

UNIVERSITÀ DEGLI STUDI DI
CAMERINO

FACOLTÀ DI SCIENZE E TECNOLOGIE

*Corso di Laurea in Matematica e
Applicazioni Gestionali e Tecnologiche*

Dipartimento di Matematica e Informatica



IRRAZIONALI QUADRATICI,
FRAZIONI CONTINUE E FLUSSI
GEODETICI CHIUSI SULLA
SUPERFICIE MODULARE

Tesi di Laurea in Fisica Matematica

Relatore:

Prof. Stefano Isola

Laureando:

Marcello Seri

ANNO ACCADEMICO 2005-2006

La matematica è l'arte di dare lo stesso nome a cose diverse.

– Henri Poincaré

Indice

1	Cenni preliminari sulle azioni di gruppo	5
2	Il flusso geodetico su \mathbb{H}	10
2.1	L'azione di $SL_2(\mathbb{R})$ su \mathbb{H}	10
2.2	La curvatura di Gauss di \mathbb{H}	14
2.3	Alcune proprietà di \mathbb{H}	15
2.4	Calcolo dell'area di un triangolo iperbolico asintotico.	16
2.5	Un modo alternativo di calcolare la curvatura di \mathbb{H}	18
2.6	Geodetiche di \mathbb{H}	20
3	La superficie modulare	23
3.1	Quozienti di \mathbb{H}	23
3.2	Isometrie nel semipiano iperbolico	27
3.3	Classificazione della trasformazioni lineari frazionarie	29
3.3.1	Trasformazioni paraboliche	30
3.3.2	Trasformazioni ellittiche	32
3.3.3	Trasformazioni iperboliche	33
3.4	Domini fondamentali	35
4	Sviluppi in frazioni continue	41
4.1	Definizioni preliminari	41

4.2	L'algoritmo di Euclide	42
4.3	Una relazione di ricorrenza	44
4.4	Frazioni continue, convergenti e proprietà	46
4.5	La differenza tra frazione continua e convergente	52
4.6	Frazioni continue infinite	54
4.7	Numeri irrazionali e frazioni continue	55
4.8	Ultimi preliminari per arrivare agli irrazionali quadratici	57
4.9	Irrazionali quadratici e frazioni continue periodiche	62
5	Dinamica simbolica e flusso geodetico su una superficie modulare	66
5.1	Il flusso geodetico su una superficie modulare	66
5.2	Dinamica simbolica del flusso geodetico per le superfici modulari	68
5.3	Irrazionali quadratici e geodetiche chiuse sulla superficie modulare	72
5.4	Calcolo del periodo delle geodetiche chiuse	73
6	Alcuni appunti sulla mappa di Gauss	75
6.1	Prime proprietà strutturali	75
6.2	Breve revisione del rapporto tra frazioni continue e mappa di Gauss	77
6.3	Ergodicità della trasformazione di Gauss	79
6.4	Il teorema ergodico di Birkhoff	81
6.5	Teoria dei numeri e mappa di Gauss	82
6.6	Un'ultima curiosità	88
7	La tassellazione di Farey e le geodetiche di scattering	91
7.1	La mappa di Farey e i numeri razionali	91

7.1.1	I convergenti di Farey	91
7.1.2	L'albero di Farey	93
7.1.3	Codifiche	96
7.2	La tassellazione di Farey e le geodetiche di scattering	98
8	Brevi note conclusive	102
	Bibliografia	104

Capitolo 1

Cenni preliminari sulle azioni di gruppo

Per cominciare è necessario riprendere alcuni concetti sulle azioni continue di gruppo. Gli spazi (localmente) compatti saranno considerati di Hausdorff. Si userà la notazione $[x]$ per indicare la classe di equivalenza contenente x .

Un gruppo G su cui è definita una topologia è un *gruppo topologico* se le mappe

$$(g, g') \mapsto gg' : G \times G \rightarrow G, \quad g \mapsto g^{-1} : G \rightarrow G$$

sono continue.

Due elementi $g_1, g_2 \in G$ gruppo, si dicono *coniugati* se $\exists h \in G$ tale che $g_2 = hg_1h^{-1}$. Si può osservare facilmente che la coniugazione è una relazione di equivalenza.

Siano G un gruppo topologico ed X uno spazio topologico. Un'azione di

G su X ,

$$(g, x) \mapsto gx : G \times X \rightarrow X$$

è *continua* se la sua mappa è continua. Allora, $\forall g \in G, x \mapsto gx : X \rightarrow X$ è un omeomorfismo (con inversa $x \mapsto g^{-1}x$); l'insieme $Gx = \{gx : g \in G\}$ delle immagini di una $x \in X$ si chiama *orbita* dell'azione.

Lo *stabilizzatore* di $x \in X$ (o *gruppo di isotropia* di x) è l'insieme

$$\text{Stab}(x) = \{g \in G : gx = x\}.$$

Se X è di Hausdorff, allora $\text{Stab}(x)$ è chiuso ¹.

Si può notare anche l'esistenza della biiezione

$$G/\text{Stab}(x) \rightarrow Gx, \quad g \cdot \text{Stab}(x) \mapsto gx;$$

in particolare, quando G agisce transitivamente² su X , questa diventa

$$G/\text{Stab}(x) \rightarrow X.$$

$G \backslash X$ è l'insieme delle orbite dell'azione³ di G su X , $G \backslash X$ è dotato naturalmente della topologia quoziente: se p denota la mappa $x \mapsto Gx : X \rightarrow G \backslash X$, allora $U \subset G \backslash X$ è aperto se e solo se $p^{-1}(U)$ è aperto in X . Si noti che $p : X \rightarrow G \backslash X$ è sia continua che aperta infatti:

- é continua per definizione avendo dato a $G \backslash X$ la più sottile topologia per cui p è una mappa continua;
- sia U un sottoinsieme aperto di $G \backslash X$, dobbiamo mostrare che $p(U)$ è aperto; ma $p^{-1}(p(U)) = \cup_{g \in G} gU$ che è chiaramente aperto.

¹In quanto controimmagine di x rispetto alla funzione $g \mapsto gx : G \rightarrow X$.

²Dunque esiste una sola orbita, infatti l'azione di un gruppo topologico G su uno spazio topologico X si dice *transitiva* se $\forall x_1, x_2 \in X, \exists g \in G$ tale che $x_1 = gx_2$.

³Ossia l'insieme delle classi di equivalenza $[x] = Gx$.

Sia H un sottogruppo di G . H agisce su G a destra e a sinistra e G/H e $H \backslash G$ sono gli spazi dei laterali destri e sinistri.

LEMMA 1.1 *Lo spazio G/H è di Hausdorff se e solo se H è chiuso in G .*

DIM. Si chiami p la mappa $G \rightarrow G/H$, $g \mapsto gH$. Se G/H è di Hausdorff, il punto eH in G/H è un chiuso dunque $H = p^{-1}(eH)$ è un chiuso (e =identità in G).

Viceversa, sia H un sottogruppo chiuso e siano $aH, bH \in G/H$ due elementi distinti. Poiché G è un gruppo topologico, la mappa

$$f : G \times G \rightarrow G, \quad (g, g') \mapsto g^{-1}g',$$

è continua, quindi $f^{-1}(H)$ è chiuso. $aH \neq bH$ dunque $(a, b) \notin f^{-1}(H)$ (altrimenti $a, b \in H$ e $aH = bH$), allora c'è un intorno aperto di (a, b) , che possiamo prendere della forma $U \times V$, disgiunto da $f^{-1}(H)$. Per quanto osservato, le immagini di U e V in G/H sono intorni aperti disgiunti di aH e bH .

□

Se G agisce transitivamente su X , c'è una biiezione

$$G/\text{Stab}(x) \rightarrow X \quad \forall x \in X.$$

Con qualche ipotesi in più si può avere un omeomorfismo.

TEOREMA 1.2 *Si supponga che G agisca transitivamente e con continuità su X . Se G ed X sono localmente compatti e di Hausdorff e G ammette una base numerabile per la topologia, allora la mappa*

$$[g] \mapsto gx : G/\text{Stab}(x) \rightarrow X$$

è un omeomorfismo.

DIM. È già noto che l'applicazione è una biiezione. Dalla definizione è ovvio che sia continua, resta da dimostrare che è aperta.

Sia U un sottoinsieme aperto di G e sia $g \in U$, si deve mostrare che gx è un punto interno di Ux .

Si consideri la mappa $G \times G \rightarrow G$, $(h, h') \mapsto gh'h'$. Tale mappa è continua e manda (e, e) in U , quindi c'è un intorno V di e , che si può considerare compatto (non necessariamente aperto), tale che $V \times V$ sia mappato in U ; quindi $gV^2 \subset U$. Dopo aver rimpiazzato V con $V \cap V^{-1}$, si può assumere $V^{-1} = V$. (Si intende $V^{-1} = \{h^{-1} : h \in V\}$; $V^2 = \{hh' : h, h' \in V\}$).

Se $e \in V$, $G = \cup_{g \in G} gV$. Ogni gV è l'unione di aperti della base numerabile, si prendano un numero sufficiente di g per fare in modo che ogni insieme della base sia contenuto in almeno un gV . A questo punto c'è un insieme numerabile di elementi $g_1, g_2, \dots \in G$ tali che $G = \cup g_n V$.

Essendo gli $g_n V$ compatti, la loro immagine $g_n Vx$ in X è compatta, ma poiché X è di Hausdorff, $g_n Vx$ è anche chiusa.

Il lemma seguente mostra che almeno uno degli $g_n Vx$ ha un punto interno. Ma $y \mapsto g_n y : X \rightarrow X$ è un omeomorfismo che manda Vx in $g_n Vx$, quindi Vx ha un punto interno, cioè \exists un punto $hx \in Vx$ ed un sottoinsieme aperto W di X tali che $x \in W \subset Vx$. Dunque basta osservare che

$$gx = gh^{-1} \cdot hx \in gh^{-1}W \subset gV^2x \subset Ux$$

per dimostrare che gx è un punto interno di Ux .

□

LEMMA 1.3 *Sia X uno spazio non vuoto e localmente compatto (e di Hausdorff) tale che $X = \cup V_n$ dove (V_n) è una famiglia numerabile di sottoinsiemi chiusi. Allora almeno uno degli V_n ha un punto interno.*

DIM. Le ipotesi implicano che X sia regolare: i punti di X sono chiusi e per ogni punto x non contenuto in un insieme chiuso A , ci sono in X due aperti disgiunti U e V che contengono x ed A rispettivamente.

Si supponga per assurdo che nessun V_n abbia un punto interno. Sia U_1 un sottoinsieme aperto non vuoto di X che abbia chiusura \bar{U}_1 compatta. Poiché V_1 non ha punti interni, U_1 non può essere contenuto in V_1 ; inoltre, poiché U_1 è regolare, c'è un sottoinsieme aperto non vuoto $U_2 \subset U_1$ tale che $\bar{U}_2 \subset U_1 - V_1$. Procedendo in questo modo si ottiene una successione di sottoinsiemi aperti non vuoti U_3, U_4, \dots tali che $\bar{U}_{n+1} \subset U_n - V_n$. Gli U_n sono una successione di compatti non vuoti, per cui $\bigcap \bar{U}_n \neq \emptyset$ che contraddice $X = \bigcup V_n$.

□

Capitolo 2

Il flusso geodetico su \mathbb{H}

Si indicherà con \mathbb{H} il semipiano superiore di Poincaré

$$\mathbb{H} = \{z \in \mathbb{C} : z = x + iy, y > 0\}$$

dotato della metrica iperbolica

$$ds^2 = \sum_{i,j=1}^2 g_{ij} dx^i dx^j = \frac{dx^2 + dy^2}{y^2} = \frac{dzd\bar{z}}{(\Im z)^2} \quad \text{con} \quad [g_{ij}] = \begin{pmatrix} \frac{1}{y^2} & 0 \\ 0 & \frac{1}{y^2} \end{pmatrix} \quad (2.1)$$

2.1 L'azione di $SL_2(\mathbb{R})$ su \mathbb{H}

Su \mathbb{H} si può far agire il gruppo

$$SL_2(\mathbb{R}) = \left\{ \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R}, \det \alpha = 1 \right\}$$

come segue:

$$\Theta : SL_2(\mathbb{R}) \times \mathbb{H} \rightarrow \mathbb{R}, \quad (\alpha, z) \mapsto \Theta(\alpha, z) = \alpha(z) = \frac{az + b}{cz + d}, \quad \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (2.2)$$

LEMMA 2.1 *L'applicazione (2.2) è effettivamente un'azione di gruppo.*

DIM. Innanzitutto c'è da dimostrare che $z \in \mathbb{H} \implies \alpha(z) \in \mathbb{H}$.

$$\Im(\alpha z) = \Im\left(\frac{az+b}{cz+d}\right) = \Im\left(\frac{(az+b)(c\bar{z}+d)}{|cz+d|^2}\right) = \frac{\Im(adz+bc\bar{z})}{|cz+d|^2} = \frac{\Im(z)}{|cz+d|^2}$$
 infatti $(az+b)(c\bar{z}+d) = ac|z| + bd + adz + bc\bar{z}$ e $ac|z| + bd \in \mathbb{R}$, inoltre $\Im(adz+bc\bar{z}) = (ad-bc) \cdot \Im(z) = \det(\alpha) \cdot \Im(z) = \Im(z)$. Dunque se $\Im(z) > 0$, di certo anche

$$\Im(\alpha z) = \frac{\Im(z)}{|cz+d|^2} > 0, \quad (2.3)$$

per cui $\Im(\alpha z) \in \mathbb{H}$.

Per essere effettivamente un'azione di gruppo deve anche accadere che $\Theta(\alpha_2, \Theta(\alpha_1, z)) = \Theta(\alpha_2\alpha_1, z)$. Siano

$$\alpha_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \quad \text{e} \quad \alpha_2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$$

si ha che

$$\begin{aligned} \Theta(\alpha_2\alpha_1, z) &= \begin{pmatrix} a_2a_1 + b_2c_1 & a_2b_1 + b_2d_1 \\ c_2a_1 + d_2c_1 & c_2b_1 + d_2d_1 \end{pmatrix} (z) = \\ &= \frac{(a_2a_1 + b_2c_1)z + a_2b_1 + b_2d_1}{(c_2a_1 + d_2c_1)z + c_2b_1 + d_2d_1} = \\ &= \frac{a_2(a_1z + b_1) + b_2(c_1z + d_1)}{c_2(a_1z + b_1) + d_2(c_1z + d_1)} = \\ &= \frac{a_2\left(\frac{a_1z+b_1}{c_1z+d_1}\right) + b_2}{c_2\left(\frac{a_1z+b_1}{c_1z+d_1}\right) + d_2} = \Theta(\alpha_2, \Theta(\alpha_1, z)). \end{aligned}$$

□

TEOREMA 2.2 *Ogni $\alpha \in SL_2(\mathbb{R})$ agisce su \mathbb{H} come un'isometria.*

DIM. Si deve dimostrare che il ds^2 si conserva.

$$dw = \frac{a(cz + d) - c(az + b)}{(cz + d)^2} dz = \frac{dz}{(cz + d)^2}, \quad (2.4)$$

per l'equazione (2.3) e per la (2.1) si ha che

$$\frac{dw d\bar{w}}{(\Im(w))^2} = \frac{dz d\bar{z}}{(\Im(z))^2} \cdot \frac{|cz + d|^4}{(cz + d)^2(c\bar{z} + d)^2} = \frac{dz d\bar{z}}{(\Im(z))^2} = ds^2 \quad (2.5)$$

□

TEOREMA 2.3 (i) *Il gruppo $SL_2(\mathbb{R})$ agisce transitivamente su \mathbb{H} , cioè $\forall z, z' \in \mathbb{H}, \exists \alpha \in SL_2(\mathbb{R})$ tale che $\alpha z = z'$.*

(ii) *L'azione di $SL_2(\mathbb{R})$ su \mathbb{H} induce un isomorfismo*

$$PSL_2(\mathbb{R}) = SL_2(\mathbb{R})/\{\pm I\} \rightarrow \text{Aut}(\mathbb{H})$$

(iii) *Lo stabilizzatore di i è $SO_2(\mathbb{R})$.*

(iv) *La mappa*

$$SL_2(\mathbb{R})/SO_2(\mathbb{R}) \rightarrow \mathbb{H}, \quad \alpha \cdot SO_2(\mathbb{R}) \mapsto \alpha(i)$$

è un omeomorfismo.

DIM. (i) Sia $z \in \mathbb{H}$, è sufficiente mostrare che esiste $\alpha \in SL_2(\mathbb{R})$ tale che $\alpha(i) = z$, infatti se z' fosse un secondo punto, $\alpha'(i) = z'$ per qualche $\alpha' \in SL_2(\mathbb{R})$ e dunque $\alpha'\alpha^{-1}(z) = z'$.

Sia dunque $z = x + iy$; allora $\alpha = \sqrt{y}^{-1} \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} \in SL_2(\mathbb{R})$ e $\alpha(i) = z$.

(ii) C'è da chiedersi quando $\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = z \quad \forall z \in \mathbb{H}$. Sviluppando il calcolo si ottiene $cz^2 + (d-a)z - b = 0$, poiché questo deve essere vero $\forall z \in \mathbb{H}$, il polinomio deve avere tutti i coefficienti nulli dunque $c = 0$, $d = a$ e $b = 0$. Dunque $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ che ha determinante 1 se e solo se $a = \pm 1$. Questo vuol dire che gli unici elementi di $SL_2(\mathbb{R})$ che agiscono banalmente (come l'identità) su \mathbb{H} sono $\pm I$.

Sia ora γ un automorfismo su \mathbb{H} . Dalla (i) si sa che esiste $\alpha \in SL_2(\mathbb{R})$ tale che $\alpha(i) = \gamma(i)$. Sostituendo γ con $\alpha^{-1} \circ \gamma$, si può assumere che $\gamma(i) = i$. Si consideri ora la mappa $\rho : \mathbb{H} \rightarrow D$ disco aperto unitario in \mathbb{C} , tale che $z \mapsto \frac{z-i}{z+i}$. Tale mappa è il famoso isomorfismo tra \mathbb{H} e il disco aperto unitario D che mappa i nello 0. Si usi ρ per trasformare γ in un automorfismo γ' di D che fissi lo 0.

LEMMA 2.4 *Gli automorfismi del cerchio unitario D che fissano lo 0 sono le mappe della forma $z \mapsto \lambda z$, $|\lambda| = 1$.*

DIM. È una conseguenza del Lemma di Schwarz (cfr. [NP69, p.143]) che dice che

LEMMA 2.5 (Schwarz) *Sia $f(z)$ una funzione ologomorfa nel disco $|z| < 1$. Si supponga che*

$$f(0) = 0, \quad |f(z)| < 1 \text{ per } |z| < 1.$$

Allora

(a) $|f(z)| \leq |z|$ per $|z| < 1$;

(b) se $|f(z_0)| = |z_0|$ per qualche $z_0 \neq 0$, allora $\exists \lambda$ tale che $f(z) = \lambda z$ (e $|\lambda| = 1$).

Sia γ un automorfismo di D che fissi lo 0. Applicando la (a) a γ e γ^{-1} si ottiene $|\gamma(z)| = |z| \quad \forall z \in D$, quindi si può applicare (b) per mostrare che la f è della forma richiesta.

□

(iii) Si ha che

$$\frac{ai + b}{ci + d} = i \iff ai + b = -c + di \iff a = d, b = -c$$

allora la matrice è della forma $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ con $a^2 + b^2 = 1$, da cui l'appartenenza ad $SO_2(\mathbb{R})$.

(iv) È una conseguenza diretta del Teorema 1.2. Si ricordi che $SL_2(\mathbb{R})$ è un gruppo ma anche uno spazio topologico.

□

2.2 La curvatura di Gauss di \mathbb{H}

Sia $g = \det([g_{ij}])$ in cui $[g_{ij}]$ è la prima forma fondamentale definita nella formula (2.1). È noto dalla geometria differenziale (cfr. [Sera, Ser94]) che la curvatura di una superficie può essere calcolata con la seguente formula:

$$K = -\frac{1}{g} \sum_{m=1}^2 \left(\frac{\partial \Gamma_{12}^m}{\partial x} g_{2m} - \frac{\partial \Gamma_{11}^m}{\partial y} + \Gamma_{12}^m \Gamma_{1m2} - \Gamma_{12}^m \Gamma_{2m2} \right) \quad (2.6)$$

in cui le Γ_{ij}^k rappresentano i simboli di Christoffel.

Per completezza $\Gamma_{ij}^k = \sum_l g^{kl} \Gamma_{ijk}$ dove $[g^{ij}] = [g_{ij}]^{-1} = [\frac{(-1)^{i+j}}{g} g_{ij}]$ ($\hat{k} = 1$ se $k = 2$ e viceversa) e $\Gamma_{ijk} = \frac{1}{2} \left(\frac{\partial g_{jk}}{\partial x_i} + \frac{\partial g_{ki}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_k} \right)$.

Bastano pochi semplici calcoli per osservare che $[g^{ij}] = \begin{pmatrix} y^2 & 0 \\ 0 & y^2 \end{pmatrix}$ e $\mathbf{g} = \frac{1}{y^4}$, per cui, sostituendo nelle espressioni, si ha:

$$\frac{\partial g_{11}}{\partial x} = 0, \quad \frac{\partial g_{12}}{\partial x} = 0, \quad \frac{\partial g_{21}}{\partial x} = 0, \quad \frac{\partial g_{22}}{\partial x} = 0 \quad (2.7)$$

$$\frac{\partial g_{11}}{\partial y} = -\frac{2}{y^3}, \quad \frac{\partial g_{12}}{\partial y} = 0, \quad \frac{\partial g_{21}}{\partial y} = 0, \quad \frac{\partial g_{22}}{\partial y} = -\frac{2}{y^3}, \quad (2.8)$$

di conseguenza

$$\Gamma_{111} = \Gamma_{122} = \Gamma_{221} = 0 \quad (2.9)$$

$$\Gamma_{112} = \Gamma_{212} = \frac{1}{y^3} \quad (2.10)$$

$$\Gamma_{222} = \Gamma_{121} = \Gamma_{211} = -\frac{1}{y^3} \quad (2.11)$$

e

$$\Gamma_{11}^1 = \Gamma_{12}^2 = \Gamma_{22}^1 = 0 \quad (2.12)$$

$$\Gamma_{11}^2 = \Gamma_{12}^1 = \frac{1}{y} \quad (2.13)$$

$$\Gamma_{22}^2 = -\frac{1}{y}. \quad (2.14)$$

A questo punto basta sostituire i valori adeguati nella formula (2.6) per ottenere $K = -1$ costante.

2.3 Alcune proprietà di \mathbb{H}

Si è potuto osservare con il Teorema 2.3 che

$$\forall z_1, z_2 \in \mathbb{H}, \exists \alpha : \mathbb{H} \rightarrow \mathbb{H} \text{ isometria t.c. } \alpha(z_1) = z_2,$$

cioè che \mathbb{H} è una superficie *omogenea*.

L'omogeneità è una proprietà difficile da trovare ed è legata ad una serie di importanti conseguenze sulla geometria intrinseca della superficie e in

qualche modo sulla sua struttura topologica.

Una superficie S è detta *geodeticamente completa* se ogni sua geodetica può essere prolungata ad una geodetica $\gamma : \mathbb{R} \rightarrow S$ definita su tutto \mathbb{R} .

Dalla geometria differenziale (cfr. [Sera, Ber02, CCL99]) si sa che:

TEOREMA 2.6 *Una superficie omogenea è sempre geodeticamente completa.*

TEOREMA 2.7 (Hopf-Rinow) *Una superficie è geodeticamente completa se e solo se è uno spazio metrico completo.*

Quindi \mathbb{H} è uno spazio metrico completo ed è geodeticamente completo!
Ma non solo:

TEOREMA 2.8 *Una superficie è a curvatura¹ costante se e solo se è (localmente) omogenea.*

Dunque basta l'omogeneità della superficie per dire che K è costante. Questo, insieme a qualche altro risultato, permette di calcolare K in un modo alternativo... ma serve ancora una piccola premessa.

2.4 Calcolo dell'area di un triangolo iperbolico asintotico.

Si calcolerà l'area di un generico triangolo iperbolico asintotico perché la formula ricavata si rivelerà utile anche nel prossimo capitolo.

Il triangolo iperbolico asintotico è un triangolo geodetico che ha per lati due geodetiche verticali² che partono da due vertici p e q e la geodetica che li unisce (Figura 2.1).

¹Si intende *curvatura di Gauss*.

²Il terzo vertice è un punto all'infinito, per questo si chiama triangolo asintotico.

