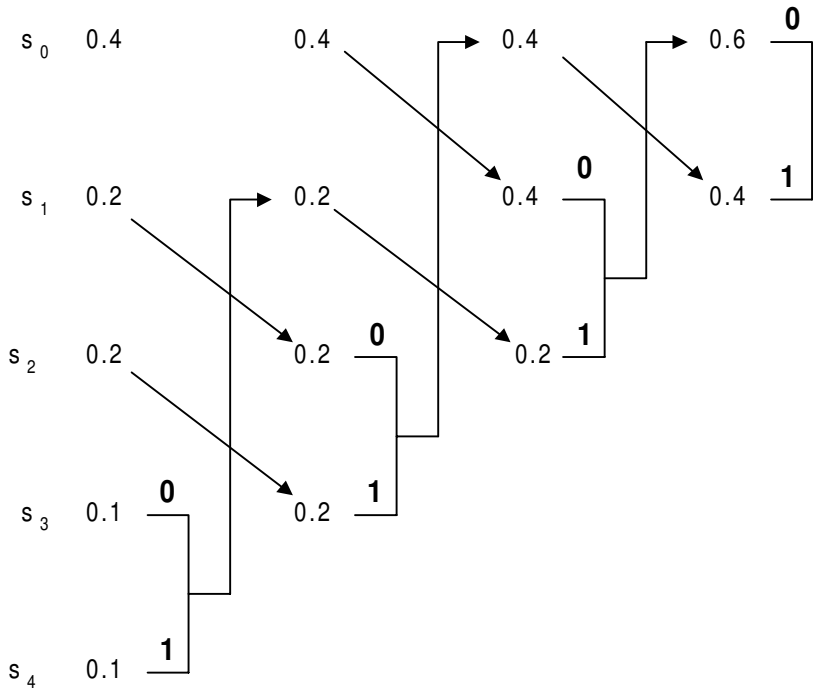

Figura 3.4: Exemplo de aplicação do algoritmo de codificação de Huffman.

$$\overline{L} = 0.4 \cdot 2 + 0.2 \cdot 2 + 0.2 \cdot 2 + 0.1 \cdot 3 + 0.1 \cdot 3 = 2.2 \; bits.$$

A entropia da fonte é:

$$H = 0.4 \cdot \log_2 \left(\frac{1}{0.4}\right) + 2 \cdot 0.2 \cdot \log_2 \left(\frac{1}{0.2}\right) + 2 \cdot 0.1 \cdot \log_2 \left(\frac{1}{0.1}\right) = 2.1219 \; bits.$$

Portanto, a eficiência deste código é:

$$\eta = \frac{H}{\overline{L}} = 96,45\%.$$

3.2.2 Codificação aritmética: os códigos de Elias

Do que foi visto anteriormente, a entropia de uma fonte é dada por $-\sum_j p_j \log p_j$, e o comprimento médio das palavras-código de um código de Huffman é dado por $\sum_j p_j l_j$. De modo a codificar a uma taxa igual a entropia da fonte, as palavras-código deverão ser escolhidas de tal forma que o seu comprimento satisfaça

$$l_j = -\log p_j. \tag{3.3}$$

Pelo fato de ser o comprimento de uma palavra-código um número inteiro, não será possível, com exceção de alguns casos especiais, satisfazer esta condição. Vimos que os códigos de Huffman são uma alternativa de realizar a quantização de $-\log p_j$ para valores inteiros.

O **código de Elias** é um código de comprimento variável usado para compactação de fontes com a seguinte característica: transforma uma seqüência muito longa de símbolos da saída da fonte em uma outra seqüência muito longa de símbolos do alfabeto do código. A operação de codificação possui uma propriedade *deslizante* no sentido de que para poucos símbolos da fonte entrando no codificador, poucos bits da palavra-código saem do codificador. O número de dígitos da palavra-código deixando o codificador por símbolo de entrada é variável, dependendo do padrão particular de símbolos produzidos pela fonte.

Suponha que desejamos compactar uma fonte binária cujo alfabeto é $\{a_1, a_2\}$ com probabilidades $\{0.7, 0.3\}$. A entropia desta fonte é:

$$H = 0.3 \cdot \log_2\left(\frac{1}{0.3}\right) + 0.7 \cdot \log_2\left(\frac{1}{0.7}\right) = 0.88129\ bits.$$

No instante zero, a fonte emite uma seqüência muito longa de símbolos de saída. Esta seqüência semi-infinita de símbolos de saída da fonte pode ser representada como uma seqüência (também semi-infinita) de bits. Por exemplo,

$$,011010111010... = \delta$$

onde os símbolos a_1 e a_2 são representados por 0 e 1, respectivamente. Note que, colocamos uma vírgula na frente do zero, com isto, tornamos esta seqüência uma expansão binária infinita de um número real. Dessa forma, podemos dizer que esta expansão representa uma expansão binária infinita no intervalo $[0, 1)$.

O processo para a geração da seqüência é o seguinte: divida o intervalo semi-aberto $[0, 1)$ em dois sub-intervalos $[0, 0.7)$ e $[0.7, 1)$. Use o primeiro símbolo da fonte, se for 0, para especificar que δ está no primeiro sub-intervalo, se for 1, para especificar que δ está no segundo sub-intervalo. Novamente, divida cada sub-intervalo na mesma proporção, isto é; $[0, 0.7)$ nos sub-intervalos $[0, 0.49)$ e $[0.49, 0.7)$ e, $[0.7, 1)$ nos sub-

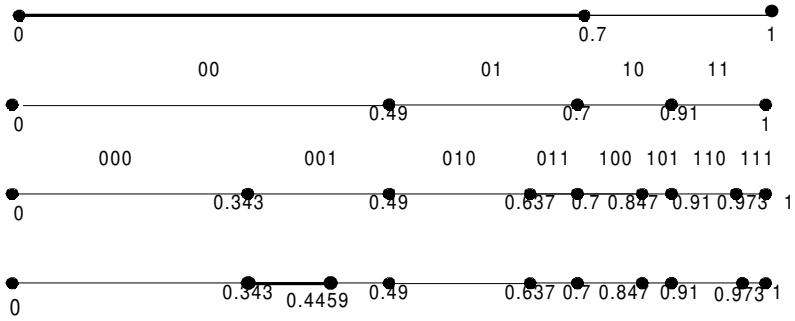

Figura 3.5: Codificação de Elias: saída da fonte.

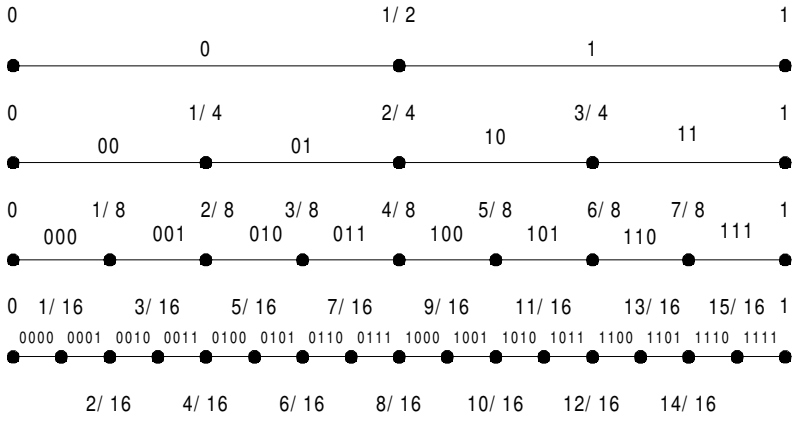

Figura 3.6: Codificação de Elias: saída do codificador da fonte.

intervalos $[0.7, 0.91)$ e $[0.91, 1)$. Use o segundo símbolo da fonte para especificar um destes sub-intervalos. Cada símbolo da fonte é usado nas subdivisões que seguem.

A regra geral de formação da seqüência é a seguinte: seja p a probabilidade do símbolo da fonte a_1 e $1-p$ a probabilidade do símbolo da fonte a_2. Após $n-1$ bits da fonte, o intervalo $[\alpha_{n-1}, \beta_{n-1})$ está especificado. Se o n-ésimo símbolo da fonte é a_1, o novo intervalo será especificado pelas extremidades

$$\begin{aligned} \alpha_n &= \alpha_{n-1} \\ \beta_n &= \alpha_{n-1} + p\left(\beta_{n-1} - \alpha_{n-1}\right). \end{aligned}$$

Se o n-ésimo símbolo da fonte é a_2, o novo intervalo será especificado pelas extremidades

$$\begin{aligned} \alpha_n &= \alpha_{n-1} + p\left(\beta_{n-1} - \alpha_{n-1}\right) \\ \beta_n &= \beta_{n-1}. \end{aligned}$$

Deste modo, o intervalo especificado por qualquer seqüência de símbolos da fonte terá um comprimento igual à probabilidade daquela seqüência. Note que, os intervalos ficam cada vez menores à medida que a seqüência cresce. No limite quando a seqüência tende a infinito, exatamente um ponto será especificado.

A palavra-código, por outro lado, é a representação binária do ponto δ, como pode ser visto na Figura 3.6. Tão logo tenha um número suficiente de símbolos da fonte, o codificador terá a condição de determinar se δ está no intervalo $[0, 0.5)$ ou no intervalo $[0.5, 1)$, e com isso o primeiro bit da palavra-código poderá ser enviado. Da mesma forma, o segundo bit da palavra-código poderá ser enviado a partir do conhecimeneto do intervalo em que δ está situado, isto é, $[0, 0.25)$ ou $[0.25, 0.5)$ ou em um dos sub-intervalos $[0.5, 0.75)$ ou $[0.75, 1)$. E o procedimento de codificação segue desta forma.

O decodificador começa a recuperar os símbolos da fonte somente após receber alguns bits da palavra-código. Por exemplo, se a seqüência binária da palavra-código inicia com 0110..., então o número real δ deve estar entre 0.375 e 0.4375; então os três primeiros símbolos da fonte serão $a_1a_1a_2$, pois o intervalo $[0.375, 0.4375]$ está contido no intervalo $[0343, 0.4459]$ na quarta linha da Figura 3.5. Se a seqüência binária da palavra-código inicia com 011..., então o número real δ deve estar entre 0.375 e 0.5; observe que na primeira linha da Figura 3.5 o intervalo $[\ 0.375, 0.5]$ está contido no intervalo $[0, 0.7]$, portanto, só temos informação do primeiro símbolo da fonte que é a_1.

Como conseqüência deste exemplo, vemos que existe uma certa imprecisão do código de Elias para seqüências de comprimento pequeno.

3.2.3 Codificação universal: métodos de Lempel e Ziv

Os métodos de Lempel e Ziv revolucionaram a codificação universal principalmente pela efeciência na utilização dos dados históricos das fontes de informação. Além disso, a simplicidade dos métodos fez com que suas implementações passassem a ser corriqueiras nas manipulações de dados do dia-a-dia.

O primeiro método, chamado $LZ1$ ou $LZ77$, foi publicado em 1977, [20]. Enquanto que o segundo, denominado $LZ2$ ou $LZ78$ foi publicado em 1978, [21]. Ambos se baseiam na idéia de armazenar o passado recente de uma fonte, e codificar cadeias de elementos através de identificadores da ocorrência anterior da cadeia nesse passado recente armazenado. As duas próximas subseções apresentam os algoritmos $LZ1$ e $LZ2$ de forma sucinta. Maiores detalhes podem ser encontrados em [20] e [21].

Método LZ1 ou LZ77

O método $LZ1$ armazena o passado recente numa janela de texto. Esta janela é composta dos elementos já gerados pela fonte e já codificados. Estes elementos são ordenados de acordo com sua ordem de geração pela fonte. Os elementos novos são armazenados numa área de trabalho antes de serem codificados. Tanto a janela de texto quanto a área de trabalho têm tamanhos máximos denotados, respectivamente, por t e s. A Figura 3.7 ilustra o processo de codificação.

A **janela de texto** pode ser entendida como um dicionário no qual as cadeias novas procurarão antigas ocorrências no histórico da fonte.

A área de trabalho vai sendo preenchida enquanto a cadeia de elementos que ela contém for encontrada em alguma posição na janela de texto, e enquanto o tamanho da cadeia for menor que s.

Tão logo uma destas condições seja violada, a cadeia da área de trabalho é codificada, seus elementos passam a fazer parte da janela de texto e a área de trabalho se renova para receber novos elementos. A codificação de uma cadeia é feita por intermédio de uma tripla contendo os seguintes campos:

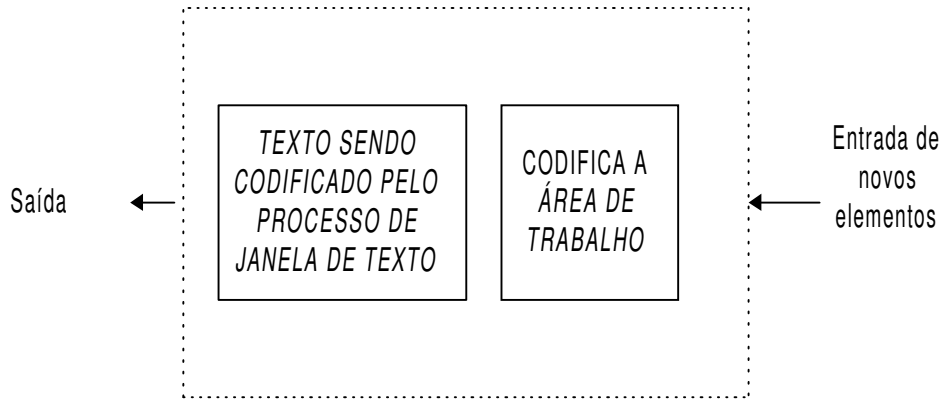

Figura 3.7: Codificação $LZ77$.

- Identificação do posicionamento da cadeia de elementos da janela de texto que coincide com a cadeia da área de trabalho;
- Número de elementos coincidentes nas cadeias da janela de texto e da área de trabalho;
- Último elemento inserido na área de trabalho.

Vamos descrever o processo de codificação com mais detalhes.

Inicialmente um elemento a gerado pela fonte é inserido na área de trabalho. Caso o elemento exista na janela de texto o método prossegue. Caso contrário, o elemento é codificado isoladamente, e é transferido da área de trabalho para a janela de texto (passa a fazer parte do histórico da fonte). Esta codificação é feita com a tripla $(0, 0, a)$.

Seja uma cadeia S de comprimento $0 < \ell(S) < s$ (ℓ é a função que atribui a uma cadeia de elementos o seu comprimento) que se encontra na área de trabalho. Pela definição do método esta cadeia existe na janela de texto. Os próximos passos do processo de codificação dependem do novo elemento gerado a pela fonte e do comprimento da cadeia $\ell(S)$:

- Caso a seja tal que $S \cup a$, isto é, a cadeia formada pela concatenação de S e a, ainda esteja na janela de texto e $\ell(S \cup a) < s$, então o elemento a é incorporado pela cadeia S e o processo continua;

- Caso $S \cup a$, ainda esteja na janela de texto e $\ell(S \cup a) = s$ então o elemento a é incorporado pela cadeia S e esta é codificada através da tripla

  $$(\text{endereço da cadeia igual a } S \text{ na janela de texto}, \ell(S), a).$$

  A janela de texto é deslocada de $\ell(S)$ elementos para que a cadeia seja a ela concatenada. A área de trabalho é esvaziada para que possa receber novos elementos;

- Caso $S \cup a$, não esteja mais na janela de texto, o novo elemento a é incorporado a S que é codificada através da tripla

  $$(\text{endereço da cadeia igual a } S \text{ na janela de texto}, \ell(S) - 1, a).$$

  A janela de texto é deslocada de $\ell(S)$ elementos para que a cadeia seja a ela concatenada. A área de trabalho fica novamente vazia após a codificação de S.

Desta forma percebe-se que a codificação é realizada de duas maneira distintas: para cadeias encontradas na janela de texto e para elementos isolados. Para que possa haver a decodificação é necessário existir um campo a mais nas palavras-código informando o tipo de código (código de cadeia ou de elemento isolado).

Os comprimentos das janelas de texto e da área de trabalho são dois parâmetros de projeto do codificador $LZ1$. Quanto maior for t, maior será o passado considerado da fonte. Portanto, maior será a chance de casamento de cadeias novas ao passado da fonte. Quanto maior for s, por sua vez, menor será a chance de se codificar uma cadeia por causa do enchimento da área de trabalho ($\ell(S) = s$, ou seja, sub-utilização do codificador).

Estes parâmetros, por outro lado, não podem ser aumentados indiscriminadamente, pois a complexidade computacional da busca de um "elemento" da área de trabalho na janela de textos cresce linearmente com s e na melhor das hipóteses, logaritmicamente com t. Além disso, os números de bits utilizados para representar a cadeia igual à codificada na janela de texto e o comprimento da cadeia dependem, respectivamente, de $\log_2 t$ e $\log_2 s$. Caso s e t cresçam de forma significativa, a própria codificação binária das duplas correspondentes às cadeias codificadas crescerá desnecessariamente.

Método LZ2 ou LZ78

Tal como o $LZ1$, o $LZ2$ procura casar cadeias novas (armazenadas na área de trabalho) às ocorrências anteriores do histórico da fonte. De forma similar, as cadeias são codificadas através de um identificador de ocorrência anterior no histórico.

A diferença entre os dois métodos reside na forma com que o histórico da fonte é armazenado. Ao invés de uma janela de texto, armazena-se um dicionário de frases ocorridas. Cada entrada do dicionário contém uma seqüência de elementos da frase e seu comprimento.

Um detalhe de implementação também contribui para diferenciar os dois métodos. O dicionário de frases é preenchido inicialmente com frases de comprimento 1, correspondendo a todos os elementos do alfabeto. Assim, não há necessidade de codificação especial para elementos isolados.

A área de trabalho vai sendo preenchida enquanto a cadeia de elementos que ela contém não for encontrada em alguma entrada do dicionário. No momento em que um novo elemento a, adicionado à área de trabalho, fizer com que a cadeia nela armazenada não seja mais encontrada no dicionário, esta cadeia é codificada através de duplas compostas por:

(posição da frase do dicionário coincidente com a cadeia anterior à adição do último elemento a, elemento a).

Exemplo 3.2.4 *Seqüência de símbolos a codificar segundo o algoritmo* $LZ78$. *Considere a seguinte seqüência* $ABBCBCABA$.

Processo de codificação

Posição	**Dicionário**	**Saída**
1	A	$(0, A)$
2	B	$(0, B)$
3	BC	$(2, C)$
4	BCA	$(3, A)$
5	BA	$(2, A)$

Uma seqüência compactada, isto é, após aplicado o método de Lempel-Ziv, consiste de uma lista de $c(n)$ pares. No exemplo em consideração, temos a seqüência de símbolos $ABBCBCABA$ separada em 5 frases, a saber, A, B, BC, BCA, BA cujos pares codificados têm a seguinte racionalidade: $(0, A)$ e $(0, B)$ significa que os símbolos A e B são frases de comprimento 1, isto é, são elementos do alfabeto. O número 0, significa que não existem frases anteriores concatenada nem com A nem com B. Para o par $(3, A)$, o símbolo A, significa o novo símbolo concatenado a frase BC e o número 3, a posição da frase anterior ao símbolo concatenado. Nos demais pares o procedimento é similar ao caso $(3, A)$.

A seguir, iremos introduzir resultados de entropia referente ao método de Lempel-Ziv, em analogia aos conceitos de entropia para os códigos livre de prefixo. Para tal, faz-se necessário as seguintes definições:

Uma **separação** S de uma seqüência binária $x_1 x_2 \cdots x_n$ é uma divisão dessa seqüência em frases, separadas por vírgulas. Uma **separação distinta** é uma separacão que não possui frases identicas.

Por exemplo, $0, 111, 1$ é uma separação distinta de 01111, mas $0, 11, 11$ é uma separação que não é distinta.

O método de Lempel-Ziv descrito fornece uma separação distinta da seqüência da fonte.

A combinação dos resultados enunciados a seguir, fornece um teorema análogo ao teorema de codificação de fonte (Teorema 3.1.4).

Lema 3.2.5 (Lempel-Ziv) *[6] O número de frases $c(n)$ numa separação distinta de uma seqüência binária $X_1, X_2, ..., X_n$ satisfaz*

$$c(n) \leq \frac{n}{(1 - \epsilon_n) \log n}$$

onde $\epsilon_n \to 0$ quando $n \to \infty$.

Um processo estacionário é **ergódico** na média com probabilidade 1 quando, no limite, a média estatística e a média temporal coincidem.

Teorema 3.2.6 *[6], Seja $\{X_n\}$ um processo estácionário ergódico, com entropia $H(X)$, e $c(n)$ o número de frases numa separação distinta de uma amostra de comprimento n deste processo. Então*

$$\limsup_{n \to \infty} \frac{c(n) \log c(n)}{n} \leq H(X)$$

com probabilidade 1.

Teorema 3.2.7 *[6], Sejam $\{X_i\}_{i \in \mathbb{Z}}$ um processo estocástico ergódico estacionário e $l(X_1, X_2, ..., X_n)$ os comprimentos das palavras-código associadas á $X_1, X_2, ..., X_n$. Então*

$$\limsup_{n \to \infty} \frac{1}{n} l(X_1, X_2, ..., X_n) \leq H(X) \tag{3.4}$$

com probabilidade 1.

Quando X é uma fonte ergódica estacionária, os resultados provenientes do Lema 3.2.5 e dos Teoremas 3.2.6 e 3.2.7, implicam que

$$\limsup_{n \to \infty} \frac{l(X_1, X_2, ..., X_n)}{n} = \limsup_{n \to \infty} \frac{c(n) \log c(n)}{n} \leq H(X). \tag{3.5}$$

Capítulo 4

Codificação de Geodésicas em Superfícies de Curvatura Negativa: Esfera 3-vezes Puncionada

Neste capítulo iremos considerar a codificação de geodésicas fechadas associadas a classes de matrizes hiperbólicas em $SL(2,\mathbb{Z})$ de duas formas diferentes. **O código geométrico**, com respeito a uma dada região fundamental, é obtido por uma construção universal para todos os grupos fuchsianos, enquanto **o código aritmético** dado por frações contínuas vem da teoria da redução de Gauss e, é específico para $SL(2,\mathbb{Z})$. Nosso objetivo é dar uma descrição completa de todas as geodésicas fechadas para as quais os dois códigos coincidem.

É importante observar que a codificação aritmética na Seção 3.2 é feita através de órbitas discretas enquanto que no presente capítulo, a mesma é feita por meio de órbitas contínuas, a saber; as órbitas das geodésicas no plano hiperbólico. As referências para este capítulo são: [2], [3], [11], [15], [16], [17], [23], [28], [29], [30] e [31].

4.1 Codificação de Geodésicas

A codificação de geodésicas no plano hiperbólico está associada ao elemento geodésico primitivo (geodésica que intersecta a região fundamental) de um dado A-ciclo. Os códigos geométricos estão associados à "geometria" da geodésica enquanto que os códigos aritméticos estão associados a "uma aritmética" (frações contínuas) dos pontos fixos de uma transformação linear fracionária (transformações de Möbius). A esta transformação está associada uma matriz hiperbólica do grupo $SL(2,\mathbb{Z})$.

4.1.1 Códigos geométricos

O grupo $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm I\} = \underset{A\in SL(2,\mathbb{R})}{\cup} \{-A, A\}$ age no semi-plano superior $\mathbb{H}^2 = \{z \in \mathbb{C} | \operatorname{Im} z > 0\}$ pelas **transformações de Möbius**:

$$T : z \to \frac{az+b}{cz+d} \tag{4.1}$$

as quais são isometrias (que preservam orientação) de $\mathbb{H}^2$ munida da métrica hiperbólica. Esta ação se estende para a fronteira de $\mathbb{H}^2$, $\mathbb{R} \cup \{\infty\}$, pela mesma fórmula. Os pontos fixos de $\gamma \in PSL(2,\mathbb{R})$ são encontrados resolvendo-se a equação: $z = \gamma(z) = \dfrac{az+b}{cz+d}$. Se a matriz

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

associada à transformação é hiperbólica, isto é, $|\operatorname{tr} A| > 2$, então γ tem dois pontos fixos em $\mathbb{R} \cup \{\infty\}$, a saber: as raízes do polinômio quadrático

$$Q_A(z) = cz^2 + (d-a)z - b = 0 \tag{4.2}$$

de descriminante

$$D = (a+d)^2 - 4 > 0.$$

Um ponto fixo, denotado por w, é dito **atrator**, isto é,

$$A'(w) = \frac{d}{dz}T(z)_{|z=w} = \frac{1}{(cw+d)^2} < 1$$

e o outro, denotado por u, é dito **repulsor**, $A'(u) > 1$. Uma geodésica em $\mathbb{H}^2$ de u para w, chamada **eixo** de γ e denotada por $\mathcal{C}(\gamma)$, fica invariante quando escolhemos a matriz $-A$ ao invés de A. Se γ pertence a um grupo fuchsiano Γ, isto é, um subgrupo discreto de $PSL(2,\mathbb{R})$, seu eixo torna-se uma **geodésica fechada** no espaço quociente $\mathbb{H}^2/\Gamma$, [17].

