

DINAMICA E GEOMETRIA IPERBOLICA

STEFANO ISOLA

1. QUALCHE RICHIAMO DI GEOMETRIA IPERBOLICA

Com'è noto, la geometria iperbolica si ottiene da quella euclidea sostituendo il quinto postulato con la sua negazione, in particolare con l'affermazione che data una retta r ed un punto P esterno ad essa vi sono esattamente due rette parallele ad r passanti per P . Da ciò si deducono una varietà di teoremi equivalenti a tale negazione. Ad esempio che la somma degli angoli interni di un triangolo è minore di due angoli retti (e non è costante), che le aree dei triangoli sono superiormente limitate, che il luogo dei punti equidistanti ad una retta non è una retta, che non vale il teorema di Pitagora, e così via. Ora, gli assiomi della geometria euclidea erano accettati senza ulteriore indagine perchè riflettevano in modo semplice e chiaro l'evidenza empirica. Non con altrettanta facilità si potevano accettare gli assiomi della geometria iperbolica e dunque si è posto fin da subito il problema della coerenza di tale teoria. In accordo con la concezione moderna dell'assiomatica, tale problema è stato risolto scoprendo 'applicazioni' della geometria iperbolica che si sono poi tradotte in veri e propri *modelli* di tale teoria. Questo metodo consiste nel far vedere che, in base ad un'opportuna interpretazione dei concetti primitivi, gli assiomi della teoria sono veri relativamente ad un certo dominio di oggetti. Nel caso della geometria iperbolica tali oggetti sono stati individuati tra gli enti della geometria euclidea (punti, rette, cerchi, angoli, ecc) ed è per questo che si parla di *modelli euclidei della geometria iperbolica*.

Ad esempio, un modello ormai standard del piano iperbolico è dato il semipiano superiore di Lobachevski

$$(1.1) \quad \mathbb{H} := \{z = x + iy : x \in \mathbb{R}, y \in \mathbb{R}^+\}$$

\mathbb{H} è una varietà riemanniana con la metrica iperbolica

$$(1.2) \quad ds^2 := \frac{dz d\bar{z}}{(\Im z)^2} = \frac{dx^2 + dy^2}{y^2}$$

La funzione distanza su \mathbb{H} è data esplicitamente da

$$(1.3) \quad d(z, w) = \log \frac{|z - \bar{w}| + |z - w|}{|z - \bar{w}| - |z - w|}$$

Ad esempio la distanza tra due punti posti sulla stessa retta verticale, $z = a + ip$ e $w = a + iq$ è data da

$$(1.4) \quad d(z, w) = \log \frac{|p + q| + |p - q|}{|p + q| - |p - q|} = \left| \log \frac{p}{q} \right|$$

Per descrivere le proprietà geometriche di \mathbb{H} risultano fondamentali le cosiddette *trasformazioni di Möbius*:

$$(1.5) \quad g(z) = \frac{az + b}{cz + d}, \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{R})$$

dove $SL(2, \mathbb{R})$ è il gruppo delle matrici reali con determinante unitario, che d'ora in poi indicheremo con G . Osserviamo innanzitutto che una trasformazione di Möbius determina la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ a meno del segno. In particolare, entrambe le matrici I e $-I$ sono associate alla trasformazione identità $g(z) = z$. Il gruppo $PSL(2, \mathbb{R}) = G/(\pm I)$ delle trasformazioni di Möbius agisce sul piano complesso compattificato $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ (la sfera di Riemann) attraverso trasformazioni conformi. Infatti, una trasformazione g come la (1.5) trasforma un cerchio (euclideo) in un altro cerchio, se adottiamo la convenzione che le rette euclidee in $\hat{\mathbb{C}}$ siano anch'esse cerchi. Naturalmente il centro del cerchio non sarà preservato perchè in generale g non è un'isometria euclidea, ad eccezione della traslazione $g = \pm \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix}$.

Osserviamo che se $g = \begin{pmatrix} * & * \\ c & d \end{pmatrix} \in G$ allora si ha

$$(1.6) \quad g(z) - g(w) = \frac{z - w}{(cz + d)(cw + d)}$$

In particolare, ciò mostra che se z e w si trovano sulla curva $C_g := \{z \in \mathbb{C} : |cz + d| = 1\}$ allora $|g(z) - g(w)| = |z - w|$, cioè g agisce su C_g come un'isometria euclidea. Se $c \neq 0$ la curva C_g è un cerchio di raggio $|c|^{-1}$ centrato in $-d/c$, detto *cerchio isometrico* di g .

In generale si ha

$$(1.7) \quad \frac{d}{dz}g(z) = (cz + d)^{-2}$$

e dunque chiameremo il fattore $|cz + d|^{-2}$ la *deformazione* di g in z . In questi termini, la parte interna di C_g consiste nei punti con deformazione più grande di 1, quella esterna con deformazione più piccola.

Osserviamo ora che si ha (provarlo!)

$$(1.8) \quad \Im g(z) = |cz + d|^{-2} \Im z$$

e dunque la sfera di Riemann $\hat{\mathbb{C}}$ si decompone in tre sottospazi G -invarianti: \mathbb{H} , $\overline{\mathbb{H}}$ (il semipiano inferiore) e $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$.