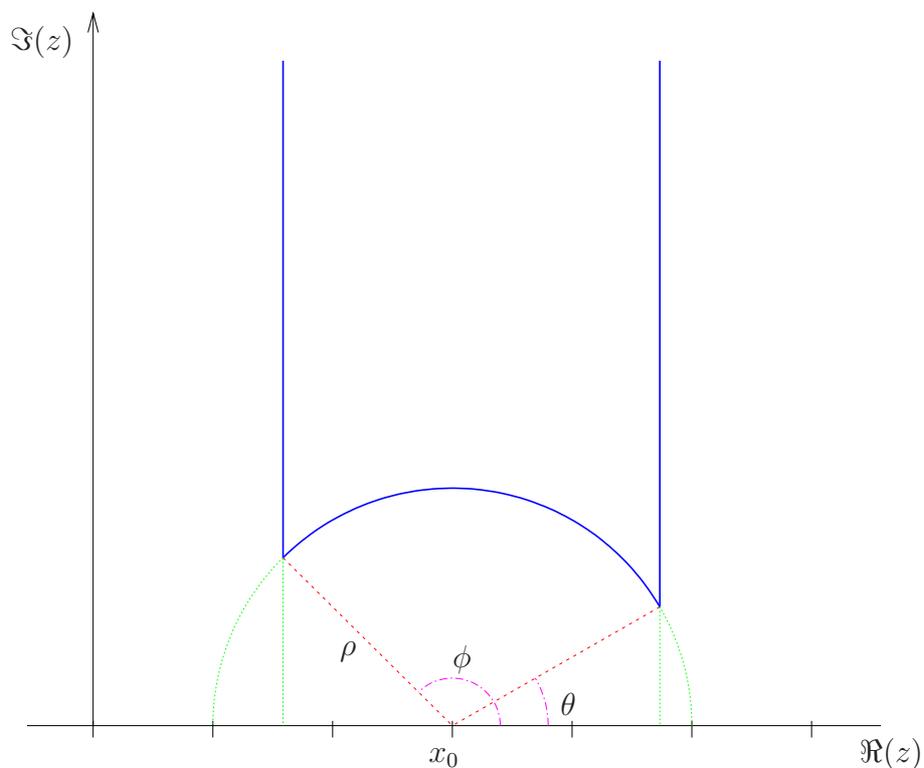


Figura 2.1: Generico triangolo iperbolico asintotico

Siano $p = (x_0 + \rho \cos \theta, \rho \sin \theta)$ e $q = (x_0 + \rho \cos \phi, \rho \sin \phi)$ due punti sul semicerchio di centro $(x_0, 0)$ e raggio ρ con $0 \leq \theta < \phi \leq \pi$. Si consideri la regione T del piano iperbolico definita da

$$T = \{(x, y) : x_0 + \rho \cos \phi \leq x \leq x_0 + \rho \cos \theta, y \geq \sqrt{\rho^2 - (x - x_0)^2}\}$$

cioè la regione di piano iperbolico compresa tra il segmento geodetico passante per p e q e le due semirette verticali uscenti da p e q .

Il differenziale d'area di \mathbb{H} è

$$dA = \sqrt{g} \, dx \, dy = \frac{dx \, dy}{y^2}$$

quindi

$$\begin{aligned}
A(T) &= \int_{x_0 + \rho \cos \phi}^{x_0 + \rho \cos \theta} \int_{\sqrt{\rho^2 - (x-x_0)^2}}^{+\infty} \frac{dx dy}{y^2} = \int_{x_0 + \rho \cos \phi}^{x_0 + \rho \cos \theta} \left\{ \int_{\sqrt{\rho^2 - (x-x_0)^2}}^{+\infty} \frac{dy}{y^2} \right\} dx = \\
&= \int_{x_0 + \rho \cos \phi}^{x_0 + \rho \cos \theta} \left[-\frac{1}{y} \right]_{\sqrt{\rho^2 - (x-x_0)^2}}^{+\infty} dx = \int_{x_0 + \rho \cos \phi}^{x_0 + \rho \cos \theta} \frac{dx}{\sqrt{\rho^2 - (x-x_0)^2}}
\end{aligned}$$

ponendo $x = x_0 + \rho \cos t$, $dx = -\rho \sin t dt$ e l'ultimo integrale diventa

$$\int_{x_0 + \rho \cos \phi}^{x_0 + \rho \cos \theta} \frac{dx}{\sqrt{\rho^2 - (x-x_0)^2}} = \int_{\phi}^{\theta} \frac{-\rho \sin t}{\rho \sin t} dt = - \int_{\phi}^{\theta} dt = \phi - \theta.$$

Ricapitolando si è mostrato che l'area di un triangolo iperbolico asintotico T di vertici $p = (x_0 + \rho \cos \theta, \rho \sin \theta)$, $q = (x_0 + \rho \cos \phi, \rho \sin \phi)$ e ∞ si calcola

$$A(T) = \phi - \theta. \tag{2.15}$$

2.5 Un modo alternativo di calcolare la curvatura di \mathbb{H}

Un importante risultato globale della teoria delle superfici (cfr. [Sera, Ber02, CCL99, Ser94]) è il seguente:

TEOREMA 2.9 (Formula di Gauss-Bonnet) *Se T è un triangolo geodetico su una superficie S di angoli interni α , β e γ , allora*

$$\iint_T K dA = (\alpha + \beta + \gamma) - \pi.$$

Si consideri il triangolo iperbolico asintotico con $q = (0, 0)$ e $p = (1, 0)$ (Figura 3.1). Nella costruzione della sezione precedente si ponga $x_0 = \frac{1}{2}$ e $\rho = \frac{1}{2}$. Banalmente si ha $\theta = 0$ e $\phi = \pi$.

L'angolo tra due curve incidenti in un punto è l'angolo tra i vettori tangenti alle curve in quel punto; si può verificare che in p , q ed ∞ le curve sono tangenti, per cui α , β e γ sono nulli³.

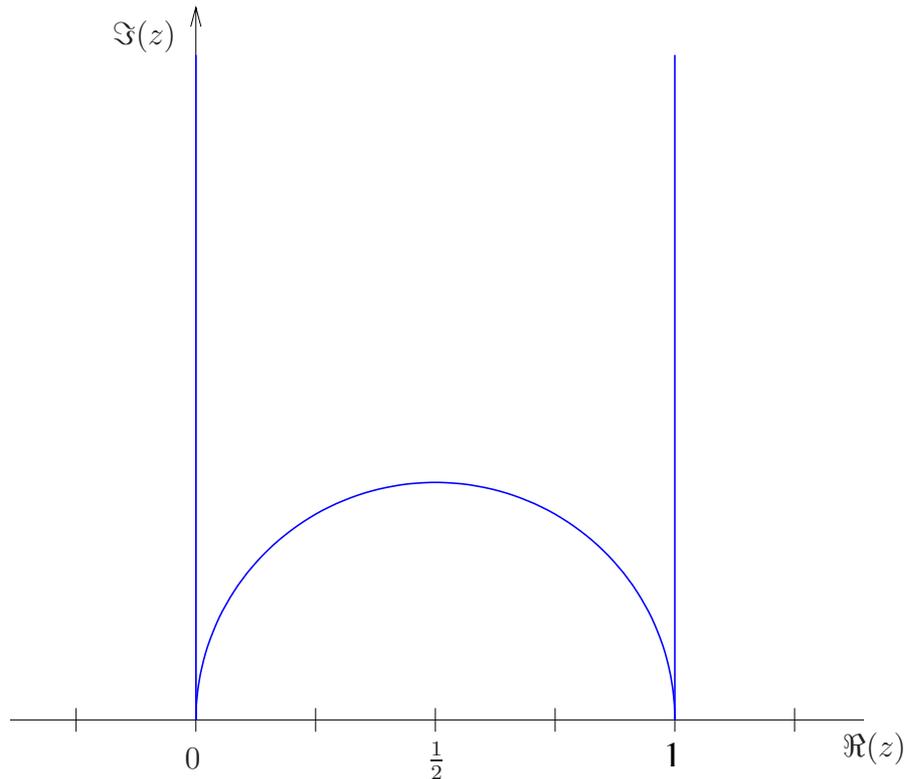


Figura 2.2: Triangolo iperbolico asintotico con vertici in 0, 1 e ∞

Inoltre, con il Teorema 2.8 si è stabilito che K deve essere costante, per cui si ha:

$$K \iint_T dA + \pi = 0$$

ma

$$\iint_T dA = \phi - \theta = \pi \Rightarrow K\pi + \pi = 0 \Rightarrow K = -1,$$

³La cosa non deve stupire, in geometria iperbolica la somma degli angoli interni di un triangolo è sempre minore di π , il fatto che sia nulla non è affatto un problema.

esattamente come ci si aspettava.

2.6 Geodetiche di \mathbb{H}

Sia $\gamma(t) = (x_1(t), x_2(t))$ una curva di estremi a e b in \mathbb{H} . Un altro importante risultato della geometria differenziale (cfr. [Sera, Ber02, CCL99, Ser94]) è la possibilità di trovare le geodetiche come soluzioni di sistemi di equazioni differenziali.

Ci sono due modi per trovare le geodetiche. Siano $\begin{cases} x_1(0) = x_0, & x_1'(0) = v_x \\ x_2(0) = y_0, & x_2'(0) = v_y \end{cases}$ le condizioni iniziali:

- si sfrutta la funzione $L(x_1, x_2, dx_1, dx_2) = L(\gamma(t), \gamma'(t)) = \frac{1}{2}ds^2$ e si risolve il sistema

$$\begin{cases} \frac{\partial L}{\partial x_1} - \frac{d}{dt} \frac{\partial L}{\partial x_1'} = 0 \\ \frac{\partial L}{\partial x_2} - \frac{d}{dt} \frac{\partial L}{\partial x_2'} = 0 \end{cases} \quad (2.16)$$

- equivalentemente, considerata γ come sopra e considerate le stesse condizioni iniziali, si può sfruttare il seguente sistema di equazioni differenziali

$$\begin{cases} \frac{d^2 x_1}{dt^2} = - \sum_{i,j=1}^2 \Gamma_{ij}^1 \frac{dx_i}{dt} \frac{dx_j}{dt} \\ \frac{d^2 x_2}{dt^2} = - \sum_{i,j=1}^2 \Gamma_{ij}^2 \frac{dx_i}{dt} \frac{dx_j}{dt} \end{cases} \quad (2.17)$$

Considerando il secondo sistema si avrebbe

$$\begin{cases} x''(t) = - \left(-\frac{2}{y(t)} x'(t) y'(t) \right) \\ y''(t) = - \left(\frac{1}{y(t)} (x'(t))^2 - \frac{1}{y(t)} (y'(t))^2 \right) \end{cases} \quad (2.18)$$

ponendo $x_0 = 0$, $v_x = 0$, $y_0 = 1$, $v_y = 1$, si ottiene

$$\begin{cases} x''(t) = 0 \\ y''(t) = \frac{(y'(t))^2}{y(t)} \end{cases} \quad (2.19)$$

da cui

$$\gamma(t) = \begin{cases} x(t) = 0 \\ y(t) = e^t \end{cases} = (0, e^t) \quad (2.20)$$

Se si considera l'isometria su \mathbb{H} rappresentata da $\alpha = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \in SL_2(\mathbb{R})$ è banale osservare che si può traslare la geodetica γ appena calcolata in modo che x_0 sia un qualsiasi punto dell'asse reale, ottenendo in questo modo una classe di geodetiche identificabile con tutte le semirette verticali che partono dall'asse reale.

Allo stesso modo si può osservare che $\forall \alpha \in SL_2(\mathbb{R})$, $\alpha\gamma$ è una geodetica di \mathbb{H} .

Si consideri ora

$$\alpha = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \in SL_2(\mathbb{R}) \quad (2.21)$$

e si calcoli $\alpha\gamma$, si otterrà

$$\alpha\gamma = \frac{1}{e^{2t} + 1}((e^{2t} - 1) + 2ie^t). \quad (2.22)$$

A questo punto, può essere interessante calcolare $\|\alpha\gamma(t)\|$:

$$\begin{aligned} x(t) &= \frac{e^{2t} - 1}{e^{2t} + 1} \Rightarrow dx = \frac{4e^{2t}}{(e^{2t} + 1)^2} \\ y(t) &= \frac{2e^t}{e^{2t} + 1} \Rightarrow dy = 2e^t \frac{1 - e^{2t}}{(e^{2t} + 1)^2} \\ \|\alpha\gamma(t)\|^2 &= \frac{(e^{2t} + 1)^2}{4e^{2t}} \cdot \frac{16e^{4t} + 4e^{2t}(1 - e^{2t})^2}{(e^{2t} + 1)^4} = \\ &= \frac{4e^{2t} + (1 - e^{2t})^2}{(e^{2t} + 1)^2} = \frac{(e^{2t} + 1)^2}{(e^{2t} + 1)^2} = 1 \end{aligned}$$

questo vuol dire che la curva così ottenuta è una circonferenza, anzi, considerate le caratteristiche dell'applicazione di $SL_2(\mathbb{R})$, deve necessariamente essere una semicirconferenza di raggio unitario in \mathbb{H} . In effetti, se si calcolano i limiti per t che tende a $\pm\infty$ si può osservare che la semicirconferenza va da -1 ad 1 ed è centrata in 0 .

Come si è già osservato⁴, l'applicazione di un qualche isomorfismo lascia invariata la natura della curva. Nel prossimo capitolo si osserverà come traslazioni, dilatazioni e rotazioni facciano parte delle isometrie di \mathbb{H} rappresentabili con matrici in $SL_2(\mathbb{R})$, questo vuol dire che tutte le semicirconferenze centrate sull'asse reale e con estremi sull'asse reale sono geodetiche.

Ma per ogni punto $z = (x, y) \in \mathbb{H}$ e per ogni direzione non verticale v , esiste sempre una semicirconferenza centrata in un qualche punto $z_0 = (x_0, 0)$ passante per z ed avente per tangente nel punto la direzione assegnata⁵. È noto dalla geometria differenziale che per un punto ed una direzione assegnata passa una ed una sola geodetica, per cui le due classi di curve descritte fino a questo punto sono tutte e sole le geodetiche di \mathbb{H} con la metrica iperbolica (2.1).

⁴ma non è difficile da provare

⁵Per trovare il centro basta considerare la perpendicolare al vettore tangente passante per il punto e la sua intersezione con l'asse reale. La distanza del punto con il centro è il raggio. Si prende $\alpha\gamma$ calcolata in precedenza e la si dilata per raggiungere il raggio calcolato, dopodiché la si trasla. La geodetica così trovata è una pregeodetica che passa per il punto con tangente parallela a quella data, si applica una dilatazione opportuna al parametro t e si ottiene la geodetica cercata.

Capitolo 3

La superficie modulare

3.1 Quozienti di \mathbb{H}

Sia X uno spazio topologico su cui agisce un qualche gruppo Γ . Se $\Gamma \backslash X$ è di Hausdorff allora le orbite sono chiuse, purtroppo il viceversa non è sempre vero.

Si dice che l'azione di Γ è *discontinua* se $\forall x \in X$ e $\forall (\gamma_i)$ successione di elementi distinti in Γ , l'insieme $\{\gamma_i x\}$ non ha punti di accumulazione; mentre si definisce *propriamente discontinua* l'azione tale che $\forall x, y \in X$, $\exists U_x, U_y$ intorno di x ed y tali per cui l'insieme $\{\gamma \in \Gamma : \gamma U_x \cap U_y \neq \emptyset\}$ è finito.

Sia Γ un gruppo topologico. Un sottogruppo Γ_0 di Γ si dice *discreto* se ogni punto di Γ_0 è isolato (nella relativa topologia in Γ_0). In particolare, questo significa che se g_1, g_2, \dots sono elementi di Γ_0 che convergono ad un qualche γ in Γ , allora $g_n = \gamma$ per ogni n sufficientemente grande (infatti $g_{n+1}g_n^{-1} \in \Gamma_0$ in quanto sottogruppo e tale successione converge all'identità I che è un elemento isolato di Γ_0).

Un sottogruppo discreto di $SL_2(\mathbb{R})$ viene chiamato *gruppo Fuchsiano* in quanto gruppo di trasformazioni di Möbius che agisce con discontinuità

su \mathbb{H} . Per la precisione, il sottogruppo discreto $SL_2(\mathbb{Z}) \subset SL_2(\mathbb{R})$ viene chiamato *gruppo modulare completo*; un sottogruppo $\Gamma \subset SL_2(\mathbb{Z})$ di indice finito ($[SL_2(\mathbb{Z}) : \Gamma] < \infty$) è chiamato *gruppo modulare*. Se $-I \in \Gamma$, quando agisce sullo spazio iperbolico \mathbb{H} identifichiamo i gruppi Γ e $P\Gamma = \Gamma/\{\pm I\}$. Lo spazio delle orbite $\Gamma \backslash \mathbb{H}$ del gruppo modulare Γ è chiamato *superficie modulare*, in tale spazio $h_1, h_2 \in \mathbb{H}$ sono identificate come segue:

$$h_1 \sim h_2 \Leftrightarrow h_2 = gh_1$$

per qualche $g \in \Gamma$.

LEMMA 3.1 *Tali superfici modulari non sono compatte ma hanno area iperbolica finita.*

DIM. Banalmente, \mathbb{H} non è un compatto, poiché si conosce dalla topologia che la compattezza si conserva per quozienti (cfr. [Pie]), $\Gamma \backslash \mathbb{H}$ non può essere un compatto.

Per calcolare l'area di $\Gamma \backslash \mathbb{H}$ si deve trovare il suo dominio fondamentale in \mathbb{H} . Si potrà effettuare il calcolo alla fine del capitolo.

□

TEOREMA 3.2 *Sia G un gruppo localmente compatto che agisce su uno spazio topologico X tale che preso un qualunque punto $x_0 \in X$, lo stabilizzatore K di x_0 in G sia un compatto e l'applicazione $gK \mapsto gx_0 : G/K \rightarrow X$ sia un omeomorfismo.*

Allora le seguenti condizioni su un sottogruppo Γ di G sono equivalenti:

- (a) *l'azione di Γ su X è discontinua;*
- (b) *l'azione di Γ su X è propriamente discontinua;*

(c) $\forall A, B \subset X$ compatti, l'insieme $\{\gamma \in \Gamma : \gamma A \cap B \neq \emptyset\}$ è finito;

(d) Γ è un sottogruppo discreto di G .

DIM. Le implicazioni (c) \Rightarrow (b) \Rightarrow (a) e (d) \Rightarrow (a) sono banali.

L'implicazione (b) \Rightarrow (c) è semplice da verificare.

(c) \Rightarrow (d): sia V un intorno di 1_G che abbia chiusura \bar{V} compatta. $\forall x \in X$, $\Gamma \cap V \subset \{\gamma \in \Gamma : \gamma x \in \bar{V} \cdot x\}$ che è finito perché sia $\{x\}$ che $\bar{V} \cdot X$ sono compatti. Dunque $\Gamma \cap V$ è discreto, quindi $e = 1_G$ è un punto isolato di Γ .

(d) \Rightarrow (c): sia p la mappa $g \mapsto gx_0 : G \rightarrow X$. Sia $A \subset X$ compatto. Si vuole mostrare che $p^{-1}(A)$ è compatto.

Si scriva $G = \cup V_i$ dove gli insiemi V_i sono aperti con chiusura \bar{V}_i compatta. Allora $A \subset \cup p(V_i)$, ed inoltre è necessario soltanto un numero finito di $p(V_i)$ per ricoprire A . Dunque $p^{-1}(A) \subset (\cup V_i K) \subset (\cup \bar{V}_i K)$ (unione finita) ed ogni $\bar{V}_i K$ è compatto (in quanto immagine di $\bar{V}_i \times K$ sotto la mappa moltiplicativa $G \times G \rightarrow G$). In altre parole $p^{-1}(A)$ è un sottoinsieme chiuso di un compatto, per cui è compatto. Allo stesso modo si può verificare che $p^{-1}(B)$ è compatto.

Si supponga ora che $\gamma A \cap B \neq \emptyset$ e $\gamma \in \Gamma$. Allora $\gamma(p^{-1}(A)) \cap p^{-1}(B) \neq \emptyset$ quindi $\gamma \in \Gamma \cap p^{-1}(B) \cdot (p^{-1}(A))^{-1}$. Ma quest'ultimo insieme è l'intersezione di un insieme discreto con uno compatto, per cui è finito.

□

Con il prossimo teorema si riesce a rendere più precisa la proposizione (c).

TEOREMA 3.3 *Siano G, K, X come nel Teorema 3.2, e sia Γ un sottogruppo discreto di G .*

(a) $\forall x \in X$, $\{g \in \Gamma : gx = x\}$ è finito;

- (b) $\forall x \in X$, esiste un intorno U di x con la seguente proprietà: se $\gamma \in \Gamma$ ed $U \cap \gamma U \neq \emptyset$, allora $\gamma x = x$;
- (c) $\forall x, y \in X$ che non sono nella stessa Γ -orbita, esistono un intorno U di x ed un intorno V di y tali che $\gamma U \cap V = \emptyset \quad \forall \gamma \in \Gamma$.

DIM. (a) Si è visto nella dimostrazione del Teorema 3.2 che $p^{-1}(\text{compatto})$ è compatta, dove $p(g) = gx$. Quindi $p^{-1}(x)$ è compatto. L'insieme che ci interessa è $p^{-1}(x) \cap \Gamma$ che, in quanto intersezione di un insieme discreto e di uno compatto, è un insieme finito.

(b) Sia V un intorno compatto di x . Per il punto (c) del Teorema 3.2, c'è un insieme finito $\{\gamma_1, \dots, \gamma_n\}$ di elementi di Γ tali che $V \cap \gamma_i V \neq \emptyset$. Siano $\gamma_1, \dots, \gamma_s$ quei γ_i che fissano x . $\forall i > s$, si scelgano due intorni disgiunti V_i di x e W_i di $\gamma_i x$ e si ponga

$$U = V \cup (\bigcap_{i>s} V_i \cap \gamma_i^{-1} W_i)$$

Per $i > s$, $\gamma_i(U) \subset W_i$ che è disgiunto da V_i che contiene U .

(c) Si scelgano due intorni compatti A di x e B di y e siano $\gamma_1, \dots, \gamma_n$ gli elementi di Γ tali che $\gamma_i A \cap B \neq \emptyset$. Per ipotesi $\gamma_i x \neq y$, quindi si possono considerare due intorni disgiunti U_i e V_i rispettivamente di $\gamma_i x$ e y . Si prendano $U = A \cap \gamma_1^{-1} U_1 \cap \dots \cap \gamma_n^{-1} U_n$, $V = B \cap V_1 \cap \dots \cap V_n$.

□

COROLLARIO 3.4 *Sotto le ipotesi del Teorema 3.3, lo spazio $\Gamma \backslash X$ è di Hausdorff.*

DIM. Siano x ed y due punti di X che non appartengono alla stessa Γ -orbita, si scelgano due intorni U e V come nel Teorema 3.3(c). Allora le

immagini di U e V in $\Gamma \backslash X$ sono intorni disgiunti di Γx e di Γy .

□

3.2 Isometrie nel semipiano iperbolico

Gli elementi di $\text{Aut}(\mathbb{H})$ sono isometrie del semipiano di Poincaré provvisto della metrica iperbolica, si vuole dimostrare che sono essenzialmente le uniche.

TEOREMA 3.5 *Ogni isometria di \mathbb{H} con la metrica iperbolica è della forma $z \mapsto g(z)$, $z \mapsto g(-\bar{z})$ per qualche g in $\text{Aut}(\mathbb{H})$.*

DIM. Sia h un'isometria di \mathbb{H} . Di certo $\exists g_1 \in \text{Aut}(\mathbb{H})$ tale che $g_1(h(i)) = i$, dunque, se $h_1 = g_1 h$, h_1 è un'isometria che fissa i . Sia $d(\cdot, \cdot)$ la distanza iperbolica su \mathbb{H} :

$$d(i, 2i) = d(h_1(i), h_1(2i)) = d(i, h_1(2i))$$

dunque $|h_1(2i) - i| = |2i - i| = |i| = 1$. Sia g_2 una rotazione attorno ad i che mappa $h_1(2i)$ in $2i$. Se $h_2 = g_2 h_1$, allora h_2 fissa sia i che $2i$. Considerando che h_2 è un'isometria è facile rendersi conto che debba fissare tutti i punti dell'asse immaginario $\Im(z) > 0$. Dunque $\forall z \in \mathbb{H}$, $h_2(z)$ deve essere necessariamente z o $-\bar{z}$. Inoltre, per lo stesso motivo, h_2 deve lasciare invariati (o scambiare) i semipiani di \mathbb{H} destro e sinistro, quindi o $\forall z$, $h_2(z) = z$ oppure $\forall z$, $h_2(z) = -\bar{z}$.

□

Come già visto, $\text{Aut}(\mathbb{H})$ è isomorfo a $PSL_2(\mathbb{R})$ dunque gli elementi di $PSL_2(\mathbb{R})$ sono in relazione biunivoca con le isometrie di \mathbb{H} .

TEOREMA 3.6 *Sia $C_\infty = \mathbb{R} \cup \{\infty\}$ il cosiddetto cerchio all'infinito per \mathbb{H} . Sia $g \in SL_2(\mathbb{R})$, $g \neq \pm I$. Allora*

- (1) *g ha un unico punto fisso in C_∞ , oppure*
- (2) *g ha due punti fissi entrambi in C_∞ , oppure*
- (3) *g ha un punto fisso in \mathbb{H} e nessuno in C_∞ .*

DIM. Sia $g \in SL_2(\mathbb{R})$, i punti fissi di g sono le due soluzioni (che possono anche includere ∞) di $g(z) = z$. È evidente che $g(z) = \frac{az+b}{cz+d}$ fissi ∞ se e solo se $c = 0$ ed in tal caso g ha al più un altro punto fisso. Poiché i punti fissi di g in \mathbb{C} sono gli zeri di un polinomio di secondo grado a coefficienti reali ($cz^2 + (d-a)z - b = 0$), può accadere soltanto che siano su \mathbb{R} o che siano complessi coniugati, in tal caso solo uno dei due apparterebbe ad \mathbb{H} .