Observação 4.1.1 *[16], Se dois elementos γ_1 e γ_2 são conjugados em Γ, isto é, $\gamma_1 = \gamma\gamma_2\gamma^{-1}$ para algum $\gamma \in \Gamma$, então γ mapeia o eixo de γ_2 sobre o eixo de γ_1. Consequentemente, γ_1 e γ_2 representam a mesma geodésica fechada orientada em $\mathbb{H}^2/\Gamma$. Reciprocamente, toda geodésica fechada e orientada em $\mathbb{H}^2/\Gamma$ representa a classe de conjugação de uma transformação hiperbólica primitiva em Γ.*

No que segue todas as geodésicas fechadas são consideradas orientadas.

Seja Γ um grupo fuchsiano do primeiro tipo. O objetivo é codificar geodésicas em relação a uma dada região fundamental de Dirichlet $\mathcal{D}$ (Figura 4.1) para Γ, [28]. Tal região sempre tem um número par de lados identificados pelos geradores de Γ e denotados por $\{\gamma_i\}$. O procedimento a ser utilizado é o de rotular os lados de $\mathcal{D}$ pelos elementos do conjunto $\{\gamma_i\}$ como segue: se um lado s de $\mathcal{D}$ é identificado em $\mathcal{D}$ com o lado $\gamma_j(s)$, rotulamos o lado s por γ_j. Por rotulamento das imagens de um lado s sobre Γ pelo mesmo gerador γ_j obtemos o rotulamento de todo reticulado $\mathcal{N}$ de imagens de lados de $\mathcal{D}$, tal que, cada lado em $\mathcal{N}$ tenha dois rótulos correspondentes às imagens de $\mathcal{D}$ compartilhadas por este lado. Qualquer geodésica orientada em $\mathbb{H}^2$ pode ser codificada por uma seqüência de geradores de Γ, tal geodésica rotula os sucessivos lados do reticulado $\mathcal{N}$ por ela cortados. Em cada corte escolhemos o rótulo correspondente a imagem da geodésica de entrada. Sem perda de generalidades , descreve-se a seqüência de codificação de uma geodésica sempre supondo que ela não passa pelos vértices do reticulado $\mathcal{N}$ (*Corolário* 4.1.23) . Assim, podemos assumir que ela intersecta $\mathcal{D}$ e escolher um ponto inicial no interior de $\mathcal{D}$. Após sair de $\mathcal{D}$ esta geodésica entra numa outra região vizinha de $\mathcal{D}$, a saber: a imagem de $\mathcal{D}$ por um elemento do conjunto $\{\gamma_j\}$ através do lado rotulado, digamos, por γ_1, conforme está ilustrado na Figura 4.1. Portanto, esta imagem é $\gamma_1(\mathcal{D})$ e o primeiro símbolo do código é γ_1. Se ela entra na segunda imagem de

$\mathcal{D}$ através do lado rotulado por γ_2, a segunda imagem é $(\gamma_1\gamma_2\gamma_1^{-1})(\gamma_1(\mathcal{D})) = \gamma_1\gamma_2(\mathcal{D})$, e o segundo símbolo do código é γ_2, e assim por diante. Deste modo, obtem-se uma seqüência de todas as imagens de $\mathcal{D}$ cortadas por essa geodésica na direção de sua orientação, a saber: $\mathcal{D}, \gamma_1(\mathcal{D}), \gamma_1\gamma_2(\mathcal{D}), \ldots$.

[16], Se a geodésica é o eixo de um elemento hiperbólico primitivo $\gamma \in \Gamma$, tem-se

$$\gamma = \gamma_1\gamma_2\cdots\gamma_n \tag{4.3}$$

para algum n. Neste caso, a seqüência é periódica com período mínimo $[\gamma_1, \gamma_2, ..., \gamma_n]$.

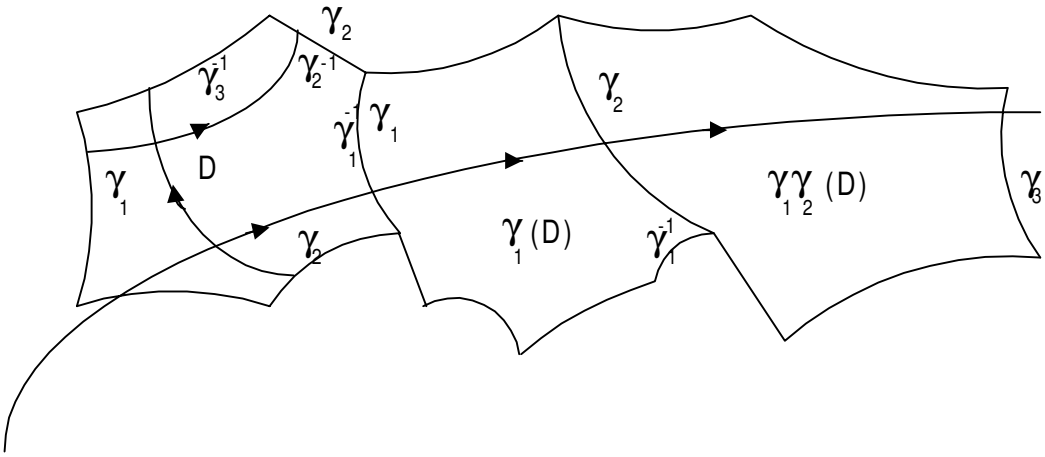

Figura 4.1: A região de Dirichlet e suas imagens.

Seguindo-se os segmentos geodésicos orientados do reticulado $\mathcal{N}$ entre todos os cortes consecutivos, chega-se novamente em $\mathcal{D}$, como pode ser visto na Figura 4.1. Desta maneira obtemos uma geodésica em $\mathcal{D}$.

Observação 4.1.2 *No contexto, a palavra levantamento significa o inverso da projeção.*

A seqüência de codificação descrita acima, pode também ser obtida tomando os inversos dos geradores de Γ, isto é, $\gamma^{-1} = \gamma_n^{-1}...\gamma_2^{-1}\gamma_1^{-1}$ conforme mostrado na Figura 4.1. O eixo de um elemento hiperbólico primitivo $\mathcal{C}(\gamma)$ torna-se uma geodésica fechada

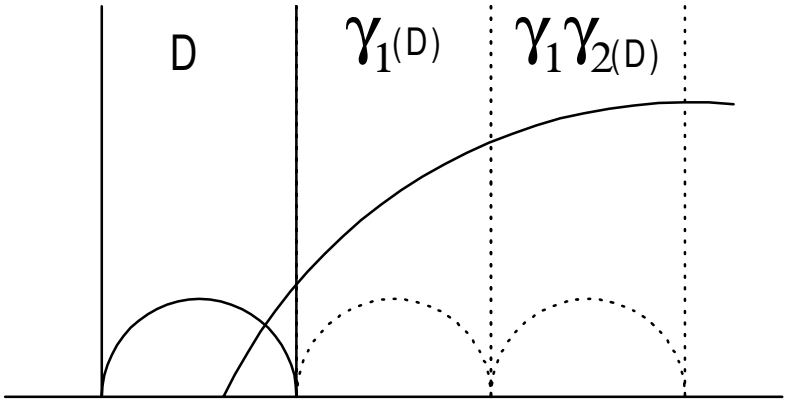

Figura 4.2: Levantamento da Figura 4.1 para o plano hiperbólico.

em $\mathcal{D}$. Podemos sempre supor que as geodésicas não passam pelos vértices de $\mathcal{D}$. Quando isto acontece, gera uma ambigüidade, que pode sempre ser removida, [16].

Se dois elementos são conjugados em Γ, suas geodésicas fechadas coincidem, e consequentemente, o período de suas seqüências de codificação diferem por um ciclo de permutação. Reciprocamente, se dois elementos primitivos hiperbólicos têm seus períodos em suas seqüências de codificação que diferem por um ciclo de permutação, então por (4.3) eles são conjugados em Γ, e consequentemente, suas geodésicas fechadas coincidem. Chamamos o período da seqüência de codificação de $\mathcal{C}(\gamma)$ em relação a uma região de Dirichlet, a menos de um ciclo de permutação, o **código de Morse** de uma geodésica fechada associada a uma classe de conjugação de γ, e denotamo-lo por $[\gamma] = [\gamma_1, \gamma_2, ..., \gamma_n]$. O eixo da transformação inversa, $\mathcal{C}(\gamma^{-1})$, é a mesma geodésica que $\mathcal{C}(\gamma)$, só que orientada na direção oposta. É claro que seu código de Morse é dado por $[\gamma^{-1}] = [\gamma_n^{-1}, \gamma_{n-1}^{-1}, ..., \gamma_1^{-1}]$. O código de Morse da matriz $A = \begin{pmatrix} 4 & -1 \\ 1 & 0 \end{pmatrix}$, denotado também por $[A]$, é o código de Morse da correspondente transformação de Möbius.

O código de Morse da matriz $A \in SL(2,\mathbb{Z})$ em relação à região fundamental

$$F = \{z \in \mathbb{H}^2; |z| \geq 1, |\mathrm{Re}\, z| \leq \frac{1}{2}\}$$

pode ser descrito por uma seqüência finita de números inteiros.

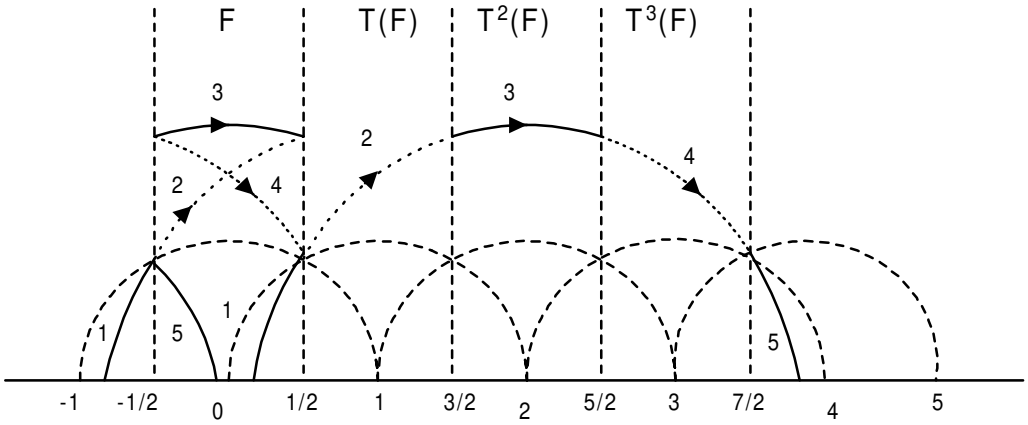

Figura 4.3: Codificação geométrica.

É conveniente considerar F como na Figura 4.4, isto é, como um quadrilátero ao invés de um triângulo, com o ponto i separando em dois lados, o arco do círculo unitário. O lado vertical esquerdo v_1 é identificado com o lado direito v_2 pela transformação $T(z) = z + 1$, e os dois arcos do circulo unitário, denotados por a_1 e a_2 respectivamente, são permutados pela transformacão $S(z) = -1/z$, a qual fixa o ponto i. De acordo com nossa convenção os lados v_1 e v_2 são rotulados por T e T^{-1} respectivamente, e ambos os lados a_1 e a_2 por S. Nesta particular situação, o código de Morse $[A]$ é uma seqüência contendo os símbolos; T, T^{-1} e S. Esta seqüência deve conter pelo menos um S, de tal forma que: um S não pode ser seguido de outro S, um T não pode ser seguido de T^{-1} e vice-versa. Portanto, $[A]$ é uma seqüência finita de blocos, definida a menos de um ciclo de permutação, consistindo de $T's$ e $T^{-1}'s$ que são separados por $S's$. Como ela não pode começar e terminar com S, vamos assumir que ela sempre termina com S. Para cada bloco de $T's$ associamos um número inteiro positivo igual ao seu comprimento, e a cada bloco de $T^{-1}'s$ associamos um inteiro negativo cujo valor absoluto (módulo) é igual ao seu comprimento. Assim, obtemos uma seqüência finita de inteiros $[n_1, n_2, ..., n_m]$, chamada o **código geométrico** de A, também denotado por $[A]$, o qual classifica geodésicas fechadas sobre a superfície modular $\mathbb{H}^2/PSL(2,\mathbb{Z})$. A seqüência de codificação de uma geodésica passando pelo vértice ρ de F no sentido

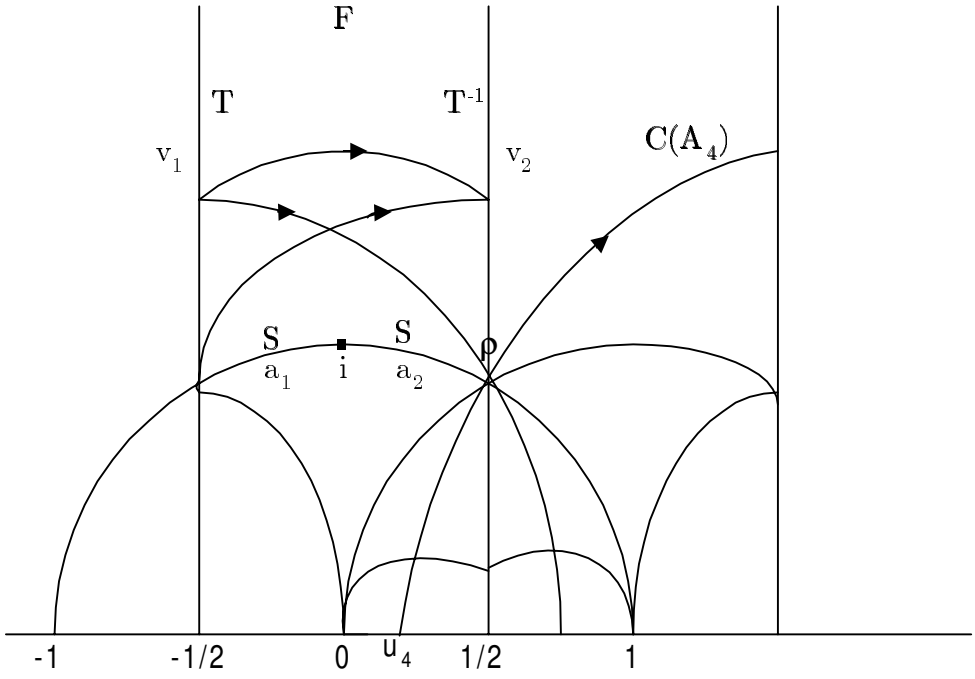

Figura 4.4: Confinamento da Figura 4.3 à região fundamental F.

horário é obtida pela convenção de que existe F em ambos os lados de v_2. O eixo de $A_4 = \begin{pmatrix} 4 & -1 \\ 1 & 0 \end{pmatrix}$, passando pelo vértice ρ, e correspondente à geodésica fechada $\mathcal{C}(A_4)$ em F, é mostrado na Figura 4.4. De acordo com a nossa convenção, o código de Morse é $[T, T, T, T, S]$, e o correspondente código geométrico é [4].

4.1.2 Códigos aritméticos

Nesta subseção faremos uso das frações contínuas na construção de códigos a serem utilizados na codificação de geodésicas sobre superfícies modulares (Proposição 4.1.11), usando a teoria da redução de Gauss. Tais códigos são seqüências de números inteiros $(n_1, n_2, ..., n_m)$, com $n_i \geq 2$, definidos a menos de um ciclo de permutação. Chama-

lo-emos de **código aritmético** da classe de conjugação de A, denotado por (A). O Teorema 4.1.21 fornece uma descrição completa de todas as geodésicas fechadas para as quais os códigos aritméticos e geométricos coincidem.

Suponha agora que temos um conjunto munido de uma relação de equivalência. Em termos gerais, a **teoria da redução** ($cf.$ [15]) é um algoritmo para achar representantes canônicos em cada classe de equivalência. Tais representantes são chamados **elementos reduzidos.** Cada classe de equivalência contém um conjunto canônico (finito e não-vazio) de elementos reduzidos, os quais formam um ciclo. Seguindo o algoritmo de redução (Proposição 4.1.14), podemos passar de um dado elemento dentro da classe de equivalência para o elemento reduzido, por um número finito de etapas. Aplicando o mesmo algoritmo para um elemento reduzido, obtemos todos os elementos reduzidos do ciclo.

De modo a tornar claro esses conceitos, considere o algoritmo de redução para grupos fuchsianos co-compactos [15]. Os elementos cujo eixo intersecta uma dada região fundamental $\mathcal{D}$ são os elementos reduzidos. O ciclo de seus elementos reduzidos Γ-conjugados são todos os elementos reduzidos com o mesmo código de Morse, e a intersecção de suas geodésicas com a região $\mathcal{D}$ compreende todas as geodésicas fechadas associadas a esta particular classe de conjugação.

Uma matriz hiperbólica em $SL(2,\mathbb{Z})$ é chamada **reduzida** se atrai e repele pontos fixos w e u respectivamente, satisfazendo:

$$w > 1, \quad 0 < u < 1 \tag{4.4}$$

O conjunto de todas as matrizes reduzidas conjugadas de uma dada matriz $A \in SL(2,\mathbb{Z})$ é chamado de A-**ciclo**.

Veremos adiante (Proposição 4.1.14) que para qualquer matriz $A \in SL(2,\mathbb{Z})$ o A-**ciclo** consistindo de todas as matrizes B tais que $(B) = (A)$ é finito e não-vazio. Decorre do Corolário 4.1.17, que a m-upla de inteiros $(n_1, n_2, ..., n_m)$, com $n_i \geq 2$, é um código aritmético de algum A-**ciclo** explícito.

Seja $\mathcal{D}$ uma região fundamental em $\mathbb{H}^2$. Uma matriz $A \in SL(2,\mathbb{Z})$ é chamada $\mathcal{D}$-**reduzida** se é reduzida e seu eixo intersecta $\mathcal{D}$.

No que segue $\mathcal{D}$ será a região fundamental F contida na Figura 4.4 ou uma de suas imagens por elementos de $SL(2,\mathbb{Z})$.

Uma matriz hiperbólica em $SL(2,\mathbb{Z})$ é chamada **totalmente F-reduzida** se todas as matrizes no A-ciclo são F-reduzidas.

Observação 4.1.3 *[16], pag.5, A direção do eixo de uma transformação hiperbólica não é invariante por conjugação, isto é, algumas geodésicas em F podem ser orientadas no sentido horário, e outras no sentido anti-horário. Este fato pode ser claramente observado na Figura* 4.5. *Um caso quando* ***todos*** *os segmentos geodésicos são orientados no sentido horário é especial e será objeto do Teorema* 4.1.20, *Subseção* 4.1.3.

Teoria das frações contínuas

Sejam $n_0, n_1, n_2, ...$ uma seqüência de números inteiros satisfazendo $n_1, n_2, ... \geq 2$. Vamos denotar por $(n_0, n_1, ..., n_s)$ a fração contínua finita

$$(n_0, n_1, ..., n_s) = n_0 - \cfrac{1}{n_1 - \cfrac{1}{n_2 - \cfrac{1}{\ddots - \cfrac{1}{n_s}}}} \tag{4.5}$$

e por $(n_0, n_1, n_2, ...)$ o limite $\lim_{s\to\infty}(n_0, n_1, ..., n_s)$. Reciprocamente, (pela unicidade do limite), todo número α tem uma única expansão em frações contínuas $\alpha = (n_0, n_1, n_2, ...)$ com $n_i \in \mathbb{Z}$ e $n_1, n_2, ... \geq 2$, fazendo;

$n_0 = [[\alpha]] + 1$ e, indutivamente;

$n_i = [[\alpha_i]] + 1$, onde $\alpha_{i+1} = \dfrac{1}{n_i - \alpha_i}$ e, $[[x]]$ denota a **parte inteira** de x.

Isto estabelece uma correspondência biunívoca entre o conjunto de números reais α e o conjunto das seqüências infinitas $(n_0, n_1, n_2, ...)$ com $n_i \in \mathbb{Z}$, e $n_1, n_2, ... \geq 2$.

Essa correspondência satisfaz as seguintes propriedades, [16]:

(1.1) α é racional se, e somente se, a partir de algum ponto todos os $n_i's$ são iguais a 2;

(1.2) α é uma irracionalidade quadrática, isto é, uma raiz de um polinômio quadrático com coeficientes em $\mathbb{Z}$ se, e somente se, sua expansão em frações contínuas é eventualmente periódica, isto é, de um certo ponto em diante se repete periodicamente, ou seja, $\alpha = (n_0, n_1, ...n_k, \overline{n_{k+1}, n_{k+2}, ..., n_{k+m}})$, onde a barra sobre $n_{k+1}, n_{k+2}, ..., n_{k+m}$ significa que esses números se repetem periodicamente, e seu período mínimo é m;

(1.3) α tem uma expansão em frações contínuas periódica pura se, e somente se, $\alpha > 1$, e $0 < \alpha' < 1$, onde α' é a raiz conjugada de α, isto é, a outra raiz do mesmo polinômio quadrático do qual α é raiz;

(1.4) Se $\alpha = (\overline{n_1, n_2, ..., n_m})$, então $1/\alpha' = (\overline{n_m, n_{m-1}, ..., n_1})$.

Observação 4.1.4 *A propriedade* (1.3) *é deveras importante. Ela dá uma definição equivalente para uma matriz reduzida: Uma matriz é reduzida se, e somente se, seu ponto fixo atrator tem uma expansão em frações contínuas periódica pura.*

A proposição a seguir é crucial para os nossos propósitos. Ela mostra que o período de uma expansão em frações contínuas é um sistema completo de $SL(2,\mathbb{Z})$-invariantes.

Proposição 4.1.5 *[16], Duas irracionalidades quadráticas são obtidas uma da outra pela aplicação de uma transformação de $SL(2,\mathbb{Z})$ se, e somente se, os períodos de suas expansões em frações contínuas são permutações cíclicas uma da outra.*

Exemplos

Exemplo 4.1.6 *Considere a matriz* $A = \begin{pmatrix} 14 & -3 \\ 5 & -1 \end{pmatrix} \in SL(2,\mathbb{Z})$. *Resolvendo-se a equação* $T(z) = \dfrac{14z-3}{5z-1} = z$, *encontramos os pontos fixos da transformação* T *associada a matriz* A, *a saber;* $z = \dfrac{15 \pm \sqrt{165}}{10}$. *Tome* $w = \dfrac{15+\sqrt{165}}{10} \backsimeq 2,7845 > 1$ *e* $0 < u = \dfrac{15-\sqrt{165}}{10} \backsimeq 0,2154 < 1$. *Representado como fração contínua* $w = (\overline{3,5})$, *consequentemente o código aritmético é* $(A) = (3,5)$.