Inoltre, le (1.7) e (1.8) implicano che

$$(1.9) \quad (\Im g(z))^{-1} |dg(z)| = (\Im z)^{-1} |dz|$$

D'altra parte la metrica iperbolica è data dal differenziale di Poincaré

$$(1.10) \quad y^{-1} |dz| = y^{-1} \sqrt{dx^2 + dy^2}$$

Vediamo dunque che le trasformazioni di Möbius sono isometrie del piano iperbolico. Un'altra (ovvia) isometria è la riflessione rispetto all'asse immaginario $z \mapsto -\bar{z}$ (che inverte l'orientazione). Si può mostrare che ciò esaurisce l'intero gruppo $\text{Isom}(\mathbb{H})$ delle isometrie di \mathbb{H} .

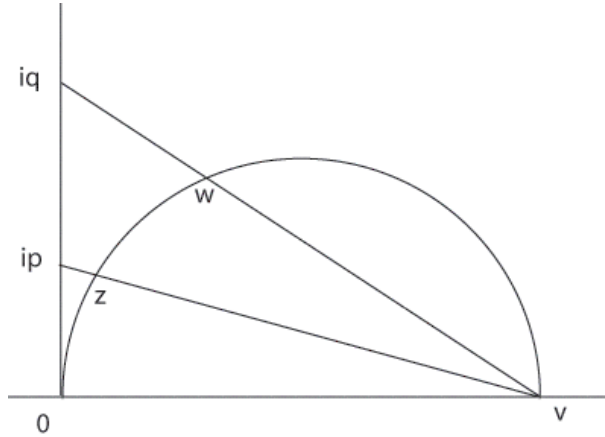
1.1. Geodetiche.

Lemma 1.1. *Le rette iperboliche, ovvero le geodetiche in \mathbb{H} , sono rappresentate da semicerchi euclidei e semirette ortogonali a \mathbb{R} .*

Dimostrazione. Consideriamo inizialmente due punti $z_0 = ip$ e $z_1 = iq$ con $p < q$. La lunghezza di una curva differenziabile $\gamma(t) = (x(t), y(t))$ in \mathbb{H} di estremi $\gamma(0) = z_0$ e $\gamma(1) = z_1$ è data da

$$\int_0^1 \frac{\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}}{y(t)} dt \geq \int_0^1 \frac{\left|\frac{dy}{dt}\right|}{y(t)} dt \geq \int_0^1 \frac{\frac{dy}{dt} dt}{y(t)} \geq \int_p^q \frac{dy}{y} = \log \frac{q}{p} \equiv d(ip, iq)$$

Pertanto la geodetica che porta da z_0 a z_1 è il tratto di semiretta verticale (semiasse immaginario) che connette i due punti. Inoltre, in riferimento alla figura si ha $w = g(iq)$ e $z = g(ip)$ con $g = v^{-1/2} \begin{pmatrix} v & 0 \\ 1 & 1 \end{pmatrix}$ e dunque $d(z, w) = d(ip, iq)$. Pertanto la funzione distanza è stazionaria sulla famiglia di curve suddette che emanano da un punto su \mathbb{R} .



Esercizio 1.2. *Provare il Lemma 1.1 facendo uso del principio variazionale di Eulero-Lagrange con lagrangiana $L = \frac{1}{2} ds^2$. Più precisamente, la curva $\gamma(t) = (x(t), y(t))$ in \mathbb{H} di estremi $z_0 = (x_0, y_0)$ e $z_1 = (x_1, y_1)$ è una geodetica se rende minimo il funzionale d'azione $\int_{t_0}^{t_1} L(\gamma(t), \gamma'(t)) dt$ e ciò è equivalente a risolvere il sistema di equazioni differenziali $\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial x'} = 0$ e $\frac{\partial L}{\partial y} - \frac{d}{dt} \frac{\partial L}{\partial y'} = 0$ con le condizioni iniziali $(x(0), y(0)) = (x_0, y_0)$ e $(x'(0), y'(0)) = (v_x, v_y)$.*

1.2. Il disco di Poincaré. La mappa

$$(1.11) \quad f(z) = \frac{z - i}{z + i}$$

è un omeomorfismo che manda \mathbb{H} sul disco unitario $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. Quest'ultimo costituisce un modello alternativo - e spesso di utilizzo più agevole - per la geometria iperbolica. La distanza su \mathbb{D} è indotta da quella su \mathbb{H} via $d(z, w) = d(f^{-1}(z), f^{-1}(w))$, $z, w \in \mathbb{D}$. Osservando che $2|f'(z)|/(1 - |f(z)|^2) = 1/\Im(z)$ vediamo che la distanza su \mathbb{D} si deriva dalla