□

Se si generalizza ancora e si considerano due matrici $\alpha, \beta \in SL_2(\mathbb{C})$, eguagliando le trasformazioni associate $\alpha(\cdot), \beta(\cdot)$ si può notare come l'uguaglianza si verifichi se e solo se $\alpha = \pm\beta$. Si può quindi definire una funzione $\tau : SL_2(\mathbb{C}) \rightarrow \mathbb{C}$ come segue:

$$\tau(\alpha) = [\text{Traccia}(\alpha)]^2 = [\text{Traccia}(-\alpha)]^2.$$

Si noti che matrici coniugate mappano trasformazioni coniugate, per l'invarianza della traccia rispetto alla coniugazione, si ha che

$$\tau(h\alpha h^{-1}) = \tau(\alpha).$$

Dunque $\tau(f) = \tau(g)$ è una condizione *necessaria* ad f e g per essere coniugate; in effetti è anche una condizione *sufficiente* (cfr. [Bea83, p.66]).

3.3 Classificazione della trasformazioni lineari frazionarie

È possibile estendere l'azione di $SL_2(\mathbb{C})$, e quindi quella di $SL_2(\mathbb{R})$ in particolare, a $\mathbb{C} \cup \{\infty\}$ definendo:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az + b}{cz + d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \infty = \frac{a}{c}$$

Queste applicazioni vengono chiamate *trasformazioni lineari frazionarie* (*linear fractional transformations*) ed hanno la proprietà di mappare cerchi e rette in \mathbb{C} in cerchi o rette in \mathbb{C} (cfr. [Con78, pag. 46]). La matrice $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ agisce come trasformazione identità.

Dalla teoria delle forme canoniche di Jordan (cfr. [Saa92, Saa03]), ogni matrice α è coniugata ad una matrice del tipo seguente:

$$(i) \quad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \quad (ii) \quad \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}, \quad \lambda \neq \mu$$

in relazione ai suoi autovalori. Se sono coincidenti α è coniugata ad una trasformazione $z \mapsto z + \lambda^{-1}$, se sono distinti a $z \mapsto cz$, $c \neq 1$. Nel caso (i) α viene detta *parabolica* mentre nel caso (ii) viene chiamata *ellittica* se $|c| = 1$, *iperbolica* se c è un numero reale positivo e *loxodromica* altrimenti.

Quando $\alpha \in SL_2(\mathbb{C})$, si vedrà tra poco che i quattro casi possono essere

distinti dalla traccia di α :

$$\alpha \text{ è } \textit{parabolica} \iff \tau(\alpha) = 4;$$

$$\alpha \text{ è } \textit{ellittica} \iff \tau(\alpha) \text{ è reale e } \tau(\alpha) < 4;$$

$$\alpha \text{ è } \textit{iperbolica} \iff \tau(\alpha) \text{ è reale e } \tau(\alpha) > 4;$$

$$\alpha \text{ è } \textit{loxodromica} \iff \tau(\alpha) \text{ non è reale.}$$

Si vogliono ora esaminare gli elementi di queste classi in $SL_2(\mathbb{C})$ ed in particolare in $SL_2(\mathbb{R})$.

3.3.1 Trasformazioni paraboliche

Si è definita parabolica la mappa del tipo (i).

LEMMA 3.7 $\alpha \in SL_2(\mathbb{C})$ è parabolica se è coniugata con la traslazione $T : z \mapsto z + 1$.

Tutte le traslazioni sono coniugate con T .

DIM. Essendo $\alpha \in SL_2(\mathbb{C})$, $\det(\alpha) = 1$, quindi $\lambda^2 = 1$. Dunque ogni trasformazione parabolica deve essere coniugata con $T_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ o con

$$T_2 = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

Le trasformazioni associate a tali matrici sono rispettivamente $T = T_1 : z \mapsto z + 1$ e $T_2 : z \mapsto z - 1$ per cui T_2 equivale a $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ che è una traslazione.

Resta dunque da dimostrare che tutte le traslazioni sono coniugate con T .

Una traslazione è una trasformazione del tipo $\alpha : z \mapsto z + k$, per cui $\alpha = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$. Si consideri $h = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{C})$, se $Th = h\alpha$, α e T

sono coniugate:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$$

$$\Updownarrow$$

$$d = \pm\sqrt{k}, \quad a = \frac{1}{d}, \quad c = 0, \quad b \text{ qualunque.}$$

Per comodità si scelga $b = 0$. La matrice $h = \begin{pmatrix} \sqrt{k} & 0 \\ 0 & \frac{1}{\sqrt{k}} \end{pmatrix} \in SL_2(\mathbb{C})$ è la matrice che coniuga α e T .

□

A questo punto, se $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ è parabolica deve avere lo stesso numero di punti fissi di T , quindi per il Teorema 3.6 ogni trasformazione parabolica ha un unico punto fisso in C_∞ .

Si chiami $\zeta \in C_\infty$ l'unico punto fisso di α . Se $\zeta = \infty$, necessariamente deve avere $c = 0$, quindi $ad = 1$, per cui $a = d = 1$ altrimenti il punto fisso sarebbe un numero reale finito ($\zeta = \frac{db}{d^2-1}$), ma allora $\tau(\alpha) = 4$.

Sia ora $\zeta \neq 0$, l'unica soluzione dell'equazione quadratica $cz^2 + (d-a)z - b = 0$ (cioè $\alpha(z) = z$), ricordando che $ad - bc = 1$, si ha $\Delta = (d-a)^2 + 4bc = (d+a)^2 - 4 = 0 \Leftrightarrow (a+d)^2 = 4 \Leftrightarrow \tau(\alpha) = 4$.

Viceversa, sia $\alpha \in SL_2(\mathbb{C})$ che abbia $\tau(\alpha) = 4$. α è coniugata ad una qualche β che fissa ∞ e quindi della forma $\begin{pmatrix} a & b \\ 0 & \frac{1}{a} \end{pmatrix}$. Ora,

$$0 = \tau(\alpha) - 4 = \tau(\beta) - 4 = \left(a + \frac{1}{a}\right)^2 - 4 = \left(a - \frac{1}{a}\right)^2,$$

dunque $a = \frac{1}{a} = \pm 1$. Questo vuol dire che β è una traslazione, per cui α è parabolica.

È appena stato dimostrato il seguente risultato:

TEOREMA 3.8 *Per ogni trasformazione di Möbius $\alpha \neq I$ sono equivalenti:*

- (1) α è parabolica;
- (2) α ha un unico punto fisso in C_∞ ;
- (3) $\tau(\alpha) = 4$;
- (4) α è coniugata (in $SL_2(\mathbb{C})$) con $z \mapsto z + 1$.

Come esempio di trasformazione parabolica non banale, si può considerare la mappa $\alpha : z \mapsto \frac{z}{\lambda z + 1}$ con $\lambda > 0$. Questa mappa è associata alla matrice $\begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$ ed, in quanto elemento di $SL_2(\mathbb{R})$, è un'isometria di \mathbb{H} . Evidentemente ha come unico punto fisso l'origine e, come ci si dovrebbe aspettare, $\tau(\alpha) = 4$.

3.3.2 Trasformazioni ellittiche

Sia $\zeta \in \mathbb{C}$ tale che $|\zeta| = 1$, allora lo si può scrivere come $\zeta = e^{i\theta}$ per qualche $\theta \in (0, 2\pi)$. Quindi α è *ellittica* se è coniugata con una rotazione euclidea non banale $z \mapsto ze^{i\theta}$ intorno all'origine.

A differenza delle trasformazioni paraboliche, non è detto che due differenti rotazioni siano necessariamente coniugate. Se però α e β sono due rotazioni coniugate di periodi p e q , necessariamente $p=q$. Di certo una trasformazione ellittica deve avere due punti fissi¹ ma non è vero il contrario, basta considerare la trasformazione $z \mapsto 2z$ che non è coniugata a nessuna rotazione ma fissa anch'essa 0 e ∞ .

¹in quanto $z \mapsto ze^{i\theta}$ è ellittica, fissa 0 ed ∞ e vale il Teorema 3.6

Sia α una trasformazione ellittica, α deve essere coniugata ad una qualche β della forma $\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$, dunque

$$\tau(\alpha) = \tau(\beta) = 4 \cos^2(\theta)$$

in particolare $\tau(\alpha) \in [0, 4)$.

Sia ora $\alpha \in SL_2(\mathbb{C})$ tale che $\tau(\alpha) \in [0, 4)$: allora esiste un qualche $\theta \in \mathbb{R}$ tale che $\tau(\alpha) = 4 \cos^2(\theta)$. Poiché $\tau(\alpha) \neq 4$, α non può essere parabolica, quindi α ha due punti fissi. Di conseguenza, α deve essere coniugata ad una qualche β che fissi 0 ed ∞ , quindi della forma $\begin{pmatrix} a & 0 \\ 0 & \frac{1}{a} \end{pmatrix}$, perciò

$$\left(a + \frac{1}{a}\right)^2 = \tau(\beta) = \tau(\alpha) = 4 \cos^2(\theta)$$

cioè $a = e^{\pm i\theta}$. In questo caso β è una rotazione e quindi α è ellittica.

Questo completa la dimostrazione del seguente teorema:

TEOREMA 3.9 $\alpha \in SL_2(\mathbb{C})$ è ellittica se e solo se $\tau(\alpha) \in [0, 4)$.

3.3.3 Trasformazioni iperboliche

Una trasformazione si dice iperbolica se è coniugata ad una dilatazione euclidea non banale $z \mapsto kz$ con $k > 0$. Di nuovo, dilatazioni differenti non sono necessariamente coniugate: si verifica facilmente che due dilatazioni $z \mapsto Kz$ e $z \mapsto Hz$ sono coniugate se e solo se $HK = 1$.

È evidente che una trasformazione iperbolica debba avere due punti fissi ma, anche in questo caso, non è vero il contrario: basta considerare $z \mapsto iz$ che non è coniugata a nessuna dilatazione.

Se α è iperbolica, è coniugata ad una qualche β della forma $\begin{pmatrix} \sqrt{k} & 0 \\ 0 & \frac{1}{\sqrt{k}} \end{pmatrix}$

quindi

$$\tau(\alpha) = \tau(\beta) = \left(\sqrt{k} + \frac{1}{\sqrt{k}}\right)^2$$

In particolare $\tau(\alpha) > 4$ (si noti che $\tau(\alpha) = 4$ significa $k = 1$, cioè $\alpha(z) = z \forall z$).

Sia ora $\alpha \in SL_2(\mathbb{C})$ tale che $\tau(\alpha) > 4$, si può trovare un $k \in \mathbb{R}$ tale che $\tau(\alpha) = \left(\sqrt{k} + \frac{1}{\sqrt{k}}\right)^2$. Poiché $\tau(\alpha) \neq 4$, α ha necessariamente due punti fissi.

Si può allora trovare β coniugata con α che fissi 0 e ∞ della forma $\begin{pmatrix} a & 0 \\ 0 & \frac{1}{a} \end{pmatrix}$

quindi

$$\left(a + \frac{1}{a}\right)^2 = \tau(\beta) = \tau(\alpha) = \left(\sqrt{k} + \frac{1}{\sqrt{k}}\right)^2$$

cioè $a = (\sqrt{k})^{\pm 1}$. Ciò implica che β sia una dilatazione dunque α sia iperbolica. Risulta provato che:

TEOREMA 3.10 $\alpha \in SL_2(\mathbb{C})$ è iperbolica se e solo se $\tau(\alpha) > 4$.

Sia Γ un sottogruppo discreto di $SL_2(\mathbb{R})$. Un punto $z \in \mathbb{H}$ è detto *ellittico* se è il punto fisso di un elemento ellittico $\alpha \in \Gamma$, un punto $s \in \mathbb{R} \cup \{\infty\}$ è detto *cuspidale* se è il punto fisso di un qualche elemento parabolico $\alpha \in \Gamma$.

LEMMA 3.11 Se z è un punto ellittico per Γ , allora $\{\alpha \in \Gamma : \alpha z = z\}$ è un gruppo ciclico finito.

DIM. È stato provato che $\exists \beta \in SL_2(\mathbb{R})$ tale che $\beta(i) = z$. L'applicazione $\alpha \mapsto \beta^{-1}\alpha\beta$ definisce un isomorfismo tra

$$\{\alpha \in \Gamma : \alpha z = z\} \approx SO_2(\mathbb{R}) \cap (\beta^{-1}\Gamma\beta)$$

e questo ultimo gruppo è finito. Le corrispondenze tra $\theta \leftrightarrow e^{2\pi i\theta} \leftrightarrow \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ sono anch'esse isomorfismi tra

$$\mathbb{R}/\mathbb{Z} \leftrightarrow \{z \in \mathbb{C} : |z| = 1\} \leftrightarrow SO_2(\mathbb{R}).$$

Infine $SO_2(\mathbb{R})_{\text{torsione}} \approx \mathbb{Q}/\mathbb{Z}$ ed ogni sottogruppo di \mathbb{Q}/\mathbb{Z} è ciclico (infatti è della forma $n^{-1}\mathbb{Z}/\mathbb{Z}$ dove n è il minimo comun denominatore degli elementi del gruppo).

□

Un altro risultato utile riguarda invece il gruppo modulare completo $SL_2(\mathbb{Z})$.

TEOREMA 3.12 *Le cuspidi di $SL_2(\mathbb{Z})$ sono esattamente i punti di $\mathbb{Q} \cup \{\infty\}$ ed ognuno è $SL_2(\mathbb{Z})$ -equivalente a ∞ .*

DIM. Di certo ∞ è un punto fisso della trasformazione parabolica associata a $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Sia $\frac{m}{n} \in \mathbb{Q}$, si possono supporre m ed n relativamente primi,

quindi $\exists r, s \in \mathbb{Z}$ tali che $rm - sn = 1$. Sia $\gamma = \begin{pmatrix} m & s \\ n & r \end{pmatrix}$, $\gamma(\infty) = \frac{m}{n}$ quindi $\frac{m}{n}$ è un punto fisso dell'elemento parabolico $\gamma T \gamma^{-1}$.

D'altra parte, ogni elemento parabolico $\alpha \in SL_2(\mathbb{Z})$ è coniugato con $\pm T$, dunque $\alpha = \pm \gamma T \gamma^{-1}$, $\gamma \in SL_2(\mathbb{Q})$. Il punto fissato da α è $\gamma(\infty)$ che appartiene a $\mathbb{Q} \cup \{\infty\}$.

□

3.4 Domini fondamentali

Sia Γ un sottogruppo finito di $SL_2(\mathbb{R})$. Un *dominio fondamentale* per Γ è un sottoinsieme aperto connesso F di \mathbb{H} tale che

$$\bigcup_{\gamma \in \Gamma} \gamma(\bar{F}) = \mathbb{H}, \quad \gamma_1(F) \cap \gamma_2(F) = \emptyset \text{ sse } \gamma_1 \neq \gamma_2$$

dove \bar{F} è la chiusura di F rispetto ad \mathbb{H} , cioè ogni punto di \mathbb{H} giace nella chiusura dell'immagine di F rispetto a una qualche $\gamma \in \Gamma$ e non esistono punti di \mathbb{H} equivalenti rispetto a Γ .

In altre parole, si è richiesto che la mappa $F \rightarrow \Gamma \backslash \mathbb{H}$ sia iniettiva e la mappa $\bar{F} \rightarrow \Gamma \backslash \mathbb{H}$ sia suriettiva. Si può dimostrare che ogni Γ ha un domino fondamentale, per quanto riguarda questa tesi sarà sufficiente poterlo dimostrare per $SL_2(\mathbb{Z})$.

Siano

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (3.1)$$

È evidente che:

$$Sz = -\frac{1}{z}, \quad Tz = z + 1$$

$$S^2 \equiv 1 \pmod{\pm I}, \quad (ST)^3 \equiv 1 \pmod{\pm I}$$

TEOREMA 3.13 *Sia $F = \{z \in \mathbb{H} : |z| > 1, |\Re(z)| < \frac{1}{2}\}$.*

(a) *F è un dominio fondamentale per $SL_2(\mathbb{Z})$; per la precisione, due elementi z e z' di \bar{F} sono equivalenti se e solo se*

(i) $\Re(z) = \pm \frac{1}{2}$ e $z' = z \pm 1$ (quindi $z' = Tz$ o $z = Tz'$), oppure

(ii) $|z| = 1$ e $z' = -\frac{1}{z} = Sz$.

(b) *Sia $z \in \bar{F}$; se lo stabilizzatore di z è diverso da $\pm I$, allora*

(i) $z = i$, e $\text{Stab}(i) = \langle S \rangle$ che ha periodo 2 in $PSL_2(\mathbb{Z})$, oppure

(ii) $z = \rho = e^{\frac{2\pi i}{6}}$ e $\text{Stab}(\rho) = \langle TS \rangle$ che ha periodo 3 in $PSL_2(\mathbb{Z})$,
oppure

(iii) $z = \rho^2$ e $\text{Stab}(\rho^2) = \langle ST \rangle$ che ha periodo 3 in $PSL_2(\mathbb{Z})$.

(c) Il gruppo $PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm I\}$ è generato da S e da T .

DIM. Sia Γ' il sottogruppo di $SL_2(\mathbb{Z})$ generato da S e da T . Si mostrerà che $\Gamma'(F) = \mathbb{H}$.

LEMMA 3.14 *Siano $z \in \mathbb{H}$ e $N \in \mathbb{N}$ fissati.*

Esiste solo un numero finito di coppie $(c, d) \in \mathbb{Z} \times \mathbb{Z}$ tali che

$$|cz + d| \leq N.$$

DIM. Sia $z = x + iy$. Se (c, d) è una delle coppie cercate,

$$|cz + d|^2 = (cx + d)^2 + c^2y^2,$$

dunque

$$c^2y^2 \leq (cx + d)^2 + c^2y^2 \leq N.$$

Poiché $z \in \mathbb{H}$, $y > 0$ quindi $|c| \leq \frac{N}{y}$, il che implica che ci sono solo un numero finito di possibilità per c .

Per ogni c ammissibile, l'equazione

$$(cx + d)^2 + c^2y^2 \leq N$$

mostra che esiste solo un numero finito di valori di d per i quali vale la tesi.

□

Si ricordi che, se $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R})$, allora $\Im(\gamma z) = \frac{\Im(z)}{|cz+d|^2}$.

Sia $z \in \mathbb{H}$ fissato. Si scelga $\gamma \in \Gamma'$ tale che $|cz + d|$ sia minimo². Dunque $\Im(\gamma z)$ è un massimo tra gli elementi dell'orbita di z .

²per il Lemma 3.14 tale γ deve esistere

Per qualche n , $z' \stackrel{def}{=} T^n(\gamma z)$ deve avere

$$-\frac{1}{2} \leq \Re(z') \leq \frac{1}{2}.$$

Necessariamente $|z'| \geq 1$. Se così non fosse,

$$\Im(Sz') = \Im\left(-\frac{1}{z'}\right) = \Im\left(\frac{-x' + iy'}{|z'|^2}\right) = \frac{\Im(z')}{|z'|^2} > \Im(z') = \Im(\gamma z),$$

che contraddice la scelta di γz . È appena stato dimostrato che $\Gamma'(F) = \mathbb{H}$.

Si supponga ora che $z, z' \in \mathbb{H}$ siano $SL_2(\mathbb{Z})$ -coniugati. Dunque $\Im(z) \geq \Im(z')$ oppure $\Im(z) \leq \Im(z')$, possiamo assumere la seconda senza perdere di generalità. Sia $z' = \gamma z$ con $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, e sia $z = x + iy$. Allora il nostro assunto implica che

$$(cx + d)^2 + c^2y^2 = |cz + d|^2 \leq 1.$$

Questo è impossibile se $c \geq 2$ (perché $y \geq \frac{1}{2}$), si debbono quindi considerare solamente i casi $c = 0, 1, -1$.

$c = 0$: dovrà essere $d = \pm 1$, quindi $\gamma = \pm \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$, cioè γ è una traslazione di b . Poiché $z, \gamma z \in \bar{F}$, questo implica che $b = \pm 1$ e si è nel caso (a(i)).

$c = 1$: se $|z + d| \leq 1$ deve essere $d = 0$, a meno che $z = \rho = \frac{1}{2} + i\frac{\sqrt{3}}{2}$ (per cui $d = 0$ o -1) oppure $z = \rho^2$ (per cui $d = 0$ o 1). Se $d = 0$, $\gamma = \pm \begin{pmatrix} a & -1 \\ 1 & 0 \end{pmatrix}$, e $\gamma z = a - \frac{1}{z}$. Se $a = 0$ si è nel caso (a(ii)). Se $a \neq 0$, allora $a = 1$ e $z = \rho^2$ oppure $a = -1$ e $z = \rho$.

$c = -1$: questo caso si tratta in maniera simile (o semplicemente cambiando i segni di a, b, c, d nel caso precedente).

Questo completa la dimostrazione dei punti (a) e (b).

Si vuole ora provare il punto (c). Sia $\gamma \in SL_2(\mathbb{Z})$. Si scelga un punto $z_0 \in F$. Poiché $SL_2(\mathbb{Z})(F) = \mathbb{H}$, esistono un elemento $\gamma' \in SL_2(\mathbb{Z})$ ed un punto $z \in \bar{F}$ tali che $\gamma'z = \gamma z_0 \in \bar{F}$. Dunque z_0 è $SL_2(\mathbb{Z})$ -equivalente a $(\gamma'^{-1})z_0 \in \bar{F}$. Siccome $z_0 \in F$, per il punto (a), $z_0 = (\gamma'^{-1}\gamma)z_0$. Quindi $\gamma'^{-1}\gamma \in \text{Stab}(z_0) \cap SL_2(\mathbb{Z}) = \pm I$, perciò γ e γ' sono equivalenti in $PSL_2(\mathbb{Z})$.

□

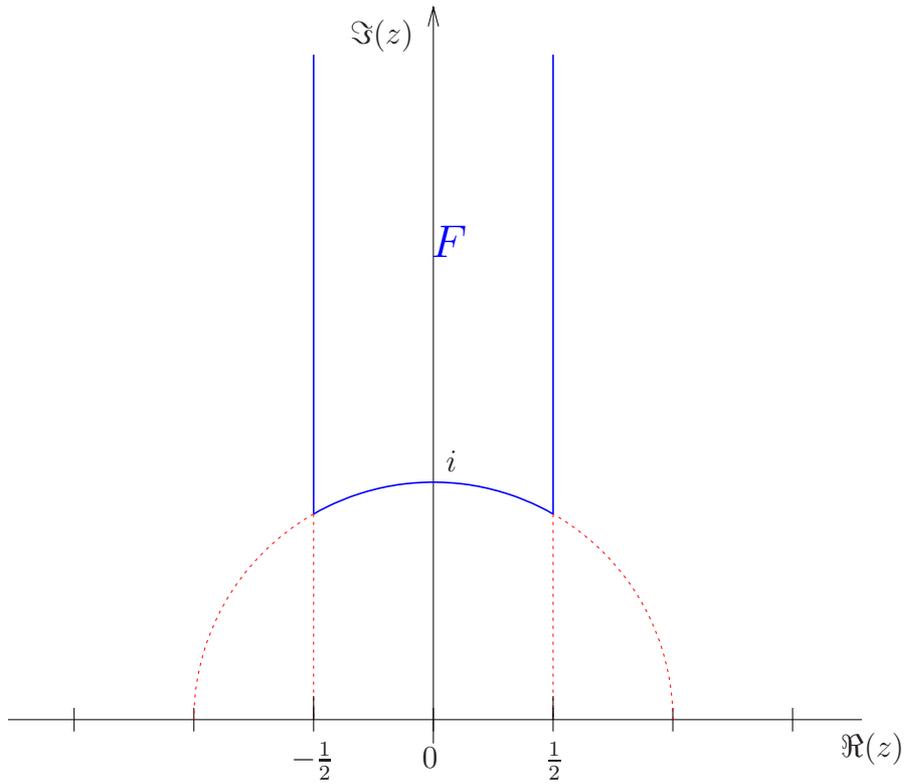


Figura 3.1: Il dominio fondamentale F

Sulla base di quanto osservato e dimostrato, la superficie modulare $SL_2(\mathbb{Z}) \backslash \mathbb{H}$ può essere identificata con lo spazio F / \sim dove i punti in F equivalenti sotto $SL_2(\mathbb{Z})$ sono identificati.

DIM. Si può finalmente completare la dimostrazione del Lemma 3.1.

Applicando la (2.15) con $p = e^{i\frac{\pi}{3}} = (0 + 1 \cdot \cos \frac{\pi}{3}, 1 \cdot \sin \frac{\pi}{3})$ e $q = e^{i\frac{2\pi}{3}} = (0 + 1 \cdot \cos \frac{2\pi}{3}, 1 \cdot \sin \frac{2\pi}{3})$ si ottiene $T \equiv F$ e dunque

$$A(\Gamma \backslash \mathbb{H}) \leq A(SL_2(\mathbb{Z}) \backslash \mathbb{H}) = A(F) = A(T) = \frac{2\pi}{3} - \frac{\pi}{3} = \frac{\pi}{3}.$$

□

Si denoti $\bar{\mathbb{H}} \stackrel{def}{=} \mathbb{H} \cup \{\infty\} \cup \mathbb{Q}$ la superficie iperbolica compattificata. I punti ∞ e $\frac{p}{q} \in \mathbb{Q}$ sono chiamati cuspidi e corrispondono a quelle che, come si è dimostrato, sono le cuspidi della superficie modulare.

Estendendo l'azione del gruppo modulare all'asse reale $\{z = x + iy : y = 0\}$, queste cuspidi decadono in alcune classi di equivalenza. Una cuspidale del gruppo modulare Γ è una di queste classi di equivalenza. Geometricamente le cuspidi rappresentano punti sulla frontiera della superficie modulare all'infinito.