Exemplo 4.1.7 *Vamos calcular* $w = (\overline{3,5})$ *usando o teoria das frações contínuas. Faça*

$$\alpha = \frac{15 + \sqrt{165}}{10} \backsimeq 2,7845,$$

daí

$$n_0 = [[\alpha_0]] + 1 = [[2,7845]] + 1 = 3.$$

$n_1 = [[\alpha_1]] + 1$, *onde*

$$\alpha_1 = \frac{1}{n_0 - \alpha_0} = 4.6404 \Rightarrow n_1 = [[4,6404]] + 1 = 5.$$

$n_2 = [[\alpha_2]] + 1$, *onde*

$$\alpha_2 = \frac{1}{n_1 - \alpha_1} = 2,7800 \Longrightarrow n_2 = [[2,7800]] + 1 = 3.$$

$n_3 = [[\alpha_3]] + 1$, *onde,*

$$\alpha_3 = \frac{1}{n_2 - \alpha_2} = 4,5454 \Longrightarrow n_3 = [[4,5454]] + 1 = 5.$$

Continuando assim, obtemos que

$$w = (3,5,3,5,\ldots) = (\overline{3,5}).$$

Exemplo 4.1.8 *A geodésica fechada em* F *correspondente à classe de conjugação de* A, *mostrada na Figura* 4.5, *consiste de* 10 *segmentos geodésicos orientados, a saber: os segmentos numerados de* 1-7 *são orientados no sentido horário, enquanto que os segmentos numerados de* 8-10 *são orientados no sentido anti-horário. Seguindo a geodésica fechada em* F *obtemos seu código geométrico* $[A] = [6,-2]$, *que é diferente de seu código aritmético* $(A) = (8,2)$. *Note que* A *não é totalmente* F*-reduzida (*F *não é reduzida) pois seu ponto fixo repulsor associado à* A *é* $u = 4 - 2\sqrt{3} \backsimeq 0,5359 > \frac{1}{2}$.

Exemplo 4.1.9 *As matrizes* $A = \begin{pmatrix} 14 & -3 \\ 5 & -1 \end{pmatrix}$ *e* $B = \begin{pmatrix} 65 & -17 \\ 23 & -6 \end{pmatrix}$ *são reduzidas, mas não são totalmente* F*-reduzidas. Seus códigos aritméticos e geométricos são respectivamente;*

$(A) = (3,5)$, mas $[A] = [-1,1,-1,3]$ e,

$(B) = (3,6,4)$, mas $[B] = [-1,1,-1,5,3]$.

O Teorema 4.1.22, Subseção 4.1.3 fornece uma condição necessária e suficiente para que uma matriz seja totalmente F-reduzida em função de seu código aritmético.

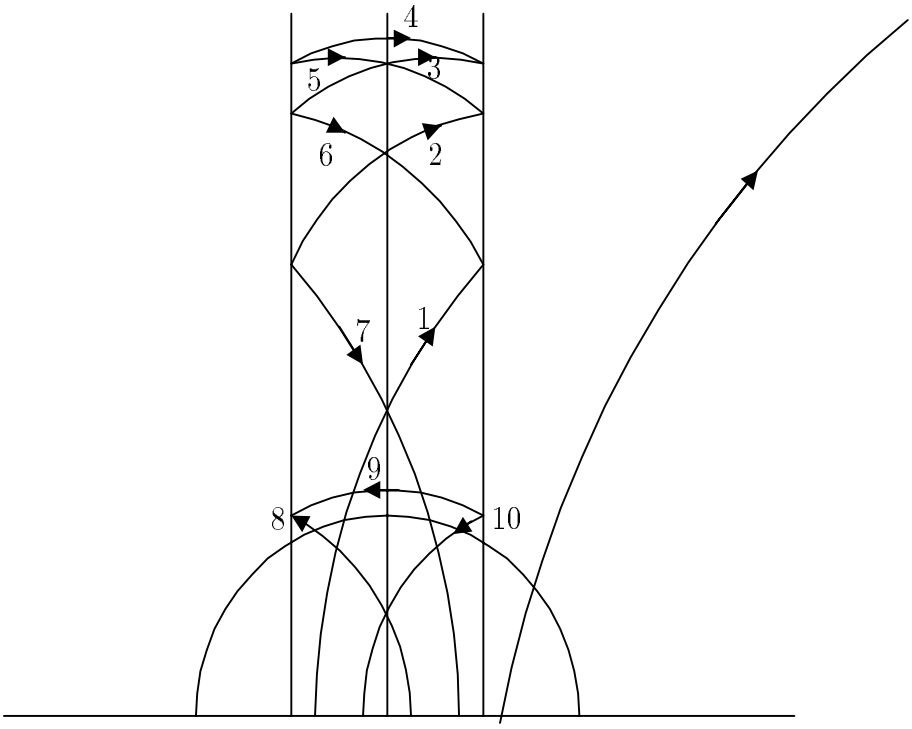

Figura 4.5: Segmentos geodésicos

Teoria da redução de Gauss e fatoração de matrizes reduzidas

Para entender a conexão entre os códigos aritméticos e geométricos apresentamos aqui uma teoria equivalente à teoria da redução de Gauss para matrizes em $SL(2,\mathbb{Z})$. Um dos resultados importantes é o estabelecido no seguinte:

Lema 4.1.10 [*16*], *Sejam* Γ *um grupo fuchsiano e* $\gamma_1, \gamma_2 \in \Gamma$ *elementos hiperbólicos tendo um ponto fixo comum. Então os segundos pontos fixos também coincidem, e consequentemente, eles têm o mesmo eixo, e ambos são potência de uma matriz primitiva com o mesmo eixo.*

Demonstração. Por conjugação podemos supor que γ_1 e γ_2 fixam ∞, assim

$$\gamma_1(z) = \lambda z \ (\lambda > 1), \quad \text{e} \quad \gamma_2(z) = \mu z + k \ (\mu \neq 1, k \neq 0).$$

Então

$$\gamma_1^{-n}\gamma_2\gamma_1^n(z) = \lambda^{-n}(\mu(\lambda^n z) + k) = \mu z + \lambda^{-n}k.$$

Portanto, $\left\|\gamma_1^{-n}\gamma_2\gamma_1^n\right\|$ é limitada quando $n \to \infty$, e consequentemente a seqüência $\{\gamma_1^{-n}\gamma_2\gamma_1^n\}$ contém uma subseqüência convergente de termos distintos, o que é uma contradição pois Γ é discreto. ■

A proposição a seguir fornece uma caracterização das matrizes conjugadas em $SL(2,\mathbb{Z})$.

Proposição 4.1.11 *[16], Duas matrizes hiperbólicas $A, B \in SL(2,\mathbb{Z})$ com o mesmo traço são conjugadas em $SL(2,\mathbb{Z})$ se, e somente se, seus pontos fixos atratores (repulsores) têm períodos nas suas expansões em frações contínuas que são permutações cíclicas um do outro.*

Demonstração. Sejam w_A e w_B os pontos fixos atratores de A e B, respectivamente, tais que os períodos das correspondentes expansões em frações contínuas diferem por um ciclo de permutação. Então, pela Proposição 4.1.5, existe $C \in SL(2,\mathbb{Z})$ tal que $w_A = Cw_B$. Daí, $CBC^{-1}(w_A) = CBC^{-1}(Cw_B) = CB(w_B) = C(w_B) = w_A = A(w_A)$, isto é, CBC e A têm o mesmo ponto fixo w_A, e pelo Lema 4.1.10 eles têm o mesmo traço, ou seja, $CBC^{-1} = A$ ou $CBC^{-1} = -A$. Como tanto w_A e w_B são atratores e w_A é atrator para CBC^{-1} e A, resulta que $CBC^{-1} = A$. Reciprocamente, suponhamos que $A, B \in SL(2,\mathbb{Z})$ são conjugadas. Então seus pontos fixos atratores w_A e w_B são obtidos um do outro por aplicação de uma matriz $C \in SL(2,\mathbb{Z})$. Daí, pela Proposição 4.1.5 os períodos de suas expansões em frações contínuas de w_A e w_B diferem por um ciclo de permutação. ■

Assim, em adição aos códigos geométricos descritos na Subseção 4.1.1, obtemos dois outros invariantes de uma geodésica fechada, também definidos a menos de um ciclo de permutação: os períodos de suas expansões em frações contínuas de seus pontos fixos atrator e repulsor. O primeiro invariante, o qual chamamos de **código aritmético** de A e denotamos por (A), coincide com o código geométrico para uma ampla classe de geodésicas fechadas identificadas no Teorema 4.1.20. O segundo é o código aritmético associado a matriz (A^{-1}) que corresponde à mesma geodésica associada a matriz A com orientação oposta. Note que, se A é reduzida, a relação (1.4) da Subseção 4.1.1 é válida.

A seguir enunciamos alguns resultados, que dizem respeito a matrizes reduzidas, necessários nas demostrações dos Teoremas 4.1.20 e 4.1.22. O lema a seguir descreve as matrizes reduzidas em função de suas entradas.

Lema 4.1.12 *[16] Seja* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *uma matriz em* $SL(2,\mathbb{Z})$ *com* $|a+d| > 2$. *A é reduzida se, e somente se,* $c > 0$, $a + b - c - d > 0$ *e* $b < 0$. *Para uma matriz reduzida A com ponto fixo atrator* $w > 1$ *também temos:*

(i) $a > 0,\ c + d > 0,\ d \leq 0$;

(ii) $\dfrac{a+b}{c+d} < w < \dfrac{a}{c}$.

Corolário 4.1.13 *[16], Seja A uma matriz reduzida com ponto fixo atrator w. Então, considerada como uma função sobre* $\mathbb{R} \cup \{\infty\}$, $A(x)$ *é crescente para* $x > 1$. *Para qualquer número fixado* $x > 1$ *a seqüência* $\{A^n(x)\}$ *(convergindo para w) é decrescente se* $x > w$ *e crescente se* $1 < x < w$.

Proposição 4.1.14 (Algoritmo de redução) *[16], [15], Existe um número finito de matrizes em* $SL(2,\mathbb{Z})$ *com um dado traço t,* $|t| > 2$. *Uma matriz hiperbólica em* $SL(2,\mathbb{Z})$ *com um dado traço t pode sempre ser reduzida por um número finito de conjugações. Aplicada a uma matriz reduzida A, esta conjugação dá uma outra matriz reduzida. Assim, qualquer matriz reduzida conjugada à matriz A, pode ser obtida de A por conjugação. Desse modo, o conjunto das matrizes reduzidas pode sempre ser decomposto em ciclos disjuntos de matrizes conjugadas.*

Proposição 4.1.15 *[16], Sejam* $n_1, n_2, ..., n_m \geq 2$ *números inteiros, e* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *uma matriz hiperbólica* ($|\mathrm{Tr}\ A| > 2$) *primitiva com traço positivo e ponto fixo atrator*

$$w = (\overline{n_1, n_2, ..., n_m}). \tag{4.6}$$

Então A pode ser representada na forma

$$A = T^{n_1} S T^{n_2} S \cdots T^{n_m} S,$$

onde $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ e $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Enunciamos agora um resultado que é usado em [16], na demonstração da Proposição 4.1.15, bem como será utilizado na demostração do Teorema 4.1.22.

Lema 4.1.16 *[16], Sejam $n \geq 2$ um número inteiro, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z})$ e w um número real satisfazendo as seguintes condições:*

(i) $c, c+d > 0$;

(*ii*) $n-1 < w < n$;

(*iii*) $\dfrac{a+b}{c+d} < w < \dfrac{a}{c}$.

Então, $n-1 < \dfrac{a}{c} < n$.

Demonstração. A desigualdade da esquerda segue de (*ii*) e (*iii*) por transitividade. Para provar a desigualdade do lado direito, suponha por absurdo que $n < \dfrac{a}{c}$. Então por (*iii*) e (*ii*),

$$\frac{a+b}{c+d} < n < \frac{a}{c}$$

daí, $ac + bc < nc(c+d) < ac + ad \Rightarrow bc < nc(c+d) < ad - bc = 1$. O que é uma contradição. Pois $n \geq 2$, $c > 0$ e $(c+d) > 0$. ■

Corolário 4.1.17 *[16], Sejam $n_1, n_2, ..., n_m$ números inteiros maiores ou iguais a 2, $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ e $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Então*

$$A = T^{n_1} S T^{n_2} S \cdots T^{n_m} S,$$

é uma matriz hiperbólica com traço positivo, reduzida, e $(A) = (n_1, n_2, ..., n_m)$.

Observação 4.1.18 *Os códigos aritméticos são denotados por $A = (n_1, n_2, ..., n_m)$ com $n_i \geq 2$, $\forall i = 1, ..., m$. Portanto, uma condição necessária para que um código geométrico seja igual a um código aritmético é que ele não tenha números negativos em sua representação. Por exemplo, o código geométrico $[6, -2]$ jamais coincidirá com um código aritmético.*

Lema 4.1.19 *Uma matriz $A \in SL(2,\mathbb{Z})$ é F-reduzida se $y(1/2) \geq \frac{\sqrt{3}}{2}$, onde $y = y(x)$ é definido pela equação*

$$c(x^2 + y^2) + (d - a)x - b = 0.$$

Observe-se que este lema caracteriza todas geodésicas F-reduzidas. Além disso, não sendo F-reduzida, uma tal geodésica não será totalmente F-reduzida, então, pelo Teorema 4.1.20 a seguir, o código aritmético associado a mesma, não coincidirá com qualquer código geométrico dado. A Figura 4.3 ilustra este fato.

4.1.3 Condições de equivalência dos códigos geométricos e aritméticos

Vamos relembrar que uma matriz A é totalmente F-reduzida se todas as matrizes no A-ciclo são F-reduzidas. O teorema a seguir fornece a condição necessária e suficiente para que uma matriz hiperbólica $A \in SL(2,\mathbb{Z})$ seja totalmente F-reduzida em função de seu código aritmético.

Teorema 4.1.20 *[16], Seja $A \in SL(2,\mathbb{Z})$ uma matriz hiperbólica. As seguintes afirmações são equivalentes:*

(1) A é totalmente F-reduzida.

(2) Os códigos aritméticos e geométricos de A coincidem.

(3) Todos os segmentos geodésicos (componentes da geodésica) em F correspondentes a classe de conjugação de A são orientados no sentido horário.

Demonstração. (1) $\Rightarrow$ (2) Suponha que A é totalmente F-reduzida. Então os eixos de todas as matrizes do A-ciclo devem entrar em F através do lado a_2. Como elas são orientadas no sentido horário, devem sair de F pelo lado v_2. Seja $(\overline{n_1, n_2, ..., n_m})$ a expansão em frações contínuas do ponto fixo atrator w de A. Então o eixo de $T^{-1}AT$ entra em F pelo lado v_1 e sai pelo lado v_2. Suponha que os eixos de $T^{-i}AT^i$ têm a

mesma propriedade para todo $1 \leq i < k$, e que o eixo de $T^{-k}AT^{k}$ entra em F por v_1 e deixa F por a_1 ou a_2. Queremos mostrar que $k = n_1$. Se $k < n_1$, então o eixo $T^{-n_1}AT^{n_1}$ não intersecta F, o que contradiz o fato de que $S^{-1}T^{-n_1}AT^{n_1}S$ é F-reduzida. Por outro lado, o eixo de $T^{-n_1+1}AT^{n_1+1}$ não intersecta F , consequentemente, $k = n_1$ e o eixo de $T^{-n_1}AT^{n_1}$ deve sair de F por a_1 e o primeiro símbolo de $[A]$ é n_1. (O eixo de uma matriz reduzida $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ não pode passar pelo ponto i. Pois se isso acontece, tomando $z = i$ na equação

$$c\,|z|^2 + (d-a)\,x - b = 0 \tag{4.7}$$

obtem-se $c = b$, o que contradiz o fato de que A é reduzida conforme Lema 4.1.12, pois no referido lema $c > 0$ e $b < 0$. O argumento ainda é válido se o eixo de A passa pelo vértice ρ de F. Neste caso, o eixo de $T^{-n_1}AT^{n_1}$ passa pelo vértice $\rho - 1$, como mostra a Figura 4.6. Mostraremos a posteriori (Corolário 4.1.23) que existem somente três matrizes primitivas F-reduzidas cujos eixos passam pelo vértice ρ). Continuando esse processo e, contando o número de $T's$ entre os $S's$ obtemos o código geométrico de A na forma $[n_1, n_2, ..., n_m]$, o qual é igual ao código aritmético.

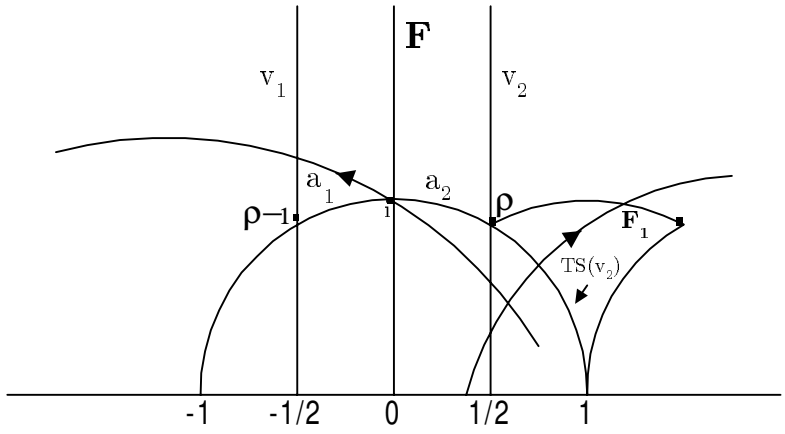

Figura 4.6: Ilustração do Teorema 4.1.20.

(1) $\Rightarrow$ (3) Note que a interseçcão dos eixos das matrizes no A-ciclo com F constitui somente uma parte da geodésica fechada em F representando a classe de conjugação de A, ou seja, a parte "reduzida", as partes restantes representam o eixo dos *shifts* das

geodésicas reduzidas. Vimos que todos os segmentos geodésicos obtidos neste processo, tornam-se orientados no sentido horário.

(2) $\Rightarrow$ (1) Suponha agora que A não é totalmente F-reduzida. Então existe pelo menos uma matriz no A-ciclo que não é F-reduzida. Sem perda de generalidades, vamos supor que A é reduzida, mas não é F-reduzida. Então ela é F_1-reduzida, onde $F_1 = TS(F)$ e seu eixo deve entrar em F_1 por $TS(v_2)$ como pode ser visto na Figura 4.6. Consequentemente, o eixo de sua conjugada $T^{-1}S^{-1}T^{-1}ATST$ intersecta F na direção anti-horária e sai no lado v_1 de F. Isto significa que o código geométrico de A contém no mínimo um T^{-1}, em outras palavras, no mínimo um dos números inteiros componentes é negativo. Desta forma, ele não pode coincidir com qualquer código aritmético.

(3) $\Rightarrow$ (1) Se A não é totalmente F-reduzida, pelo argumento anterior um dos segmentos geodésicos componentes da geodésica associada à classe de conjugação de A, é orientada no sentido anti-horário, o que é uma contradição. ■

O teorema a seguir, é o resultado principal do presente trabalho. O mesmo estabelece condições de existência de um código aritmético coincidente com um código geométrico dado. Foi baseado na Observação 4.1.18, que notamos a existência de tais condições que geraram o presente resultado.

Teorema 4.1.21 *Sejam $n_1, n_2, ..., n_m$ números inteiros maiores ou iguais a 2 e $[A] = [n_1, n_2, ..., n_m]$ um código geométrico. Então existe um código aritmético (A) tal que o código geométrico $[A]$ coincide com (A).*

Demonstração. Sendo $n_1, n_2, ..., n_m$ números inteiros maiores ou iguais a 2, consideremos $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ e $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Então pelo Corolário 4.1.17, a matriz $A = T^{n_1}ST^{n_2}S...T^{n_m}S$ *é uma matriz hiperbólica com traço positivo, reduzida, e* $(A) = (n_1, n_2, ..., n_m)$, onde $T^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$. ■

O teorema a seguir identifica todas as matrizes reduzidas em função de seus códigos aritméticos.

Teorema 4.1.22 [16], *Seja $(A) = (n_1, n_2, ..., n_m)$. A é totalmente F-reduzida se, e*

somente se, $\frac{1}{n_i} + \frac{1}{n_{i+1}} \leq \frac{1}{2}$ *para todo* $i \pmod m$, *isto é, o código aritmético* (A) *não pode conter* 2, *nem os seguintes pares:* $\{3,3\}, \{3,4\}, \{4,3\}, \{3,5\}$ *e* $\{5,3\}$.

Demonstração. A demostração é feita por absurdo. Seja $(A) = (n_1, n_2, ..., n_m)$. Primeiramente, mostra-se que se A contém uma das combinações proibidas, então A não é totalmente F-reduzida.

Para provar a recíproca do teorema, mostra-se que qualquer código não contendo pares proibidos está associado a uma matriz totalmente F-reduzida. ■

Corolário 4.1.23 [16], *As únicas matrizes primitivas totalmente* F*-reduzidas cujos eixos passam pelo vértice* ρ *da região fundamental* F *são* A_4, A_3A_6 *e* A_6A_3, *onde* $A_i = T^iS = \begin{pmatrix} i & -1 \\ 1 & 0 \end{pmatrix}$.

4.2 Superfície Modular e Frações Contínuas

Nesta seção iremos considerar a codificação de geodésicas no plano hiperbólico. Todavia, existem duas diferenças em relação à codificação apresentada na Seção 4.1. A primeira é que nesta seção, a tesselação do semi-plano superior $\mathbb{H}^2 = \{z \in \mathbb{C}; \operatorname{Im} z > 0\}$ é feita por triângulos cujos vértices estão na fronteira do disco de Poincaré, mais pecisamente, um vértice no infinito e os outros dois em $\mathbb{R} \cup \{\infty\}$, enquanto que na Seção 4.1, o plano foi tesselado por triângulos com um vértice no infinito, e os outros dois em $\mathbb{H}^2$. A segunda diferença está relacionada com os elementos da seqüência de codificação. Enquanto na Seção 4.1, os elementos da seqüência são potências associadas às transformações T e T^{-1} separados por S, desde que sejam satisfeitas as hipóteses do Teorema 4.1.22, nesta seção tais elementos são segmentos geodésicos rotulados por D (direita) e E (esquerda), também com a restrição que tais segmentos não coincidam com *cusps*. É notório portanto que, tal como às técnicas de codificação de fontes discretas, existem certas restrições as codificações ora apresentadas. A Figura 4.7 ilustra o comentário acima.

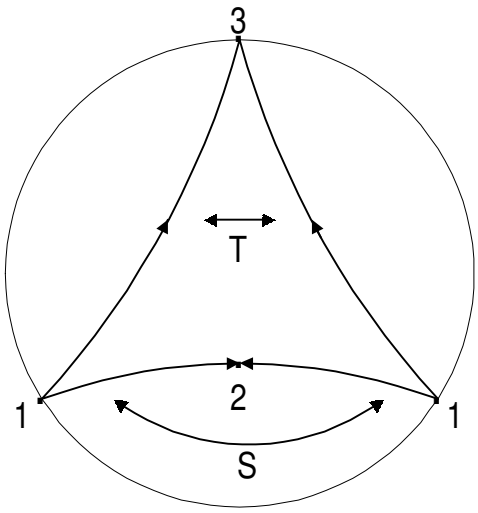

Figura 4.7: Triângulo ideal.

4.2.1 Geodésicas em superfícies modulares e frações contínuas

O objetivo é estabelecer uma conexão entre geodésicas em uma superfície modular M, (resultante do quociente entre o grupo que atua transitivamente no semi-plano superior e o correspondente subgrupo) e as frações contínuas. Esta conexão foi observada por Artin [3] quando, aplicando naturalmente frações contínuas deduziu a existência de uma geodésica densa sobre M.

A idéia que a seqüência na qual uma geodésica γ corta certas linhas fixadas sobre M (ou sobre seu levantamento para $\mathbb{H}^2$) foi objeto de muitas pesquisas. Na Seção 4.1, usamos as linhas da tesselação canônica $\mathcal{T}$ de $\mathbb{H}^2$ por cópias da região fundamental $F = \{z \in \mathbb{C}; |\mathrm{Re}\, z| \leq \frac{1}{2}, |z| \geq 1\}$, onde achamos uma relação entre a seqüência cortante de γ e as expansões em frações contínuas de pontos terminais de adequados levantamentos de γ. Outras duas possíveis soluções, nenhuma delas totalmente simples, são encontradas em [2] e [29].

Nesta seção substituimos $\mathcal{T}$ pela **tesselação de Farey** $\mathbb{F}$, descrita na Subseção 4.2.2 e mostrada na Figura 4.8. Tal tesselação, é uma tesselação de $\mathbb{H}^2$ por triângulos ideais. O conjunto de vértices é precisamente $\mathbb{Q} \cup \{\infty\}$.