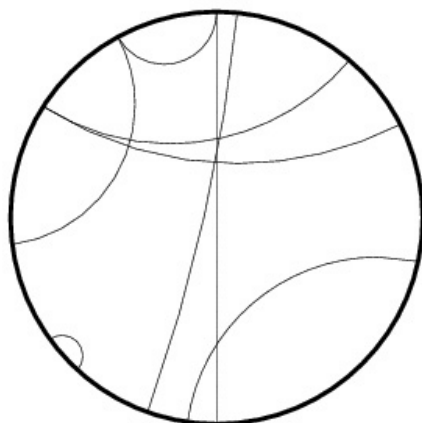
metrica

$$(1.12) \quad ds = \frac{2|dz|}{1 - |z|^2}$$

Le geodetiche in \mathbb{D} sono archi di cerchio e diametri ortogonali alla frontiera euclidea di \mathbb{D} , ossia il cerchio unitario. Le isometrie di \mathbb{D} sono le coniugate di quelle su \mathbb{H} , cioè della forma $\tilde{g} = f \circ g \circ f^{-1}$, $g \in PSL(2, \mathbb{R})$. Si può mostrare che le isometrie che preservano l'orientazione hanno la forma

$$z \mapsto \frac{az + \bar{c}}{cz + \bar{a}}, \quad a, c \in \mathbb{C}, \quad a\bar{a} - c\bar{c} = 1$$

mentre la trasformazione che corrisponde alla riflessione $z \mapsto -\bar{z}$ è anch'essa la riflessione di \mathbb{D} attorno al diametro verticale.



Ricorreremo talvolta a questo modello per trattare alcuni aspetti specifici.

1.3. \mathbb{H} come spazio omogeneo. Iniziamo con una definizione generale.

Definizione 1.3. Uno spazio topologico X su cui agisce un gruppo G si dice *spazio omogeneo* se l'azione di G è *transitiva*, cioè se per ogni coppia di punti $z, w \in X$ esiste $g \in G$ tale che $g(z) = w$.

Lemma 1.4. Il gruppo G agisce transitivamente su \mathbb{H} .

Dimostrazione. È sufficiente mostrare che dato $z \in \mathbb{H}$ esiste $g_1 \in G$ t.c. $g_1(i) = z$, perchè in tal caso basta prendere $g_2 \in G$ tale che $g_2(i) = w$ e porre $g = g_2g_1^{-1}$ per avere $g(z) = w$.

D'altra parte, se $z = x + iy$ si ha $g_1 = y^{-1/2} \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}$.

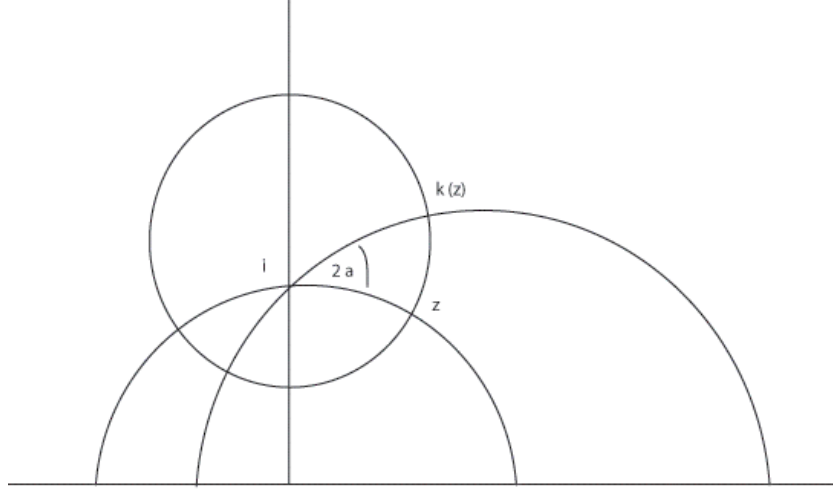
Dunque in particolare \mathbb{H} è uno spazio omogeneo e pertanto si può rappresentare come l'orbita di un punto:

$$(1.13) \quad \mathbb{H} = Gz = \{gz : g \in G\}$$

In questa prospettiva il punto $z = i$ ha un carattere speciale. Il suo gruppo di stabilità è il sottogruppo ortogonale

$$(1.14) \quad K := \{k \in G : k(i) = i\} = \left\{ k_\alpha = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} : \alpha \in \mathbb{R} \right\} \cong SO(2, \mathbb{R})$$

dove k_α agisce come la rotazione intorno al punto i di angolo 2α . Infatti è evidente che $d(i, z) = d(k_\alpha(i), k_\alpha(z)) = d(i, k_\alpha(z))$ e inoltre $k_\pi(z) = z$.