Capitolo 4

Sviluppi in frazioni continue

Prima di parlare delle dinamiche simboliche dei flussi geodetici sulla superficie modulare, è necessario fare una pausa ed introdurre una serie di concetti preliminari che si riveleranno essenziali per il lavoro successivo. Si tratta degli *sviluppi in frazioni continue*.

Le frazioni continue sono state introdotte da Bombelli nel 1572 in uno studio per il calcolo approssimato delle radici quadrate di numeri che non siano quadrati perfetti. In effetti tali serie numeriche giocano un ruolo fondamentale nelle approssimazioni dei numeri per mezzo dei numeri razionali.

4.1 Definizioni preliminari

Ogni numero $x \in \mathbb{R}$ può essere rappresentato, come si potrà osservare, mediante uno *sviluppo in frazioni continue* dato da

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}} = [a_0; a_1, a_2, a_3, \dots] \in \mathbb{N}^{\mathbb{N}}, \quad (4.1)$$

dove $a_0 \in \mathbb{Z}$ e $a_i \in \mathbb{N} \forall i \geq 1$.

Per lavorare più agevolmente con tali costruzioni è utile definire le due funzioni che seguono:

- si chiama *parte intera* di x la funzione $[\cdot] : \mathbb{R} \rightarrow \mathbb{Z}$ che associa ad x il più grande intero k tale che $k \leq x < k + 1$;
- si chiama *parte frazionaria* di x la funzione $\{\cdot\} : \mathbb{R} \rightarrow [0, 1)$ definita $\{x\} = x - [x]$.

Poiché $\{x\} = [0; a_1, a_2, \dots]$, si può focalizzare l'attenzione sui numeri $x \in [0, 1)$. In altre parole da ora in poi si considererà che, quando non è indicato, $a_0 = 0$, ragion per cui la scrittura $[a_1, a_2, \dots]$ starà ad indicare $[0; a_1, a_2, \dots]$.

4.2 L'algoritmo di Euclide

Sia $T_G : (0, 1) \rightarrow (0, 1)$ la funzione definita $T_G x = \frac{1}{x} \bmod 1 = \left\{ \frac{1}{x} \right\}$ se $x \neq 0$ e $T_G 0 = 0$, T_G è chiamata *mappa di Gauss*. La seguente proprietà di shift sarà cruciale per la nostra analisi.

TEOREMA 4.1 $T_G[a_1, a_2, a_3, \dots] = [a_2, a_3, \dots]$. Cioè la mappa di Gauss $T_G : (0, 1) \rightarrow (0, 1)$ corrisponde allo shift su $\mathbb{N}^{\mathbb{N}}$, $\sigma[a_1, a_2, a_3, \dots] = [a_2, a_3, \dots]$.

DIM.

$$\begin{aligned}
 T_G[a_1, a_2, a_3, \dots] &= T_G \left[\frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}}} \right] = \\
 &= \left\{ a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}} \right\} = \\
 &= \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}} = \\
 &= [a_2, a_3, \dots].
 \end{aligned}$$

□

Posto $x_0 = x \in \mathbb{R}$, si scriva:

$$\begin{aligned} a_0 &= [x_0], & x_1 &= \{x_0\}, \\ a_1 &= \left[\frac{1}{x_1} \right], & x_2 &= \left\{ \frac{1}{x_1} \right\} = T_G x_1, \\ a_2 &= \left[\frac{1}{x_2} \right], & x_3 &= \left\{ \frac{1}{x_2} \right\} = T_G x_2, \\ &\dots, & &\dots \end{aligned}$$

Se $\exists N \geq 0$ tale che $x_{N+1} = 0$, l'algoritmo si ferma. Evidentemente lo sviluppo in frazioni continue di 0 è la frazione $[0;]$.

TEOREMA 4.2 *Ogni numero razionale può essere rappresentato da uno sviluppo in frazioni continue finito.*

DIM. Se x è un intero, è banale che $a_0 = x$ e l'algoritmo si ferma.

Se x non è intero,

$$x = \frac{h}{k}$$

dove $h, k \in \mathbb{Z}$ e $k > 1$. Poiché

$$\frac{h}{k} = a_0 + x_1, \quad h = a_0 k + x_1 k,$$

a_0 e $k_1 = x_1 k$ sono rispettivamente il quoziente ed il resto¹ della divisione $\frac{h}{k}$.

Se $x_1 \neq 0$ (come deve essere visto che x non è intero) si ha

$$\frac{1}{x_1} = \frac{k}{k_1} = a_1 + x_2, \quad k = a_1 k_1 + x_2 k_1,$$

¹Il resto, qui ed in futuro, è considerato non negativo. Se $a_0 \geq 0$ allora x ed h sono positivi e k_1 è il classico resto aritmetico; d'altra parte, se $a_0 < 0$, x e k sono negativi ed il resto diverrebbe $(x - [x])k$.

da cui a_1 e $k_2 = x_2 k_1$ sono rispettivamente il quoziente ed il resto della divisione $\frac{k}{k_1}$. Si ottengono quindi una serie di equazioni

$$h = a_0 k + k_1, \quad k = a_1 k_1 + k_2, \quad k_1 = a_2 k_2 + k_3, \quad \dots$$

che continuano fino a che $x_n \neq 0$, o equivalentemente $k_{n+1} \neq 0$.

Gli interi non negativi k, k_1, k_2, \dots formano una serie strettamente decrescente², per cui $\exists N$ tale che $k_{N+1} = 0$ e quindi $x_N = 0$, cioè l'algoritmo si ferma.

□

L'algoritmo rappresentato dal sistema di equazioni

$$\left\{ \begin{array}{ll} h = a_0 k + k_1 & (0 < k_1 < k) \\ k = a_1 k_1 + k_2 & (0 < k_2 < k_1) \\ \dots & \dots \\ k_{N-2} = a_{N-1} k_{N-1} + k_N & (0 < k_N < k_{N-1}) \\ k_{N-1} = a_N k_N & \end{array} \right. \quad (4.2)$$

è conosciuto come *algoritmo di Euclide*.

4.3 Una relazione di ricorrenza

Per il momento lo studio sarà rivolto solo agli sviluppi in frazioni continue finiti anche se tutte le proprietà che si dimostreranno valgono anche per quelli infiniti.

Generalmente $[a_1, \dots, a_n]$ con $0 \leq n \leq N$ viene chiamato *n-esimo convergente* ed a_n viene chiamato *n-esimo quoziente parziale*. È possibile trovare una formula ricorsiva per il calcolo di tali convergenti.

² $k_{i+1} = \theta k_i$ con $\theta \in (0, 1)$

Innanzitutto si noti che, secondo la definizione, è evidente che

$$\begin{aligned} [a_0; a_1, a_2, \dots, a_{n-1}, a_n] &= [a_0; a_1, a_2, \dots, a_{n-1} + \frac{1}{a_n}] = \dots = \\ &= [a_0; a_1, \dots, a_{m-1}, [a_m, a_{m+1}, \dots, a_n]] = [a_0; [a_1, \dots, a_n]]. \end{aligned} \quad (4.3)$$

LEMMA 4.3 *Sia N l'indice dell'ultimo quoziente parziale dello sviluppo di x . Se p_n e q_n sono definiti come segue*

$$\begin{aligned} p_0 &= a_0, & p_1 &= a_1 a_0 + 1, & p_n &= a_n p_{n-1} + p_{n-2} \quad (2 \leq n \leq N), \\ q_0 &= 1, & q_1 &= a_1, & q_n &= a_n q_{n-1} + q_{n-2} \quad (2 \leq n \leq N), \end{aligned} \quad (4.4)$$

allora

$$[a_0; a_1, a_2, \dots, a_n] = \frac{p_n}{q_n} \quad (4.5)$$

DIM. Applicando la definizione, è facile calcolare che

$$\begin{aligned} [a_0;] &= \frac{a_0}{1} = \frac{p_0}{q_0} \\ [a_0; a_1] &= a_0 + \frac{1}{a_1} = \frac{a_0 a_1 + 1}{a_1} = \frac{p_1}{q_1} \end{aligned}$$

Quindi il Lemma risulta verificato per $n = 0$ ed $n = 1$.

Si supponga ora vero per $n \leq m$, dove $m < N$. Allora

$$[a_0; a_1, a_2, \dots, a_{m-1}, a_m] = \frac{p_m}{q_m} = \frac{a_m p_{m-1} + p_{m-2}}{a_m q_{m-1} + q_{m-2}}$$

e p_{m-1} , p_{m-2} , q_{m-1} , q_{m-2} dipendono solo da a_0, a_1, \dots, a_{m-1} . In base all'uguaglianza (4.3) si ottiene

$$[a_0; a_1, a_2, \dots, a_{m-1}, a_m, a_{m+1}] = \left[a_0; a_1, a_2, \dots, a_{m-1}, a_m + \frac{1}{a_{m+1}} \right] \quad (4.6)$$

$$= \frac{\left(a_m + \frac{1}{a_{m+1}} \right) p_{m-1} + p_{m-2}}{\left(a_m + \frac{1}{a_{m+1}} \right) q_{m-1} + q_{m-2}} = \quad (4.7)$$

$$= \frac{a_{m+1}(a_m p_{m-1} + p_{m-2}) + p_{m-1}}{a_{m+1}(a_m q_{m-1} + q_{m-2}) + q_{m-1}} = \quad (4.8)$$

$$= \frac{a_{m+1} p_m + p_{m-1}}{a_{m+1} q_m + q_{m-1}} = \frac{p_{m+1}}{q_{m+1}} \quad (4.9)$$

da cui, per induzione, la tesi.

□

Si noti, inoltre, che

$$\frac{q_{n-1}}{q_n} = \frac{q_{n-1}}{a_n q_{n-1} + q_{n-2}} = \frac{1}{a_n + \frac{q_{n-1}}{q_{n-2}}},$$

per cui

$$\frac{q_{n-1}}{q_n} = [a_n, \dots, a_1]. \quad (4.10)$$

4.4 Frazioni continue, convergenti e proprietà

Segue direttamente dalla (4.4) che

$$\frac{p_n}{q_n} = \frac{a_n p_{n-1} + p_{n-2}}{a_n q_{n-1} + q_{n-2}} \quad (4.11)$$

e che

$$\begin{aligned} p_n q_{n-1} - p_{n-1} q_n &= (a_n p_{n-1} + p_{n-2}) q_{n-1} - p_{n-1} (a_n q_{n-1} + q_{n-2}) = \\ &= -(p_{n-1} q_{n-2} - p_{n-2} q_{n-1}). \end{aligned}$$

Ripetendo il procedimento con $n - 1, n - 2, \dots, 2$ al posto di n , si ottiene

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1} (p_1 q_0 - p_0 q_1) = (-1)^{n-1}$$

e

$$\begin{aligned} p_n q_{n-2} - p_{n-2} q_n &= (a_n p_{n-1} + p_{n-2}) q_{n-2} - p_{n-2} (a_n q_{n-1} + q_{n-2}) = \\ &= a_n (p_{n-1} q_{n-2} - p_{n-2} q_{n-1}) = (-1)^n a_n. \end{aligned}$$

Risulta quindi dimostrato che

TEOREMA 4.4 *Le funzioni p_n e q_n definite sopra soddisfano*

$$\begin{aligned} p_n q_{n-1} - p_{n-1} q_n &= (-1)^{n-1} && \text{(Formula di Lagrange)} \\ p_n q_{n-2} - p_{n-2} q_n &= (-1)^n \end{aligned} \quad (4.12)$$

in altre parole

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n-1}}{q_{n-1} q_n} \quad (4.13)$$

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{(-1)^n a_n}{q_{n-2} q_n} \quad (4.14)$$

Per dimostrare i seguenti teoremi si deve osservare che $\forall i > 0$ si ha $a_i \geq 0^3$.

Si definiscano $\xi_n = \frac{p_n}{q_n}$ e $x = \xi_N$. In questo modo il valore delle frazioni continue è x o ξ_N . Segue dalla (4.3) che

$$\begin{aligned} x &= [a_0; a_1, a_2, \dots, a_N] = [a_0; a_1, \dots, a_{n-1}, [a_n, a_{n+1}, \dots, a_N]] = \\ &= \frac{[a_n, a_{n+1}, \dots, a_N] p_{n-1} + p_{n-2}}{[a_n, a_{n+1}, \dots, a_N] q_{n-1} + q_{n-2}} \end{aligned}$$

TEOREMA 4.5 (a) *Al crescere di n , i convergenti pari ξ_{2n} crescono strettamente mentre quelli dispari ξ_{2n+1} decrescono strettamente.*

(b) $\xi_{2i+1} > \xi_{2j} \quad \forall i, j \in \mathbb{N}$

(c) *Il valore di una frazione continua è maggiore di ogni suo convergente pari e minore di ogni suo convergente dispari (fatta eccezione per l'ultimo convergente).*

DIM.

(a) Ogni q_n è positivo⁴, per la (4.14) $\xi_n - \xi_{n-2}$ ha segno $(-1)^n$. Da cui la tesi.

³ a_0 può anche essere negativo

⁴Per induzione si dimostra che è una somma di interi positivi

(b) Dalla (4.13), si evince che $\xi_n - \xi_{n-1}$ ha segno $(-1)^{n-1}$ per cui

$$x_{2m+1} > x_{2m}. \quad (4.15)$$

Se la tesi (b) fosse falsa, dovrebbe essere $x_{2m+1} \leq x_{2\mu}$ per qualche coppia (m, μ) . Se $\mu < m$, per il punto (a), $x_{2m+1} < x_{2m}$, mentre se $\mu > m$, $x_{2\mu+1} < x_{2\mu}$. Ma entrambe le disuguaglianze contraddicono la (4.15).

(c) Infine, per i punti precedenti, ξ_N è il maggiore dei convergenti pari, o il minore dei dispari, ragion per cui il punto (c) risulta verificato in entrambi i casi.

□

È evidente dalla definizione che p_n e q_n sono interi e che q_n è positivo. Se

$$[a_0; a_1, a_2, \dots, a_N] = \frac{p_N}{q_N} = x,$$

si dice che il numero x (che è necessariamente razionale) è rappresentato dalla frazione continua. Si mostrerà presto che con una piccola riserva, la rappresentazione è unica.

TEOREMA 4.6

- $q_n \geq q_{n-1}$ per $n \geq 1$, con la disuguaglianza stretta per $n > 1$.
- $q_n \geq n$ con la disuguaglianza stretta per $n > 3$.

DIM. Nel primo caso, $q_0 = 1$ e $q_1 = a_1 \geq 1$. Se $n \geq 2$, allora

$$q_n = a_n q_{n-1} + q_{n-2} \geq q_{n-1} + 1,$$

dunque $q_n > q_{n-1}$ e $q_n \geq n$. Se $n > 3$, allora

$$q_n \geq q_{n-1} + q_{n-2} > q_{n-1} + 1 \geq n,$$

per cui $q_n > n$.

□

Si può ormai dimostrare un'altra proprietà molto importante dei convergenti:

TEOREMA 4.7 *I convergenti ad una frazione continua sono nei loro termini minori.*

DIM. Sia d tale che $d|p_n$ e $d|q_n$, per il Teorema 4.4, $d|(-1)^{n-1}$ quindi $d|1$.

□

Banalmente, ogni frazione continua finita rappresenta un numero razionale. Si è visto anche come ogni numero razionale sia rappresentabile con una frazione continua finita. Si vuole ora mostrare che tale rappresentazione è, a parte per un'ambiguità, unica.

TEOREMA 4.8 *Se x è rappresentabile mediante una frazione continua con un numero dispari (pari) di quozienti parziali, è anche rappresentabile mediante una che ne ha in numero pari (dispari).*

DIM. Sia $x = [a_0; a_1, \dots, a_N]$. Se $a_N \geq 2$, $[a_0; a_1, \dots, a_N] = [a_0; a_1, \dots, a_N - 1, 1]$. Se invece $a_N = 1$, $[a_0; a_1, \dots, a_{N-1}, 1] = [a_0; a_1, \dots, a_{N-2}, a_{N-1} + 1]$.

□

Spesso scegliere una tra le due rappresentazioni alternative è utile.

Si chiamerà

$$a'_n = [a_n, a_{n+1}, \dots, a_N] \quad (0 \leq n \leq N)$$

n -esimo quoziente completo della frazione continua $[a_0, a_1, \dots, a_N]$.

Per cui

$$x = a'_0, \quad x = \frac{a'_1 a_0 + 1}{a'_1}$$

e per $2 \leq n \leq N$

$$x = \frac{a'_n p_{n-1} + p_{n-2}}{a'_n q_{n-1} + q_{n-2}}. \quad (4.16)$$

TEOREMA 4.9 $a_n = [a'_n]$, parte intera di a'_n , eccezione fatta per

$$a_{N-1} = [a'_{N-1}] - 1$$

quando $a_N = 1$.

DIM. Se $N = 0$, $a_0 = a'_0 = [a'_0]$.

Se $N > 0$,

$$a'_n = a_n + \frac{1}{a'_{n+1}} \quad (0 \leq n \leq N-1).$$

Ora $\frac{1}{a'_{n+1}} < 1$ per $0 \leq n \leq N-1$, ad eccezione del caso $a'_{n+1} = 1$ quando $n = N-1$ ed $a_N = 1$. Per cui

$$a_n < a'_n < a_n + 1 \quad (0 \leq n \leq N-1)$$

e

$$a_n = [a'_n] \quad (0 \leq n \leq N-1)$$

ad eccezione del caso specificato. In ogni caso $a_N = a'_N = [a'_N]$.

□

TEOREMA 4.10 Se due frazioni continue $[a_0; a_1, \dots, a_N]$ e $[b_0; b_1, \dots, b_M]$ con $a_N, b_M > 1$ corrispondono allo stesso valore x , allora $M = N$ e le frazioni continue sono identiche⁵.

⁵Due frazioni continue si dicono *identiche* se sono formate dalla stessa sequenza di quozienti parziali.

DIM. Per il Teorema 4.9, $a_0 = [x] = b_0$. Si supponga che i primi n quozienti parziali siano identici e che a'_n e b'_n siano i quozienti completi n -esimi. Allora $[a_0; a_1, \dots, a_{n-1}, a'_n] = [a_0; a_1, \dots, a_{n-1}, b'_n]$.

Se $n = 1$, allora $a_0 + \frac{1}{a'_1} = a_0 + \frac{1}{b'_1}$, per cui $a'_1 = b'_1$ e quindi per il Teorema 4.9, $a_1 = b_1$.

Se $n > 1$, allora sfruttando la (4.16) si ha $\frac{a'_n p_{n-1} + p_{n-2}}{a'_n q_{n-1} + q_{n-2}} = \frac{b'_n p_{n-1} + p_{n-2}}{b'_n q_{n-1} + q_{n-2}}$, cioè

$$(a'_n - b'_n)(p_{n-1}q_{n-2} - p_{n-2}q_{n-1}) = 0.$$

Ma è stato provato nel Teorema 4.4 che $p_{n-1}q_{n-2} - p_{n-2}q_{n-1} = (-1)^n$ e quindi $a'_n = b'_n$. Segue dal Teorema 4.9 che $a_n = b_n$.

Si su supponga ora, senza perdere di generalità, che $N \leq M$. Si è dimostrato fin qui che $a_n = b_n$ per $n \leq N$. Se fosse $M > N$, si avrebbe per la (4.16)

$$\frac{p_N}{q_N} = [a_0; a_1, \dots, a_N] = [a_0; a_1, \dots, a_N, b_{N+1}, \dots, b_M] = \frac{b'_{N+1}p_N + p_{N-1}}{b'_{N+1}q_N + q_{N-1}};$$

o anche $p_N q_{N-1} - q_N p_{N-1} = 0$ che è falso. Dunque $M = N$ e le frazioni sono identiche.

□

Confrontando i Teoremi 4.2, 4.8 e 4.10 si può finalmente affermare che

TEOREMA 4.11 *Un numero razionale può essere espresso come una frazione continua finita solamente in due modi, uno con un numero pari e l'altro con un numero dispari di convergenti. In una forma l'ultimo quoziente parziale è 1, nell'altra è maggiore di 1.*

4.5 La differenza tra frazione continua e convergente

In questa sezione si supporrà $N > 1$ ed $n > 0$. Riprendendo la (4.16), si avrebbe

$$x = \frac{a'_{n+1}p_n + p_{n-1}}{a'_{n+1}q_n + q_{n-1}}$$

per $1 \leq n \leq N - 1$, e dunque

$$x - \frac{p_n}{q_n} = -\frac{p_n q_{n-1} - q_n p_{n-1}}{q_n(a'_{n+1}q_n + q_{n-1})} = \frac{(-1)^n}{q_n(a'_{n+1}q_n + q_{n-1})},$$

ma anche

$$x - \frac{p_0}{q_0} = x - a_0 = \frac{1}{a'_1}.$$

Scrivendo

$$q'_1 = a'_1, \quad q'_n = a'_n q_{n-1} + q_{n-2} \quad (1 < n \leq N)$$

(quindi in particolare $q'_N = q_N$), si ottiene

TEOREMA 4.12 *Se $1 \leq n \leq N - 1$, allora*

$$x - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n q'_{n+1}}.$$

Questa formula fornisce una ulteriore dimostrazione del Teorema 4.5(c).

Si è visto nella dimostrazione del Teorema 4.9 che

$$a_{n+1} < a'_{n+1} < a_{n+1} + 1$$

per $n \leq N - 2$, eccezion fatta per $a'_{N-1} = a_{N-1} + 1$ quando $a_N = 1$. Dunque, se per il momento si ignora questo caso eccezionale, si ha

$$q_1 = a_1 < a'_1 < a_1 + 1 \leq q_2 \tag{4.17}$$

e

$$q'_{n+1} = a'_{n+1}q_n + q_{n-1} > a_{n+1}q_n + q_{n-1} = q_{n+1}, \quad (4.18)$$

$$q'_{n+1} < a_{n+1}q_n + q_{n-1} + q_n = q_{n+1} + q_n \leq a_{n+2}q_{n+1} + q_n = q_{n+2} \quad (4.19)$$

per $1 \leq n \leq N - 2$. Ne consegue che

$$\frac{1}{q_{n+2}} < |p_n - q_n x| < \frac{1}{q_{n+1}} \quad (n \leq N - 2) \quad (4.20)$$

mentre

$$|p_{N-1} - q_{N-1}x| = \frac{1}{q_N}, \quad p_N - q_N x = 0. \quad (4.21)$$

Nel caso eccezionale, la (4.19) sarebbe sostituita da

$$q'_{N-1} = (a_{N-1} + 1)q_{N-2} + q_{N-3} = q_{N-1} + q_{N-2} = q_N$$

e la prima disuguaglianza della (4.20) da una uguaglianza. In ogni caso la (4.20) mostrerebbe che $|p_n - q_n x|$ decresce strettamente al crescere di n . Inoltre poiché q_n cresce strettamente, $\left|x - \frac{p_n}{q_n}\right|$ decresce strettamente.

Si possono recuperare le conclusioni più importanti ottenendo

TEOREMA 4.13 *Se $N > 1$, $n > 0$, allora le differenze*

$$x - \frac{p_n}{q_n}, \quad q_n x - p_n$$

decrescono strettamente in valore assoluto al crescere di n . Inoltre

$$q_n x - p_n = \frac{(-1)^n \delta_n}{q_{n+1}},$$

dove $0 < \delta_n < 1$, $1 \leq n \leq N - 2$ e $\delta_{N-1} = 1$ e

$$\left|x - \frac{p_n}{q_n}\right| \leq \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2} \quad (4.22)$$

per $n \leq N - 1$, con due disuguaglianze strette ad eccezione del caso $n = N - 1$.

4.6 Frazioni continue infinite

Fin qui si è parlato quasi esclusivamente di frazioni continue finite e si è visto come rappresentino i numeri razionali. I maggiori interessi riguardo alle frazioni continue, però, riguardano le loro applicazioni nell'approssimazione dei numeri irrazionali, ma per questo è necessario che siano infinite.

Si suppongano $\{a_1, a_2, \dots, a_n\}$ un insieme di interi positivi ed a_0 un intero tali che

$$x_n = [a_0; a_1, \dots, a_n]$$

sia, per ogni n , una frazione continua che rappresenti un numero razionale x_n . Se, come si proverà tra breve, x_n tende ad un limite x finito quando $n \rightarrow \infty$, risulta naturale affermare che la frazione continua $[a_0; a_1, a_2, \dots]$ converga al valore x e scrivere

$$x = [a_0; a_1, a_2, \dots]. \quad (4.23)$$

TEOREMA 4.14 *Sia $\{a_n\}_{n \in \mathbb{Z}}$ tale che $a_i \geq 0 \forall i > 0$. Allora $x_n = [a_0; a_1, \dots, a_n]$ tende ad un limite finito x per $n \rightarrow \infty$.*

O equivalentemente

TEOREMA 4.15 *Tutte le frazioni continue sono convergenti.*

Si chiameranno $x_n = \frac{p_n}{q_n} = [a_0; a_1, \dots, a_n]$ i *convergenti* a (4.23). Si deve dimostrare che i convergenti tendono ad un certo limite finito.

DIM. Se $n \leq N$, il convergente x_n è anche un convergente ad $[a_0; a_1, \dots, a_N]$, quindi per il Teorema 4.5(a), i convergenti pari formano una serie crescente mentre quelli dispari una decrescente.

Per il Teorema 4.5(b) ogni convergente pari è minore di x_1 , per cui la serie crescente è superiormente limitata; d'altra parte ogni convergente dispari è

maggiore di x_0 , per cui la serie decrescente è limitata inferiormente. Allora i convergenti pari dovranno tendere ad un limite ζ_1 finito e quelli dispari ad un limite ζ_2 finito, tali che $\zeta_1 \leq \zeta_2$.