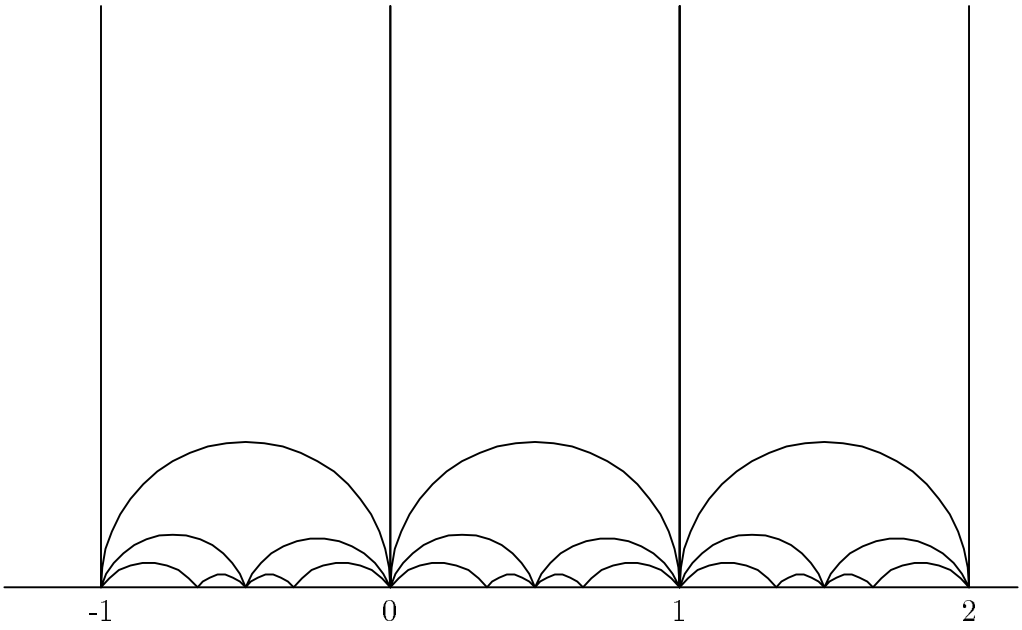

Figura 4.8: A tesselação de Farey.

Observação 4.2.1 *Se os números racionais irredutíveis* $\frac{p}{q}$ *e* $\frac{p'}{q'}$ *são vértices terminais de um lado de um triângulo em* $\mathbb{F}$*, então* $pq\prime + p\prime q = \pm 1$*. Os lados de* $\mathbb{F}$ *são as imagens do eixo imaginário por* $SL(2,\mathbb{Z})$*.*

Uma geodésica orientada em $\mathbb{H}^2$ é dividida em segmentos quando ela corta transversalmente os triângulos que compõem $\mathbb{F}$. No corte de um tal triângulo Δ, um segmento s corta dois lados que se encontram num vértice na fronteira $(\mathbb{R} \cup \{\infty\})$. Esses segmentos (orientados) são rotulados por D ou E de acordo com a posição que o vértice se situa à direita ou à esquerda do lado s como é mostrado na Figura 4.9.

Este rotulamento é invariante pela ação do grupo $SL(2,\mathbb{Z})$. Consequentemente, para qualquer geodésica $\overline{\gamma}$ em M, podemos associar uma seqüência cortante formada de $D's$ e $E's$, a saber; $\ldots D^{n_0}E^{n_1}D^{n_2}\ldots$, onde $n_i \in \mathbb{N}$. Se $x \in \overline{\gamma}$ encontra-se no fim de um segmento rotulado por D, iremos mostrar (*Teorema* 4.2.4) que existe um único levantamento γ de $\overline{\gamma}$ tal que o levantamento ξ de x é da forma $\xi = iy$, e os pontos terminais positivo e negativo de γ sobre $\mathbb{R}$ são dados por:

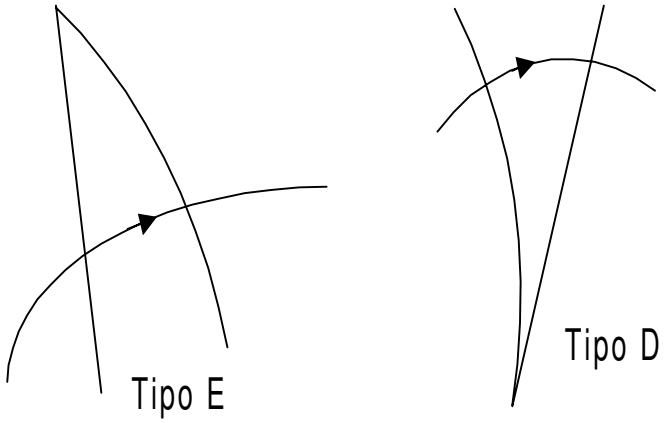

Figura 4.9: Segmentos geodésicos dos tipos E e D.

$$\gamma_\infty = n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + ...}}, \qquad \gamma_{-\infty} = \cfrac{-1}{n_0 + \cfrac{1}{n_{-1} + \cfrac{1}{n_{-2} + ...}}},$$

respectivamente.

Como na Seção 4.1, consideremos $\mathbb{H}^2$ munido da métrica hiperbólica. Geodésicas em $\mathbb{H}^2$ são semi-círculos centrados em pontos de $\mathbb{R}$ ou retas verticais. O grupo modular $SL(2,\mathbb{Z})$ age sobre $\mathbb{H}^2$ por isometrias, a saber; pelas transformações de Möbius

$$T : z \to \frac{(az+b)}{(cz+d)}.$$

O semi-plano $\mathbb{H}^2$ é projetado na superfície modular $M = \mathbb{H}^2/SL(2,\mathbb{Z})$ pela função $\pi : \mathbb{H}^2 \to \mathbb{H}^2/SL(2,\mathbb{Z})$ chamada **função projeção**. Geodésicas sobre M são exatamente as imagens das geodésicas de $\mathbb{H}^2$ pela transformação π. No que segue todas as geodésicas são consideradas orientadas.

Iremos usar a mesma notação $(n_1, n_2, ...)$ para representar a fração contínua $n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + ...}}$ e também $[[x]]$ para a **parte inteira** de x, $x > 0$.

4.2.2 Seqüências cortantes e frações contínuas

A Tesselação de Farey

A região fundamental $|\mathrm{Re}\, z| \leq \frac{1}{2}, |z| \geq 1$ para $SL(2,\mathbb{Z})$ é dividida ao meio pelo eixo imaginário. Movendo-se o lado esquerdo (denotado por A) da região na Figura 4.10 usando a transformação $z \to z+1$ e colando-se as duas metades, obtem-se assim, uma nova região fundamental para $SL(2,\mathbb{Z})$, a saber: o quadrilátero superior ilustrado na Figura 4.11.

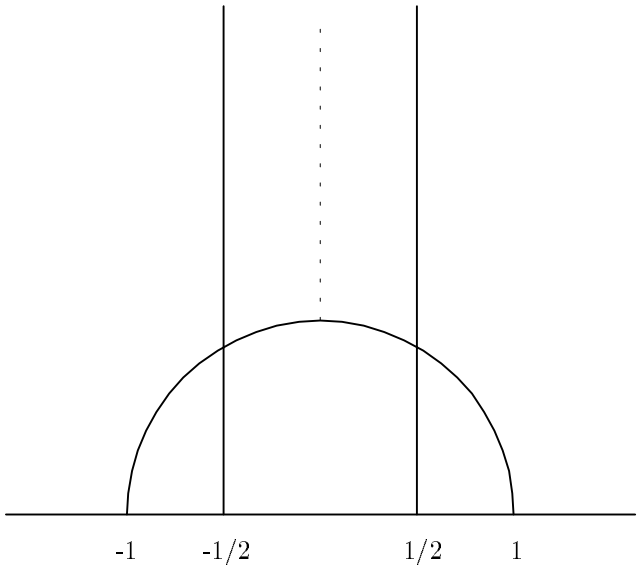

Figura 4.10: Região fundamental consistindo das arestas v_1, a_1, a_2 e v_2.

Se S denota a matriz $\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \in SL(2,\mathbb{Z})$, então as três imagens deste quadrilátero por I, S e S^2 dá exatamente o triângulo ideal Δ cujos vértices estão em $0, 1$, e ∞, como pode ser observado na Figura 4.12. As imagens de Δ por $SL(2,\mathbb{Z})$ tessela $\mathbb{H}^2$. Essa tesselação é chamada **tesselação de Farey**, e será denotada por $\mathbb{F}$. Note que $\mathbb{F}$ pode ser considerada como as imagens do eixo imaginário por $SL(2,\mathbb{Z})$.

Não é difícil ver que as imagens de $\{0, 1, \infty\}$ por $SL(2,\mathbb{Z})$ são exatamente os pontos

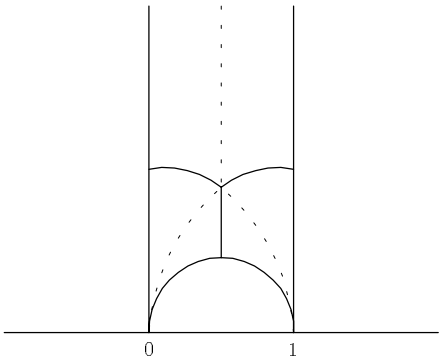

Figura 4.11: Região fundamental transladada.

$\mathbb{Q} \cup \{\infty\}$ e, que dois pontos $\frac{p}{q}$ e $\frac{p'}{q'}$ são ligados por um lado de $\mathbb{F}$ se, e somente se,

$$\begin{pmatrix} p & p' \\ q & q' \end{pmatrix} \in SL(2, \mathbb{Z}).$$

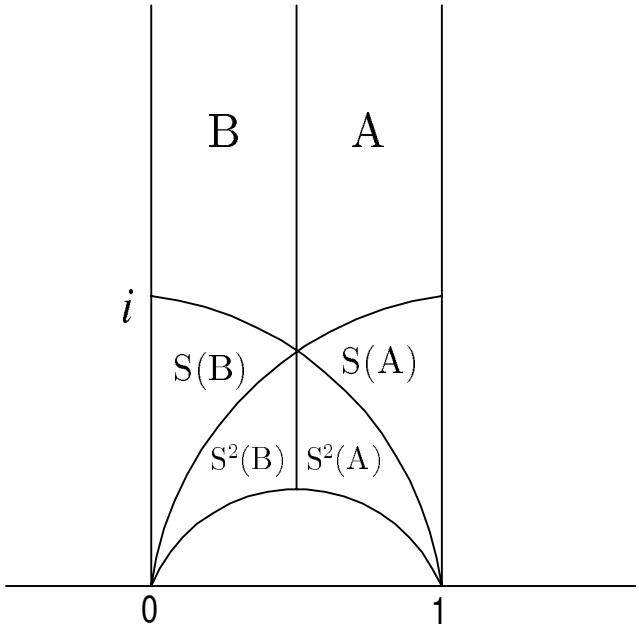

Figura 4.12: Ilustração das imagens se I, S e S^2 sobre a região F.

Existe uma descrição de $\mathbb{F}$ em função das seqüências de Farey, isto é,

Definição 4.2.2 *A* n***-ésima seqüência de Farey,*** $\mathbb{F}_n$*, é o conjunto dos números racionais* $\frac{p}{q}$ *com* $|p|, |q| \leq n$ *arranjados em ordem crescente.*

Da Definição 4.2.2, resulta que

$$\begin{aligned}
F_1 &= \{-\infty, -1, 0, 1, \infty\} \\
F_2 &= \{-\infty, -2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, \infty\} \\
F_3 &= \{-\infty, -3, -2, -3/2 - 1, -2/3 - 1/2, -1/3, 0, 1/3, 1/2, 2/3, 1, 3/2, 2, 3, \infty\} \\
&\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots
\end{aligned}$$

e, assim por diante.

Observação 4.2.3 *Dois números racionais* $\frac{p}{q}$ *e* $\frac{p'}{q'}$ *são adjacentes em alguma seqüência de Farey se, e somente se,*

$$\begin{pmatrix} p & p' \\ q & q' \end{pmatrix} \in SL(2, \mathbb{Z}), (cf. [11]) .$$

Da observação acima concluímos que $\mathbb{F}$ pode ser obtida pelo traço da linha vertical passando pelo ponto 0 e posteriormente ligando os pontos adjacentes em cada seqüência de Farey, como pode ser visto na Figura 4.13.

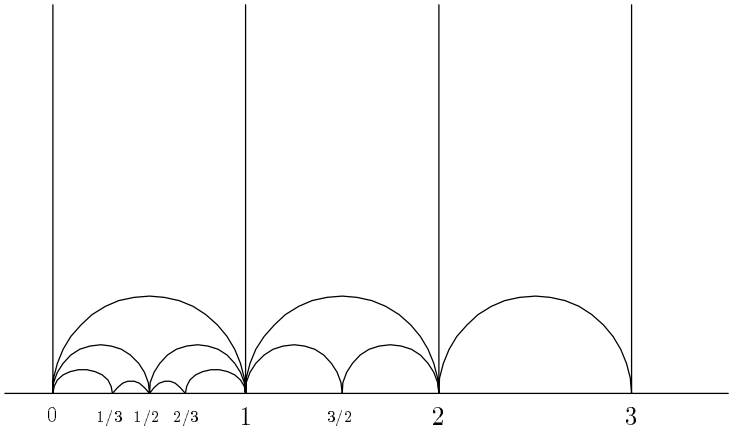

Figura 4.13: Seqüências de Farey.

Seqüências cortantes

Topologicamente, a superfície modular M é a esfera 3-vezes puncionada, com singularidades nas imagens de i, $\frac{1}{2}\left(1+i\sqrt{3}\right)$ e ∞, respectivamente. As linhas da tesselação de Farey são projetadas para a linha singular que vai da imagem $\pi(\infty)$ para $\pi(i)$ e volta novamente. Se tomarmos qualquer geodésica $\overline{\gamma} \in M$, outra que não seja $S = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$, podemos levanta-la para uma geodésica $\gamma \in \mathbb{H}^2$ e obter uma seqüência cortante $\cdots E^{n_{-1}} D^{n_0} E^{n_1} \cdots$ como descrita na Subseção 4.2.1. Como diferentes levantamentos de $\overline{\gamma}$ diferem por translações (que deixam $\mathbb{F}$ invariante) e preservam orientação, os rótulos de um segmento e, consequentemente a seqüência cortante obtida, independem do levantamento escolhido. A seqüência cortante termina se, e somente se, γ começa e termina na ima- gem $\pi(\infty)$. Neste caso, o rótulo do segmento inicial (ou final) pode ser D (ou E). Tais segmentos são chamados inicial ou final e denota-lo-emos por D_∞ ou E_∞. Se $x \in \overline{\gamma} \cap S$ então podemos indicar a posição de x na seqüência cortante escrevendo $\cdots D^{n_0} x E^{n_1} \cdots$. Dizemos que uma seqüência **muda de tipo** em $x \in S$, se os segmentos mudam o rótulo de D para E ou vice-versa, incluindo os pontos imediatamente precedentes (ou seguintes) ao segmento inicial (ou final).

Seja A o conjunto das geodésicas em $\mathbb{H}^2$ com pontos terminais satisfazendo $|\gamma_\infty| \geq 1$ e $0 < \left|\gamma_{-\infty}\right| \leq 1$. Tal geodésica intersecta o eixo imaginário $i\mathbb{R}$ no ponto ξ_γ. Note que, a seqüência cortante de tal geodésica γ muda de tipo em ξ_γ.

Como as geodésicas sobre M cortam repetidamente a linha singular S, isto gera naturalmente uma seção transversal X formada por um conjunto de vetores tangentes $T_1 M$, a saber, o conjunto dos vetores tangentes unitários com base no ponto $x \in S$ com pontos ao longo das geodésicas cujas seqüências cortantes mudam de tipo em x.

Se $\gamma \in A$, então o vetor unitário tangente u_γ para γ em ξ_γ é projetado para um elemento em X. Esta identificação de A com X é "quase" um homeomorfismo.

Teorema 4.2.4 *[30] A função $i : A \to X$, definida por $i(x) = \pi(u_\gamma)$ é sobrejetiva (isto é, $\pi(A) = X$), contínua e aberta. Ela é quase injetiva, isto é, é injetiva exceto nas duas geodésicas orientadas opostamente ligando $+1$ à -1 tendo a mesma imagem. Além*

disso, se $u_x \in X$ *define uma geodésica com seqüência cortante* $\cdots D^{n_0} x E^{n_1} \cdots$, *então* $\gamma = i^{-1}(u_x)$ *tem pontos terminais dados por*

$$\gamma_\infty = (n_1, n_2, ...), \qquad \frac{-1}{\gamma_{-\infty}} = (n_0, n_{-1}, n_{-2}, ...),$$

onde, se a seqüência cortante termina em um ou outro (dos pontos terminais), temos a correspondente fração contínua.

Se na seqüência cortante D *e* E *são permutados, então*

$$\gamma_\infty = -(n_1, n_2, ...), \qquad \frac{1}{\gamma_{-\infty}} = (n_0, n_{-1}, n_{-2}, ...).$$

Observação 4.2.5 *Note que* γ_∞ *é independente de* $\gamma_{-\infty}$ *e da parte da seqüência cortante que precede* ξ *e vice-versa.*

Observação 4.2.6 *Como* $n_r + 1 = n_r + \dfrac{1}{1+0}$, *as seqüências* $\cdots E^{n_r+1}$ *e* $\cdots E^{n_r} D$ *geram a mesma expansão no ponto terminal, causada pela ambigüidade do rotulamento de um ponto terminando num* **cusp**. *A mesma observação também é válida para uma seqüência começando com* $ED^{n-r} \cdots$ *ou* $D^{n-r+1} \cdots$.

Demonstração. (*do* Teorema 4.2.4) Seja γ uma geodésica em $\mathbb{H}^2$ que intersecta $i\mathbb{R}$. Como Δ é convexo, $\gamma_\infty \geq 1$ se, e somente se, o segmento $\gamma \cap \Delta$ é do tipo E, e $-1 \leq \gamma_{-\infty} < 0$ se, e somente se, o segmento imediatamente precedente a Δ é do tipo D. Observação análoga se aplica quando $\gamma_\infty \leq -1$ e $0 < \gamma_{-\infty} \leq 1$. Como qualquer geodésica sobre M cortando S em x pode ser levantada para uma geodésica em $\mathbb{H}^2$ intersectando $i\mathbb{R}$ no levantamento ξ de x, vimos que, i mapeia A sobre X. Além disso, soponha que $\gamma, \gamma\prime \in A$ com $i(\gamma) = i(\gamma\prime)$. Como a única identificação de pares de pontos em $i\mathbb{R}$ por $SL(2,\mathbb{Z})$ é dada por $Q : z \to -1/z$ vemos que, se $\gamma \neq \gamma\prime$, então $Q(\gamma) = \gamma\prime$. As únicas geodésicas $\gamma \in A$ tais que $Q(\gamma) \in A$ são as geodésicas ligando $+1$ à -1 e sua inversa. Isto prova a primeira parte do teorema.

Suponha, agora, que $\gamma \in A$ e $\gamma_\infty > 1$. Seja $\gamma_\infty = [n_1, n_2, ...]$. Tome $p_1 = n_1$ se $\gamma_\infty > n_1$ e $p_1 = n_1 - 1$ se $\gamma_\infty = n_1$. Seja $\eta_\gamma = \gamma \cap (\mathrm{Re}\, z = p_1)$. Entre ξ_γ e η_γ a tesselação de $\mathbb{F}$ particiona γ em p_1 segmentos nas linhas verticais $\mathrm{Re}\, z = 0, 1, ..., p_1$. Assim, a

seqüência cortante de γ entre ξ_γ e η_γ é E^{p_1}. Finalmente, se $\gamma_\infty = 1$, então a seqüência começando em ξ_γ é D^∞ ou E^∞.

Considere a função $\rho_1 = -1/(z - p_1)$. Note que $\rho_1 \in SL(2,\mathbb{Z})$; além disso $\rho_1(\gamma)_\infty \leq -1$, $0 < \rho_1(\gamma)_{-\infty} \leq 1$ e $\rho_1(\eta_\gamma) = \xi_{\rho_1(\gamma)}$. Como $\rho_1 \in SL(2,\mathbb{Z})$, $\rho_1(\gamma)$ é também um levantamento de $\pi(\gamma)$ e, assim, a seqüência cortante de γ começando em η_γ é a mesma seqüência cortante de $\rho_1(\gamma)$ começando em $\xi_{\rho_1(\gamma)}$. Se $\rho_1(\gamma_\infty) = -1$, esta seqüência termina com um segmento ambíguo D^∞ ou E^∞; caso contrário, fixe $p_2 = n_2$ se $-\rho_1(\gamma_\infty) \notin \mathbb{N}$ e $p_2 = n_2 - 1$ se $\rho_1(\gamma_\infty) \in \mathbb{N}$, a seqüência cortante começa com p_2 segmentos do tipo D. Aplicando $\rho_2 = -1/(z - p_2)$, o argumento se repete. Argumento análogo se aplica se $\gamma_\infty \leq -1$.

Note que, se $\gamma_\infty \in \mathbb{N}$, $\gamma_\infty > 1$, então essa seqüência é ambiguamente $E^{n_1}D$ ou E^{n_1+1}, consistente com a ambigüidade nas expansões em frações contínuas.

Para estudar o ponto terminal negativo $\gamma_{-\infty}$ aplicamos a função $Q : z \to -1/z$. Se γ tem seqüência cortante $\cdots E^{n_{-1}} D^{n_0} \xi_\gamma E^{n_1} D^{n_2} \cdots$, então como $Q \in SL(2,\mathbb{Z})$, a geodésica $Q(\gamma)^{-1}$ ligando $Q(\gamma_\infty)$ à $Q(\gamma_{-\infty})$ tem seqüência cortante $\cdots E^{n_1} Q(\xi_\gamma) D^{n_0} E^{n_{-1}} \cdots$. Claramente, $Q(\gamma)^{-1} \in A$ e $\xi_{Q(\gamma)^{-1}} = Q(\xi_\gamma)$. Assim $Q(\gamma_{-\infty}) = [n_0, n_{-1}, ...]$, o que prova o resultado. ■

Corolário 4.2.7 *[30], Duas geodésicas com a mesma seqüência cortante coincidem.*

Demonstração. Fixe os pontos iniciais x, x' nos mesmos pontos de divisão em cada seqüência, e levante os pontos $\xi, \xi' \in i\mathbb{R}^+$ como no Teorema 4.2.4. Como os pontos terminais no infinito dos levantamentos das duas geodésicas têm a mesma expansão em frações contínuas, as geodésicas coincidem. ■

Capítulo 5

Codificação de Geodésicas em Superfícies de Curvatura Negativa: Toro 1-vez Puncionado

Seja Σ uma superfície (possivelmente puncionada, isto é, uma superfície de onde foi retirado um ou mais pontos) com característica de Euler negativa, e seja $\mathcal{C}(\Sigma)$ o conjunto de classes de famílias isotópicas (curvas homotópicas que preservam orientação) de curvas fechadas disjuntas. Nosso estudo é baseado na construção da função de Markov feita em [5] e [31] para grupos fuchsianos, onde foi generalizada a relação entre o $PSL(2,\mathbb{Z})$ (pensado como um grupo fuchsiano que age sobre o plano hiperbólico) e as frações contínuas (pensadas como pontos no conjunto limite de $PSL(2,\mathbb{Z})$), para uma ampla classe de grupos fuchsianos.

Embora envolva conceitos matemáticos não muito elementares, esta teoria torna-se deveras simples quando nos restringimos ao estudo de uma particular superfície.

Nosso objetivo é construir o **grupo de classes de funções** $\mathcal{MCG}(\Sigma)$ e a **máquina de estados finitos**, quando $\Sigma = \Sigma_1$ é o toro 1-vez puncionado.