\mathbb{H} può dunque essere identificato con il quoziente G/K (lo spazio delle orbite) così che un punto $z \in \mathbb{H}$ corrisponda al *laterale (coset)* gK di K in G , cioè l'insieme di tutti i movimenti che portano i in z . L'elemento g , che dovrà evidentemente verificare $g(i) = z$, può essere preso come elemento rappresentativo di gK (nel senso della classe di equivalenza). In questa realizzazione di \mathbb{H} il gruppo G agisce su se stesso attraverso moltiplicazione di matrici. A questo proposito risulta di grande utilità la *decomposizione di Iwasawa*:

$$(1.15) \quad G = NAK$$

dove N e A sono i sottogruppi

$$(1.16) \quad N = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}, \quad A = \left\{ \begin{pmatrix} \sqrt{p} & 0 \\ 0 & \sqrt{p}^{-1} \end{pmatrix} : p \in \mathbb{R}^+ \right\}$$

In altre parole, ogni $g \in G$ ammette un'unica fattorizzazione

$$(1.17) \quad g = nak, \quad n \in N, \quad a \in A, \quad k \in K$$

In effetti, dati $c, d \in \mathbb{R}$ scegliamo u ed α in modo che $p = 1/(c^2 + d^2)$ e $-\sin \alpha = c\sqrt{p}$, $\cos \alpha = d\sqrt{p}$. Ciò assicura che

$$\begin{pmatrix} * & * \\ c & d \end{pmatrix} = \begin{pmatrix} \sqrt{p} & * \\ 0 & \sqrt{p}^{-1} \end{pmatrix} \begin{pmatrix} * & * \\ -\sin \alpha & \cos \alpha \end{pmatrix}$$

A questo punto si può agire con un'opportuna traslazione $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ per arrivare ad ottenere la riga superiore (a, b) desiderata senza alterare quella inferiore (c, d) . Tale procedura è univoca ad ogni passo.

In questo modo, il punto $z = x + iy$ in \mathbb{H} corrisponde al coset gK in G/K dove $g = nak$ con

$$n = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad a = \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}$$

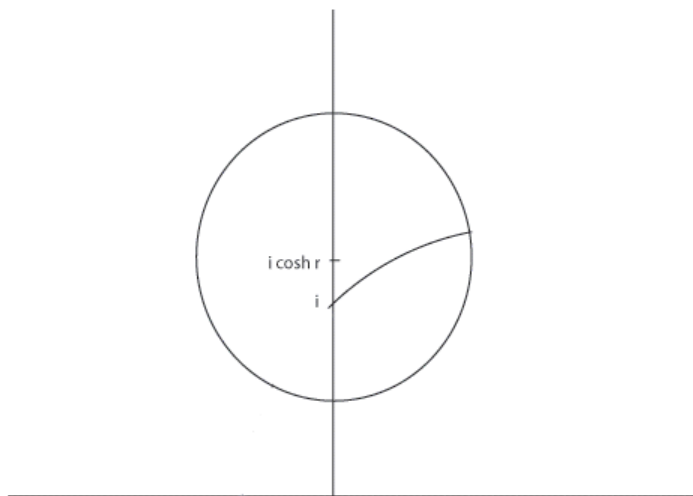
1.4. **Curvatura.** Per un teorema di geometria differenziale, il fatto che \mathbb{H} sia una varietà riemanniana omogenea è equivalente al fatto che curvatura gaussiana¹ κ di \mathbb{H} è costante. Per calcolare la costante κ possiamo usare un formula dovuta a Gauss (vedremo più avanti un metodo alternativo): se $C(r)$ è la lunghezza della circonferenza di un (qualunque) disco iperbolico di raggio r in \mathbb{H} , allora

$$(1.18) \quad \kappa = \lim_{r \rightarrow 0} (2\pi r - C(r)) \frac{3}{\pi r^3}$$

Calcoliamo la lunghezza della circonferenza del disco di raggio r centrato in i . I punti di tale circonferenza hanno la forma $z = g(i)$ con $g = ka = ak$ con $k = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix}$ e $a = \begin{pmatrix} e^{-r/2} & 0 \\ 0 & e^{r/2} \end{pmatrix}$. Le coordinate (r, φ) si chiamano *coordinate polari geodetiche* del punto z . In termini di queste coordinate, l'elemento di linea si scrive

$$(1.19) \quad ds^2 = dr^2 + (2 \sinh r)^2 d\varphi^2$$

Integrando ds con r fissato e $\varphi \in [0, \pi]$ (ricordiamo che k ruota attorno al punto i di un angolo 2φ) si ottiene che $C(r) = 2\pi \sinh r$. Infine, essendo $\sinh r = r + \frac{r^3}{6} + o(r^3)$, $r \rightarrow 0$, otteniamo $\kappa = -1$.



1.5. **Differenziale e spazio tangente.** Un'altra quantità interessante è il *differenziale* dg di una trasformazione di Möbius g , visto come mappa lineare che manda lo spazio tangente $T_z\mathbb{H}$ in $T_{g(z)}\mathbb{H}$ e definita dalla matrice 2×2

$$(1.20) \quad dg = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

¹La curvatura gaussiana $\kappa(z)$ in un punto z di una superficie riemanniana è definita dal prodotto delle due curvatures principali in z , ovvero il massimo e il minimo della curvatura di una curva contenuta nella superficie e passante per il punto.

dove si è posto $z = x + iy$ e $g(z) = w = u + iv$. Essa agisce su un vettore tangente $\zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$ come

$$dg(\zeta) = \begin{pmatrix} \frac{\partial u}{\partial x}\xi + \frac{\partial u}{\partial y}\eta \\ \frac{\partial v}{\partial x}\xi + \frac{\partial v}{\partial y}\eta \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x}\xi - \frac{\partial v}{\partial x}\eta \\ \frac{\partial v}{\partial x}\xi + \frac{\partial u}{\partial x}\eta \end{pmatrix}$$

dove abbiamo usato le classiche equazioni di Cauchy-Riemann (di facile verifica in questo caso)

$$(1.21) \quad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad , \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$$