Finalmente, per i Teoremi 4.4 e 4.6,

$$\left| \frac{p_{2n}}{q_{2n}} - \frac{p_{2n-1}}{q_{2n-1}} \right| = \frac{1}{q_{2n}q_{2n-1}} \leq \frac{1}{2n(2n-1)} \rightarrow 0,$$

ma allora $\zeta_1 = \zeta_2 = x$ e la frazione continua $[a_0; a_1, \dots, a_n]$ converge ad x .

□

Accidentalmente si è anche dimostrato che

TEOREMA 4.16 *Una frazione continua infinita è minore di ogni suo convergente dispari e maggiore di ogni suo convergente pari.*

Da questo momento in poi si parlerà di *frazione continua* come abbreviazione di *valore della frazione continua*.

4.7 Numeri irrazionali e frazioni continue

Si chiama *n-esimo quoziente completo* della frazione continua $x = [a_0; a_1, a_2, \dots]$

la frazione continua $a'_n = [a_n, a_{n+1}, \dots]$.

Chiaramente

$$\begin{aligned} a'_n &= \lim_{N \rightarrow \infty} [a_n, a_{n+1}, \dots, a_N] = \\ &= a_n + \lim_{N \rightarrow \infty} \frac{1}{[a_{n+1}, \dots, a_N]} = a_n + \frac{1}{a'_{n+1}}, \end{aligned}$$

ed in particolare $x = a'_0 = a_0 + \frac{1}{a'_1}$. Inoltre, è banale osservare che

$$a'_n > a_n, \quad a'_{n+1} > a_{n+1} > 0, \quad 0 < \frac{1}{a'_{n+1}} < 1,$$

per cui $a_n = [a'_n]$.

TEOREMA 4.17 Se $[a_0; a_1, a_2, \dots] = x$, allora $\forall n \geq 0$

$$a_0 = [x], \quad a_n = [a'_n].$$

Da cui si può dedurre, esattamente come per il Teorema 4.10, che

TEOREMA 4.18 Due frazioni continue infinite che hanno lo stesso valore sono identiche.

Si riprenda l'algoritmo della sezione §4.2. Se x è irrazionale, il processo non può terminare, per cui definisce una sequenza infinita di interi a_0, a_1, \dots

Si ricordi che

$$x = [a_0; a'_1] = [a_0; a_1, a'_2] = \dots = [a_0; a_1, a_2, \dots, a_n, a'_{n+1}]$$

dove $a'_{n+1} = a_{n+1} + \frac{1}{a'_{n+2}} > a_{n+1}$.

Recuperando la (4.16), si ha

$$x = \frac{a'_{n+1}p_n + p_{n-1}}{a'_{n+1}q_n + q_{n-1}},$$

per cui

$$x - \frac{p_n}{q_n} = \frac{p_{n-1}q_n - p_nq_{n-1}}{q_n(a'_{n+1}q_n + q_{n-1})} = \frac{(-1)^n}{q_n(a'_{n+1}q_n + q_{n-1})},$$

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n(a_{n+1}q_n + q_{n-1})} = \frac{1}{q_nq_{n+1}} \leq \frac{1}{n(n+1)} \rightarrow 0$$

per $n \rightarrow \infty$. Dunque

$$x = \lim_{n \rightarrow \infty} \frac{p_n}{q_n} = [a_0; a_1, a_2, \dots],$$

e l'algoritmo converge alla frazione continua di valore x , che è unica per il Teorema 4.18.

TEOREMA 4.19 Ogni numero irrazionale può essere espresso in un unico modo come frazione continua infinita.

Naturalmente, il valore di una frazione continua infinita è necessariamente irrazionale, conseguenza del fatto che l'algoritmo termina solo nel caso di x razionali.

Se si definisce $q'_n = a'_n q_{n-1} + q_{n-2}$ come nella sezione §4.5, ripetendo lo stesso esatto procedimento, si dimostra che

TEOREMA 4.20 *I risultati dei Teoremi 4.12 e 4.13 valgono anche nel caso delle frazioni continue infinite (eccezion fatta per il riferimento ad N). In particolare*

$$\left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2} \quad (4.24)$$

4.8 Ultimi preliminari per arrivare agli irrazionali quadratici

Innanzitutto è necessario dimostrare un lemma fondamentale in cui si comincia a vedere qualche legame con i capitoli precedenti.

LEMMA 4.21 *Se*

$$x = \frac{P\zeta + R}{Q\zeta + S}$$

dove $\zeta > 1$ e P, Q, R ed S sono interi tali che $Q > S > 0$ e $PS - QR = \pm 1$, allora $\frac{R}{S}$ e $\frac{P}{Q}$ sono due convergenti successivi alla frazione continua di valore x . Se $\frac{R}{S}$ è l' $(n-1)$ -esimo convergente e $\frac{P}{Q}$ l' n -esimo, allora ζ è l' $(n+1)$ -quoziente completo.

DIM. Si può sviluppare $\frac{P}{Q}$ in frazioni continue

$$\frac{P}{Q} = [a_0; a_1, \dots, a_n] = \frac{p_n}{q_n}.$$

Per il Teorema 4.8, si può supporre n indifferentemente pari o dispari. Si scelga n tale che

$$PS - QR = \pm 1 = (-1)^{n-1}.$$

Ora $\text{mcd}(P, Q) = 1$ e $Q > 0$, lo stesso vale per p_n e q_n . Quanto osservato implica che $P = p_n$, $Q = q_n$ e

$$p_n S - q_n R = PS - QR = (-1)^{n-1} = p_n q_{n-1} - q_n p_{n-1},$$

in altre parole

$$p_n(S - q_{n-1}) = q_n(R - p_{n-1}).$$

Questo, accanto al fatto che $\text{mcd}(p_n, q_n) = 1$, implica che

$$q_n | (S - q_{n-1}). \quad (4.25)$$

Ma $q_n = Q > S > 0$ e $q_n \geq q_{n-1} > 0$, per cui $S - q_{n-1} < q_n$ che non è consistente con la (4.25) a meno che $S - q_{n-1} = 0$. Quindi

$$S = q_{n-1}, \quad R = p_{n-1}$$

e perciò

$$x = \frac{p_n \zeta + p_{n-1}}{q_n \zeta + q_{n-1}}$$

o anche

$$x = [a_0; a_1, \dots, a_n, \zeta].$$

Se si sviluppa ζ come frazione continua si ottiene $\zeta = [a_{n+1}, a_{n+1}, \dots]$ dove $a_{n+1} = [\zeta] \geq 1$. Quindi $x = [a_0; a_1, \dots, a_n, a_{n+1}, \dots]$ è una frazione continua. In tal caso $\frac{p_{n-1}}{q_{n-1}}$ e $\frac{p_n}{q_n}$, che coincidono con $\frac{R}{S}$ e $\frac{P}{Q}$ sono due convergenti consecutivi di questa frazione continua e ζ è l'(n+1)-esimo quoziente parziale.

□

Se ξ ed η sono due numeri tali che

$$\xi = \frac{a\eta + b}{c\eta + d},$$

dove a, b, c, d sono interi tali che $ad - bc = \pm 1$, allora si dice che ξ è *equivalente* ad η . In particolare ξ è equivalente a se stesso⁶.

Se ξ è equivalente ad η , allora

$$\eta = \frac{-d\xi + b}{c\xi - a}, \quad (-d)(-a) - bc = ad - bc = \pm 1,$$

per cui η è equivalente a ξ . Per cui l'equivalenza così definita è simmetrica.

TEOREMA 4.22 *Se ξ ed η sono equivalenti ed η e ζ sono equivalenti, allora ξ e ζ sono equivalenti.*

DIM. Siano

$$\begin{aligned} \xi &= \frac{a\eta + b}{c\eta + d}, & ad - bc &= \pm 1; \\ \eta &= \frac{a'\zeta + b'}{c'\zeta + d'}, & a'd' - b'c' &= \pm 1. \end{aligned}$$

allora

$$\xi = \frac{A\zeta + B}{C\zeta + D}$$

dove

$$a = aa' + bc', \quad B = ab' + bd', \quad C = ca' + dc', \quad D = cb' + dd',$$

$$\text{e } AD - BC = (ad - bc)(a'd' - b'c') = \pm 1.$$

□

⁶ $a = d = 1, b = c = 0.$

In altre parole la relazione appena definita è anche transitiva. Si è dunque definita una relazione di equivalenza a tutti gli effetti, il che permette di riarrangiare gli irrazionali in classi di irrazionali equivalenti.

Se h e k sono interi primi tra loro, allora devono esistere due interi h' e k' tali che $hk' - k'h = 1$; e dunque

$$\frac{h}{k} = \frac{h' \cdot 0 + h}{k' \cdot 0 + k} = \frac{a \cdot 0 + b}{c \cdot 0 + d},$$

con $ad - bc = -1$. Dunque ogni numero razionale è equivalente a 0 e, per il Teorema 4.22, ad ogni altro numero razionale.

TEOREMA 4.23 *Due qualunque numeri razionali sono equivalenti.*

Da questo momento in poi l'attenzione sarà rivolta ai numeri irrazionali rappresentati mediante uno sviluppo in frazioni continue infinito.

TEOREMA 4.24 *Due numeri irrazionali ξ ed η sono equivalenti se e solo se*

$$\xi = [a_0; a_1, \dots, a_m, c_0, c_1, c_2, \dots], \quad \eta = [b_0; b_1, \dots, b_n, c_0, c_1, c_2, \dots], \quad (4.26)$$

cioè se la sequenza dei quozienti parziali di ξ da un certo m -esimo in poi è la stessa dei quozienti parziali di η da un certo n -esimo in poi.

DIM. Si suppongano ξ ed η come in (4.26) e si scriva $\omega = [c_0; c_1, c_2, \dots]$.

Dunque

$$\xi = [a_0; a_1, \dots, a_m, \omega] = \frac{p_m \omega + p_{m-1}}{q_m \omega + q_{m-1}}$$

e $p_m q_{m-1} - p_{m-1} q_m = \pm 1$, il che implica che ξ ed ω siano equivalenti. Esattamente nello stesso modo si prova che η ed ω sono equivalenti. La condizione è dunque sufficiente. Resta da mostrare che è necessaria.

Siano ξ ed η due numeri equivalenti, di certo

$$\eta = \frac{a\xi + b}{c\xi + d}$$

con $ad - bc = \pm 1$. Si può supporre che $c\xi + d > 0$ ⁷. Sviluppando ξ con l'algoritmo delle frazioni continue, si ottiene

$$\begin{aligned}\xi &= [a_0; a_1, a_2, \dots, a_k, a_{k+1}, \dots] = \\ &= [a_0; a_1, \dots, a_{k-1}, a'_k] = \frac{p_{k-1}a'_k + p_{k-2}}{q_{k-1}a'_k + q_{k-2}}.\end{aligned}$$

Quindi

$$\eta = \frac{Pa'_k + R}{Qa'_k + S}$$

dove

$$P = ap_{k-1} + bq_{k-1}, \quad R = ap_{k-2} + bq_{k-2},$$

$$Q = cp_{k-1} + dq_{k-1}, \quad S = cp_{k-2} + dq_{k-2}.$$

Ma allora P, Q, R ed S sono interi e $PS - QR = (ad - bc)(p_{k-1}q_{k-2} - p_{k-2}q_{k-1}) = \pm 1$. Per il Teorema 4.20,

$$p_{k-1} = \xi q_{k-1} + \frac{\delta}{q_{k-1}}, \quad p_{k-2} = \xi q_{k-2} + \frac{\delta'}{q_{k-2}}$$

dove $|\delta|, |\delta'| < 1$. Quindi

$$Q = (c\xi + d)q_{k-1} + \frac{c\delta}{q_{k-1}}, \quad S = (c\xi + d)q_{k-2} + \frac{c\delta'}{q_{k-2}}.$$

Ora $c\xi + d > 0$, $q_{k-1} > q_{k-2} > 0$ e q_{k-1} e q_{k-2} divergono a ∞ , quindi $Q > S > 0$ per k sufficientemente grande. Per tale k

$$\eta = \frac{P\zeta + R}{Q\zeta + S}$$

con $PS - QR = \pm 1$, $Q > S > 0$, $\zeta = a'_k > 1$. Per il Lemma 4.21,

$$\eta = [b_0; b_1, b_2, \dots, b_h, \zeta] = [b_0; b_1, b_2, \dots, b_h, a_k, a_{k+1}, \dots],$$

per qualche b_0, b_1, \dots, b_h . Risulta così provata anche la necessità della condizione.

□

⁷In caso contrario basta moltiplicare per -1 il numeratore e il denominatore.

4.9 Irrazionali quadratici e frazioni continue periodiche

Una *frazione continua periodica* è una frazione continua nella quale $a_l = a_{l+k}$ per un $k > 0$ fissato e tutti gli $l \geq L$. L'insieme dei quozienti parziali $\{a_L, a_{L+1}, \dots, a_{L+k-1}\}$ è chiamato *periodo* e la frazione continua può essere scritta

$$[a_0; a_1, \dots, a_{L-1}, \overline{a_L, a_{L+1}, \dots, a_{L+k-1}}].$$

TEOREMA 4.25 *Una frazione continua periodica è un irrazionale quadratico, cioè la radice irrazionale di un'equazione quadratica a coefficienti interi.*

DIM. Sia a'_L l' L -esimo quoziente completo della frazione continua periodica x , si avrà

$$\begin{aligned} a'_L &= [a_L, a_{L+1}, \dots, a_{L+k-1}, a_L, a_{L+1}, \dots] = \\ &= [a_L, a_{L+1}, \dots, a_{L+k-1}, a'_L], \\ \Rightarrow a'_L &= \frac{p'a'_L + p''}{q'a'_L + q''} \end{aligned}$$

da cui

$$q'a_L'^2 + (q'' - p')a'_L - p'' = 0, \quad (4.27)$$

dove $\frac{p''}{q''}$ e $\frac{p'}{q'}$ sono gli ultimi due convergenti a $[a_L, a_{L+1}, \dots, a_{L+k-1}]$. Ma

$$x = \frac{p_{L-1}a'_L + p_{L-2}}{q_{L-1}a'_L + q_{L-2}} \Rightarrow a'_L = \frac{p_{L-2} - q_{L-2}x}{q_{L-1}x - p_{L-2}}.$$

Sostituendo a'_L in (4.27) e rimuovendo le frazioni si ottiene

$$ax^2 + bx + c = 0, \quad (4.28)$$

un'equazione quadratica a coefficienti interi⁸. Poiché x deve essere irrazionale, $b^2 - 4ac \neq 0$. Segue che x è un irrazionale quadratico.

□

Il teorema inverso è altrettanto vero, ma la sua dimostrazione mostra qualche difficoltà in più.

TEOREMA 4.26 *La frazione continua che rappresenta un irrazionale quadratico è periodica.*

DIM. Un irrazionale quadratico soddisfa un'equazione quadratica a coefficienti interi, che può essere scritta nella forma (4.28). Se $x = [a_0; a_1, a_2, \dots, a_n, \dots]$, di certo

$$x = \frac{p_{n-1}a'_n + p_{n-2}}{q_{n-1}a'_n + q_{n-2}};$$

sostituendo x nella (4.28), si ottiene

$$A_n a_n'^2 + B_n a_n' + C_n = 0 \tag{4.29}$$

dove

$$\begin{aligned} A_n &= ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2, \\ B_n &= 2ap_{n-1}p_{n-2} + b(p_{n-1}q_{n-2} + p_{n-2}q_{n-1}) + 2cq_{n-1}q_{n-2}, \\ C_n &= ap_{n-2}^2 + bp_{n-2}q_{n-2} + cq_{n-2}^2. \end{aligned}$$

Se $A_n = ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2 = 0$, la (4.28) avrebbe la radice razionale $\frac{p_{n-1}}{q_{n-1}}$, il che è impossibile perché x è irrazionale. Quindi $A_n \neq 0$ e

$$A_n y^2 + B_n y + C_n = 0$$

⁸Infatti $a = q'q_{L-2}^2 - (q'' - p')q_{L-1}q_{L-2}$, $b = (q'' - p')(p_{L-2}q_{L-1} + p_{L-1}q_{L-2}) - 2q'p_{L-2}q_{L-2} - p''q_{L-1}$ e $c = q'p_{L-2}^2 + (q'' - p')p_{L-1}p_{L-2} + p''p_{L-1}$, cioè sono prodotti, somme e differenze di numeri interi

è un'equazione che ha a'_n come una delle radici. Con un breve calcolo si può osservare che

$$B_n^2 - 4A_nC_n = (b^2 - 4ac)(p_{n-1}q_{n-2} - p_{n-2}q_{n-1})^2 = b^2 - 4ac. \quad (4.30)$$

Per il Teorema 4.20,

$$p_{n-1} = xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}} \quad (|\delta_{n-1}| < 1).$$

Per cui

$$\begin{aligned} A_n &= a \left(xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}} \right)^2 + bq_{n-1} \left(xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}} \right) + cq_{n-1}^2 = \\ &= (ax^2 + bx + c)q_{n-1}^2 + 2ax\delta_{n-1} + a\frac{\delta_{n-1}^2}{q_{n-1}^2} + b\delta_{n-1} = \\ &= 2ax\delta_{n-1} + a\frac{\delta_{n-1}^2}{q_{n-1}^2} + b\delta_{n-1} \end{aligned}$$

che implica $|A_n| < 2|ax| + |a| + |b|$, inoltre, visto e considerato che $C_n = A_{n-1}$, $|C_n| < 2|ax| + |a| + |b|$.

Finalmente, sfruttando la (4.30), si può osservare che

$$\begin{aligned} B_n^2 &\leq 4|A_nC_n| + |b^2 - 4ac| < \\ &< 4(2|ax| + |a| + |b|)^2 + |b^2 - 4ac|. \end{aligned}$$

Quindi i moduli di A_n , B_n e C_n sono minori di numeri indipendenti da n .

Segue che c'è solo un numero finito di triplette differenti (A_n, B_n, C_n) , perciò da un certo n in poi si potrà trovare una tripletta (A, B, C) che compare almeno tre volte, si indichi $(A_{n_1}, B_{n_1}, C_{n_1})$, $(A_{n_2}, B_{n_2}, C_{n_2})$ e $(A_{n_3}, B_{n_3}, C_{n_3})$. Ovviamente a'_{n_1} , a'_{n_2} e a'_{n_3} debbono essere tutte e tre radici della stessa equazione

$$Ay^2 + By + C = 0,$$

e naturalmente due di loro non possono che coincidere. Ma se, per esempio, fosse $a'_{n_1} = a'_{n_2}$, si avrebbe

$$a_{n_2} = a_{n_1}, \quad a_{n_2+1} = a_{n_1+1}, \quad \dots, \quad a_{n_2+k} = a_{n_1+k}, \quad \dots$$

e la frazione continua sarebbe periodica.

□

Capitolo 5

Dinamica simbolica e flusso geodetico su una superficie modulare

Si può finalmente tornare al discorso iniziale per riprendere il flusso geodetico su \mathbb{H} dotato della metrica iperbolica e trasportarlo su $SL_2(\mathbb{Z})\backslash\mathbb{H}$.

5.1 Il flusso geodetico su una superficie modulare

Come si è dimostrato in precedenza, le geodetiche dello spazio iperbolico \mathbb{H} sono le semicirconferenze centrate sull'asse reale e le semirette¹ che partono dall'asse reale parallele all'asse immaginario.

Sfruttando la proprietà per cui per due punti passa una ed una sola geodetica, si può identificare ogni geodetica γ orientata con i suoi due punti

¹Si potrebbe vederle come semicirconferenze di raggio infinito.

base $\gamma_{-\infty}, \gamma_{+\infty} \in \mathbb{R} \cup \{\infty\}$, dunque si può scrivere $\gamma = (\gamma_{-\infty}, \gamma_{+\infty})$. Sia $\Gamma = SL_2(\mathbb{Z})$.

Sia \mathcal{A} il seguente insieme dei geodetiche:

$$\mathcal{A} \stackrel{def}{=} \{\gamma = (\gamma_{-\infty}, \gamma_{+\infty}) : 0 < |\gamma_{-\infty}| \leq 1 \leq |\gamma_{+\infty}|, \quad \gamma_{+\infty}\gamma_{-\infty} < 0\}. \quad (5.1)$$

Non è difficile osservare che per ogni geodetica γ in \mathbb{H} esiste un elemento $g \in \Gamma$ tale che $g\gamma \in \mathcal{A}$. Questo significa che ogni geodetica in $\Gamma \backslash \mathbb{H}$ ha un rappresentante nell'insieme \mathcal{A} .

Certamente è interessante osservare che le geodetiche dell'insieme \mathcal{A} , quando sono semicirconferenze², devono necessariamente essere incidenti all'asse immaginario; considerando che, fissati un punto z_0 ed un vettore unitario v_0 applicato al punto z_0 , esiste una sola geodetica orientata passante per z_0 e tangente nel punto con pendenza v_0 , si potrebbero identificare tutte queste geodetiche incidenti con coppia (z, v_z) nel prodotto cartesiano tra l'insieme dei punti dell'asse immaginario e l'insieme dei vettori unitari applicati ad ogni punto.

Si denoti con $\pi^\Gamma : \mathbb{H} \rightarrow \Gamma \backslash \mathbb{H}$ il rivestimento³ $\pi^\Gamma(z) = \Gamma z$.

Ovviamente, per ogni geodetica γ in \mathbb{H} , $\pi^\Gamma \gamma$ denota una geodetica in $\Gamma \backslash \mathbb{H}$: se $\gamma_1 = \gamma_2$ per qualche $g \in \Gamma$, allora $\pi^\Gamma \gamma_1 = \pi^\Gamma \gamma_2$. Il flusso geodetico ϕ_t su \mathbb{H} è il flusso lungo le geodetiche con velocità v tale che $|v| = 1$.

Se $T^1\mathbb{H}$ denota il fibrato tangente unitario⁴ su \mathbb{H} , allora il flusso geodetico $\phi_t : T^1\mathbb{H} \rightarrow T^1\mathbb{H}$ è dato da

$$\phi_t(z, v) = (z_t, v_t) \quad (5.2)$$

²Di raggio finito

³Che esiste per il seguente Teorema (cfr. [Ber02, CCL99, Ser94]): *Ogni varietà connessa possiede un rivestimento universale*

⁴Cioè l'unione di tutti gli spazi tangenti unitari corrispondenti ai diversi punti di una varietà differenziabile

dove z_t è un punto a distanza iperbolica t da z lungo la geodetica passante per z e con vettore unitario v tangente in z , mentre v_t è il vettore unitario tangente alla geodetica nel punto z_t .

Naturalmente il flusso $\pi_*^\Gamma \phi_t (\pi_*^\Gamma)^{-1} : T^1(\Gamma \backslash \mathbb{H}) \rightarrow T^1(\Gamma \backslash \mathbb{H})$ indotto dalla proiezione $\pi_*^\Gamma : T^1\mathbb{H} \rightarrow T^1(\Gamma \backslash \mathbb{H})$ è il flusso geodetico sul fibrato tangente unitario della superficie modulare $\Gamma \backslash \mathbb{H}$. Si denoterà questo flusso geodetico con ϕ_t^Γ .

5.2 Dinamica simbolica del flusso geodetico per le superfici modulari

Si vuole ora determinare la dinamica simbolica del flusso $\phi_t : T^1M \rightarrow T^1M$, dove $M \stackrel{def}{=} SL_2(\mathbb{Z}) \backslash \mathbb{H}$.

Per trovare una *sezione di Poincaré*⁵ per questo flusso, si consideri l'insieme

$$S = \{z = iy : y \geq 1\} \subset \mathbb{H}. \quad (5.3)$$

Per ogni $z \in S$ si definisca

$$C_z = \{v \in T_z\mathbb{H} : |v| = 1 \text{ t.c. } \gamma(z, v) \in \mathcal{A} \text{ oppure } S\gamma(z, v) \in \mathcal{A}\}$$

insieme di vettori tangenti dove $\gamma(z, v)$ denota la geodetica determinata da z e v , ed S è l'inversione definita in (3.1). In questo modo, se la geodetica γ' intersecasse l'asse immaginario in un punto con $y < 1$, la geodetica $S\gamma'$ lo intercetterebbe in un punto con $y' = \frac{1}{y} > 1$.

⁵Anche chiamata *sezione globale*, è una sottovarietà dello spazio considerato di codimensione 1 in cui il flusso, che resta sempre incidente rispetto alla sottovarietà, ritorna infinite volte nel passato e nel futuro della sua evoluzione

Allora l'insieme

$$X = \bigcup_{\substack{z \in S \\ v \in C_z}} \pi_*^{SL_2(\mathbb{Z})}(z, v) \subset T^1M$$

è una sezione di Poincaré per il flusso geodetico su T^1M . Effettivamente quasi tutte le orbite di ϕ_t incrociano X trasversalmente. La mappa $\varrho : X \rightarrow \mathcal{A}$ definita

$$\begin{aligned} \varrho(\pi_*^{SL_2(\mathbb{Z})}(z, v)) &= \gamma(z, v), \\ \varrho(\pi_*^{SL_2(\mathbb{Z})}(z, v)) &= S\gamma(z, v) \end{aligned}$$

lascia rispettivamente invariate o scambia le geodetiche γ_{\pm} con

$$\gamma_{\pm} = (\pm 1, \mp 1)$$

che quindi sono coincidenti in X .