Essa idéia está suportada na analogia entre a ação de um grupo fuchsiano Γ sobre o plano hiperbólico e a ação do grupo de classes de funções $\mathcal{MCG}(\Sigma)$ sobre o espaço de Teichmüller $\mathcal{T}(\Sigma)$, [8], *pag*.159. Nesta analogia, a fronteira S^1 do plano hiperbólico (ou conjunto limite de Γ) é substituída por uma fronteira adequada de $\mathcal{T}(\Sigma)$. Usamos

aqui a **fronteira de Thurston,** a saber, o espaço de medidas das laminações projetivas $\mathcal{PML}(\Sigma)$ descrita na Subseção 5.1.3. O grupo de classes de funções $\mathcal{MCG}(\Sigma)$ construído na Seção 5.2 age em ambos $\mathcal{T}(\Sigma)$ e $\mathcal{PML}(\Sigma)$. Em anologia com a construção feita em [5], definimos a função de Markov f sobre $\mathcal{PML}(\Sigma)$ cuja ação de $\mathcal{MCG}(\Sigma)$ sobre $\mathcal{PML}(\Sigma)$ está diretamente relacionada com a ação de Γ sobre $\Lambda(\Gamma) = S^1$.

Um tratado da teoria dos *train tracks* pode ser encontrado em [26]. Aqui usamos uma classe particular de *train tracks*, a saber: os π_1-*train tracks* introduzidos por Birman and Series em [4]. Essa classe especial de *train tracks* tem sua definição condicionada a escolha de um domínio fundamental fixo, e associados a geradores geométricos de $\pi_1(\Sigma)$, de tal forma que uma pesagem inteira resulte não somente numa curva fechada simples ou numa malha, mas simultaneamente permita-nos ler inteiramente um menor representante como um ciclo de palavras em $\pi_1(\Sigma)$. Assim, o espaço $\mathcal{ML}(\Sigma)$ é particionado em um número finito de células maximais correspondendo às pesagens sobre os possíveis (finitos) *train tracks* associados a um domínio fundamental de Σ_1.

Neste capítulo analisamos o caso especial do toro 1-vez puncionado Σ_1. Usamos seu grupo de classes de funções para construir a função de Markov f sobre $\mathcal{ML}(\Sigma_1)$. A partição de Markov consiste do conjunto de células maximais e restrição de f a cada célula, onde, cada restrição, é um elemento específico de $\mathcal{MCG}(\Sigma_1)$, o qual age linearmente sobre o conjunto de pesos. Rotulando as células pelos elementos correspondentes de $\mathcal{MCG}(\Sigma_1)$, mostraremos que as f-expansões (isto é, os trajetos das órbitas rotuladas de f) fornecem uma única forma normal para os elementos de $\mathcal{MCG}(\Sigma_1)$. Em particular, os rótulos são um conjunto de geradores de $\mathcal{MCG}(\Sigma_1)$. Comparando as formas normais para elementos próximos, podemos determinar uma presentação para $\mathcal{MCG}(\Sigma_1)$. Sendo f uma função Markoviana, as f-expansões pertencem a um subshift do tipo finito o qual é de fato uma geodésica fechada com relação ao conjunto de geradores em questão.

Na linguagem dos grupos automáticos (Subseção 5.1.1), esta forma normal permite-nos construir a máquina de estados. Se mostrarmos que esta forma normal satisfaz a propriedade de companheirismo (Subseção 5.1.1), então ela fornece uma estrutura automática para $\mathcal{MCG}(\Sigma_1)$. Concluiremos este capítulo mostrando que de fato este é o

caso. A dimensão de $\mathcal{PML}(\Sigma_1)$ é 1, já que $\mathcal{PML}(\Sigma_1) \backsim S^1$. No nosso caso particular, o espaço de Teichmüller é o semi-plano superior e o grupo de classes de funções é o $PSL(2,\mathbb{Z})$.

Na Seção 5.1, introduzimos as definições básicas para o desenvolvimento deste capítulo. Na Seção 5.2, construiremos o grupo de classes de funções para o toro Σ_1, a máquina de estados finitos e a estrutura automática para o grupo de classes de funções $\mathcal{MCG}(\Sigma_1)$.

5.1 Conceitos e Definições

5.1.1 Funções de Markov e grupos automáticos

A **função de Markov** sobre um conjunto X é uma função $f : X \longrightarrow X$ juntamente com uma partição (em geral finita) de X em conjuntos $I_j,\ j \in \mathbb{N}$, tal que $f(I_i)$ é uma união exata de conjuntos I_j.

Diz-se que f satisfaz a **propriedade de Markov** se $f(I_i^\circ) \cap I_j^\circ \neq \varnothing$ implicar que $I_j \subset f(I_i)$. Para $\xi \in X$, considere a função $p : X \to \mathbb{N}$, definida por: $p(\xi) = j$ se $\xi \in I_j$. A seqüência $p(\xi), p(f(\xi)), p(f^2(\xi)), ...$ é chamada a f-**expansão** de ξ. Associada a f existe uma matriz de $0's$ e $1's$ informando quais transições podem ocorrer. Como f é Markov, todas as seqüências infinitas que são permissívies ocorrem. Os blocos finitos que ocorrem nestas expansões são chamados **admissíveis.**

A construção de Bowen-Series [5], é baseada na relação entre o $PSL(2,\mathbb{Z})$, agindo no plano hiperbólico $\mathbb{H}^2$, e as transformações frações contínuas agindo sobre o conjunto dos números reais estendido, $\mathbb{R} \cup \{\infty\}$.

Exemplo 5.1.1 *A transformação fração contínua*

$$f(x) = \begin{cases} x - 1 \ se\ x \geq 1 \\ x + 1 \ se\ x \leq -1 \\ \dfrac{-1}{x} \ se\ |x| \leq 1 \end{cases} \tag{5.1}$$

pode ser considerada como um exemplo de uma função de Markov com partição $\Pi = \{[\infty, -1], [-1, 0], [0, 1], [1, \infty]\}$.

A f-expansão de um ponto $\xi \in \mathbb{R}$ é essencialmente a mesma que sua expansão em frações contínuas. Note que a restrição de f a cada elemento da partição Π pertence ao subcunjunto $\Gamma_0 = \{x \longrightarrow x - 1,\ x \to x + 1,\ x \to -1/x\} \subset PSL(2, \mathbb{Z})$. (Um resultado bastante conhecido na geometria hiperbólica é que $\langle \Gamma_0 \rangle = PSL(2, \mathbb{Z})$, [31]).

Na construção da função de Markov f sobre a fronteira infinita, [5] e [31], a saber, o círculo unitário S^1, os elementos da partição são intervalos I_j, e , para cada j, a restrição de $f|_{I_j}$ pertence a Γ_0.

Um **alfabeto** $\mathcal{A}$ é um conjunto finito. Uma **linguagem** $\mathcal{L}$ sobre um alfabeto $\mathcal{A}$ é uma coleção de seqüências finitas de $\mathcal{A}$ chamadas **palavras.** O **comprimento** de uma palavra $w = (a_1, a_2, ..., a_n)$ será denotado por $|w| = n$. Uma **máquina de estados** sobre um alfabeto $\mathcal{A}$ é um grafo finito, orientado, cujos vértices são os estados ligados por arestas orientadas e rotuladas. Existe um estado chamado **estado inicial** e uma partição dos estados em dois conjuntos disjuntos, os **estados aceitos** e os **estados não aceitos.** Toda aresta saindo de um estado é rotulada com um símbolo de $\mathcal{A}$ e não existem duas arestas saindo de um mesmo estado com o mesmo rótulo. Dada qualquer palavra $w = (a_1, a_2, ..., a_n)$ sobre $\mathcal{A}$ e qualquer estado s existe no máximo um caminho começando em s tal que a *j-ésima* aresta é rotulada por a_j. Este estado s termina num estado s'. Neste caso diz-se que w vai de s para s'.

A linguagem **aceita** por esta máquina é uma coleção de palavras w sobre $\mathcal{A}$ que vai de um estado inicial para algum outro estado. Uma linguaguem $\mathcal{L}$ sobre $\mathcal{A}$ é chamada **regular** se ela é aceita por alguma máquina de estados finitos sobre $\mathcal{A}$ e diz-se que esta máquina **reconhece** a linguagem $\mathcal{L}$. Seja G um grupo com identidade e. Considere a função $f : \mathcal{A} \longrightarrow G$ definida por $a \to \overline{a}$. A definição desta função se estende para uma coleção de palavras de $\mathcal{A}$ para G por $w = (a_1, a_2, ..., a_n) \to \overline{w} = \overline{a_1} \cdots \overline{a_n}$, o produto das imagens das letras de w. Se todos os elementos de G podem ser descritos desta forma diz-se que $\mathcal{A}$ é um conjunto finito **gerador** de G. Uma linguagem $\mathcal{L} = \mathcal{L}(G)$ sobre $\mathcal{A}$ é chamada uma **estrutura automática** para G se satisfaz as seguintes propriedades. Primeira, $\mathcal{L}(G)$ é uma linguagem regular sobre G, isto é, existe uma máquina de

estados finito $\mathcal{W}(G)$ tal que todo elemento de G pode ser descrito por (no mínimo) um caminho nesta máquina. A segunda propriedade é conhecida como propriedade de "**companheirismo**" como descrita abaixo. Se um grupo tem uma estrutura automática, então G é chamado um **grupo automático**.

Seja G um grupo com um conjunto gerador finito $\mathcal{A}$. O **comprimento** de uma palavra $g \in G$, denotado por $|g|$ em relação a $\mathcal{A}$, é por definição o menor comprimento entre todas as palavras que representam g. A **métrica** sobre G é definida por $d(g,h) = |g^{-1}h|$. Dada uma palavra $w = (a_1, a_2, ..., a_n)$ sobre $\mathcal{A}$, para cada $0 \leq t \leq n$, denota-se por $w(t) = (a_1, a_2, ..., a_t)$ o **prefixo** de w de comprimento t, e para inteiros $t \geq n$ denota-se $w(t) = w$. Dada uma constante k, duas palavras w, v sobre $\mathcal{A}$ são k-**companheiras** se $d(\overline{w}(t), \overline{v}(t)) \leq k$ para todo $t \geq 0$. A constante k é chamada **constante de companheirismo** para w e v. O grupo G satisfaz a **propriedade de companheirismo** se existe uma constante k tal que para quaisquer palavras w, v sobre $\mathcal{L}$ com $d(\overline{w}, \overline{v}) \leq 1$ então w e v são k-**companheiras**.

Sejam w e v duas palavras e $a \in \mathcal{A} \cup \{\mathbf{e}\}$ tal que $\overline{wa} = \overline{v}$. Se w e v são k-companheiras, então $\psi(t) = w(t)^{-1}v(t)$ tem comprimento no máximo k, para todo $t \geq 0$.

Assim, para qualquer escolha de w e v com $d(\overline{w}, \overline{v}) \leq 1$, $\psi(t)$ pertence a um conjunto finito $\mathcal{D}$, a saber; a coleção das **palavras-diferença** e $\mathcal{A} \cup \{\mathbf{e}\} \subset \mathcal{D} \subset \mathcal{L}$. O conhecimento das palavras-diferença permite-nos reconstruir a estrutura multiplicativa de G de maneira automatizada.

Mais precisamente, a propriedade de companheirismo é equivalente a existência de uma **máquina multiplicativa** $\mathcal{M}_a$, $a \in \mathcal{A} \cup \{\$\}$, para G. Cada $\mathcal{M}_a$ é uma automata de comprimento 2, cujo alfabeto é $\mathcal{A}' \times \mathcal{A}'$, onde $\mathcal{A}'$ é o **alfabeto estendido** $\mathcal{A} \cup \{\$\}$. Tal máquina aceita o par estendido (w^+, v^+) das palavras (w, v) sobre $\mathcal{A}$ sempre que w, v são estados aceitos da máquina de estados finito $\mathcal{W}(G)$ e $\overline{wa} = \overline{v}$. Os símbolos w^+ e v^+ indicam o acréscimo do símbolo \$ ao alfabeto $\mathcal{A}$. Tal símbolo mapeia a identidade em G e o mesmo pode ser adicionado à menor das palavras entre w e v, de modo a torna-las com o mesmo comprimento. A máquina $\mathcal{M}_\$$ reconhece a identidade de G, substituindo a condição $\overline{wa} = \overline{v}$ por $\overline{w} = \overline{v}$.

As máquinas multiplicadoras $\mathcal{M}_a$, $a \in \mathcal{A} \cup \{\$\}$, podem ser construídas por meio de

uma **máquina de palavras-diferença**. Isto, na realidade é uma coleção de novas máquinas, todas tendo o mesmo espaço de estados, ou seja, o conjunto das ternas (s_1, s_2, ψ) tal que s_1 e s_2 são estados de $\mathcal{L}(G)$ e $\psi \in \mathcal{D}$. O estado inicial é (s_0, s_0, e) onde s_0 é o estado inicial de $\mathcal{L}(G)$ e e, é a identidade de G. Para $a, b \in \mathcal{A}$, existe uma aresta de (s_1, s_2, ψ) para (s_1', s_2', ψ') se, e somente se, existem arestas de $s_1 \xrightarrow{a} s_1'$ e $s_2 \xrightarrow{b} s_2'$ na máquina de estados e se $\overline{\psi'} = \overline{x^{-1}\psi y}$. Na máquina $\mathcal{M}_a$, o estado (s_1, s_2, ψ) é um estado de sucesso (aceito) se s_1 e s_2 pertencem a $\mathcal{L}$ e se $\overline{\psi} = \overline{a}$.

Existe uma técnica extra para negociar com o símbolo adicionado. A saber, temos que adicionar um estado extra na máquina de estados $\mathcal{W}(G)$ o qual é alcançado quando $\mathcal{W}(G)$ está num estado aceito e o símbolo adicionado é lido. Se s_1 ou s_2 é este estado extra , então x ou y acima definido, será substituído pelo símbolo adicionado \$ e a condição $\overline{\psi'} = \overline{x^{-1}\psi y}$ será substituída por $\overline{\psi'} = \overline{\psi y}$ ou $\overline{\psi'} = \overline{x^{-1}\psi}$. Frequentemente, pensaremos nessas condições como triângulos ou quadrados comutativos relacionados por elementos num grupo (Seção 5.2.5). Se pudermos construir uma máquina de palavras-diferença usando um conjunto finito de palavras-diferença $\mathcal{D}'$, então teremos verificado a propriedade de companheirismo e com isso, podemos usar o processo descrito para construir simultaneamente todas as máquinas multiplicativas $\mathcal{M}_a$. Esse processo pode ser visto como uma concatenação de quadrados ou triângulos resultando numa coleção de caminhos cruzados em $\mathcal{D}$ entre todos os prefixos $w(t)$ e $wa(t)$ ocorrendo nas formas normais de quaisquer dois caminhos w e wa para $a \in \mathcal{A}$. A coleção de todos os quadrados e triângulos estabelece uma presentação para o grupo.

5.1.2 Malha simples e π_1-*train tracks*

Considere uma curva fechada sobre uma superfície Σ (superfície com característica de Euler negativa, possivelmente puncionada). Tal curva é chamada **simples** se não tem auto interseções e, é dita **paralela à fronteira** ou **periférica** se é homotópica (se pode ser deformada na curva em questão) a uma curva em torno do ponto puncionado da superfície. Uma **malha simples** é uma coleção de curvas fechadas simples duas a duas disjuntas nenhuma das quais é homotòpicamente trivial ou paralela a fronteira.

A definição a seguir considera uma estrutura hiperbólica sobre Σ. Seja $R \subset \mathbb{H}^2$ a região fundamental de Σ cujos vértices são ideais, isto é, encontram-se na fronteira do disco de Poincaré[1].

Suponha que R tenha lados σ_k e função emparelhamento de lados $\mu_k : \sigma_k \longrightarrow \sigma_{k'}$ onde $\sigma_{k'} = \sigma_k^{-1}$ para cada k. Seja $\overline{R}$ o fecho de R em $\mathbb{H}^2$. Um π_1-**train track** τ é uma coleção de arcos $\alpha_j : [0,1] \longrightarrow \overline{R}$ dois a dois disjuntos tais que:

(i) $\alpha_j(0) \in \sigma_k$ e $\alpha_j(1) \in \sigma_l$;

(ii) $\alpha_j(\lambda) \in R^o$ para $\lambda \in (0,1)$;

(iii) No máximo um arco liga cada par de lados;

(iv) Nenhum arco liga um lado a si mesmo, isto é, se k e l são como em (i) então $k \neq l$.

Um arco de τ é chamado um **arco quina** se o mesmo liga lados adjacentes de R. Cada arco quina vê um vértice particular de R e, para cada ciclo de vértices no emparelhamento de lados existe um correspondente ciclo de quinas consistindo de todas as bifurcações das quinas correspondentes ao mesmo ponto (retirado da superfície).

Uma **pesagem** ω sobre um π_1-*train track* τ é uma designação de um número não-negativo $\omega(\alpha_j)$ para cada arco α_j de τ. Uma pesagem é **inteira** se cada peso é um inteiro não-negativo. Define-se o **peso** de ω, e denotado por $|\omega|$, como $|\omega| = \sum \omega(\alpha_j)$ onde a soma é sobre todos os arcos α_j de τ. Iremos considerar, a partir de um malha simples γ, como obter um π_1-*train track* com pesagem inteira. Iniciaremos com a realização do levantamento da malha simples γ para a região fundamental R. A malha simples γ torna-se uma coleção de arcos, chamados **cordões,** ligando os lados de R. Diz-se que um malha simples γ está **suportada** por um π_1-*train track* τ se, para todo cordão de γ existe um arco de τ ligando o mesmo par de lados. Se γ está suportado por τ então podemos estabelecer à τ uma pesagem inteira ω_γ fixando para cada arco de τ o número de cordões de γ ligando os mesmos pares de lados. Esta pesagem tem as seguintes propriedades:

[1] A teoria dos train tracks envolve geometria hiperbólica, portanto faz-se necessário a escolha de uma estrutura hiperbólica sobre Σ.

(i) Para cada emparelhamento de lados $\mu_k : \sigma_k \longrightarrow \sigma_{k'}$, a soma dos pesos dos arcos com pontos terminais sobre σ_k é igual a soma dos pesos dos arcos com pontos terminais em $\sigma_{k'}$.

(ii) Pelo menos um arco em cada ciclo de quinas deve ter peso zero.

Uma pesagem (não-negativa mas não necessariamente inteira) é dita ser uma **pesagem formal** se satisfaz (i) e (ii) acima. Reciprocamente, toda pesagem inteira formal dá origem a uma malha simples γ. Isto significa que estudar as malhas simples é equivalente a estudar as pesagens inteiras formais sobre os π_1-*train tracks.* Denota-se por $\mathbf{W}(\tau)$ a coleção de todas as pesagens formais sobre o π_1-*train track* τ e por $\mathbf{W}_{\mathcal{O}}(\tau)$ a coleção de todas as pesagens inteiras formais sobre τ.

5.1.3 Curvas fechadas irredutíveis

Um π_1-*train track* τ é dito ser **recorrente** se existe uma pesagem inteira formal $\omega \in W_{\mathcal{O}}(\tau)$ tal que $\omega(\alpha_j) \neq 0$ para todos os ramos α_j de τ. Um π_1-*train track* recorrente τ é dito **maximal** se não existe outro π_1-*train track* recorrente τ' tal que $\tau \nsubseteq \tau'$. Se τ é um *train track* maximal recorrente então a dimensão de $W(\tau)$ é $6g-6+2p$, onde Σ é uma superfície de gênero g com p pontos removidos, $(cf.\,[25])$. A coleção de todas as pesagens formais $W(\tau)$ sobre um π_1-*train track* recorrente maximal τ é chamada uma **célula maximal.**

Qualquer curva fechada simples γ define um π_1-*train track* recorrente $\tau = \tau(\gamma)$ com pesagem $\omega(\gamma)$ como acima. Uma curva fechada γ é dita ser **irredutível** quando $\omega(\gamma) \neq \omega_1 + \omega_2$, quaisquer que sejam $\omega_1, \omega_2 \in W_{\mathcal{O}}(\tau(\gamma))$ e $\omega_j \neq 0$ para $j = 1, 2$.

A coleção de todas as classes homotópicas de malhas simples sobre Σ não-paralelas à fronteira é denotada por $\mathcal{ML}_{\mathcal{O}}(\Sigma)$ e, a coleção de todas as medidas de laminações geodésicas sobre Σ por $\mathcal{ML}(\Sigma)$, [25]. Os conjuntos $\mathcal{ML}_{\mathcal{O}}(\Sigma)$ e $\mathcal{ML}(\Sigma)$ podem ser identificados [4], como as coleções das pesagens formais inteiras e pesagens formais, respectivamente, sobre os π_1-*train tracks* sobre Σ. Para $\mathcal{ML}_{\mathcal{O}}(\Sigma)$ a prova segue do comentário feito acima. Se $\omega \in \mathcal{ML}(\Sigma)$, então ω está contido em alguma célula máxima $W(\tau)$. Assim, $\mathcal{ML}(\Sigma)$ é a união de células máximas $W(\tau_i)$ onde τ_i percorre todos os

π_1-*train tracks* recorrentes maximais sobre Σ. Isto dá a $\mathcal{ML}(\Sigma)$ uma estrutura natural de célula, como estabelece o próximo resultado.

Proposição 5.1.2 [25], *Para $p = 1, 2$, seja Σ_p o toro com p pontos removidos. Existe um número finito de curvas fechadas irredutíveis $\mathbf{e}_1, ..., \mathbf{e}_k$ sobre Σ_p tal que para cada π_1-train track maximal τ, a correspondente célula máxima $W(\tau)$ é a combinação linear positiva de $2p$ curvas fechadas irredutíveis: $W(\tau) = sp^+\{\mathbf{e}_{i_1}, ..., \mathbf{e}_{i_{2p}}\}$ onde $i_j \in \{1, ..., k\}$, e a interseção de duas células é dada por:*

$$W(\tau) \cap W(\tau') = sp^+\left(\{\mathbf{e}_{i_1}, ..., \mathbf{e}_{i_{2p}}\} \cap \{\mathbf{e}_{i'_1}, ..., \mathbf{e}_{i'_{2p}}\}\right),$$

onde

$$W(\tau) = sp^+\{\mathbf{e}_{i_1}, ..., \mathbf{e}_{i_{2p}}\} \text{ e } W(\tau') = sp^+\{\mathbf{e}_{i'_1}, ..., \mathbf{e}_{i'_{2p}}\}.$$

O espaço $W(\tau)$ pode ser projetado de uma maneira natural para obter-se o cone $\mathbb{P}W(\tau)$. Analogamente, $W_{\mathcal{O}}(\tau)$ pode ser projetado para obter-se o conjunto das pesagens racionais $\mathbb{P}W_{\mathcal{O}}(\tau)$. O espaço $\mathcal{PML} = \mathcal{PML}(\Sigma)$ é a união sobre todos os τ dos correspondentes cones $\mathbb{P}W(\tau)$, os quais são chamados π_1-cones. Da mesma forma define-se $\mathcal{PML}_{\mathcal{O}}(\Sigma)$ como a união de todas as pesagens racionais $\mathbb{P}W_{\mathcal{O}}(\tau)$. Usando a identificação de Birman e Series [4], o espaço $\mathcal{PML}(\Sigma)$ foi mostrado por Thurston [26], ser a esfera de dimensão $6g - 6 + 2p$. Assim, no caso do toro Σ_1 tem-se o círculo unitário S^1 ($g = 1$ e $p = 1$).

5.1.4 Torções de Dehn e o grupo de classes de funções $\mathcal{MCG}(\Sigma)$

O **grupo de classes de funções** $\mathcal{MCG} = \mathcal{MCG}(\Sigma)$ de Σ (isometrias que preservam orientação), é o grupo de todas as classes isotópicas de automorfismos de Σ (que preservam orientação), isto é, um elemento de $\mathcal{MCG}$ é um homeomorfismo (que preserva orientação) de Σ sobre si mesmo e, dois tais homeomorfismos representam o mesmo elemento de $\mathcal{MCG}$ se um pode ser deformado no outro isotopicamente ao longo de um caminho contínuo de homeomorfismos de Σ sobre si mesmo. Existe uma ação natural do grupo de classes de funções sobre o espaço de Teichmüller de Σ como o grupo modular

de Teichmüller. Esta ação pode ser estendida para $\mathcal{PML}(\Sigma)$, e será considerada no que segue.