D'altra parte, esprimendo x e y in termini di z e \bar{z} si trova

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) \quad , \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right)$$

e dunque usando ancora le (1.21) si vede che

$$(1.22) \quad g'(z) = \frac{dw}{dz} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}$$

Pertanto, se identifichiamo lo spazio tangente $T_z\mathbb{H} \simeq \mathbb{R}^2$ con il piano complesso \mathbb{C} per mezzo della mappa $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \mapsto \xi + i\eta = \zeta$ si ha

$$(1.23) \quad g'(z)\zeta = \left(\frac{\partial u}{\partial x}\xi - \frac{\partial v}{\partial x}\eta \right) + i \left(\frac{\partial v}{\partial x}\xi + \frac{\partial u}{\partial x}\eta \right)$$

e confrontando otteniamo

$$(1.24) \quad dg(\zeta) = g'(z)\zeta$$

dove a sinistra c'è una trasformazione lineare di $T_z\mathbb{H} \simeq \mathbb{R}^2$ e a destra la moltiplicazione di due numeri complessi.

Lemma 1.5. *Il differenziale dg preserva la norma nello spazio tangente $T_z\mathbb{H}$.*

Dimostrazione. Per quanto visto sopra si ha

$$\|dg(\zeta)\| := \frac{|dg(\zeta)|}{\Im(g(z))} = \frac{|g'(z)||\zeta|}{\Im(g(z))} = \frac{|\zeta|}{\Im(z)} = \|\zeta\|$$

1.6. Area. Vediamo ora come si calcola l'area di una data regione in \mathbb{H} . Per far questo ci serve una misura. La misura riemanniana derivata dal differenziale di Poincarè $ds = y^{-1}|dz|$ si può esprimere per mezzo della misura di Lebesgue come

$$(1.25) \quad d\mu(z) = \frac{dxdy}{y^2}$$

Lemma 1.6. *La misura μ è G -invariante: per ogni regione $A \subset \mathbb{H}$ tale che $\mu(A)$ esiste, si ha $\mu(A) = \mu(g(A))$ per ogni $g \in G$.*

Dimostrazione Il confronto tra la misura delle regioni A e $g(A)$ si ottiene dalla comparazione tra due integrali doppi per mezzo di cambiamento di variabili del tipo $w = g(z)$ in cui compare il determinante jacobiano di g , cioè il determinante del differenziale dg , e questo, usando nuovamente le (1.21) e le osservazioni fatte sopra, è dato da

$$\frac{\partial(u, v)}{\partial(x, y)} := \det \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 = |g'(z)|^2 = |cz + d|^{-4}$$

Pertanto, usando la (1.9) troviamo

$$\mu(g(A)) = \int_{g(A)} \frac{dudv}{v^2} = \int_A \frac{\partial(u, v)}{\partial(x, y)} \frac{dxdy}{v^2} = \int_A \frac{1}{|cz + d|^4} \frac{|cz + d|^4}{y^2} dxdy = \mu(A)$$

In coordinate polari geodetiche (r, φ) la misura μ si esprime come

$$(1.26) \quad d\mu(z) = 2 \sinh r \, dr d\varphi$$

Così, ad esempio, l'area $A(r)$ del disco riportato sopra è data da

$$A(r) = \int_0^\pi \int_0^r 2 \sinh r' \, dr' d\varphi = 2\pi (\cosh r - 1) = 4\pi \left(\sinh \frac{r}{2} \right)^2$$

Osserviamo a questo punto che se riguardiamo il disco come euclideo allora il suo centro è in $i \cosh r$ e dunque il suo raggio è $\sinh r$. Pertanto la circonferenza misura $2\pi \sinh r$ (la stessa di quella iperbolica) mentre l'area è $\pi(\sinh r)^2$, che è molto più grande di quella iperbolica per grandi r !

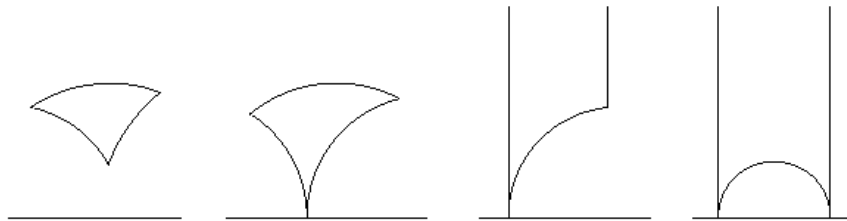
Ciò esemplifica una proprietà generale che discende dalla cosiddetta *disuguaglianza isoperimetrica*, secondo la quale se Ω è l'area di un certo dominio su una varietà riemanniana e L la lunghezza della sua frontiera allora (assumendo la curvatura κ costante) si ha

$$(1.27) \quad 4\pi \Omega - \kappa \Omega^2 \leq L^2$$

Nel piano euclideo, dove $\kappa = 0$, la (1.27) si riduce a $4\pi \Omega \leq L^2$ e dunque l'area può essere molto più grande della lunghezza della frontiera. Al contrario sul piano iperbolico, dove $\kappa = -1$, si ha $4\pi \Omega + \Omega^2 \leq L^2$ e dunque $\Omega \leq L$, ovvero l'area è comparabile con la lunghezza della frontiera.