Un'orbita $\phi_t(z, v)$ in T^1M è periodica se e solo se le geodetiche $\gamma(z, v)$ in \mathbb{H} sono fissate da un qualche $g \in SL_2(\mathbb{Z})$ diverso dall'identità, questo vuol dire per $\gamma = (\gamma_{-\infty}, \gamma_{+\infty})$ che

$$g\gamma_{\pm\infty} = \gamma_{\pm\infty}.$$

Dunque se $\pi_*^{SL_2(\mathbb{Z})}(z, v) \in X$ è un punto di quest'orbita periodica, è anche un punto periodico della mappa di Poincaré $P : X \rightarrow X$ con

$$Px = \phi_{r(x)}x$$

dove $r : X \rightarrow \mathbb{R}_+$ denota il tempo di ricorrenza per i punti nella sezione di Poincaré X , cioè

$$r(x) = \min\{t > 0 : \phi_t(x) \in X \text{ per } x \in X\}.$$

Per determinare tale mappa di Poincaré si consideri per $x = \pi_*^{SL_2(\mathbb{Z})}(z, v) \in X$ la geodetica $\gamma = (\gamma_{-\infty}, \gamma_{+\infty}) \in \mathcal{A}$ tale che $\varrho(\pi_*^{SL_2(\mathbb{Z})}(z, v)) = \gamma$.

Consideriamo una geodetica $\gamma = (\gamma_{-\infty}, \gamma_{+\infty}) \in \mathcal{A}$ tale che $\gamma_{+\infty} > 3$. Capire cosa succeda su M è più facile se si riesce a visualizzarlo su \mathbb{H} . L'azione di $SL_2(\mathbb{Z})$ su \mathbb{H} non fa altro che tassellare lo spazio con triangoli geodetici equivalenti. Poiché in $SL_2(\mathbb{Z})$ sono presenti la traslazione unitaria e tutte le sue potenze, ogni punto del dominio fondamentale è equivalente al suo corrispondente per traslazioni a coefficiente intero⁶. In altre parole tutta una fascia di \mathbb{H} è tassellata da triangoli geodetici asintotici con i lati verticali di coordinate $\frac{k}{2}$, con k intero.

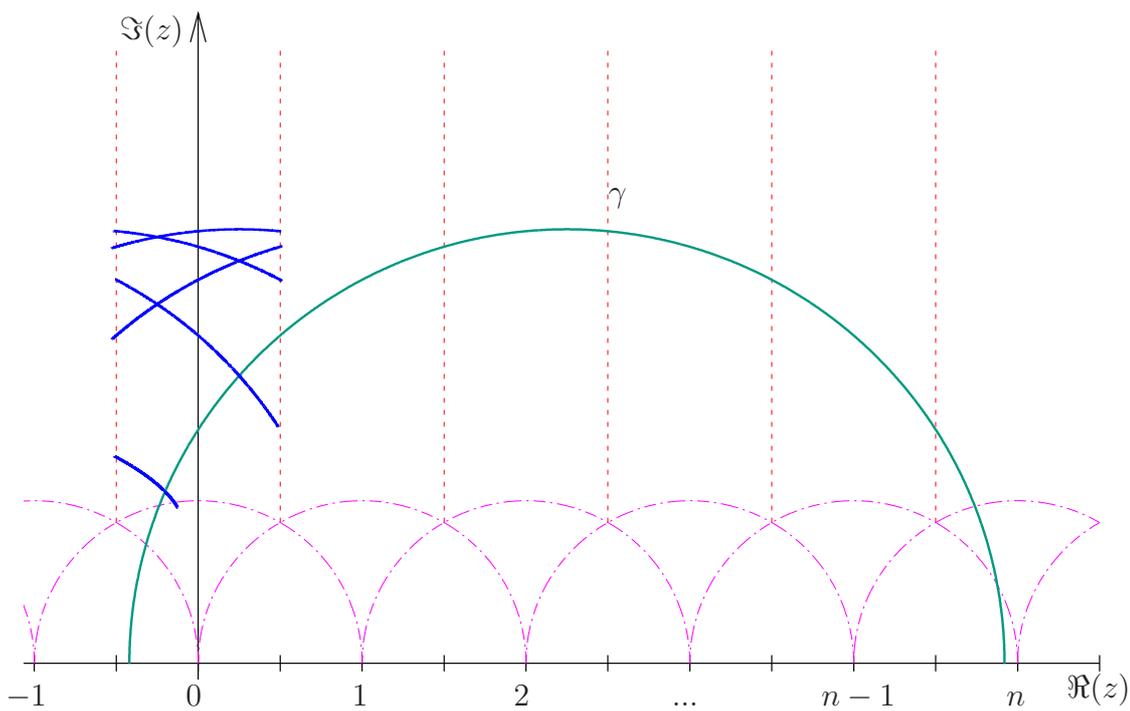


Figura 5.1: Una geodetica di \mathcal{A} in \mathbb{H}

Questo vuol dire che dopo aver intersecato X , la geodetica considerata,

⁶Graficamente, traslazioni verso destra e verso sinistra.

arriva (necessariamente) ad intersecare $\Re z = \frac{1}{2}$ e qui viene traslata di -1 in una geodetica equivalente γ' , a questo punto però $\gamma'_{-\infty}$ e $\gamma'_{+\infty}$ hanno entrambi modulo maggiore di 1 per cui γ' non appartiene più ad \mathcal{A} .

La nuova geodetica attraversa il dominio fondamentale fino ad intersecare $\Re z = \frac{1}{2}$ e di nuovo viene traslata di -1 in un'altra geodetica equivalente γ'' . Questo si ripete fino a che non si ottiene una qualche geodetica $\gamma^{(h)}$ tale che $[\gamma^{(h)}] = 0$, a questo punto la geodetica ha un estremo tra -1 ed 1 . Purtroppo non appartiene ancora ad \mathcal{A} perché tale estremo è $\gamma_{+\infty}$, ma $|\gamma_{+\infty}| < 1$ implica che il flusso geodetico debba attraversare la circonferenza $|z| = 1$, quindi $\gamma^{(h)}$ viene ribaltata da $S = -\frac{1}{z}$ che ne cambia l'orientazione.

Questa volta, finalmente, la geodetica risultante appartiene ad \mathcal{A} e quindi la sua intersezione con $\Re z = 0$ fornisce il primo punto di ritorno in X .

Un'obiezione riguarda il caso in cui $\gamma_{+\infty} \approx n + 1$, infatti si potrebbe pensare di operare un'altra traslazione di -1 , ma questo porterebbe entrambi gli estremi $\gamma_{\pm\infty}$ ad essere negativi, a quel punto l'unico modo per poter riportare in \mathcal{A} la geodetica è traslarla di 1 , annullando di fatto l'ultima trasformazione.

Sia $n < \gamma_{+\infty} < n + 1$, in base a quanto osservato per la $\gamma^{(h)}$ è necessaria una traslazione di $-n$, allora la geodetica γ' definita

$$\gamma' = ST^{-n}\gamma,$$

dove S e T sono le trasformazioni associate alle matrici definite in (3.1), appartiene all'insieme \mathcal{A} in quanto $\gamma' = (\gamma'_{-\infty}, \gamma'_{+\infty})$ è tale che $0 < \gamma'_{-\infty} < 1$ e $\gamma'_{+\infty} < -1$. Come è già stato detto, questa geodetica γ' determinerà un punto $x' \in X$ e $x' = Px$.

Sia $\epsilon = 1$ se $\gamma_{-\infty} < 0 < \gamma_{+\infty}$ mentre $\epsilon = -1$ se $\gamma_{+\infty} < 0 < \gamma_{-\infty}$. Si può

dunque identificare la geodetica orientata $\gamma = (\gamma_{-\infty}, \gamma_{+\infty}) \in \mathcal{A}$ con

$$\gamma = (x_1, x_2, \epsilon)$$

dove $(x_1, x_2) \in \overline{I^2} = [0, 1] \times [0, 1]$ ed $\epsilon \in \{-1, 1\}$, in modo che $\gamma_{-\infty} = -\epsilon x_2$ e $\gamma_{+\infty} = \epsilon x_1^{-1}$.

Nella parametrizzazione $\gamma = (x_1, x_2, \epsilon)$ degli elementi $\gamma \in \mathcal{A}$ si può calcolare esplicitamente la $\gamma' = (x'_1, x'_2, \epsilon')$ come segue:

$$(x'_1, x'_2, \epsilon') = P(x_1, x_2, \epsilon) = \left(\frac{1}{x_1} - n, \frac{1}{x_2 + n}, -\epsilon \right) \text{ dove } n = \left[\frac{1}{x_1} \right].$$

Non è difficile convincersi che per $\gamma_{+\infty} < -1$ vale la stessa formula, basta ripetere esattamente il ragionamento di prima.

5.3 Irrazionali quadratici e geodetiche chiuse sulla superficie modulare

Se si recupera dalla sezione §4.2 la mappa di Gauss $T_G : I \rightarrow I$, allora la mappa di Poincaré per il flusso geodetico su T^1M può essere scritta come segue:

$$P(x_1, x_2, \epsilon) = \left(T_G x_1, \frac{1}{x_2 + \left[\frac{1}{x_1} \right]}, -\epsilon \right). \quad (5.4)$$

Quindi si potrebbe considerare P come una naturale estensione della mappa di Gauss.

Ovviamente le orbite chiuse di $\phi_t : T^1M \rightarrow T^1M$ possono essere caratterizzate ora con i punti periodici della mappa di Poincaré (5.4). A causa dell'inversione nella variabile ϵ , possono esistere solo punti periodici di periodo pari $2k$, quindi la coordinata x_1 di un punto x della sezione di Poincaré X deve essere un punto periodico della mappa di Gauss di periodo $2k$. Ma

$T_G^{2k}x_1 = x_1$ se e solo se x_1 ha uno sviluppo in frazioni continue periodico di periodo $2k$, cioè se

$$x_1 = [\overline{a_1, \dots, a_{2k}}]$$

con $a_i \in \mathbb{N} \forall i \in [1, 2k]$.

Applicando $2k$ volte la mappa P , si deve dunque ottenere di nuovo (x_1, x_2, ϵ) . Per quanto riguarda x_1 ed ϵ non ci sono problemi. Sia, ora, $x_2 = [b_1, b_2, \dots]$,

$$(x_1, [b_1, b_2, \dots], \epsilon) = P(x_1, x_2, \epsilon) = (x_1, [a_{2k}, a_{2k-1}, \dots, a_1, b_1, b_2, \dots], \epsilon).$$

Quindi $[b_1, b_2, \dots] = [a_{2k}, a_{2k-1}, \dots, a_1, b_1, b_2, \dots]$. Quindi, $x = (x_1, x_2, \epsilon)$ è un punto periodico di P di periodo $2k$ se e solo se $x_2 = [\overline{a_{2k}, a_{2k-1}, \dots, a_1}]$.

Ma allora, i punti periodici della mappa di Poincaré P del flusso geodetico $\phi_t : T^1M \rightarrow T^1M$ sono in corrispondenza biunivoca con l'insieme

$$Y = \{([\overline{a_1, \dots, a_{2k}}], \epsilon), a_i \in \mathbb{N}, k \in \mathbb{N}, \epsilon = \pm 1\}.$$

5.4 Calcolo del periodo delle geodetiche chiuse

Per determinare il periodo della geodetica chiusa corrispondente ad un certo punto $([\overline{a_1, \dots, a_{2k}}], \epsilon)$, c'è da determinare il suo tempo di ricorrenza rispetto alla sezione di Poincaré X . Tale tempo di ricorrenza è dato dalla distanza tra due intersezioni successive dell'orbita con la sezione.

Se $\gamma = (\gamma_{-\infty}, \gamma_{+\infty})$ con $n \leq \gamma_{+\infty} < n+1$, si denoti con $z_1 = i\xi$ il punto di intersezione di γ con la semiretta $\Re z = 0$ e con $z_2 = n + i\eta$ l'intersezione di γ con la semiretta $\Re z = n$. Si dimostra che la distanza tra questi due punti è data da

$$r(x) = \log \frac{|\gamma_{-\infty} - i\eta||\gamma_{+\infty} - i\xi|}{|\gamma_{-\infty} - i\xi||\gamma_{+\infty} - i\eta|} \quad (5.5)$$

dove γ incontra X nel punto $x = (z_1, v_1)$ e v_1 è il vettore tangente a γ in z_1 (cfr. [MS93]).

Se i punti base della geodetica $\gamma = (\gamma_{-\infty}, \gamma_{+\infty})$ sono dati da

$$\gamma_{-\infty} = [a_0, a_{-1}, a_{-2}, \dots], \quad \gamma_{+\infty} = [a_1; a_2, a_3, \dots] \quad (5.6)$$

si ricava con una quantità incredibile di calcoli che

$$r(x) = -\frac{1}{2} \log([a_1, a_2, \dots] \cdot [a_0, a_{-1}, a_{-2}, \dots] \cdot [a_2, a_3, \dots] \cdot [a_1, a_0, a_{-1}, \dots]).$$

Per una geodetica periodica γ con $\gamma_{+\infty}^{-1} = [\overline{a_1, a_2, \dots, a_{2k}}]$ e dunque con $\gamma_{-\infty} = -[\overline{a_{2k}, a_{2k-1}, \dots, a_1}]$, si ricava che il periodo $l(\gamma)$, dato dalla somma dei tempi di ricorrenza, è

$$l(\gamma) = -\frac{1}{2} \sum_{r=1}^{2k} \log [\overline{a_r, a_{r+1}, \dots, a_{2k}, a_1, \dots, a_{r-1}}] \cdot [\overline{a_{r-1}, a_{r-2}, \dots, a_1, a_{2k}, a_{2k-1}, \dots, a_r}] \cdot [\overline{a_{r+1}, a_{r+2}, \dots, a_{2k}, a_1, a_2, \dots, a_r}] \cdot [\overline{a_r, a_{r-1}, \dots, a_1, a_{2k}, a_{2k-1}, \dots, a_{r+1}}].$$

Sfruttando un risultato di Legendre, per la precisione

$$\prod_{r=1}^{2k} [\overline{a_r, a_{r+1}, \dots, a_{2k}, a_1, \dots, a_{r-1}}] = \prod_{r=1}^{2k} [\overline{a_r, a_{r-1}, \dots, a_1, a_{2k}, \dots, a_r, a_{r+1}}]$$

si ottiene che il periodo $l(\gamma)$ di una geodetica periodica γ è:

$$l(\gamma) = \log \prod_{r=1}^{2k} [\overline{a_r, a_{r+1}, \dots, a_{2k}, a_1, \dots, a_{r-1}}].$$

Capitolo 6

Alcuni appunti sulla mappa di Gauss

La *trasformazione di Gauss*, già definita nella sezione §4.2, ha inaspettatamente messo in luce una enorme quantità di proprietà interessanti riguardo agli ambiti più disparati. In particolare ce ne sono alcune che riguardano le funzioni continue e le geodetiche chiuse che vale la pena cercare.

6.1 Prime proprietà strutturali

Si noti che T_G mappa monotonamente e con continuità gli intervalli $(\frac{1}{n+1}, \frac{1}{n}]$ in $[0, 1)$ ed è discontinua in $\frac{1}{n} \forall n \in \mathbb{N}$ (Figura 6.1).

Un'altra particolarità è che i punti fissi di T_G sono 0 e tutti i numeri che hanno lo sviluppo in frazioni continue di periodo 1.

Il punto chiave sta nella misura μ trovata da Gauss che è invariante per T_G . Si consideri un intervallo $A = (a, b) \subset (0, 1)$, la misura di Gauss μ per

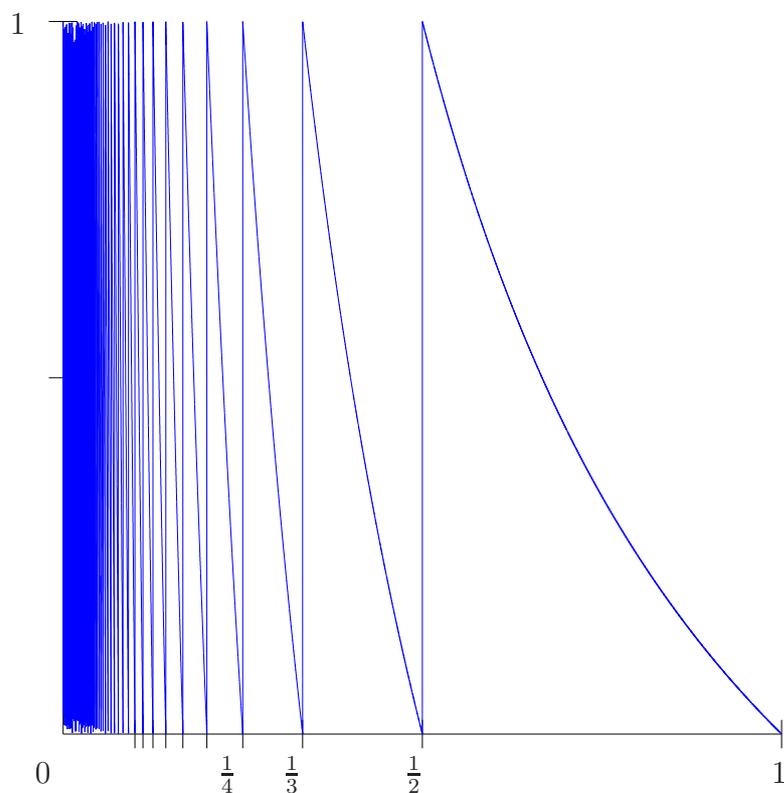


Figura 6.1: La mappa di Gauss

tale intervallo è

$$\mu(A) = \frac{1}{\log(2)} \int_a^b \frac{dx}{1+x} = \frac{1}{\log(2)} \log \left(\frac{1+b}{1+a} \right). \quad (6.1)$$

È chiaro che si tratta di una misura e che gli insiemi di misura 0 della misura di Gauss sono insiemi di misura 0 anche per la misura di Lebesgue λ , quindi tutto ciò che è definito μ -q.o. è definito λ -q.o..

TEOREMA 6.1 *La misura μ definita in (6.1) è T_G -invariante, cioè $\mu(T_G^{-1}(A)) = \mu(A)$ per ogni intervallo aperto $A \subset (0, 1)$.*

DIM. Per provare l'invarianza si noti che la preimmagine di (a, b) è data dall'unione di un numero infinito di intervalli disgiunti.

Nell'intervallo $(\frac{1}{n+1}, \frac{1}{n})$, la preimmagine è $(\frac{1}{n+b}, \frac{1}{n+a})$. Quindi si ha

$$\begin{aligned}\mu(T_G^{-1}((a, b))) &= \mu\left(\bigcup_{n=1}^{+\infty}\left(\frac{1}{n+b}, \frac{1}{n+a}\right)\right) = \\ &= \frac{1}{\log(2)} \sum_{n=1}^{+\infty} \log\left(\frac{n+a+1}{n+a}, \frac{n+b}{n+b+1}\right) = \mu((a, b)).\end{aligned}$$

□

Si noti che in generale $\mu(T_G(A)) \neq \mu(A)$.

6.2 Breve revisione del rapporto tra frazioni continue e mappa di Gauss

È evidente, riprendendo quanto visto nel Capitolo 4 e nel precedente, lo stretto rapporto che lega gli sviluppi in frazioni continue e la mappa di Gauss.

In effetti, per ogni $x \in (0, 1]$, si ha $x = \frac{1}{[\frac{1}{x}] + T_G x}$.

LEMMA 6.2 $[a_1, a_2, \dots, a_n + t] = \frac{p_n + t p_{n-1}}{q_n + t q_{n-1}}$

DIM. Sia $n = 1$, poiché $p_1 = 1, p_0 = 0, q_1 = a_1, q_0 = 1$, si ha:

$$\frac{p_1 + t p_0}{q_1 + t q_0} = \frac{1 + t \cdot 0}{a_1 + t \cdot 1}.$$

Si consideri ora il caso $n + 1$ supponendo vera l'ipotesi induttiva. Se si definisce $s = \frac{1}{a_{n+1} + t}$, è ovvio che

$$\frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots + a_n + \frac{1}{a_{n+1} + t}}}} = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots + a_n + s}}},$$

cioè $[a_1, a_2, \dots, a_n, a_{n+1} + t] = [a_1, a_2, \dots, a_n + s] = \frac{p_n + sp_{n-1}}{q_n + sq_{n-1}}$ per l'ipotesi induttiva. Sostituendo si ottiene

$$\frac{p_n + sp_{n-1}}{q_n + sq_{n-1}} = \frac{p_n + \left(\frac{1}{a_{n+1}+t}\right)p_{n-1}}{q_n + \left(\frac{1}{a_{n+1}+t}\right)q_{n-1}} = \frac{a_{n+1}p_n + p_{n-1} + tp_n}{a_{n+1} + q_{n-1} + tq_n} = \frac{p_{n+1} + tp_n}{q_{n+1} + tq_n}.$$

□

Più in generale, si è già visto che se $T_G^{n-1}x \neq 0$, definendo $a_i = \left\lceil \frac{1}{T_G^{i-1}x} \right\rceil \geq 1$ per $1 \leq i \leq n$, si ha $x = [a_1, \dots, a_{n-1}, a_n + T_G^n x]$. Dunque,

$$x = \frac{p_n + T_G^n(x)p_{n-1}}{q_n + T_G^n(x)q_{n-1}}.$$

Per dimostrare il prossimo importante teorema, è necessario introdurre alcune nuove definizioni.

Data una successione $\{b_k\}_{k=1, \dots, n} \subset \mathbb{N}$, sia

$$\Delta_{b_1, \dots, b_n} = \{x \in (0, 1] : a_k(x) = b_k, k = 1, \dots, n\}.$$

L'intervallo Δ_{b_1, \dots, b_n} è l'immagine dell'intervallo $[0, 1)$ rispetto alla mappa ψ_{b_1, \dots, b_n} definita

$$\psi_{b_1, \dots, b_n}(t) = [b_1, \dots, b_{n-1}, b_n + t].$$

Ovviamente, se n è dispari, ψ_{b_1, \dots, b_n} è decrescente; se n è pari è crescente.

Fissato $x \in \Delta_{b_1, \dots, b_n}$, $\exists t \in [0, 1)$ tale che

$$x = \psi_{b_1, \dots, b_n}(t) = \frac{p_n + tp_{n-1}}{q_n + tq_{n-1}}.$$

Quindi, se n è pari,

$$\Delta_{b_1, \dots, b_n} = \left[\frac{p_n}{q_n}, \frac{p_n + p_{n-1}}{q_n + q_{n-1}} \right),$$

mentre se n è dispari

$$\Delta_{b_1, \dots, b_n} = \left(\frac{p_n + p_{n-1}}{q_n + q_{n-1}}, \frac{p_n}{q_n} \right].$$

Se λ è la misura di Lebesgue, per il Teorema 4.13,

$$\lambda(\Delta_{b_1, \dots, b_n}) = (q_n(q_n + q_{n-1}))^{-1} \leq (2n^2 - n)^{-1} \Rightarrow 0.$$

TEOREMA 6.3 $p_n(x) = q_{n-1}(T_G x)$ per ogni $n \geq 1$ e per $x \in [0, 1]$.

DIM. Se $n = 1$, si ha $1 = p_1(x) = q_0(T_G x) = 1$, vero.

Se $n = 2$, si ha $a_2(x) = p_2(x) = q_1(T_G x) = a_1(T_G x) = a_2(x)$, vero.

Si supponga la tesi vera per $n - 2$ ed $n - 1$. Sfruttando l'ipotesi induttiva, si ha che

$$\begin{aligned} p_n(x) &= a_n(x)p_{n-1}(x) + p_{n-2} = \\ &= a_n(x)q_{n-2}(T_G x) + q_{n-3}(T_G x) = \\ &= a_{n-1}(T_G x)q_{n-2}(T_G x) + q_{n-3}(T_G x) = q_{n-1}(T_G x). \end{aligned}$$

□

6.3 Ergodicità della trasformazione di Gauss

TEOREMA 6.4 *La trasformazione di Gauss è ergodica rispetto alla misura di Gauss μ .*

DIM. Per una misura ν e due insiemi misurabili A e B tali che $\nu(B) \neq 0$, si definisce $\nu(A|B) = \frac{\nu(A \cap B)}{\nu(B)}$ la misura condizionata.

Sia (b_1, \dots, b_n) una n -upla di interi positivi fissati. Siano, poi, $\Delta_n = \Delta_{b_1, \dots, b_n}$ e $\psi_n = \psi_{b_1, \dots, b_n}$.

La lunghezza di Δ_n è $\pm(\psi_n(1), \psi_n(0))$, e per $0 \leq x < y \leq 1$,

$$\lambda(\{z : x \leq T_G^n(z) < y\} \cap \Delta_n) = \pm(\psi_n(y) - \psi_n(x)),$$

dove il segno dipende dalla parità o dalla disparità di n . A questo punto, si ottiene

$$\begin{aligned} \lambda(T_G^{-n}([x, y])|\Delta_n) &= \frac{\psi_n(y) - \psi_n(x)}{\psi_n(1) - \psi_n(0)} = \\ &= (y - x) \cdot \frac{q_n(q_n + q_{n-1})}{(q_n + xq_{n-1})(q_n + yq_{n-1})}. \end{aligned}$$

Indubbiamente $(y - x) = \lambda([x, y])$.