Seja w_0 uma curva fechada sobre uma superfície Σ, parametrizada pelo comprimento de arco $\xi \in [0, l_{w_0}]$ onde l_{w_0} é o comprimento de w_0. Considere uma vizinhança tubular em torno de w_0 em Σ denotada por $N_{w_o} = [0,1] \times w_o$. Uma **torção de Dehn,** em torno de w_0 (a esquerda), é um homeomorfismo sobre Σ denotado por δ_{w_0} e definido como sendo a identidade em $\Sigma - N_{w_o}$ e, se $(\eta, \xi) \in [0,1] \times w_0$ então $\delta_{w_0}(\eta, \xi) = (\eta, \xi - \eta l_{w_0})$, onde $\eta \in [0,1]$ e $\xi - \eta l_{w_0}$ é definido $\mathrm{mod}\, l_{w_0}$. Observe que, se $\eta = 0$ ou 1, então δ_{w_0} é a identidade.

5.2 O Grupo de Classes de Funções para Σ_1

Nesta seção construiremos os π_1-*train tracks*, a função de Markov e uma estrutura automática para o toro Σ_1. Nas Subseções 5.2.1 e 5.2.2 mostraremos como as torções elementares de Dehn agem sobre Σ_1 da mesma maneira como as frações contínuas agem sobre $\mathbb{R} \cup \{\infty\}$. Na Subseção 5.2.3 construiremos a função de Markov e, na Subseção 5.2.4 construiremos a forma normal e a máquina de estados.

5.2.1 π_1-*train tracks* e a estrutura de células maximais

Como foi observado na Subseção 5.1.2, fixaremos agora, como pode ser visto na Figura 5.1, uma estrutura hiperbólica para Σ_1 e um domínio fundamental $R \subset \mathbb{H}^2$ para a ação de $\pi_1(\Sigma)$ sobre o plano hiperbólico.

O domínio fundamental R que escolhemos é o padrão retangular com lados opostos identificados por emparelhamentos das arestas. A região fundamental tem quatro vértices ideais, todos contidos na fronteira do disco de Poincaré, veja Figura 5.1. Os vértices são rotulados por $v_1, ..., v_4$ no sentido horário. Denotando por $v_i v_j$ o lado ligando v_i a v_j, o emparelhamento de lados levando $v_1 v_2$ para $v_4 v_3$ e $v_1 v_4$ para $v_2 v_3$ serão denotados por S e T, respectivamente. As funções S e T correspondem às classes

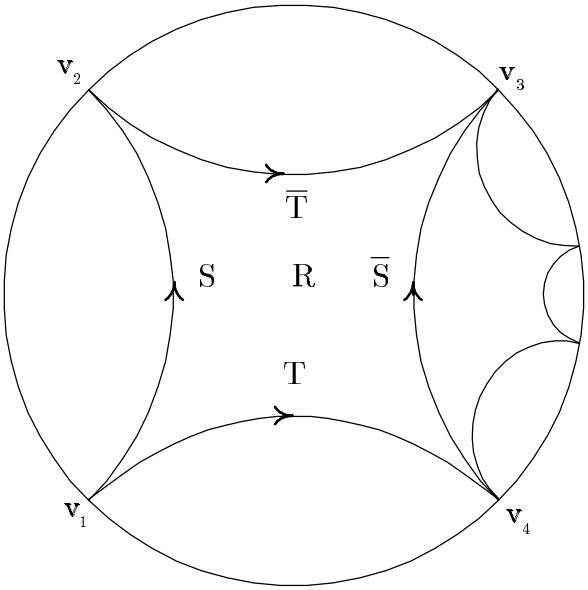

Figura 5.1: Domínio fundamental hiperbólico R para Σ_1, onde $\overline{S}$, $\overline{T}$ denotam S^{-1}, T^{-1}.

homotópicas de curvas fechadas simples as quais geram livremente o grupo fundamental.

Vamos introduzir, agora, as curvas fechadas irredutíveis $\mathbf{e_0}$, $\mathbf{e}_1$, $\mathbf{e}_\infty$, e $\mathbf{e}_{-1}$ formando uma base para $\mathcal{ML}(\Sigma_1)$, conforme estabelece a Proposição 5.1.2. Tais curvas fechadas irredutíveis, ilustradas na Figura 5.2, são definidas como

- $\mathbf{e_0}$ consiste de um arco simples ligando v_1v_2 a v_3v_4;
- $\mathbf{e}_1$ consiste de um arco ligando v_1v_2 a v_2v_3 e um arco ligando v_3v_4 a v_4v_1;
- $\mathbf{e}_\infty$ consiste de um arco simples ligando v_2v_3 a v_4v_1;
- $\mathbf{e}_{-1}$ consiste de um arco ligando v_2v_3 a v_3v_4 e um arco ligando v_4v_1 a v_1v_2.

Do que foi visto no Capítulo 4, podemos definir as curvas fechadas irredutívies como os termos das seqüências cortantes: $\mathbf{e_0} = S$, $\mathbf{e}_1 = ST$, $\mathbf{e}_\infty = T$, $\mathbf{e}_{-1} = S^{-1}T$. Como as curvas fechadas são não-orientadas podemos escrever $\mathbf{e_0} = S$ ou $\mathbf{e_0} = S^{-1}$.

A seguir definimos as células em $\mathcal{ML}(\Sigma_1)$ e mostraremos que elas são maximais. As células são:

$$I_0 = sp^+\{\mathbf{e_0}, \mathbf{e}_1\}, I_1 = sp^+\{\mathbf{e_0}, \mathbf{e}_{-1}\}, I_2 = sp^+\{\mathbf{e}_\infty, \mathbf{e}_{-1}\}, I_3 = sp^+\{\mathbf{e}_\infty, \mathbf{e}_1\}$$

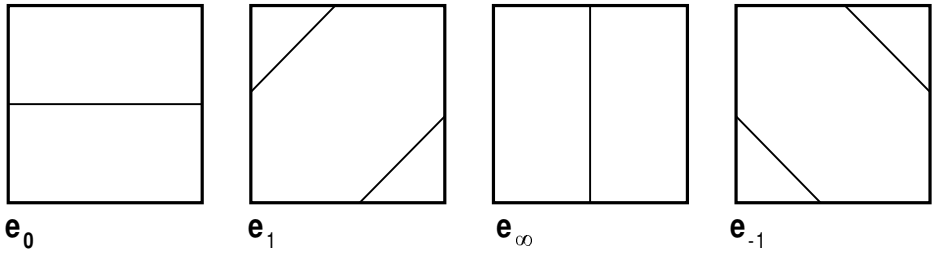

Figura 5.2: π_1-*train tracks* elementares.

Proposição 5.2.1 [25], *As células* I_0, I_1, I_2, I_3 *são maximais e sua união é* $\mathcal{ML}(\Sigma_1)$.

Demonstração. É suficiente provar que nenhum outro arco pode ser adicionado a esses quatro π_1-*train tracks* e que qualquer curva fechada está suportada por um deles. Observe que, qualquer π_1-*train track* sobre Σ_1 deve ter uma das duas formas ilustradas na Figura 5.3. Somando-se os pesos sobre os dois pares de lados identificados e cancelando-se a obtemos as duas equações: $b + e = c + d$, $b + c = e + d$. Disso obtemos que $b = d$ e $c = e$. Do fato de que um ciclo de quinas não pode ter pesos todos nulos, segue que $b = 0$ ou $c = 0$. Como todos os pesos são não negativos, significa que existem somente quatro configurações de π_1-*train tracks* correspondentes às curvas fechadas simples sobre Σ_1 não paralelas à fronteira. ■

Pela Proposição 5.2.1 qualquer curva fechada simples γ pode ser representada por $a\mathbf{e}_i + b\mathbf{e}_j$, onde $w(\gamma) \in I_k$ para algum $k = 0, 1, 2, 3$.

Iremos usar a notação (a, b) para representar os seguintes fatos

$$(a,b) = \begin{cases} a\mathbf{e_0} + b\mathbf{e}_1, \text{ se } w(\gamma) \in I_0 \\ a\mathbf{e_0} + b\mathbf{e}_{-1}, \text{ se } w(\gamma) \in I_1 \\ a\mathbf{e}_\infty + b\mathbf{e}_{-1}, \text{ se } w(\gamma) \in I_2 \\ a\mathbf{e}_\infty + b\mathbf{e}_1, \text{ se } w(\gamma) \in I_3. \end{cases} \tag{5.2}$$

A notação para os $\mathbf{e}_j$ tem a seguinte racionalidade. Considere R como sendo o quadrado com vértice v_1 no canto inferior esquerdo, e o emparelhamento de lados por translações Euclidianas. Pela Proposição 5.2.1, qualquer curva fechada simples γ sobre Σ_1 está suportada por uma das quatro células maximais I_k. Assim, a menos de

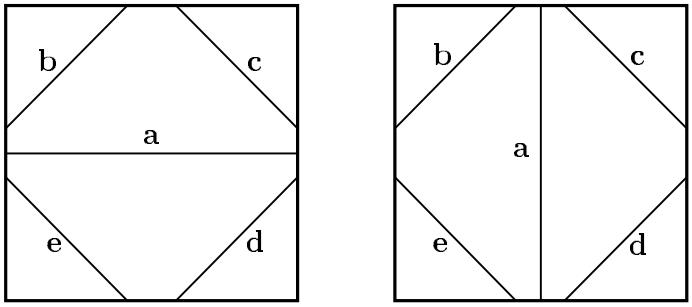

Figura 5.3: As duas possíveis configurações para um π_1-*train track* maximal ponderado.

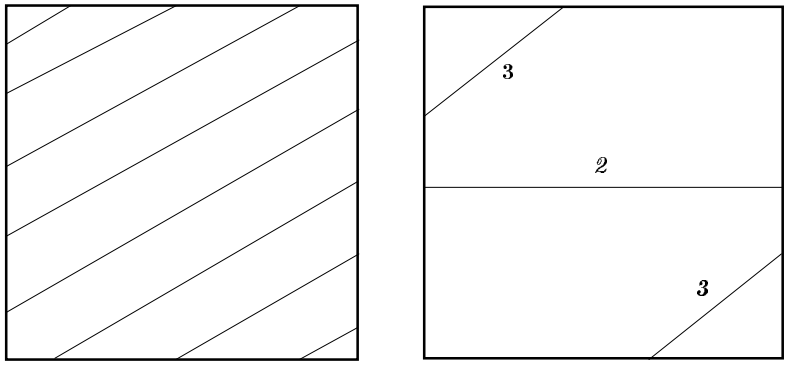

Figura 5.4: As linhas de inclinação $3/5$ traçadas sobre R e seu respectivo π_1-*train track* ponderado.

homotopia, γ é equivalente a uma família de retas cortando R. O levantamento de γ para o plano Euclidiano (recobrimento universal de R), isto é, $\mathbb{R}^2$, tal família de retas se unem para formar uma reta de inclinação racional sobre o plano. Com esta identificação a curva rotulada por $\mathbf{e}_j$ tem inclinação j.

De uma maneira geral, obtem-se uma identificação de $\mathcal{PML}(\Sigma_1)$ com os reais estendidos, $\mathbb{R}\cup\{\infty\}$, mapeando $(a,b) \in I_0$ no ponto $\dfrac{b}{a+b}$, $(a,b) \in I_1$ no ponto $\dfrac{-b}{a+b}$, $(a,b) \in I_2$ no ponto $\dfrac{-(a+b)}{b}$, $(a,b) \in I_3$ no ponto $\dfrac{a+b}{b}$. Um exemplo de uma curva representada por $(2,3) = 2\mathbf{e_0}+3\mathbf{e}_1 \in I_0$ é mostrada na Figura 5.4. Esta reta corresponde a reta de inclinação $\dfrac{3}{5}$ em $\mathbb{R}^2$.

As células máximas I_0, I_1, I_2, I_3, têm suas fronteiras identificadas pela Proposição 5.1.2. Neste caso é fácil ver que $I_0 \cap I_1 = sp^+\{\mathbf{e_0}\}, I_1 \cap I_2 = sp^+\{\mathbf{e}_{-1}\}, I_2 \cap I_3 =$

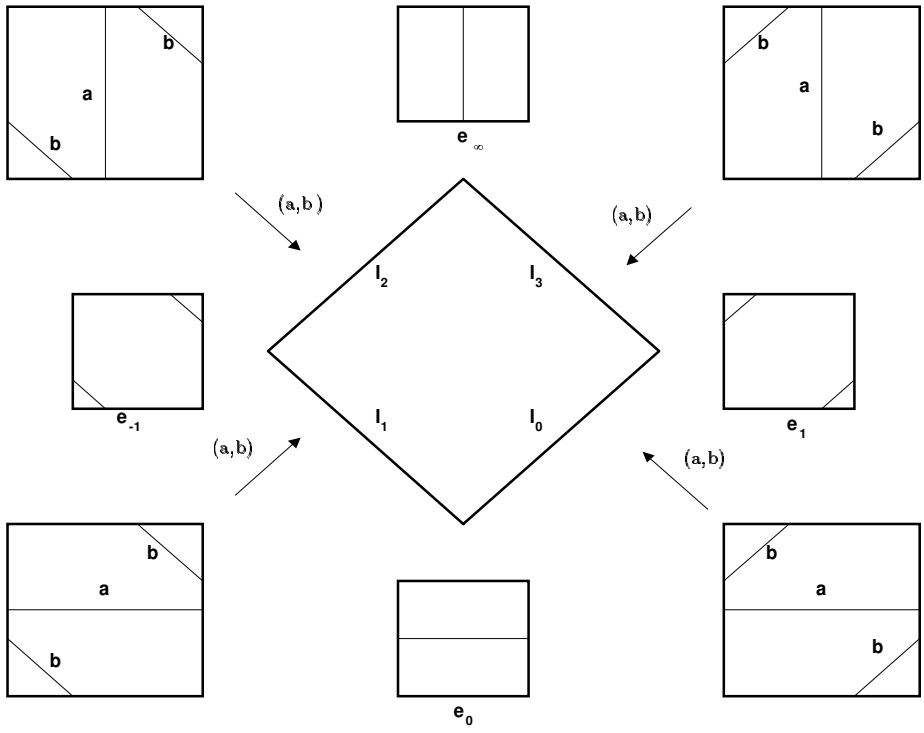

Figura 5.5: A partição de $\mathcal{ML}(\Sigma_1)$ em células maximais.

$sp^+\{\mathbf{e}_\infty\}, I_3 \cap I_0 = sp^+\{\mathbf{e}_1\}, I_0 \cap I_2 = I_1 \cap I_3 = \emptyset$. Isto está ilustrado na Figura 5.5, na qual nota-se claramente que $\mathcal{PML}(\Sigma_1) \sim S^1$.

5.2.2 Torções de Dehn e o grupo de classes de funções $\mathcal{MCG}(\Sigma_1)$

Vamos denotar as torções de Dehn em torno de $\mathbf{e}_\infty$ e $\mathbf{e_0}$ por δ_0 e δ_1, respectivamente. O objetivo é enterder a ação destas torções sobre a estrutura projetiva do espaço dos *train tracks* construídos na subseção anterior.

Para simplificar, usaremos algumas simetrias de R e os π_1-*train tracks*. Essas simetrias são definidas como segue:

- ι_1 intercambia os pares (v_1, v_4) e (v_2, v_3);
- ι_2 permuta ciclicamente os vértices v_1 para v_2, v_2 para v_3 e assim por diante;
- ι_3 fixa v_1 e v_3 e intercambia (v_2, v_4).

Quando necessário for escreveremos ι_0 para denotar a identidade.

Observe que aplicando ι_2 duas vezes obtemos uma rotação de 180° em R a qual permuta v_1, v_3 e v_2, v_4. Embora não seja a identidade sobre R ela age como a tal sobre cada I_j. (Esta função é exatamente a função que leva uma curva sobre si mesma com orientação oposta). Os ι_j agem sobre $\mathcal{PML}(\Sigma_1)$ como o grupo de Klein. Sobre as curvas fechadas irredutíveis sua ação é dada por:

$$\begin{aligned}
&\iota_1 : \mathbf{e_0} \to \mathbf{e_0}, \quad \mathbf{e_1} \to \mathbf{e_{-1}}, \quad \mathbf{e_\infty} \to \mathbf{e_\infty}, \quad \mathbf{e_{-1}} \to \mathbf{e_1}, \\
&\iota_2 : \mathbf{e_0} \to \mathbf{e_\infty}, \quad \mathbf{e_1} \to \mathbf{e_{-1}}, \quad \mathbf{e_\infty} \to \mathbf{e_0}, \quad \mathbf{e_{-1}} \to \mathbf{e_1}, \\
&\iota_3 : \mathbf{e_0} \to \mathbf{e_\infty}, \quad \mathbf{e_1} \to \mathbf{e_1}, \quad \mathbf{e_\infty} \to \mathbf{e_0}, \quad \mathbf{e_{-1}} \to \mathbf{e_{-1}}.
\end{aligned}$$

Esta ação se estende de modo natural para os I_j, como pode ser visto abaixo:

$$\begin{aligned}
\iota_1(I_0) &= I_1,\ \iota_1(I_1) = I_0,\ \iota_1(I_2) = I_3,\ \iota_1(I_3) = I_2, \\
\iota_2(I_0) &= I_2,\ \iota_2(I_1) = I_3,\ \iota_2(I_2) = I_0,\ \iota_2(I_3) = I_1, \\
\iota_3(I_0) &= I_3,\ \iota_3(I_1) = I_2,\ \iota_3(I_2) = I_1,\ \iota_3(I_3) = I_0.
\end{aligned}$$

A vantagem de aplicar essas simetrias é que somente precisamos considerar a ação de δ_0 sobre $\mathcal{ML}(\Sigma_1)$. A ação de δ_0^{-1} e $\delta_1^{\pm 1}$ segue por simetria;

$$\begin{aligned}
\iota_1\delta_0\iota_1 &= \delta_0^{-1}, \iota_2\delta_0\iota_2 = \delta_1, \iota_3\delta_0\iota_3 = \delta_1^{-1} \\
\iota_1\delta_1\iota_1 &= \delta_1^{-1}, \iota_2\delta_1\iota_2 = \delta_0, \iota_3\delta_1\iota_3 = \delta_0^{-1}.
\end{aligned}$$

Isto acontece porque ι_1 e ι_3 revertem orientação, portanto, conjugam torções à direita para torções à esquerda, enquanto ι_2 preserva orientação mas troca $\mathbf{e_0}$ com $\mathbf{e_\infty}$.

Aplicando-se uma torção de Dehn a um π_1-*train track* algumas vezes resulta num π_1-*train track* não reduzido, isto é; um π_1-*train track* que satisfaz as condições (i)-(iii) da Subseção 5.1.2, mas não satisfaz (iv). Existe um processo para converter um π_1-*train track* não reduzido num π_1-*train track* reduzido ilustrado na Figura 5.6, que consiste no seguinte: suponha que o π_1-*train track* τ não reduzido tenha um arco α ligando um lado σ_k a si mesmo e, que este arco tenha peso $w(\alpha)$. Suponha também que τ tenha uma

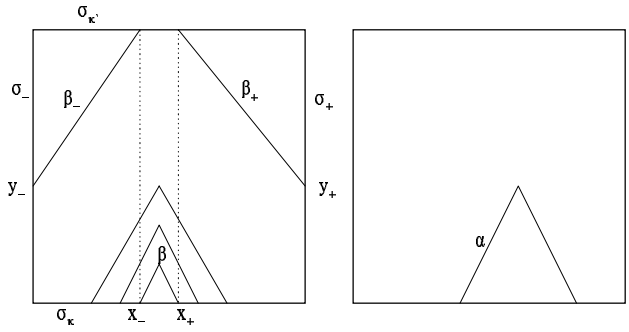

Figura 5.6: Processo de redução de um π_1-*train track* não reduzido.

pesagem inteira formal. Começamos convertendo-o em uma malha simples sobre R. Isto significa que substituímos cada arco α_j com peso $w(\alpha_j)$ por $w(\alpha_j)$ cordões ligando os mesmos pares de lados que os ligados por α_j. Em particular temos $w(\alpha)$ cordões de σ_k sobre si mesmo. Realiza-se então uma homotopia de Σ_1 que removerá as interseções de todos os cordões com σ_k. Isto é feito como segue: seja β o cordão mais interno com pontos terminais x_- e x_+ sobre σ_k. Considere, agora, as imagens de x_- e x_+ sobre o emparelhamento de lados σ_k. Esses pontos estão em σ_k' e são terminais dos cordões β_- e β_+, respectivamente. Os outros pontos terminais de β_- e β_+, são pontos y_- e y_+ sobre os lados σ_- e σ_+. Agora, substitui-se β, β_+ e β_- por um cordão simples indo de y_- para y_+. Com isso, o número de cordões foi reduzido em dois. Este processo termina após um número finito de aplicações, culminando com uma malha simples que não tem cordões com ambos pontos terminais em σ_k. Repetindo o processo para todos os valores de k obtem-se um π_1-*train track* reduzido sobre R com uma pesagem inteira formal.

Proposição 5.2.2 *[25], Seja $(a,b) \in I_0$. As torções de Dehn agem sobre I_0 como segue:*

$$\begin{aligned}
\delta_0(a,b) &= (b, a+b) \in I_3 \\
\delta_0^{-1}(a,b) &= (b,a) \in I_1 \\
\delta_1(a,b) &= \begin{cases} (a-b,b) \in I_0 \ se\ a \geq b \\ (b-a,a) \in I_3 \ se\ a \leq b \end{cases} \\
\delta_1^{-1}(a,b) &= (a+b,b) \in I_0.
\end{aligned}$$

Demonstração. Esta demonstração serve mais como ilustração de como usar os π_1-*train tracks*, como mostrada na Figura 5.7. Seja $(a,b) \in I_{\mathbf{0}}$. Queremos realizar as torções $\delta_j^{\pm 1}$, para $j = 0, 1$, sobre a curva γ_j caracterizada por $\mathbf{e}_\infty$ ou $\mathbf{e_0}$, respectivamente. Vamos considerar uma vizinhança tubular em torno de γ_j como sendo uma faixa indo de um lado para o lado oposto. Esta faixa é limitada por linhas pontilhadas como pode ser observado na Figura 5.7. Como visto anteriormente, a torção de Dehn é a identidade fora desta faixa e dentro dela fixa uma componente da fronteira do cilindro enquanto rotaciona a outra num giro total. Para melhor entendimento estamos interpolando um estágio intermediário, onde mostra de que forma um arco do *train track* foi diretamente rotacionado uma vez ao redor do cilindro antes de emergir do outro lado. Este *train track* ainda possui o mesmo peso representando o número de cordões na correspondente malha simples. Nos casos ilustrados no diagrama da Figura 5.7, considerando o primeiro e quarto casos, o que temos a fazer é juntar os arcos cujos pontos terminais estão no mesmo lado e somar seus pesos. Nos casos intermediários, a imagem dos *train tracks* são *train tracks* não reduzidos e, portanto, precisamos reduzi-los conforme descrito. ■

As torções de Dehn $\delta_{\mathbf{0}}$ e δ_1 geram o grupo de classes de funções $\mathcal{MCG}(\Sigma_1)$. De fato: tem-se $\iota_2 = \delta_1\delta_{\mathbf{0}}\delta_1$, $\iota_2\delta_1\iota_2 = \delta_{\mathbf{0}}$ e somando-se o fato de que ι_2 tem ordem 2 segue que:

$$\delta_1\delta_{\mathbf{0}}\delta_1 = \delta_{\mathbf{0}}\delta_1\delta_{\mathbf{0}}, (\delta_{\mathbf{0}}\delta_1)^3 = e.$$

Isto conduz a presentação do grupo $\mathcal{MCG}(\Sigma_1)$ como sendo

$$\mathcal{MCG}(\Sigma_1) = \langle \delta_{\mathbf{0}}, \delta_1 : \delta_1\delta_{\mathbf{0}}\delta_1 = \delta_{\mathbf{0}}\delta_1\delta_{\mathbf{0}}, (\delta_{\mathbf{0}}\delta_1)^3 = e \rangle.$$

Para obter a identificação de $\mathcal{MCG}(\Sigma_1)$ com $PSL(2,\mathbb{Z})$, considere a identificação de $\mathcal{PML}(\Sigma_1)$ com $\mathbb{R}\cup\{\infty\}$ dada na Subseção 5.2.1.