Esercizio 1.7. *Verificare che sia per il disco iperbolico visto sopra ($\kappa = -1$), sia per il suo corrispondente euclideo ($\kappa = 0$), vale il segno di uguaglianza in (1.27) (che pertanto è ottimale).*

Regioni di particolare interesse sono i *triangoli geodetici*, cioè figure limitate da tre archi di geodetica.



Per tali regioni valgono le seguenti formule di trigonometria iperbolica: dato un triangolo iperbolico con lati di lunghezza (iperbolica) a, b e c ed angoli opposti α, β e γ si ha

- (1) *teorema del seno*: $\sinh a/\sin \alpha = \sinh b/\sin \beta = \sinh c/\sin \gamma$
- (2) *teorema del coseno I*: $\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma$
- (3) *teorema del coseno II*: $\cosh c = (\cos \alpha \cos \beta + \cos \gamma)/\sin \alpha \sin \beta$

Due osservazioni: A) se nel teorema del coseno I poniamo $\gamma = \pi/2$ otteniamo $\cosh c = \cosh a \cosh b$, che è la versione iperbolica del Teorema di Pitagora. B) Il teorema del coseno II non ha equivalenti in geometria euclidea. Esso implica che se due triangoli geodetici hanno gli stessi angoli, allora esiste un'isometria che trasforma un triangolo nell'altro. Vale inoltre il

Teorema 1.8. (*Gauss-Bonnet*). *Se Δ è un triangolo geodetico su una superficie riemanniana (cioè i cui lati sono formati da archi di geodetica) di angoli interni α, β e γ allora*

$$\int_{\Delta} \kappa d\mu = (\alpha + \beta + \gamma) - \pi$$

Applicando questo risultato al triangolo geodetico di vertici $-1/2, 1/2$ e ∞ , i cui angoli interni sono tutti nulli (vedi l'ultimo esempio nella figura sopra) e ricordando che la curvatura deve essere costante troviamo

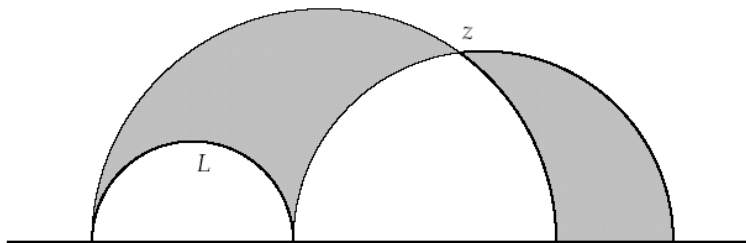
$$\kappa \mu(\Delta) + \pi = 0$$

e inoltre

$$\mu(\Delta) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\sqrt{1/4-x^2}}^{+\infty} \frac{dx dy}{y^2} = \pi$$

Ritroviamo così il risultato $\kappa = -1$.

Osservazione 1.9. Osserviamo che in questa geometria i cinque postulati di Euclide sono tutti validi eccetto il quinto, quello delle parallele: c'è più di una geodetica che passa per il punto z , esterno alla geodetica L , e che non interseca L stessa.



1.8. Classificazione dei moti. Le trasformazioni di Möbius sono moti rigidi del piano iperbolico, ed agiscono in vari modi. Essendo le azioni di moti coniugati simili tra loro, una classificazione di tali moti dovrà essere invariante per coniugazione. Data $g \in PSL(2, \mathbb{R})$ indichiamo la sua classe di coniugazione con

$$(1.28) \quad \{g\} := \{\tau g \tau^{-1} : \tau \in PSL(2, \mathbb{R})\}$$

Ovviamente l'identità forma una classe da sola. Un importante invariante geometrico della coniugazione è dato dal numero di punti fissi, ed ogni elemento

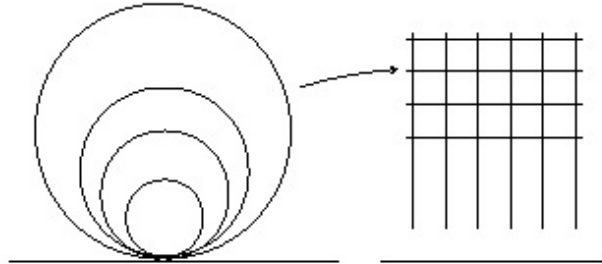
$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq \pm I$$

ne può avere uno o due in $\hat{\mathbb{C}}$: essi sono $b/(d-a)$ se $c=0$ e $(a-d \pm \sqrt{(a+d)^2 - 4})/2c$ se $c \neq 0$. Dunque vi sono tre possibilità:

- g ha un punto fisso su $\hat{\mathbb{R}}$;
- g ha due punti fissi distinti su $\hat{\mathbb{R}}$;
- g ha un punto fisso in \mathbb{H} e il suo complesso coniugato in $\overline{\mathbb{H}}$.