Inoltre, per il Teorema 4.6,

$$\frac{q_n(q_n + q_{n-1})}{(q_n + xq_{n-1})(q_n + yq_{n-1})} \leq \frac{q_n(q_n + q_{n-1})}{q_n q_n} = \frac{q_n + q_{n-1}}{q_n} = 1 + \frac{q_{n-1}}{q_n} \leq 2$$

e

$$\frac{q_n(q_n + q_{n-1})}{(q_n + xq_{n-1})(q_n + yq_{n-1})} \geq \frac{q_n(q_n + q_{n-1})}{(q_n + q_{n-1})(q_n + q_{n-1})} = \frac{q_n}{q_n + q_{n-1}} = \frac{1}{1 + \frac{q_{n-1}}{q_n}} \geq \frac{1}{2}.$$

Quindi il secondo fattore a destra è compreso tra $\frac{1}{2}$ e 2. Che implica

$$\frac{1}{2}\lambda([x, y]) \leq \lambda(T_G^{-n}([x, y])|\Delta_n) \leq 2\lambda([x, y]).$$

Dunque gli intervalli $[x, y)$ generano la σ -algebra

$$\frac{1}{2}\lambda(A) \leq \lambda(T_G^{-n}(A)|\Delta_n) \leq 2\lambda(A) \tag{6.2}$$

per ogni insieme misurabile $A \subset [0, 1]$.

Poiché la densità della misura di Gauss μ è compresa tra $\frac{1}{2\log(2)}$ ed $\frac{1}{\log(2)}$,

$$\frac{1}{2\log(2)}\lambda(A) \leq \mu(A) \leq \frac{1}{\log(2)}\lambda(A),$$

quindi, per la (6.2), si ha

$$\frac{1}{4}\mu(A) \leq \mu(T_G^{-n}(A)|\Delta_n) \leq 4\mu(A)$$

per ogni $A \subset [0, 1]$ misurabile.

Sia A un insieme T_G -invariante con $\mu(A) > 0$. Allora $\frac{1}{4}\mu(A) \leq \mu(A|\Delta_n)$, o, equivalentemente, $\frac{1}{4}\mu(\Delta_n) \leq \mu(\Delta_n|A)$. Poiché gli intervalli Δ_n generano la σ -algebra, $\frac{1}{4}\mu(B) \leq \mu(B|A)$ per ogni insieme misurabile B . Scegliendo $B = [0, 1] \setminus A$ si ottiene che $\mu(A) = 1$.

□

6.4 Il teorema ergodico di Birkhoff

Prima di arrivare a capire l'importanza di questo teorema nella teoria dei numeri, è necessaria un'ultima premessa.

Un importante risultato della Teoria Ergodica dei Sistemi Dinamici (cfr. [Isoa, SKF82]) afferma che :

TEOREMA 6.5 (Teorema Ergodico di Birkhoff) *Sia (X, \mathcal{F}, μ) uno spazio di probabilità, sia $T : X \rightarrow X$ una trasformazione che conserva la misura e sia $f \in L_1(X, \mathcal{F}, \mu)$.*

Allora,

1. *il limite $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$ esiste μ -q.o.*
2. *se $f \in L_p(X, \mathcal{F}, \mu)$, $1 \leq p < \infty$, il suddetto limite definisce una funzione $\bar{f} \in L_p(X, \mathcal{F}, \mu)$ che soddisfa*

$$\begin{aligned} \bar{f}(Tx) &= \bar{f}(x), \quad \mu\text{-q.o.} \\ \lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k - \bar{f} \right\|_p &= 0. \end{aligned}$$

3. Se $A \in \mathcal{F}$ soddisfa $T^{-1}A = A$ allora $\int_A f d\mu = \int_A \bar{f} d\mu$ (e pertanto si ha anche $\int_X f d\mu = \int_X \bar{f} d\mu$).

6.5 Teoria dei numeri e mappa di Gauss

L'ergodicità della mappa di Gauss comporta alcuni importanti risultati di teoria dei numeri.

TEOREMA 6.6 Per quasi ogni $x \in [0, 1]$ (rispetto alla misura di Gauss μ o alla misura di Lebesgue), si ha che:

- ogni intero $k \in \mathbb{N}$ compare nella sequenza $a_1(x), a_2(x), \dots$ con frequenza asintotica

$$\frac{1}{\log(2)} \log \left(\frac{k+1}{k} \right);$$

- $\lim_{n \rightarrow \infty} \frac{1}{n} (a_1(x) + \dots + a_n(x)) = \infty$;
- $\lim_{n \rightarrow \infty} \sqrt[n]{a_1(x) a_2(x) \dots a_n(x)} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2 + 2k} \right)^{\frac{\log(k)}{\log(2)}}$;
- $\lim_{n \rightarrow \infty} \frac{\log(q_n(x))}{n} = \frac{\pi^2}{12 \log(2)}$, in altre parole $q_n \sim e^{\frac{n\pi^2}{12 \log(2)}}$.

DIM.

- Sia f la funzione caratteristica dell'intervallo semiaperto $\left[\frac{1}{k}, \frac{1}{k+1} \right)$. Allora $a_n(x) = k$ se e solo se $f(T_G^n(x)) = 1$. Per il Teorema Ergodico di Birkhoff, per quasi ogni x ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T_G^i(x)) = \int_0^1 f d\mu = \mu \left(\left[\frac{1}{k}, \frac{1}{k+1} \right) \right) = \frac{1}{\log(2)} \log \left(\frac{k+1}{k} \right).$$

- Sia $f(x) = \lfloor \frac{1}{x} \rfloor$, cioè $f(x) = a_1(x)$. Si noti che $\int_0^1 \frac{f(x)}{1+x} dx = \infty$, infatti $f(x) = \frac{1}{x} - \{ \frac{1}{x} \} > \frac{1-x}{x}$ e $\int_0^1 \frac{1-x}{x(1+x)} dx = \int_0^1 \frac{2x+1-3x}{x(1+x)} dx = \infty$. Per $N > 0$, si definisca

$$f_N(x) = \begin{cases} f(x) & \text{se } f(x) \leq N, \\ 0 & \text{altrimenti.} \end{cases}$$

Allora, per ogni $N > 0$ e per quasi ogni x ,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T_G^k(x)) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f_N(T_G^k(x)) = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f_N(T_G^k(x)) \\ &= \frac{1}{\log(2)} \int_0^1 \frac{f_N(x)}{1+x} dx. \end{aligned}$$

La conclusione segue dal fatto che $\lim_{N \rightarrow \infty} \int_0^1 \frac{f_N(x)}{1+x} dx \rightarrow \infty$.

- Sia $f(x) = \log(a_1(x)) = \log(\lfloor \frac{1}{x} \rfloor)$.

Innanzitutto $f \in L_1([0, 1], \mu)$, infatti

$$\int_0^1 \frac{\log(\lfloor \frac{1}{x} \rfloor)}{1+x} dx \leq \int_0^1 \frac{\log(\frac{1}{x})}{1+x} dx = - \int_0^1 \frac{\log(x)}{1+x} dx$$

che, come si vedrà nel prossimo punto, è finito.

Per il Teorema Ergodico di Birkhoff,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log(a_k(x)) &= \frac{1}{\log(2)} \int_0^1 \frac{f(x)}{1+x} dx = \\ &= \frac{1}{\log(2)} \sum_{k=1}^{\infty} \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{\log(k)}{1+x} dx = \\ &= \sum_{k=1}^{\infty} \frac{\log(k)}{\log(2)} \cdot \log \left(1 + \frac{1}{k^2 + 2k} \right). \end{aligned}$$

Elevando alla “e” a destra e a sinistra si ottiene la tesi.

- Grazie al Teorema 6.3 si può affermare che

$$\frac{1}{q_n(x)} = \frac{p_n(x)}{q_n(x)} \frac{p_{n-1}(T_G x)}{q_{n-1}(T_G x)} \cdots \frac{p_1(T_G^{n-1} x)}{q_1(T_G^{n-1} x)}.$$

Quindi

$$\begin{aligned} -\frac{1}{n} \log(q_n(x)) &= \frac{1}{n} \sum_{k=0}^{n-1} \log \left(\frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} \right) = \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \log(T_G^k(x)) + \frac{1}{n} \sum_{k=0}^{n-1} \left(\log \frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} - \log(T_G^k(x)) \right). \end{aligned}$$

Segue dal Teorema Ergodico di Birkhoff che il primo termine converge quasi ovunque a $\frac{1}{\log(2)} \int_0^1 \frac{\log(x)}{1+x} dx$.

$$\begin{aligned} \frac{1}{\log(2)} \int_0^1 \frac{\log(x)}{1+x} dx &= \frac{1}{\log(2)} \left[\log(x) \log(1-x) - \int \frac{\log(1+x)}{x} dx \right]_0^1 = \\ &= -\frac{1}{\log(2)} \int_0^1 \frac{\log(1+x)}{x} dx = \\ &= -\frac{1}{\log(2)} \int_0^1 \frac{1}{x} \left(x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \right) dx = \\ &= -\frac{1}{\log(2)} \int_0^1 \left(1 - \frac{x}{2} + \frac{x^2}{3} - \frac{x^3}{4} + \cdots \right) dx = \\ &= -\frac{1}{\log(2)} \left[x - \frac{x^2}{4} + \frac{x^3}{9} - \frac{x^4}{16} + \cdots \right]_0^1 = \\ &= -\frac{1}{\log(2)} \left(1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \cdots \right) = \\ &= -\frac{1}{\log(2)} \left(\sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{n=1}^{\infty} \frac{1}{2n^2} \right) = \\ &= -\frac{1}{\log(2)} \left(\sum_{n=1}^{\infty} \frac{1}{n^2} + \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n^2} \right) = \\ &= -\frac{1}{\log(2)} \left(\frac{\pi^2}{6} - \frac{\pi^2}{12} \right) = -\frac{\pi^2}{12 \log(2)}. \end{aligned}$$

Resta da verificare che il secondo termine tende a 0. Innanzitutto si noti che, applicando il teorema del valor medio, se $x \leq y \in (0, 1)$, si ha

$$|\log(x) - \log(y)| \leq \frac{1}{x} |y - x| = \left| \frac{y}{x} - 1 \right|.$$

Se si applica tale ragionamento si ottiene

$$\begin{aligned} \left| \log \frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} - \log(T_G^k(x)) \right| &= \left| \log(T_G^k(x)) - \log \frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} \right| \leq \\ &\leq \left| \frac{T_G^k(x)}{\frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))}} - 1 \right|. \end{aligned}$$

Ora, $\frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} \cdot \left| \frac{T_G^k(x)}{\frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))}} - 1 \right| = \left| T_G^k(x) - \frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} \right|$ ma dal Teorema 4.20, si ricava che $\left| x - \frac{p_n(x)}{q_n(x)} \right| \leq \frac{1}{q_n^2} \leq \frac{1}{n^2} \rightarrow 0$.

Poiché $\frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} \rightarrow T_G^k(x) \in (0, 1)$, necessariamente

$$\left| \frac{T_G^k(x)}{\frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))}} - 1 \right| \rightarrow 0.$$

In conclusione $\left(\log \frac{p_{n-k}(T_G^k(x))}{q_{n-k}(T_G^k(x))} - \log(T_G^k(x)) \right)$ è una quantità limitata e le somme sono assolutamente convergenti ad un numero H finito; moltiplicando H per $\frac{1}{n}$ e mandando $n \rightarrow \infty$ si ottiene la tesi.

- Per quest'ultimo punto del teorema c'è un'altra dimostrazione, proposta dal professor Isola, decisamente più bella.

Si ha

$$x = \frac{p_n + T_G^n(x)p_{n-1}}{q_n + T_G^n(x)q_{n-1}}, \quad (6.3)$$

pertanto

$$T_G^n(x) = -\frac{q_n x - p_n}{q_{n-1} x - p_{n-1}} = -\frac{f_n}{f_{n-1}} \quad (6.4)$$

dove si sono introdotti i numeri $f_n = q_n x - p_n$. Allora

$$f_n = \prod_{k=0}^n T_G^k(x),$$

per il Teorema Ergodico di Birkhoff, si ha che

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log f_n = \int_0^1 \log(x) \mu(dx) = -\frac{\pi^2}{12 \log(2)}$$

μ -quasi ogni $x \in [0, 1]$ (che significa quasi ovunque in quanto μ è assolutamente continua rispetto alla misura di Lebesgue)¹.

Essendo $[(T_G^n(x))^{-1}] = a_{n+1}$ e quindi $a_{n+1} \leq (T_G^n(x))^{-1} \leq a_{n+1} + 1$, un'altra conseguenza² della (6.3), è che

$$\frac{1}{a_{n+1} + 1} < \frac{q_n}{q_n + q_{n+1}} < q_n f_n < \frac{q_n}{q_{n+1}} < \frac{1}{a_{n+1}}$$

e quindi, usando la (6.4),

$$\frac{1}{2} < q_n f_n < 1$$

da cui segue la tesi. □

Un'interessante conseguenza di questo teorema è la determinazione della velocità con cui i convergenti di una frazione continua approssimano il loro limite x .

Sia $x = [a_1, a_2, \dots, a_n + t]$ con $t \in (0, 1)$. Ovviamente

$$x - \frac{p_n}{q_n} = \frac{p_n + t p_{n-1}}{q_n + t q_{n-1}} - \frac{p_n}{q_n} = t \frac{(-1)^n}{q_n(q_n + t q_{n-1})} = \frac{(-1)^n}{q_n(\frac{1}{t} q_n + q_{n-1})},$$

si noti che $a_{n+1} \leq \frac{1}{t} \leq a_{n+1} + 1$, dunque

$$\frac{1}{q_n((a_{n+1} + 1)q_n + q_{n-1})} \leq \left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n(a_{n+1}q_n + q_{n-1})},$$

¹La convergenza dell'integrale è già stata dimostrata in precedenza.

²Si verifica come la (6.5)

da cui, sfruttando la ormai nota relazione di ricorrenza, si ottiene

$$\frac{1}{q_n(q_{n+1} + q_n)} \leq \left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n q_{n+1}}. \quad (6.5)$$

COROLLARIO 6.7 $\frac{1}{n} \log \left| x - \frac{p_n(x)}{q_n(x)} \right| \rightarrow -\frac{\pi^2}{6 \log(2)} \quad \mu\text{-q.o.}$

DIM. Per quanto osservato fin qui, si ha

$$-\log \left| x - \frac{p_n(x)}{q_n(x)} \right| \geq \log(q_n) + \log(q_{n+1} + q_n)$$

e

$$-\log \left| x - \frac{p_n(x)}{q_n(x)} \right| \geq \log(q_n) + \log(q_{n+1}).$$

Ma $\log(q_n + q_{n+1}) \in (\log(2q_n), \log(2q_{n+1}))$, da cui segue direttamente il risultato.

□

Un'altra interessante applicazione di quanto osservato è quella che fornisce la probabilità che un numero scelto a caso abbia uno sviluppo in frazione continua che comincia in un certo modo.

La misura di Gauss μ è una misura di probabilità, integrando nell'intervallo Δ_{a_1, \dots, a_n} si conosce la misura (di probabilità) dell'intervallo che contiene tutti i numeri che hanno lo sviluppo in frazione continua che comincia per $[a_1, \dots, a_n]$. Che è esattamente quello che si sta cercando.

Banalmente, se si vuole la probabilità che $a_1(x) = m \in \mathbb{N}$, basta calcolare

$$\begin{aligned} \frac{1}{\log(2)} \int_{\frac{1}{m+1}}^{\frac{1}{m}} \frac{dx}{1+x} &= \frac{1}{\log(2)} \left(\log \left(1 + \frac{1}{n} \right) - \log \left(1 + \frac{1}{n+1} \right) \right) = \\ &= \frac{1}{\log(2)} \log \left(\frac{(n+1)^2}{n(n+2)} \right) = \\ &= \frac{1}{\log(2)} \log \left(1 + \frac{1}{n(n+2)} \right). \end{aligned}$$

Si supponga ora che $x \in \Delta_{a_1, \dots, a_n}$. Si supponga n dispari, l'integrale da calcolare, in base alle osservazioni fatte in questo capitolo, è

$$\begin{aligned} \frac{1}{\log(2)} \int_{\psi_{a_1, \dots, a_n}(1)}^{\psi_{a_1, \dots, a_n}(0)} \frac{dx}{1+x} &= \frac{1}{\log(2)} \left(\log \left(1 + \frac{p_n}{q_n} \right) - \log \left(1 + \frac{p_n + p_{n-1}}{q_n + q_{n-1}} \right) \right) = \\ &= \frac{1}{\log(2)} \log \left(\frac{q_n(q_n + p_n + q_{n-1}) + p_n q_{n-1}}{q_n(q_n + p_n + q_{n-1}) + q_n p_{n-1}} \right). \end{aligned}$$

In caso di n pari basta un meno davanti alla formula precedente.

Le proprietà della trasformazione di Gauss non si esauriscono in quelle mostrate in questa breve panoramica. Ce ne sono molte altre che entrano fin nel profondo della distribuzione dei numeri primi, andando ad intrecciarsi con la ζ di Riemann, e degli irrazionali quadratici, in cui entrano anche nel merito del periodo $r(x)$ delle geodetiche chiuse.

Il limite $l(x) = \lim_{n \rightarrow \infty} \frac{\log(q_n(x))}{n}$, che si è visto convergere μ -q.o. a $\frac{\pi^2}{12 \log(2)}$, è chiamato *costante di Lévy*. Per quanto riguarda gli irrazionali quadratici, si può dimostrare che, se $r(x)$ è il periodo della geodetica chiusa corrispondente ad x ,

$$\sum_{r(x) < X} l(x) \text{ è asintoticamente equivalente a } \frac{1}{4} e^X$$

quando $X \rightarrow \infty$.

Si vedrà nel prossimo capitolo come la mappa di Gauss permetta in qualche modo di instaurare un legame tra le geodetiche chiuse e quindi gli irrazionali quadratici e le geodetiche aperte³ ed i numeri razionali.

6.6 Un'ultima curiosità

Fin qui si è messo in luce come ci sia un profondo legame tra la mappa di Gauss e la teoria dei numeri. In effetti la sua centralità nello studio e

³Dette *di scattering*.

nell'analisi delle frazioni continue la rende uno strumento indispensabile per indagare le leggi che si nascondono dietro la struttura stessa dei numeri.

C'è però un risultato, molto semplice da ricavare, ma strabiliante, forse inaspettato, che rende definitivamente palese che questa trasformazione, così apparentemente innocua può essere un punto di vista privilegiato per l'analisi delle proprietà dei numeri, arrivando addirittura ad entrare in merito ad uno dei problemi che stanno mettendo in difficoltà ormai da moltissimo tempo le menti dei più brillanti matematici.

Non si può quindi evitare di mostrare il seguente risultato.

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s} = \sum_{n=1}^{+\infty} s \int_n^{+\infty} \frac{dy}{y^{s+1}} = \textcircled{*}$$

Si definisca

$$\varphi_n(y) = \begin{cases} 0 & \text{se } y \leq n \\ 1 & \text{altrimenti} \end{cases}$$

dunque si può scrivere

$$\textcircled{*} = s \sum_{n=1}^{+\infty} \int_1^{+\infty} \frac{\varphi_n(y)}{y^{s+1}} dy = \textcircled{*}\textcircled{*},$$

poiché gli integrali sono limitati, ed il limite per $n \rightarrow \infty$ è 0, è vero che

$$\begin{aligned} \textcircled{*}\textcircled{*} &= s \int_1^{+\infty} \left(\sum_{n=1}^{+\infty} \frac{\varphi_n(y)}{y^{s+1}} \right) dy = \\ &= s \int_1^{+\infty} \left(\sum_{n=1}^{+\infty} \varphi_n(y) \right) \frac{1}{y^{s+1}} dy = \\ &= s \int_1^{+\infty} \left(\sum_{n \leq y} 1 \right) \frac{1}{y^{s+1}} dy = \textcircled{*}\textcircled{*}\textcircled{*} \end{aligned}$$

A questo punto

$$\begin{aligned} \textcircled{*}\textcircled{*}\textcircled{*} &= s \int_1^{+\infty} \frac{[y]}{y^{s+1}} dy = s \int_1^{+\infty} \frac{y - \{y\}}{y^{s+1}} dy = \\ &= \frac{s}{s-1} - s \int_1^{+\infty} \frac{\{y\}}{y^{s+1}} dy = \textcircled{*}\textcircled{*}\textcircled{*}\textcircled{*} \end{aligned}$$

Sia $y = \frac{1}{x}$, l'integrale diventa

$$\begin{aligned} \circledast \circledast \circledast \circledast &= \frac{s}{s-1} - s \int_1^0 \frac{\left\{\frac{1}{x}\right\}}{\left(\frac{1}{x}\right)^{s+1}} \cdot \frac{-dx}{x^2} = \\ &= \frac{s}{s-1} - s \int_0^1 T_G(x) x^{s-1} dx. \end{aligned}$$

Dunque

$$\zeta(s) = 1 + \frac{1}{s-1} - s \int_0^1 T_G(x) x^{s-1} dx$$

cioè la mappa di Gauss interviene in uno degli sviluppi analitici della funzione *Zeta di Riemann* che è legata in maniera indissolubile alla distribuzione dei numeri primi.

Capitolo 7

La tassellazione di Farey e le geodetiche di scattering

Prima di poter arrivare al dunque è necessario introdurre altri costrutti che faranno da tramite tra mappa di Gauss e geodetiche chiuse e mappa di Farey e geodetiche di scattering.

7.1 La mappa di Farey e i numeri razionali

7.1.1 I convergenti di Farey

Ormai è ben chiaro che per ogni $n \geq 1$ e per ogni $r \geq 1$ si ha

$$\left[a_1, a_2, \dots, a_n + \frac{1}{r} \right] = \frac{rp_n + p_{n-1}}{rq_n + q_{n-1}}.$$

È evidente che se $r = a_{n+1}$ dalla formula precedente si ottiene $\frac{p_{n+1}}{q_{n+1}}$.

Si supponga, allora, che $1 \leq r \leq a_{n+1}$. Per $n \geq 1$, gli insiemi

$$\left\{ \frac{rp_n + p_{n-1}}{rq_n + q_{n-1}} \right\}_{1 \leq r \leq a_{n+1}}$$

si chiamano *n-esimi convergenti di Farey* per il numero reale $x \in [0, 1)$.

Per poter approfondire il discorso, è utile definire una nuova operazione. Si considerino due razionali $\frac{a}{b} < \frac{a'}{b'}$, si chiama *somma di Farey* il mediente dato da

$$\frac{a}{b} \oplus \frac{a'}{b'} = \frac{a + a'}{b + b'} = \frac{a''}{b''}.$$

È questione di un semplice calcolo osservare che $\frac{a''}{b''}$ appartiene all'intervallo aperto $(\frac{a}{b}, \frac{a'}{b'})$.

Se $a'b - ab' = \pm 1$, $\frac{a}{b}$ e $\frac{a'}{b'}$ si chiamano *intorni di Farey*. Due intorni di Farey definiscono un *intervallo di Farey*, ognuno di questi intervalli può essere etichettato in maniera univoca dal mediente degli intorni.

Riorganizzando quanto formulato, si può fare un parallelismo con i convergenti precedentemente definiti. Si definiscano

$$\frac{t_{n,r}}{s_{n,r}} = \frac{rp_n + p_{n-1}}{rq_n + q_{n-1}},$$

per $n \geq 1$ ed $1 \leq r \leq a_{n+1}$ si può osservare che

$$\begin{aligned} \frac{t_{n,r+1}}{s_{n,r+1}} &= \frac{rp_n + p_{n-1}}{rq_n + q_{n-1}} = \\ &= \frac{(r+1)p_n + p_{n-1}}{(r+1)q_n + q_{n-1}} = \\ &= \frac{rp_n + p_{n-1} + p_n}{rq_n + q_{n-1} + q_n} = \\ &= \frac{rp_n + p_{n-1}}{rq_n + q_{n-1}} \oplus \frac{p_n}{q_n} = \\ &= \frac{t_{n,r}}{s_{n,r}} \oplus \frac{p_n}{q_n}. \end{aligned}$$

Dunque non c'è da stupirsi se

$$q_n t_{n,r} - p_n s_{n,r} = q_n p_{n-1} - p_n q_{n-1} = (-1)^n. \quad (7.1)$$

Questo vuol dire, però, che per ogni $n \geq 1$, ogni convergente di Farey $\frac{t_{n,r}}{s_{n,r}}$ è un intorno di Farey di $\frac{p_n}{q_n}$ i cui intervalli di Farey corrispondenti decrescono al crescere di r .

Più precisamente, sfruttando di nuovo la formula di Lagrange, si ottiene facilmente¹ che

$$\left| \frac{p_n}{q_n} - \frac{rp_n + p_{n-1}}{rq_n + q_{n-1}} \right| = \frac{1}{q_n(rq_n + q_{n-1})} \quad (7.2)$$

Ma allora i convergenti di Farey $\frac{t_{n,r}}{s_{n,r}}$ sono la migliore approssimazione razionale “unilaterale”² di x il cui denominatore non superi $s_{n,r}$ ³.

Quando r , crescendo, raggiunge a_{n+1} , si ottiene un nuovo convergente $\frac{p_{n+1}}{q_{n+1}}$.

L’algoritmo che genera la sequenza di convergenti di Farey di un numero reale dato è detto *algoritmo lento per le frazioni continue*.

7.1.2 L’albero di Farey

Si supponga di voler costruire un “albero” \mathcal{F} procedendo in questo modo:

1. Ogni colonna contiene un solo elemento.
2. La riga “0” contiene $\frac{0}{1}$ e $\frac{1}{1}$.
3. La prima riga contiene solo $\frac{1}{2}$.
4. Ogni riga successiva definisce il doppio degli elementi in colonna della riga precedente.