5.2.3 A função de Markov e os pares de Farey

Definiremos a seguir a função de Markov f sobre $\mathcal{ML}(\Sigma_1)$ da qual construiremos a estrutura automática para o grupo $\mathcal{MCG}(\Sigma_1)$. A partição de Markov consistirá das

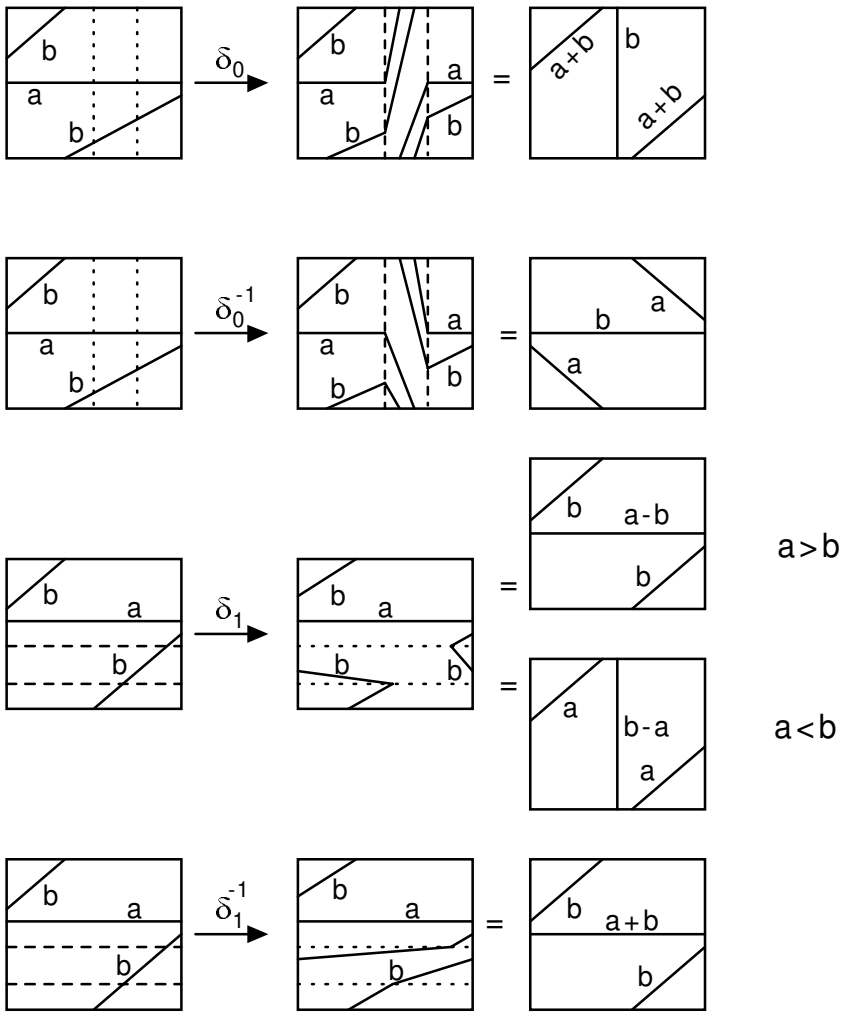

Figura 5.7: A ação das torções de Dehn nos pontos de I_0.

células I_0, I_1, I_2 e I_3. A função de Markov $f|_{I_j}$ será escolhida dentre os elementos do conjunto $\left\{\delta_0^{\pm 1}, \delta_1^{\pm 1}\right\}$ de tal forma que satisfaça a propriedade de Markov.

Lema 5.2.3 *[25],* $\delta_1(I_0) = I_0 \cup I_3$.

Demonstração. Pela Proposição 5.2.2, $\delta_1(I_0) \subset I_0 \cup I_3$. Por outro lado, $\delta_1^{-1}(I_0) \subset I_0$ e $\delta_1^{-1}(I_3) = \iota_3\delta_0\iota_3(\iota_3 I_0) = \iota_3\delta_0(I_0) \subset \iota_3(I_3) = I_0$. ■

Definamos agora a função de Markov como:

$$\begin{aligned} f|_{I_0} &= \delta_1 : I_0 \longmapsto I_0 \cup I_3 \\ f|_{I_1} &= \delta_1^{-1} : I_1 \longmapsto I_1 \cup I_2 \\ f|_{I_2} &= \delta_0 : I_2 \longmapsto I_1 \cup I_2 \\ f|_{I_3} &= \delta_0^{-1} : I_3 \longmapsto I_0 \cup I_3 \end{aligned}$$

Observe que sobre a fronteira $I_i \cap I_j$ a função f tem dois valores e pelo Lema 5.2.3 ela satisfaz a propriedade de Markov vista na Subseção 5.1.1.

Lema 5.2.4 *[25], Seja* $w \in \mathcal{ML}_{\mathcal{O}}(\Sigma_1)$ *uma pesagem inteira formal sobre um* π_1*-train track* τ*. Então* $|f(w)| \leq |w|$ *com igualdade válida se, e somente se,* $w = a\mathbf{e_0}$ *ou* $a\mathbf{e}_\infty$.

Demonstração. Basta lembrar a definição de comprimento dada na Subseção 5.1.2. No nosso caso (para $\Sigma = \Sigma_1$) podemos ver diretamente que se $w = (a, b) \in I_j$, então $|w| = a + 2b$ para $j = 0, 1, 2, 3$, e em seguida usar a Proposição 5.2.2. ■

Uma proposta simplificada de construção de uma máquina de estados para $\mathcal{MCG}(\Sigma_1)$ é usar as quatro células I_0, I_1, I_2, I_3 como estados e definir as arestas usando a matriz de transição associada à função de Markov f. Entretanto, existe um problema com esta proposta, pois o grupo $\mathcal{MCG}(\Sigma_1)$ não age livremente sobre $\mathcal{ML}(\Sigma_1)$, isto é, $\phi(\gamma) = \gamma$ para $\phi \in \mathcal{MCG}(\Sigma_1)$, onde γ é uma curva fechada simples sobre Σ_1, não implica que ϕ é a identidade. Portanto, precisamos considerar a ação do grupo $\mathcal{MCG}(\Sigma_1)$ sobre um espaço mais adequado no qual sua ação seja livre. Para este fim, introduziremos a seguir os pares de Farey.

Duas (classes homótopicas de) curvas fechadas simples sobre Σ_1 são chamadas **vizinhanças de Farey** se elas (têm representantes que) se intersectam exatamente

uma vez. Note que, esta condição por si só implica que estas curvas não dividem o toro puncionado, assim sendo elas não podem ser paralelas à fronteira ou homotopicamente triviais. Chamaremos **pares de Farey** aos pares ordenados de vinzinhanças de Farey (γ_1, γ_2)**.** É claro que $(\mathbf{e_0}, \mathbf{e}_\infty)$ e $(\mathbf{e}_\infty, \mathbf{e_0})$ são ambos pares de Farey. O conjunto de todos os pares de Farey será denotado por $\mathcal{F}$.

Note que $(\mathbf{e_0}, \mathbf{e}_\infty)$ tem estabilidade trivial em $\mathcal{MCG}(\Sigma_1)$ e que, dado qualquer outro par de Farey (γ_1, γ_2), existe um elemento $\phi \in \mathcal{MCG}(\Sigma_1)$ tal que $\phi(\gamma_1, \gamma_2) = (\mathbf{e_0}, \mathbf{e}_\infty)$. Como $(\mathbf{e_0}, \mathbf{e}_\infty)$ tem estabilidade trivial este elemento é único. Em particular ι_2 leva $(\mathbf{e_0}, \mathbf{e}_\infty)$ em $(\mathbf{e_0}, \mathbf{e}_\infty)$. Felizmente, a noção de pares de Farey é compatível com a estrutura de células de $\mathcal{MCG}(\Sigma_1)$ no seguinte sentido:

Proposição 5.2.5 *[25], Seja (γ_1, γ_2) um par de Farey. Se $\{\gamma_1, \gamma_2\} \neq \{e_0, e_\infty\}$, então γ_1 e γ_2 estão ambos contidos na mesma célula I_j para algum $j = 0, 1, 2, 3$.*

Demonstração. Usando a identificação de $\mathcal{PML}(\Sigma_1)$ com $\mathbb{R} \cup \{\infty\}$ (ver Subseção 5.2.1) um par de vizinhanças de Farey corresponde a um par de números racionais $\frac{p}{q}$ e $\frac{r}{s}$, pontos terminais dos triângulos da tesselação vista no Capítulo 4, ou seja, $ps - qr = \pm 1$. Disso segue que $\left\{\pm\frac{p}{q}, \pm\frac{r}{s}\right\} \neq \left\{0 = \frac{0}{1}, \infty = \frac{1}{0}\right\}$, o que mostra que $\frac{p}{q}$ e $\frac{r}{s}$ estão ambos contidos em um dos intervalos $[-\infty, -1], [-1, 0], [0, 1]$ ou $[1, \infty]$. O resultado segue da discussão feita na Subseção 5.2.1. ∎

Sobre uma célula I_j a funcão f é constante e igual a um elemento fixo $\alpha_j \in \mathcal{MCG}(\Sigma_1)$ com possíveis ambigüidades nos pontos terminais. A Proposição 5.2.5 permite-nos estender a ação de f para $\mathcal{F} - \{(\mathbf{e_0}, \mathbf{e}_\infty), (\mathbf{e}_\infty, \mathbf{e_0})\}$. Isto é feito como segue: $f(\gamma_1, \gamma_2) = (\alpha_j(\gamma_1), \alpha_j(\gamma_2))$ sempre que γ_1 e γ_2 estiverem na mesma célula I_j. Como o grupo de classes de funções preserva as vizinhanças de Farey, segue que $(\alpha_j(\gamma_1), \alpha_j(\gamma_2)) \in \mathcal{F}$. Com isso, podemos continuar a interação de f até que $f^n(\gamma_1, \gamma_2) \in \{(\mathbf{e_0}, \mathbf{e}_\infty), (\mathbf{e}_\infty, \mathbf{e_0})\}$. O lema a seguir mostra que esse processo é finito.

Lema 5.2.6 *[25], Suponha que $(\gamma_1, \gamma_2) \in \mathcal{F}$ e $\{\gamma_1, \gamma_2\} \neq \{\mathbf{e_0}, \mathbf{e}_\infty\}$. Então*

$$|f(\gamma_1)| + |f(\gamma_2)| < |\gamma_1| + |\gamma_2|.$$

Demonstração. Consequência do Lema 5.2.4. ■

Suponha, agora, que $\phi \in \mathcal{MCG}(\Sigma_1)$. Como a condição de ser vizinhança de Farey é topológica, o par $(\phi(\mathbf{e_0}), \phi(\mathbf{e}_\infty))$ é ainda uma vizinhança de Farey. A forma normal resulta da seguinte proposição:

Proposição 5.2.7 *[25], Seja $\phi \in \mathcal{MCG}(\Sigma_1)$. Então existe um inteiro não-negativo n tal que $\iota_2^\varepsilon f^n(\phi(\mathbf{e_0}), \phi(\mathbf{e}_\infty)) = (\mathbf{e_0}, \mathbf{e}_\infty)$, onde $\varepsilon = 0$ ou 1.*

Demonstração. Segue da argumentação anterior e do Lema 5.2.6. ■

Observação 5.2.8 *A Proposição 5.2.7, mostra que a ação de f e $\mathcal{MCG}(\Sigma_1)$ sobre o espaço de vizinhaças de Farey são* ***órbitas equivalentes.*** *Em outra palavras, para quaisquer pares de vizinhanças de Farey $\{\gamma_1, \gamma_2\}$ e $\{\gamma_1', \gamma_2'\}$ temos $\{\gamma_1', \gamma_2'\} = \{\phi(\gamma_1), \phi(\gamma_2)\}$ para algum $\phi \in \mathcal{MCG}(\Sigma_1)$ se, e somente se, existem inteiros não negativos m, n tais que $f^n\{\gamma_1, \gamma_2\} = f^m\{\gamma_1', \gamma_2'\}$.*

5.2.4 A Forma normal e a máquina de estados

Denotaremos os pares de Farey $(\mathbf{e_0}, \mathbf{e}_\infty)$ e $(\mathbf{e}_\infty, \mathbf{e_0})$ por K_0 e K_2, respectivamente. Com isso, podemos estender a definição de f fixando: $f_{|_{K_0}} = \iota_0 = e$ e $f_{|_{K_2}} = \iota_2 = \delta_1\delta_0\delta_1$. Assim, o conjunto $\mathcal{F}$ pode ser escrito como $\mathcal{F} = I_0 \cup I_1 \cup I_2 \cup I_3 \cup K_0 \cup K_2$.

O forte da Proposição 5.2.7 é que permite definir a forma normal para elementos de $\mathcal{MCG}(\Sigma_1)$ como segue: Para cada $\phi \in \mathcal{MCG}(\Sigma_1)$, o par $(\phi(\mathbf{e_0}), \phi(\mathbf{e}_\infty))$ pertence a alguma célula U_n, onde U_n é um dos elementos do conjunto $\{I_0, I_1, I_2, I_3, K_0, K_2\}$. Quando da aplicação da função f resulta em mover a seqüência conforme o esquema abaixo;

$$U_n \to U_{n-1} \to \cdots \to U_1 \to U_0 = K_0.$$

Cada célula U_j, $j \geq 2$, é uma das quatro células I_0, I_1, I_2, I_3, enquanto que U_1 é uma das cinco células I_0, I_1, I_2, I_3, K_2. Em cada estágio, $f_{|_{U_j}} = \alpha_j$ é um elemento fixo do conjunto $\left\{\delta_0^{\pm 1}, \delta_1^{\pm 1}\right\}$ (ou possivelmente ι_2 se $U_1 = K_2$). Assim temos

$$f_1^n(\phi(\mathbf{e_0}), (\phi(\mathbf{e}_\infty)) = (\alpha_1\alpha_2\cdots\alpha_n\phi(\mathbf{e_0}), \alpha_1\alpha_2\cdots\alpha_n\phi(\mathbf{e}_\infty) = (\mathbf{e_0}, \mathbf{e}_\infty).$$

Como $\mathcal{MCG}(\Sigma_1)$ age livremente sobre o espaço $\mathcal{F}$, a expressão $\alpha_1\alpha_2\cdots\alpha_n\phi = e$ nos fornece a **forma normal** $\phi = \alpha_n^{-1}\cdots\alpha_1^{-1}$de um elemento arbitrário $\phi \in \mathcal{MCG}(\Sigma_1)$, o que mostra ser $\mathcal{MCG}(\Sigma_1)$ gerado por $\left\{\delta_0^{\pm 1}, \delta_1^{\pm 1}\right\}$.

Exemplo 5.2.9 *Seja $\phi = \delta_0^2\delta_1$ um elemento de $\mathcal{MCG}(\Sigma_1)$. Vamos mostrar que a forma normal para ϕ é $\delta_0\delta_1^{-1}\iota_2$. Da Proposição 5.2.2 temos que $\phi(\mathbf{e_0}) = (1,1) \in I_3$ e $\phi(\mathbf{e}_\infty) = (0,1) \in I_3$. Assim, precisamos achar a f-expansão para o par de Farey $(\phi(\mathbf{e_0}), \phi(\mathbf{e}_\infty)) \in I_3$. Aplicando $f|I_3 = \delta_0^{-1}$ obtemos $f\phi(\mathbf{e_0}) = (0,1) \in I_0$ e $f\phi(\mathbf{e}_\infty) = (1,0) \in I_0$. Aplicando $f|I_0 = \delta_1$ obtemos K_2 e, aplicando $f|K_2 = \iota_2$ voltamos para K_0. Disso segue que*

$$f^3(\phi(\mathbf{e_0}), \phi(\mathbf{e}_\infty)) = (\iota_2\delta_1\delta_0^{-1}\phi(\mathbf{e_0}), \iota_2\delta_1\delta_0^{-1}\phi(\mathbf{e}_\infty)) = (\mathbf{e_0}, \mathbf{e}_\infty),$$

donde concluímos que $\iota_2\delta_1\delta_0^{-1}\phi = e$, o que acarreta $\phi = \delta_0\delta_1^{-1}\iota_2$.

O próximo passo é construir uma máquina de estados que reconheça a forma normal. Para isso precisamos estender a definição da função de Markov para o conjunto $\mathcal{F}$ dos pares de Farey:

$$\begin{aligned}
f|_{I_0} &= \delta_1 : I_0 \longmapsto I_0 \cup I_3 \cup K_0 \cup K_2 \\
f|_{I_1} &= \delta_1^{-1} : I_1 \longmapsto I_1 \cup I_2 \cup K_0 \cup K_2 \\
f|_{I_2} &= \delta_0 : I_2 \longmapsto I_1 \cup I_2 \cup K_0 \cup K_2 \\
f|_{I_3} &= \delta_0^{-1} : I_3 \longmapsto I_0 \cup I_3 \cup K_0 \cup K_2 \\
f|_{K_0} &= e : K_0 \longmapsto K_0 \\
f|_{K_2} &= \iota_2 : K_2 \longmapsto K_0.
\end{aligned}$$

Para construir a máquina de estados definimos os seis estados $I_0, I_1, I_2, I_3, K_0, K_2$. Uma aresta orientada sairá de um estado U para um estado V rotulada por α se $U \subset f(V)$ e $f|_V = \alpha^{-1}$. O estado inicial é K_0. Note que existe no máximo uma aresta rotulada com um dado rótulo saindo de cada estado e que todas as arestas terminando em um determinado estado têm o mesmo rótulo. Esta é a propriedade que toda fonte

deve satisfazer para ser unifilar. Qualquer caminho neste grafo começando em K_0 e seguindo as arestas numa dada direção fornece a forma normal para algum elemento de $\mathcal{MCG}(\Sigma_1)$, bastando, para isso, ler os rótulos das arestas do caminho. Além disso, qualquer $\phi \in \mathcal{MCG}(\Sigma_1)$ tem uma única forma normal dada neste caminho. Para o entendimento do processo, a Figura 5.8, mostra a máquina de estados associada.

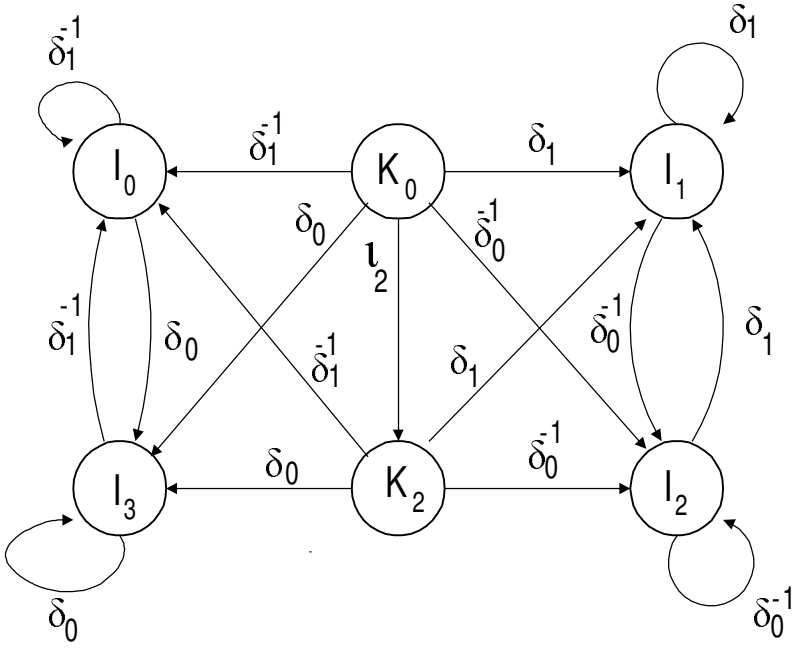

Figura 5.8: A máquina de estados finito.

Considerando também a teoria de composição de funções, as sequências são lidas da direita para a esquerda. Uma observação importante é que as formas normais produzidas por esta máquina são da forma

$$W\left(\delta_0, \delta_1^{-1}\right) \iota_2^{\varepsilon} \quad \text{ou} \quad W\left(\delta_0^{-1}, \delta_1\right) \iota_2^{\varepsilon},$$

onde $W(\alpha, \beta)$ é uma palavra que só contém as letras α e β e o expoente $\varepsilon = 0$ ou 1.

Exemplo 5.2.10 *O elemento* $\phi = \delta_0^2\delta_1 \in \mathcal{MCG}(\Sigma_1)$ *corresponde ao seguinte caminho na máquina de estados.*

$$K_0 \xrightarrow{\iota_2} K_2 \xrightarrow{\delta_1^{-1}} I_0 \xrightarrow{\delta_0} I_3.$$

5.2.5 A estrutura automática para $\mathcal{MCG}(\Sigma_1)$

A fim de produzir uma estrutura automática para o grupo de classes de funções do toro 1-vez puncionado $\mathcal{MCG}(\Sigma_1)$, passamos a explicar como construir a máquina de palavras-diferença. Nosso objetivo é encontrar um conjunto finito $\mathcal{D} \subset \mathcal{MCG}(\Sigma_1)$, chamado **conjunto de palavras-diferença**. Cada um de seus elementos é dito ser uma **palavra-diferença.** Este conjunto deverá ter as seguintes propriedades:

Primeira, o conjunto $\mathcal{D}$ deverá conter a identidade $e \in \mathcal{MCG}(\Sigma_1)$ e todas as letras do alfabeto $\mathcal{A}$. A segunda propriedade é menos simples. Suponha que $\psi \in \mathcal{D}$ e que U, V são dois subconjuntos de I_j para $j = 0, 1, 2, 3$ ou K_j para $j = 0, 2$ com a propriedade de que $\psi(U) = V$. Sejam $f_{|U} = \alpha$ e $f_{|V} = \beta$ as restições da função f à U e V, respectivamente, onde α e β são elementos do alfabeto $\mathcal{A}$. Então $\beta\psi\alpha^{-1}(\alpha(U)) = \beta\psi(U) = \beta(V)$. Daí, tomando $\psi' = \beta\psi\alpha^{-1}$, temos que ψ' leva $\alpha(U)$ em $\beta(V)$. A segunda exigência sobre $\mathcal{D}$ é que para todo $\psi \in \mathcal{D}$ tenhamos $\psi' \in \mathcal{D}$ independente da escolha de U e V. Isto significa que obtivemos um diagrama comutativo, como pode ser visto na Figura 5.9. Tal diagrama será daqui em diante chamado de **quadrado**.

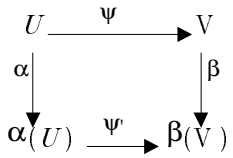

Figura 5.9: Diagrama 1.

Este quadrado corresponde à relação $\psi' = \beta\psi\alpha^{-1}$ em $\mathcal{MCG}(\Sigma_1)$.

Desejamos concatenar vários quadrados verticalmente. Em geral, $\alpha(U)$ e $\beta(V)$ conterão pontos em diversos elementos da partição de $\mathcal{F}$ em I_j e K_j. Isto significa que f pode não ser um elemento fixo de $\mathcal{MCG}(\Sigma_1)$ sobre $\alpha(U)$ e $\beta(V)$. Sejam U' e

V' subconjuntos de U e V, respectivamente, cada um deles pertencente a uma única partição e satisfazendo $\psi'(U') = V'$. Sempre podemos particionar $\alpha(U)$ e $\beta(V)$ em um número finito de sub-intervalos nos quais a propriedade vale. Como U' e V' estão cada um contido num único conjunto da partição, a restrição de f à cada um desses conjuntos é um elemento fixo de $\mathcal{MCG}(\Sigma_1)$. Dessa forma, podemos construir diversos quadrados conforme o modelo mostrado na Figura 5.10, cada um dos quais podendo ser colocado

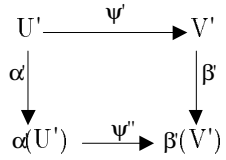

Figura 5.10: Diagrama 2.

um abaixo do outro. Em outras palavras, podemos concatenar quadrados verticalmente.

Um caso especial pode ser observado no Diagrama 3, ilustrado na Figura 5.11, que é quando a palavra-diferença é a identidade e.

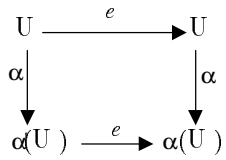

Figura 5.11: Diagrama 3.

Por exemplo, se $U = I_0$, $V = I_1$ e $\psi = \delta_0^{-1}$, pela Proposição 5.2.2 e pelo comentário feito na Subseção 5.2.4, podemos construir o quadrado correspondente ao diagrama contido na Figura 5.12.