Diciamo allora che g è rispettivamente *parabolico*, *iperbolico* o *ellittico* (e tale classificazione si estende alle classi di coniugazione). Ogni classe di coniugazione $\{g\}$ contiene un rappresentante in uno (e solo in uno) dei sottogruppi N , A , K . Gli elementi di $\pm N$, $\pm A$, K diversi da $\pm I$ sono rispettivamente parabolici, iperbolici ed ellittici. Ricordando le (7.7) e (1.16) essi agiscono come

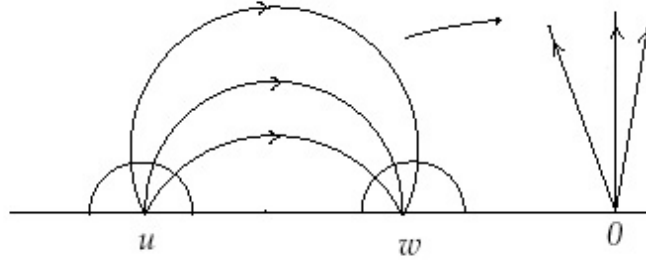
- TRASLAZIONE (p.to fisso ∞): $z \mapsto z + x$, $x \in \mathbb{R}$. Ogni elemento coniugato muove i punti lungo *orocicli*, cioè cerchi in \mathbb{H} tangenti a $\hat{\mathbb{R}}$. La distanza iperbolica tra z e $g(z)$ è $|x|$;



- DILATAZIONE (p.ti fissi 0 e ∞): $z \mapsto pz$, $p \in \mathbb{R}^+$. Ogni elemento coniugato muove i punti lungo *ipercicli*, cioè segmenti in \mathbb{H} di cerchi in $\hat{\mathbb{C}}$ che passano per i punti fissi u e w su $\hat{\mathbb{R}}$, di cui uno è attrattivo e uno repulsivo. Uno di questi ipercicli è la geodetica che unisce u e w e la distanza iperbolica tra un punto z su di essa e la sua immagine $g(z)$ è $|\log p| = 2 \log |\lambda_g|$, dove λ_g è l'autovalore di g di modulo > 1 (cioè in questo caso $\lambda = \sqrt{p}^{\pm 1}$ a seconda se $p < 1$ o $p \geq 1$). La formula

$$(1.29) \quad d(z, g(z)) = 2 \log |\lambda_g|$$

resta ovviamente valida per ogni elemento iperbolico g di $PSL(2, \mathbb{R})$;



- ROTAZIONE (p.to fisso i): $z \mapsto k_\alpha(z)$, $\alpha \in \mathbb{R}$. Ogni elemento coniugato muove i punti lungo cerchi centrati nel suo p.to fisso in \mathbb{H} .

Un importante invariante algebrico per la coniugazione è (il valore assoluto de) la traccia, nei cui termini le classi suddette si caratterizzano come segue: $g = \begin{pmatrix} a & a \\ c & d \end{pmatrix}$ è parabolico sse $|a + d| = 2$, iperbolico sse $|a + d| > 2$, ellittico sse $|a + d| < 2$.

2. I FLUSSI GEODETICI E OROCICLICI SU $S\mathbb{H}$

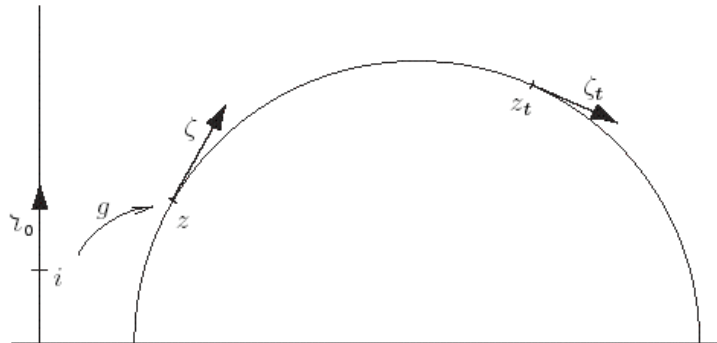
Indichiamo con $z(t)$ la geodetica con punto iniziale $z = z(0)$ e vettore tangente unitario $\zeta = z'(0) \in T_z\mathbb{H}$. Tale geodetica incontra $\partial\mathbb{H}$ in due punti (o entrambi reali o uno di essi all' ∞)

$$w := \lim_{t \rightarrow \infty} z(t) \quad u := \lim_{t \rightarrow -\infty} z(t)$$

Il *flusso geodetico* su \mathbb{H} è definito come il gruppo ad un parametro g_t che agisce sul fibrato tangente unitario $S\mathbb{H}$ come segue

$$(2.1) \quad g_t : S\mathbb{H} \rightarrow S\mathbb{H} \quad g_t(z, \zeta) = (z(t), \zeta(t))$$

Sia $v = (z, \zeta) \in S\mathbb{H}$ con $z \in \mathbb{H}$ e $\zeta \in T_z\mathbb{H} \simeq \mathbb{C}$. Osserviamo che $S\mathbb{H}$ può essere identificato con $PSL(2, \mathbb{R})$ facendo corrispondere a v l'unico elemento $g \in PSL(2, \mathbb{R})$ tale che $z = g(i)$ e $\zeta = dg(\zeta_0) = g'(z)\zeta_0$ dove ζ_0 è il vettore unitario tangente all'asse immaginario.