¹Esattamente con lo stesso procedimento utilizzato per le frazioni continue.

²Nel senso che restano sempre minori o sempre maggiori del valore che approssimano.

³Naturalmente, essendo $r < a_{n+1}$, può esserci un convergente “standard” con denominatore minore di $s_{n,r}$ e più vicino ad x ma “dall’altro lato rispetto ad x ”

5. Nessuna riga è definita finché non lo è quella che la precede. Cioè l' n -esima riga dipende solamente dalle colonne definite dalle $n - 1$ righe precedenti.
6. Ogni elemento che definisce una colonna ha come numeratore (risp. denominatore) la somma dei numeratori (risp. denominatori) delle frazioni che definiscono le colonne immediatamente a sinistra e destra dell'elemento stesso.
7. Tutti gli elementi sono nei loro termini ridotti⁴.

La stessa macchinosa costruzione può essere formalizzata come segue.

Fissato $l \geq 1$, sia \mathcal{F}_l la sequenza crescente di frazioni irriducibili tra 0 ed 1 costruite qui sotto:

- si ponga $\mathcal{F}_1 = \left(\frac{0}{1}, \frac{1}{1}\right)$,
- allora \mathcal{F}_l si ottiene da \mathcal{F}_{l-1} inserendo tra ogni coppia di intorni di Farey $\frac{a}{b}$ ed $\frac{a'}{b'}$, la loro somma di Farey $\frac{a''}{b''}$.

Quindi

$$\mathcal{F}_2 = \left(\frac{0}{1}, \frac{1}{2}, \frac{1}{1}\right) \quad \mathcal{F}_3 = \left(\frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}\right) \quad \mathcal{F}_4 = \left(\frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}\right)$$

e così via. Gli elementi di \mathcal{F}_l sono chiamati *frazioni di Farey*. Fissato $l \geq 1$, si dice che la frazione di Farey x ha *rango* l se $x \in \mathcal{F}_{l+1} \setminus \mathcal{F}_l$. Si definisce $\text{rango}\left(\frac{0}{1}\right) = \text{rango}\left(\frac{1}{1}\right) = 0$.

Per ogni $l \geq 1$ ci sono esattamente 2^{l-1} frazioni di Farey di rango l e la loro somma è pari a 2^{l-2} (cfr. [Lag01]).

L'albero \mathcal{F} di cui si parlava sopra si chiama *Albero di Farey* e può essere costruito come segue:

⁴Il numeratore ed il denominatore sono primi tra loro.

- ogni colonna in \mathcal{F} contiene solamente un nodo⁵;
- per ogni $l \geq 1$, la riga l è $\mathcal{F}_l \setminus \mathcal{F}_{l-1}$;
- il nodo $\frac{a+a'}{b+b'}$, che rappresenta l'intervallo di Farey $[\frac{a}{b}, \frac{a'}{b'}]$, è collegato da un arco ai due figli che genera, rispettivamente $\frac{2a+a'}{2b+b'}$ a sinistra e $\frac{a+2a'}{b+2b'}$ a destra.

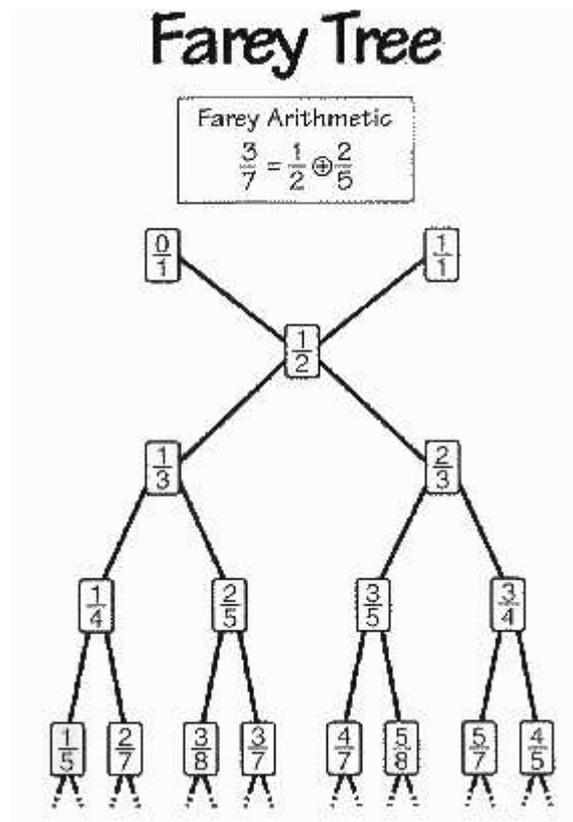


Figura 7.1: L'albero di Farey

Ogni numero razionale in $(0, 1)$ appare esattamente una volta in questa costruzione ed ogni numero reale in $(0, 1)$ corrisponde ad un percorso⁶ lungo nodi collegati.

⁵elemento

⁶finito o infinito

7.1.3 Codifiche

Quando nella sezione §4.3 si sono introdotti i convergenti delle frazioni continue, definendo

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{e} \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (7.3)$$

ed osservando che

$$BA^{k-1} = \begin{pmatrix} k & 1 \\ 1 & 0 \end{pmatrix},$$

si sarebbe potuto notare che

$$\left\{ \begin{array}{l} \begin{pmatrix} p_1 & p_0 \\ q_1 & q_0 \end{pmatrix} = A^{a_1} \\ \begin{pmatrix} p_{n+1} & p_n \\ q_{n+1} & q_n \end{pmatrix} = A^{a_1} B A^{a_2-1} \dots B A^{a_{n+1}-1}, \quad n \geq 1 \end{array} \right. \quad (7.4)$$

A questo punto si può notare anche che i convergenti di Farey non sono altro che i prodotti intermedi delle matrici A , per essere più chiari

$$\begin{pmatrix} t_{n,r} & p_n \\ s_{n,r} & q_n \end{pmatrix} = A^{a_1} B A^{a_2-1} \dots B A^{r-1}.$$

D'altra parte, se si definiscono

$$C = A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{e} \quad D = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad (7.5)$$

Etichettando gli archi con C e D , con D sull'arco che collega al figlio con denominatore maggiore, i prodotti tra tali matrici appena definite, con la prima D soppressa, determinano un cammino lungo l'albero di Farey.

Inoltre

$$\begin{pmatrix} p_{n+1} & p_n \\ q_{n+1} & q_n \end{pmatrix} = D C^{a_1-1} D C^{a_2-1} \dots D C^{a_n-1}.$$

La mappa di Gauss T_G è uno shift per lo sviluppo in frazioni continue. Quale è il significato dello shift lungo un intero percorso sull'albero di Farey?

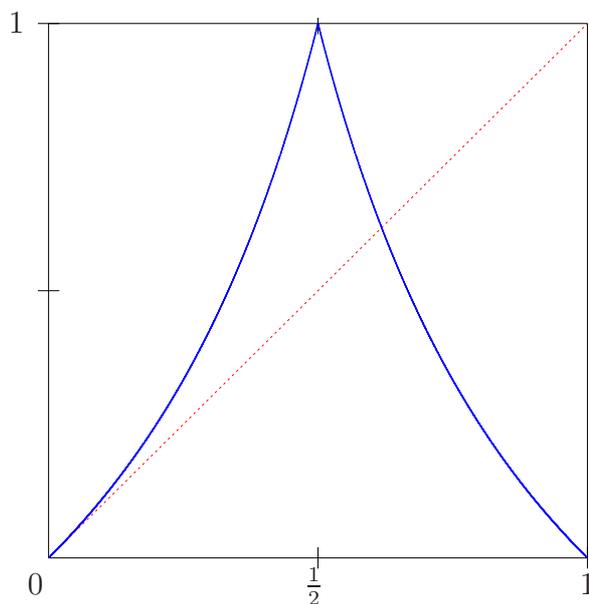


Figura 7.2: La mappa di Farey

Si può già osservare che T_G è lo “shift alla D successiva” nello sviluppo in forma matriciale. Lo shift su $\{C, D\}^{\mathbb{N}}$, chiamato *mappa di Farey*, corrisponde ad una funzione $F : [0, 1] \rightarrow [0, 1]$ definita

$$U(x) = \begin{cases} \frac{x}{1-x}, & \text{se } 0 \leq x \leq \frac{1}{2} \\ \frac{1-x}{x}, & \text{se } \frac{1}{2} \leq x \leq 1 \end{cases} \quad (7.6)$$

Con l'albero di Farey, si viene a creare la corrispondenza

$$\{C, D\}^{\mathbb{N}}, \sigma \leftrightarrow [0, 1], U$$

tramite la relazione

$$C^{a_1-1}DC^{a_2-1}D \dots \leftrightarrow x = [a_1, a_2, \dots].$$

Quindi, se $x = [a_1, a_2, a_3, \dots]$, $F(x) = [a_1 - 1, a_2, a_3, \dots]$. Così, se $x \in [\frac{1}{2}, 1]$, si ha $a_1 = 1$ e quindi $x \sim DC^{a_2-1}D \dots$, cioè se $x \in [\frac{1}{2}, 1]$ la prima etichetta di Farey per x è D .

In altre parole, la mappa di Gauss T_G è la mappa di primo ritorno $F^*(x)$ di $F(x)$ in $[\frac{1}{2}, 1]$!

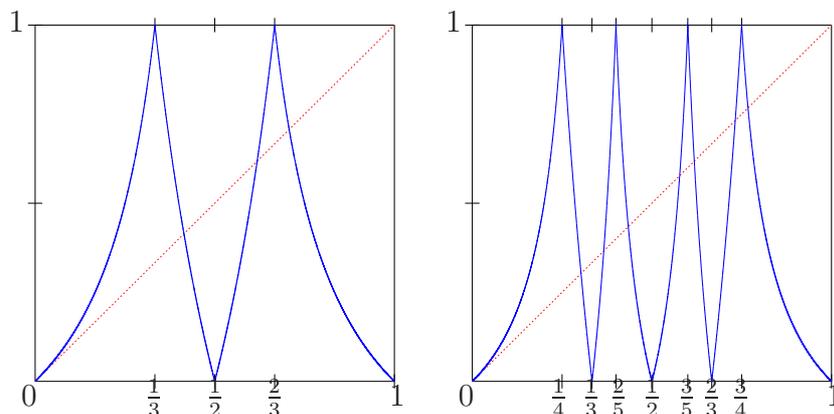


Figura 7.3: Due iterate della mappa di Farey

Il nome *mappa di Farey* può essere ricondotto anche all'osservazione (facilmente verificabile) per cui l'insieme delle preimmagini $\cup_{k=0}^l F^{-k}(0)$ coincide con \mathcal{F}_l per ogni $l \geq 1$. Si noti anche che l' l -esima riga dell'albero di Farey è precisamente $F^{-(l-1)}(\frac{1}{2})$, che implica, in particolare, che $\cup_{k=0}^{\infty} F^{-k}(0) = \mathbb{Q} \cap [0, 1]$.

7.2 La tassellazione di Farey e le geodetiche di scattering

È ormai chiaro che identificando i punti corrispondenti nella frontiera di \bar{F} , definito nel Teorema 3.13, si ottiene la superficie di Riemann $SL_2(\mathbb{Z}) \backslash \mathbb{H}$ che ha volume finito.

Ma l'azione del gruppo modulare su \mathbb{H} non è libera, ad esempio il punto $i \in \mathbb{H}$ è un punto fisso dell'ormai famosa inversione S tale che $z \mapsto -\frac{1}{z}$.

Si consideri ora la cosiddetta *Tassellazione di Farey*, il cui dominio fondamentale $G = \{z \in \mathbb{H} : 0 < \Re(z) < 1, |z - \frac{1}{2}| > \frac{1}{2}\}$ è l'interno del triangolo geodetico di vertici $0, 1$ ed ∞ già visto nella sezione §2.5. Ovviamente l'area di G è il triplo di quella di F .

Un fatto interessante è che la trasformazione associata a $Q = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ costituisce una permutazione ciclica delle tre cuspidi in $0, 1$ e ∞ .

Se si concentra l'attenzione sulle matrici dell'insieme

$$\mathcal{U} = \{I\} \cup \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : 0 \leq a \leq c, 0 \leq b \leq d \right\}$$

si può osservare che mappano i tre vertici $(0, 1, \infty)$ rispettivamente in se stessi e in

$$\left(\frac{b}{d}, \frac{a+b}{c+d}, \frac{a}{c} \right) \quad \text{dove} \quad 0 \leq \frac{b}{d} \leq \frac{a+b}{c+d} \leq \frac{a}{c} \leq 1.$$

È evidente che in questo modo le cuspidi possono essere trasformate in un qualunque numero razionale minore di 1 e che quindi l'immagine di G rispetto alle matrici in \mathcal{U} forma una tassellazione della striscia $\{z \in \mathbb{H} : 0 \leq \Re(z) \leq 1\}$ (Figura 7.4).

Per il Teorema 3.12, $(\mathbb{Q} \cap [0, 1]) \cup \{\infty\}$ è l'insieme delle cuspidi di $SL_2(\mathbb{Z})$ se consideriamo soltanto la striscia $\{z \in \mathbb{H} : 0 \leq \Re(z) \leq 1\}$.

Si prendano quindi in considerazione le geodetiche di scattering γ_q^p in \mathbb{H} associate alle matrici $M \in \mathcal{U}$ che da ∞ raggiungono la cuspidi centrale di $M(G)$. Tali geodetiche sono verticali ed hanno parte reale $\frac{p}{q}$, dove $p = a + b$ e $q = c + d$.

Si consideri γ_1 come la prima geodetica ottenuta in questo modo, tale curva ha parte reale 1. L'obiettivo è di confrontare la lunghezza delle geodetiche γ_q^p con quella di γ_1 .

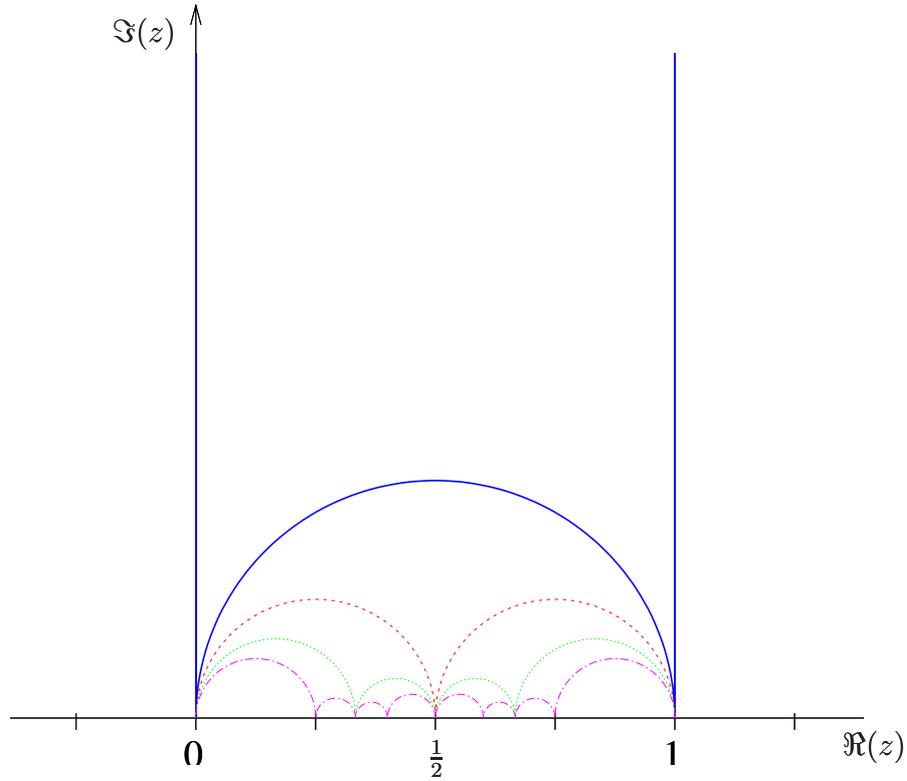


Figura 7.4: Tassellazione di Farey della striscia $\{z \in \mathbb{H} : 0 \leq \Re(z) \leq 1\}$

Ovviamente, la lunghezza di tali geodetiche è infinita; stranamente la loro differenza sarà una quantità finita ben definita.

Prima di entrare nel dettaglio, è necessario recuperare alcuni concetti (cfr. [Kna99, Ber02]). Sia, in generale, $\Phi^t : X \rightarrow X$ il flusso geodetico su un fibrato tangente unitario X di una superficie Riemanniana, con la distanza d . La varietà (in)stabile associata ad $x \in X$ è data da

$$W^\pm(x) = \left\{ y \in X : \lim_{t \rightarrow \pm\infty} d(\Phi^t(y), \Phi^t(x)) = 0 \right\}.$$

Nel caso del flusso geodetico per $SL_2(\mathbb{Z}) \backslash \mathbb{H}$, se x è un punto sulla linea del flusso geodetico corrispondente a γ_1 , ogni altra geodetica di scattering

$\gamma_{\frac{p}{q}}$ (che in \mathbb{H} è una semiretta con parte reale $\frac{p}{q}$ costante), ha due unici punti $y^\pm \in W^\pm(x)$.

Si definisce *tempo di ritardo* $T(\frac{p}{q})$ della geodetica di scattering $\gamma_{\frac{p}{q}}$, il valore definito da $\Phi^{T(\frac{p}{q})}(y^-) = y^+$.

Si dimostra che tale definizione non dipende dalla scelta del punto x sulla geodetica di riferimento γ_1 .

Il risultato più interessante, che verrà solo citato (cfr. [Kna99]), è che

TEOREMA 7.1 *Se $\text{mcd}(p, q) = 1$, allora il tempo di ritardo è $T(\frac{p}{q}) = 2\ln(q)$.*

In altre parole, il tempo di ritardo è pensabile come una sorta di equivalente del tempo di primo ritorno per le geodetiche chiuse. Il che permette poi di poter studiare relazioni simili a quelle osservate nei capitoli precedenti.

Capitolo 8

Brevi note conclusive

Lo studio delle geodetiche sulla superficie modulare è legato alla struttura dei numeri ed alle funzioni zeta molto più di quanto non si evinca da questo lavoro.

A partire dal tempo di primo ritorno e dal tempo di ritardo, ad esempio, si possono definire delle funzioni zeta che sono dei paralleli della zeta di Riemann ma sono in relazione alla distribuzione delle geodetiche chiuse (e quindi degli irrazionali quadratici) e di quelle di scattering (e quindi dei numeri razionali).

C'è ancora molto da dire, da approfondire e da scoprire sull'argomento, questa tesi vuole essere una sorta di breve e sommaria premessa, un lavoro preparatorio per potersi avvicinare allo strano mondo dei numeri, in cui geometria, analisi, topologia, algebra, logica e fisica si intrecciano e lavorano insieme.

Le applicazioni e il significato profondo di tutta la teoria che è stata recuperata in relazione all'argomento non sono ancora ben evidenti e comprensi-

bili, la speranza è che in futuro tutto questo lavoro possa essere ampliato e completato fino a diventare una panoramica il più possibile completa e ricca di spunti, adatta magari per tirare fuori qualcosa di nuovo.

Lo stupore che può nascere da alcuni dei risultati scoperti fin'ora, il misterioso intreccio di leggi che si cela dietro l'apparente naturalezza del regno dei numeri e la curiosità di cercare di comprenderlo, di certo saranno degli ottimi compagni di viaggio.

Bibliografia

- [Adl98] Roy L Adler. Symbolic dynamics and markov partitions. *Bullettin of the A.M.S.*, 35, 1998.
- [Ahl66] Lars V Ahlfors. *Complex Analysis*. McGraw-Hill Book Company, 1966.
- [Apo76] Tom M Apostol. *Introduction to Analytic Number Theory*. Springer Verlag, 1976.
- [Bea83] Alan F Beardon. *Geometry of Discrete Groups*. Springer, 1983.
- [Bea01] Alan F Beardon. Continued fractions, discrete groups and complex dynamics. *Computational Methods and Function Theory*, 1, 2001.
- [Ber02] Marcel Berger. *A panoramic view of Riemannian geometry*. Springer, 2002.
- [BKS91] Tim Bedford, Michael Keane, and Caroline Series, editors. *Ergodic theory, symbolic dynamics and hyperbolic spaces*. Oxford University Press, 1991.
- [Bow72] Rufus Bowen. The equidistribution of closed geodesics. *American Journal of Mathematics*, 94, 1972.

- [Bow73] Rufus Bowen. Symbolic dynamics for hyperbolic flows. *American Journal of Mathematics*, 95, 1973.
- [BS02] Michael Brin and Garret Stuck. *Introduction to Dynamical Systems*. Cambridge University Press, 2002.
- [CCL99] Shiing-Shen Chern, W H Chen, and K S Lan. *Lectures on Differential Geometry*. World Scientific, 1999.
- [Con78] John B Conway. *Functions of one complex variable*. Springer, 1978.
- [Cor96] Robert M Corless. Continued fractions and chaos. Lecture Notes, 1996.
- [Dav02] Wells David. *Personaggi e paradossi della matematica*. Oscar Saggi Mondadori, 2002.
- [HW80] Godfrey H Hardy and E M Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 1980.
- [IKE06a] Stefano Isola, Andreas Knauf, and Mirko Degli Esposti. Generalized farey trees, transfer operators and phase transitions. Preprint, 2006.
- [IKE06b] Stefano Isola, Andreas Knauf, and Mirko Degli Esposti. Notes on number theory, dynamical systems and statistical mechanics, 2006.
- [Isoa] Stefano Isola. Aspetti della teoria ergodica. Lecture Notes.
- [Isob] Stefano Isola. Funzioni speciali e trasformate integrali. Lecture Notes.
- [Isoc] Stefano Isola. Introduzione alla dinamica discreta (unidimensionale). Lecture Notes.

- [Iso02] Stefano Isola. On the spectrum of farey and gauss maps. *Nonlinearity*, 15, 2002.
- [Kat96] Svetlana Katok. Coding of closed geodesics after gauss and morse. *Geometriae Dedicata*, 63, 1996.
- [KG01] Svetlana Katok and Boris Gurevich. Arithmetic coding and entropy for the positive geodesic flow on the modula surface. *Moscow Mathematical Journal*, 1, 2001.
- [Kna99] Andreas Knauf. Number theory, dynamical systems and statistichal mechanics. *Reviews in Mathematical Physics*, 11, 1999.
- [Kne] Kyle Kneisl. The continued fraction system (and related systems). Lecture Notes.
- [KU05a] Svetlana Katok and Ilie Ugarcovici. Geometrically markov geodesics on the modular surface. *Moscow Mathematical Journal*, 5, 2005.
- [KU05b] Svetlana Katok and Ilie Ugarcovici. Symbolic dynamics for the modular surface and beyond. *Bulletin of the AMS*, 2005.
- [Lag91] Jeffrey C Lagarias. Number theory and dynamical systems. Preliminary Notes, 1991.
- [Lag01] Jeffrey C Lagarias. The farey shift and the minkowski ζ -function. Preprint, 2001.
- [LG01] Jeffrey C Lagarias and David Grabiner. Cutting sequences for geodesic flow on the modular surface and continued fractions. *Monatshefte für Mathematik*, 133, 2001.

- [May] Dieter H Mayer. Transfer operators, the selberg zeta function and the lewis-zagier theory of period functions. Lecture Notes.
- [Mil90] J S Milne. Modular functions and modular forms. Lecture Notes, 1990.
- [MS93] W De Melo and S Van Strien. *One-dimensional dynamics*. Springer Verlag, 1993.
- [NP69] Rolf Nevanlinna and V Paatero. *Introduction to Complex Analysis*. Addison-Wesley Publishing Company, 1969.
- [Pie] Riccardo Piergallini. Dispense del corso “elementi di topologia”. Lecture Notes.
- [PM] Planet Math (<http://planetmath.org>). Web site.
- [Pol86] Mark Pollicott. Distribution of closed geodesics on the modular surface and quadratic irrationals. *Bullettin de la S.M.F.*, 114, 1986.
- [RM] Rudi mathematici (<http://www.rudimathematici.com>). E-zine.
- [Rue89] David Ruelle. *Chaotic Evolution and Strange Attractors*. Cambridge University Press, 1989.
- [Rue02] David Ruelle. Dynamical zeta functions and transfer operators, 2002.
- [Saa92] Yousef Saad. Numerical methods for large eigenvalue problems. On-Line Book, 1992.
- [Saa03] Yousef Saad. *Iterative methods for sparse linear systems*. Manchester University Press, 2003.

- [Sar03] Peter Sarnak. Spectra of hyperbolic surfaces. *Bullettin of the A.M.S.*, 40, 2003.
- [Sera] Marcello Seri. Appunti del corso “curve e superfici” tenuto dal professor Andrea Spiro. Lecture Notes.
- [Serb] Marcello Seri. Appunti del corso “teoria della misura”. Lecture Notes.
- [Ser94] Edoardo Sernesi. *Geometria 2*. Bollati Boringhieri, 1994.
- [Ser06] Marcello Seri. La funzione zeta di Riemann. Approfondimento per esame, 2006.
- [SKF82] Iakov Grigorevich Sinai, I P Kornfeld, and S V Fomin. *Ergodic Theory*. Springer, 1982.
- [TFT] The Farey Tree (<http://www.uwm.edu/dept/math/farey.html>). Web site.
- [Tul] Sauro Tulipani. Dispense di algebra. Lecture Notes.
- [War83] Frank Wilson Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer, 1983.
- [Wei03] Eric W Weisstein. *CRC Concise Encyclopedia of Mathematics*. CRC Press, 2003.
- [WP] Wikipedia (<http://www.wikipedia.org>). Web site.