A fim de concatenar os quadrados verticalmente, precisamos subdividir os conjuntos das linhas inferiores deste quadrado. Este quadrado pode ser seguido por outros quadrados cujas linhas superiores sejam uma das descritas a seguir.

$I_0 \xrightarrow{\iota 2} I_2, \quad I_3 \xrightarrow{\iota 2} I_1, \quad K_0 \xrightarrow{\iota 2} K_2, \quad K_2 \xrightarrow{\iota 2} K_0.$

Por exemplo, tal quadrado poderia ser seguido pelo quadrado com $U' = I_0$ e $V' = I_2$, conforme Diagrama 5.

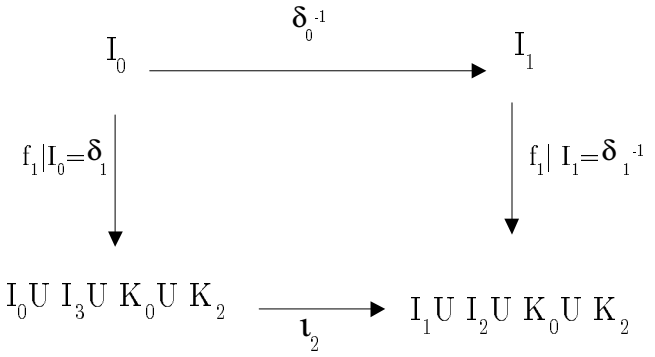

Figura 5.12: Diagrama 4.

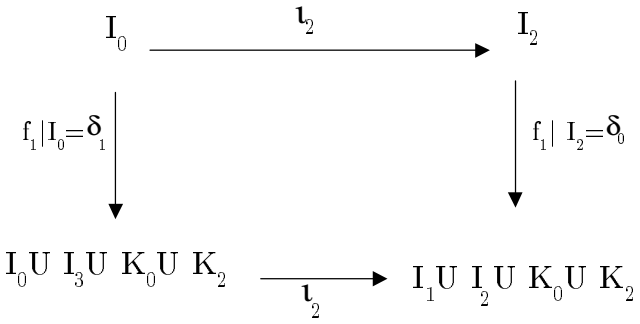

Diagrama 5.

A fim de simplificar tais diagramas, consideremos

$$Q_0 = I_0 \cup I_3 \cup K_0 \cup K_2, \quad \text{e} \quad Q_2 = I_1 \cup I_2 \cup K_0 \cup K_2.$$

Se $\psi = \alpha$ ou $\psi = \beta^{-1}$ definamos os quadrados degenerados ou triângulos como segue: no caso onde $\psi = \alpha$, não aplicamos f à V. Isto significa que V pode conter pontos de diversos conjuntos na partição. Portanto, podemos tomar $V = \alpha(U)$. De modo análogo, se $\psi = \beta^{-1}$ definimos um triângulo não aplicando f à U. Em ambos os casos, ψ' é a função identidade e. A Figura 5.13. ilustra este comentário.

Vamos

mostrar que podemos tomar o conjunto de palavras-diferença como sendo

$$\mathcal{D} = \{\mathbf{e}, \boldsymbol{\delta_0}, \boldsymbol{\delta_0^{-1}}, \boldsymbol{\delta_1}, \boldsymbol{\delta_1^{-1}}, \iota_2\}.$$

Primeiramente, indicamos todos os quadrados para os quais $U \subset I_0$. Nos casos

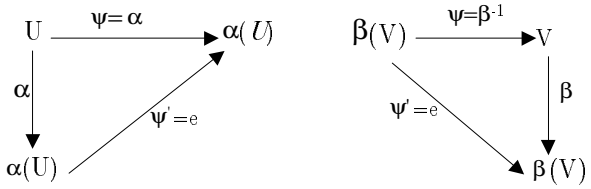

Figura 5.13: Diagrama 6.

dos triângulos onde $\psi = \beta^{-1}$ substituímos U por $\beta(V)$ e incluímos os casos nos quais $I_0 \subset \beta(V)$ na nossa relação.

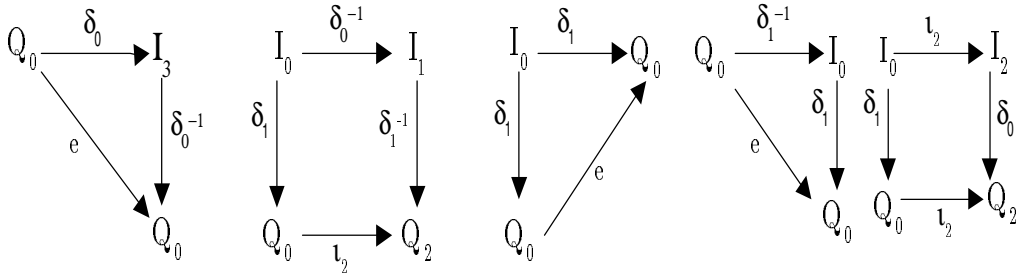

Figura 5.14: Diagrama 7.

Os quadrados e triângulos onde $U \in I_1, I_2$ ou I_3 ou $\beta(V) = Q_2$ podem ser obtidos a partir dessas simetrias.

Finalmente, suponhamos que $U = K_0$ ou K_2. Se a palavra-diferença ψ é $\delta_j^{\pm 1}$ para $j = 0, 1$, os triângulos e quadrados relevantes já estão incluídos na lista acima. Além disso, para tais palavras-diferença, a nova palavra-diferença é e. Por outro lado, se a palavra-diferença é ι_2, então obtemos os triângulos mostrados nos diagramas da Figura 5.15.

A Figura 5.16, nos dá uma idéia da máquina de palavras-diferença.

Vamos explicar como construir a máquina de palavras-diferença referente a esses quadrados e triângulos. Seguindo o esboço feito na Subseção 5.1.1, os estados da máquina deveriam ser uma tripla (U, V, ψ) que aparece na linha superior de um dos nossos quadrados ou triângulos. De qualquer modo, como a única função da escolha de U e V é determinar o valor da função f, podemos tomar os estados da máquina

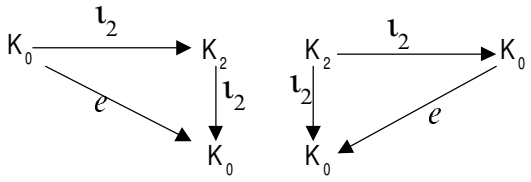

Figura 5.15: Diagrama 8.

de palavras-diferença como sendo os elementos de $\mathcal{D}$. As arestas serão pares ordenados $(x, y) \in \mathcal{A} \cup \{-\}$, onde x e y são essencialmente os inversos de α e β. Em outras palavras, dado um quadrado da forma ilustrada nas Figuras 5.9 e 5.10, traçamos uma aresta de ψ' para ψ e rotulamo-la por $(\alpha^{-1}, \beta^{-1})$. Analogamente, os triângulos ilustrados na Figura 5.13, correspondem às arestas de e para ψ são rutuladas por $(\alpha^{-1}, -)$ e $(-, \beta^{-1})$, respectivamente. A Figura 5.16 ilustra estes fatos. Em adição a estas, deve haver arestas rotuladas (α, α) de e para si mesmo, para todo $\alpha \in \mathcal{A}$.

É automático da construção que para qualquer caminho na máquina de palavras-diferença com arestas rotuladas por $(\alpha_j^{-1}, \beta_j^{-1})$, as arestas rotuladas por α^{-1} e β^{-1} são arestas da máquina de estados finito ilustrada na Figura 5.8.

Um resultado desta construção é que verificamos a presentação de $\mathcal{MCG}(\Sigma_1)$ (Subseção 5.2.2). Para ver isto, observe que qualquer caminho fechado no grafo de Cayley pode ser decomposto numa união de triângulos cada um dos quais tendo um lado de comprimento no máximo um e os outros dois lados são caminhos na forma normal voltando para a identidade. Isto forma um diagrama (conhecido como diagrama de Van Kampen) para os caminhos fechados cobrindo-o com triângulos e quadrados da forma que construímos. Não é difícil verificar que a cada um desses triângulos e quadrados corresponde uma relação que pode ser derivada de $\delta_1\delta_0\delta_1 = \delta_0\delta_1\delta_0$ ou $(\delta_1\delta_0)^3 = e$.

Esse procedimento possibilitou a construção do automata de comprimento 2 para a máquina de palavras-diferença. Existe ainda um problema técnico a ser superado. A saber, a máquina é assíncrona. Isto porque, alguns de seus rótulos têm a forma $(\alpha^{-1}, -)$ ou $(-, \beta^{-1})$. De fato, isto ocorrerá exatamente uma vez quando estamos trabalhando com pares de elementos do grupo que diferem por uma palavra de comprimento exatamente

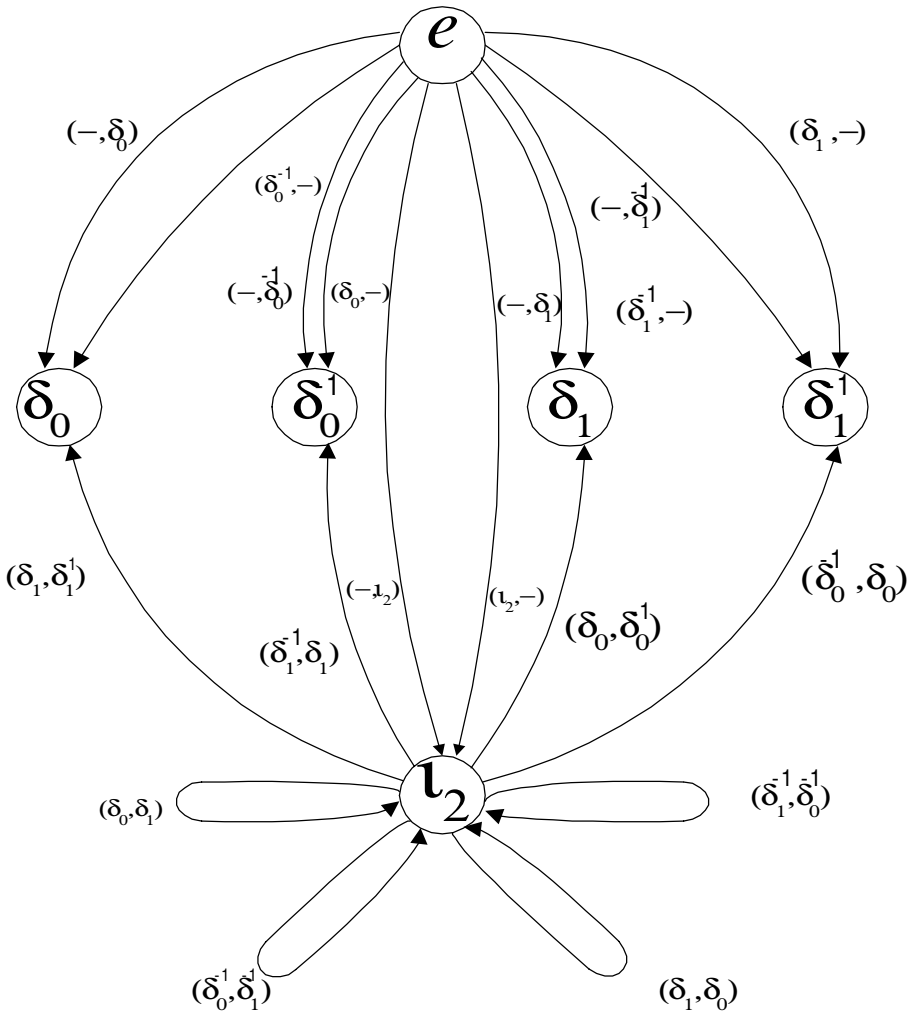

Figura 5.16: Máquina de palavras-diferença.

um. Especificamente, a princípio todas as formas normais dos prefixos da máquina de estados diferem pelo símbolo "$-$" em uma das arestas na máquina de palavras-diferença. Isto porque, todas as arestas saindo de e para qualquer outro estado tem esta forma e nenhuma outra tem.

Para contornar essa dificuldade precisamos sincronizar a máquina de palavras-diferença. Isto é feito como segue. Na definição do automata de comprimento 2, precisamos colocar um símbolo adicional \$ no final de uma das palavras para assegurar que todas as palavras tenham o mesmo comprimento. Assim, precisamos mover o símbolo "$-$" no meio da palavra para um símbolo \$ no fim da palavra. Isto é conseguido

adicionando ao conjunto de palavras-diferença as diagonais em cada quadrado. Isto é, para quadrados do tipo ilustrado nas Figuras 5.9 e 5.10, as palavras-diferença diagonais $\beta^{-1}\psi' = \psi'\alpha^{-1}$ e $\beta\psi = \psi'\alpha$. Observe que, por inspeção, essa nova palavra-diferença pode ser rearranjada na forma $\iota_2\delta_j^{\pm 1}$ para $j = 0, 1$. Isto tem o seguite efeito: suponha que a forma normal para a máquina de estados diferindo por ψ são $\alpha_1^{-1}\alpha_2^{-1}\alpha_3^{-1}\alpha_4^{-1}$ e $\beta_1^{-1}\beta_2^{-1}\alpha_4^{-1}$. Os diagramas mostrados na Figura 5.17, fornecem os caminhos na máquina de palavras-diferença (lidos de baixo para cima), os correspondentes quadrados, os quadrados emendados e o caminho na máquina diferença sincronizada.

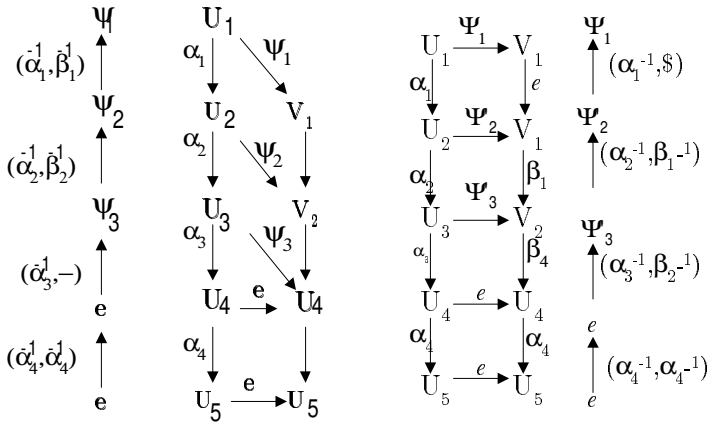

Figura 5.17: Diagrama 9.

Aqui, ψ_2, ψ_3, ψ_2' e ψ_3' são escolhidos de tal forma que os dois diagramas comutem. Está claro como mudar a máquina de palavras-diferença à luz desse exemplo. Naturalmente, existem agora mais alguns estados e arestas na máquina de palavras-diferença sincronizada. Os novos estados são

$$\mathcal{D}' = \{e, \boldsymbol{\delta}_0, \boldsymbol{\delta}_0^{-1}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_1^{-1}, \iota_2, \iota_2\boldsymbol{\delta}_0, \iota_2\boldsymbol{\delta}_0^{-1}, \iota_2\boldsymbol{\delta}_1, \iota_2\boldsymbol{\delta}_1^{-1}\}.$$

Capítulo 6

Conclusões e Sugestões para Trabalhos Futuros

6.1 Conclusões

Geodésicas fechadas sobre superfícies modulares associadas à classes de conjugação de matrizes hiperbólicas em $SL(2,\mathbb{Z})$ podem ser codificadas de duas formas distintas: os códigos aritméticos e os códigos geométricos. Neste trabalho é feita uma descrição completa das geodésicas fechadas para as quais tais códigos coincidem. Para tal o plano hiperbólico foi tesselado por triângulos de duas formas distintas, diferenciando portanto as classes de geodésicas codificáveis.

De modo a fornecer suporte à extensão desta pesquisa, foram construídos: o grupo de classes de funções $\mathcal{MCG}(\Sigma_1)$ do toro 1-vez puncionado associado à tesselação do plano hiperbólico agora por quadrados; uma máquina de estados finitos, que mostra de que forma o grupo de classes de funções desta particular superfície age sobre o conjunto $\mathcal{F}$ das células maximais; e a máquina de palavras-diferença que estabelece sua estrutura automática do grupo $\mathcal{MCG}(\Sigma_1)$.

Na seqüência, passamos mencionar os fundamentos e contribuições.

No Capítulo 1, foi feita uma introdução do trabalho, onde foi mostrada a forma de como o mesmo se insere no contexto da dinâmica simbólica, [14].

No Capítulo 2, foram fornecidos os fundamentos matemáticos necessários para o desenvolvimento do presente trabalho.

No Capítulo 3, foram revisados os seguintes conceitos de teoria da informação: informação e incerteza, entropia de uma fonte, o teorema de codificação de fontes discreta sem memória e algumas técnicas de codificação de fontes discretas, a saber: códigos de Huffman, Elias e os métodos de Lempel-Ziv.

No Capítulo 4, foram introduzidas técnicas de codificação de geodésicas fechadas associadas a classes de matrizes hiperbólicas em $SL(2,\mathbb{Z})$ em duas formas diferentes. A codificação geométrica e a codificação aritmética. Baseado na Observação 4.1.18, notamos a existência de um código aritmético coincidente com um código geométrico dado. Essa coincidência foi estabelecida no Teorema 4.1.21, (*pág.*62). Este resultado fomenta a busca de códigos ótimos no contexto hiperbólico.

A codificação geométrica: para este modelo de codificação foram fornecidas duas técnicas. Tais técnicas consistem em rotular segmentos geodésicos quando estes cortam os polígonos da tesselação do plano hiperbólico. A primeira, foi obtida com a tesselação canônica, isto é, por triângulos com um vértice no infinito e os outros dois em $\mathbb{H}^2$. Para esta técnica, foram caracterizadas todas as geodésicas codificáveis. A segunda, foi obtida tesselando-se o plano hiperbólico através da tesselação de Farey, isto é, através de triângulos cujos vértices estão na fronteira do disco de Poincaré.

A codificação aritmética foi obtida considerando a expansão em frações contínuas do ponto fixo atrator da transformação de Möbius associada a uma dada matriz $A \in SL(2,\mathbb{Z})$.

Ainda neste capítulo, foi feita uma ampla discussão sobre frações contínuas, onde foi inserida um seção de exemplos com o objetivo de explicitar como se calcula a expansão em frações contínuas de um número real a qual está diretamente relacionada o código aritmético.

Tal como as técnicas tradicionais de codificação de fontes discretas, as técnicas em consideração também apresentam uma certa imprecisão quando consideramos seqüências de comprimento pequeno.

No Capítulo 5, foi estentido o estudo feito no Capítulo 4, considerando o plano

hiperbólico tesselado por quadrados cujos vértices se situam na fronteira do disco de Poincaré, com a particularidade de que tais vértices são equivalentes e consistindo de um ponto excepcional, a saber: o infinito. A codificação foi realizada usando-se a dinâmica de uma máquina de estados cujos estados são os elementos da partição de Markov, ligados por arestas rotuladas pelos geradores do grupo de classes de funções da superfície modular representada pela região fundamental (quadrado hiperbólico).

A fim de produzir uma estrutura automática para o grupo de classes de funções do toro 1-vez puncionado $\mathcal{MCG}(\Sigma_1)$, foi construída a máquina de palavras-diferença.

6.2 Sugestões para Trabalhos Futuros

- Concatenar a presente teoria com a solução (de sistemas) de equações diferenciais parciais (a modelagem matemática de determinados fenômenos físicos, ondas, calor, etc) para transmissão de dados através de (por exemplo) linhas de transmissão.

- Extensão das técnicas de codificação ora propostas, tesselando-se o plano hiperbólico por polígonos de n lados, para $n \geq 5$.

- Buscar condições para garantir a otimalidade desses códigos utilizando-se da coincidência ("unicidade") entre os códigos aritméticos e geométricos.

Referências Bibliográficas

[1] Gonçalves, A. , *Introdução à Álgebra,* Projeto Euclides, 1982.

[2] Alder, R. and Flato, L., *Cross section maps for geodesic flows*, Ergodic theory and dynamical systems, Progress in Mathematics 2 (ed. A. Katok, Birkhäuser, Boston, 1980).

[3] Artin, E., "*Ein Mechanisches System mit quasiergodischen Bahnen*", *Collected papers* (Addison Wesley, Reading, Mass., 1965), pp. 499-501.

[4] Birman, J. S., Series, C., *Algebraic linearity for an automorphism of a surface group*, Journal of Pure and Applied Algebra **52** (1988), 227−275.

[5] Bowen, R. & Series, C. *Markov maps associated to fuchsian groups,* Inst. Hautes Études Sci. Publ. Math. **50** (1979), 153-170.

[6] Cover, T. M., *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.

[7] Felipe, M. S., *Alguns Problemas de Contornos para o Modelo de Kirchhof-Carrier*, Dissertação de Mestrado, IME-UFRJ, Julho, 1995. Rio de Janeiro RJ.

[8] Firer, M., *Grupos Fuchsianos,* Imecc-Unicamp, Fevereiro de 1999.

[9] Fraleigh, J. B., *A First Course in Abstract Algebra.* Fifth edition, Addison Wesley Publishing Company.

[10] Gersting, J. L., *Fundamentos Matemáticos para a Ciência da Computação. 3ª* Edição, Editora LTC.

[11] Hardy, G. and Wright, E.M., *An Introduction to the Theory of Numbers* (university Press, Oxford, 1975.

[12] Hedlund, G. D., *A metrically transitive group defined by the modular group,* Amer. J. Math. 57 (1935), 668-678.

[13] Herstein, I. N., *Topics in Algebra* , 2nd. Edition. New York, Wiley, 1975.

[14] Lind D. and Briam M., *Symbolic Dynamics and Coding,* Cambridge University Press, 1995.

[15] Katok, S., *Reduction theory for fuchsian groups*, Mathematische Annalen, 273, 461-470(1986).

[16] Katok, S., *Coding of geodesics after Gauss and Morse*, Geometriae Dedicata, **63:** 123-145, 1996.

[17] Katok, S., *Fuchsian Groups,* The University of Chicago Press, Chicago and London, 1992.

[18] Komolgorov, A.N., *Three approaches to the quantitative definition of information.* Probl. Inform. Transm., 1(1):1-7, 1965.

[19] Lazari, H., *Uma Contribuição à Teoria dos Códigos Geometricamente Uniformes Hiperbólicos*, Unicamp, Feec, DT, Tese de Doutorado, Fevereiro, 2000.

[20] Lempel, A., and Ziv, J., *A universal algorithm for sequential data compression.* IEEE Transactions on Information Theory, IT-23(3) : 337-343, Maio 1977.

[21] Lempel, A., and Ziv, J., *Compression of individual sequences via variable-rate coding.* IEEE Transactions on Information Theory, IT-25(5) : 530-536, Setembro de 1978.

[22] Lima, E.L., *Espaços Métricos,* Rio de Janeiro, IMPA, 1977

[23] Morse, M., *Symbolic Dynamics,* (unpublished), Institute for Advanced Study Notes, Princeton (1966).

[24] Palazzo, R.Jr., *Teoria da Informação e Codificação: Fundamentos e Aplicações*, UNICAMP-FEEC-DT, Campinas, 2000.

[25] Parker, J.R., Series, C. *The mapping class group of the twice punctured torus.* Preprint.

[26] Penner, R.C., Harer, J.L., *Combinatorics of train tracks*, Annals of Maths. Studies **125**, Princeton University Press, 1992.

[27] Series, C., *Geometrical Markov coding of geodesic on surfaces of constant negative curvature,* Ergod. Th. & Dynam. Sys. **6** (1986), 601-625.

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

[29] Series, C., *On coding geodesics with continued fractions,* Enseign. Math., 29 (1980), 67-76. (*On coding geodesics,* Enseign. Math., 29 (1983), 67-76 ou Ergodic theory (Sem, Les Plans-Sur-Bex, 1980) pag. 67-76, Monograph. Enseign. Math., 29 Univ. Genève - Geneva, 1981)

[30] Series, C., *The modular surface and continued fractions,* J. London Math. Soc. (2), 31 (1985), 69-80.

[31] Series, C., *Geometrical methods of symbolic coding,* in "Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces", ed. T. Bedford, M. Keane & C. Series, Oxford University Press, (1991), 125-151.

[32] Shannon, C.E., *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, Jully, October, 1948.