È comodo poi scrivere il vettore tangente unitario nella forma $\zeta = \Im(z)e^{i(\theta + \frac{\pi}{2})}$ dove θ è l'angolo formato da ζ con la verticale misurato in senso antiorario. Identificando quindi ζ con

θ otteniamo la parametrizzazione $v = (z, \theta)$ per i punti di $S\mathbb{H}$ e determina la corrispondenza

$$(2.2) \quad (z, \theta) = (g(i), \beta_g(i))$$

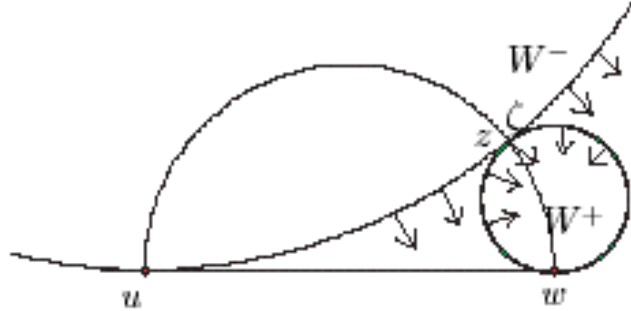
dove $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ è data da

$$(2.3) \quad z = g(i) = \frac{b + ia}{d + ic}, \quad \theta = \beta_g(i) = -2 \arg(d + ic) = -2 \tan^{-1}(c/d)$$

Con questa identificazione l'azione di g_t su $PSL(2, \mathbb{R})$ corrisponde alla moltiplicazione da destra per il sottogruppo ad un parametro delle matrici

$$(2.4) \quad a_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, \quad g_t \leftrightarrow g a_t$$

In modo analogo si possono definire i flussi orociclici. Dato un punto $v = (z, \zeta) \in S\mathbb{H}$ l'*orociclo positivo* $H^+(z, \zeta)$ è definito dal cerchio euclideo in \mathbb{H} che passa per z ed è tangente all'asse reale nel punto w . Può essere ottenuto come il limite dei cerchi iperbolici $B(z(t), t) = \{w \in \mathbb{H} : d(w, z(t)) = t\}$ quando t cresce². Osserviamo che $z \in B(z(t), t)$ per ogni t e che ζ è ortogonale a $H^+(z, \zeta)$ e rivolto verso l'interno. Nel caso in cui $w = \infty$ l'orociclo positivo è la retta orizzontale passante per z . Si può definire in modo analogo l'*orociclo negativo* $H^-(z, \zeta)$ come il limite dei cerchi iperbolici $B(z(-t), t)$. In questo caso si ha che $z \in B(z(-t), t)$ per ogni t e che ζ è ortogonale a $H^-(z, \zeta)$ ed è rivolto verso l'esterno.



Se indichiamo con $W^+(v)$ il campo vettoriale unitario che contiene v ed è normale all'orociclo H^+ , allora ciascun vettore di $W^+(v)$ definisce una geodetica che converge verso w per $t \rightarrow \infty$. Inoltre, ogni vettore appartenente al campo vettoriale unitario W^- che contiene v ed è normale all'orociclo H^- definisce una geodetica che converge verso u per $t \rightarrow -\infty$.

Il *flusso orociclico* su \mathbb{H} è definito dall'azione del gruppo ad un parametro h_t^+ su $S\mathbb{H}$ che muove i vettori normali a $H^+(z, \zeta)$ verso destra lungo $H^+(z, \zeta)$ con velocità unitaria. Dunque

$$(2.5) \quad h_t^+ : S\mathbb{H} \rightarrow S\mathbb{H} \quad h_t^+(z, \zeta) = (z', \zeta') \text{ con } H^+(z', \zeta') = H^+(z, \zeta).$$

Similmente, il flusso orociclico $h_t^- : S\mathbb{H} \rightarrow S\mathbb{H}$ muove i vettori normali esterni a $H^-(z, \zeta)$ verso destra lungo $H^-(z, \zeta)$ a velocità unitaria, ovvero

$$(2.6) \quad h_t^- : S\mathbb{H} \rightarrow S\mathbb{H} \quad h_t^-(z, \zeta) = (z', \zeta') \text{ con } H^-(z', \zeta') = H^-(z, \zeta).$$

²Si tratta dunque di un cerchio limite con centro infinitamente lontano e raggio infinito. È una nozione dovuta a Lobatchevski la cui etimologia deriva dal greco "horos" = orizzonte.

Con l'identificazione fatta sopra l'azione di h_t^+ e h_t^- su $PSL(2, \mathbb{R})$ corrisponde alla moltiplicazione da destra per i sottogruppi ad un parametro delle matrici

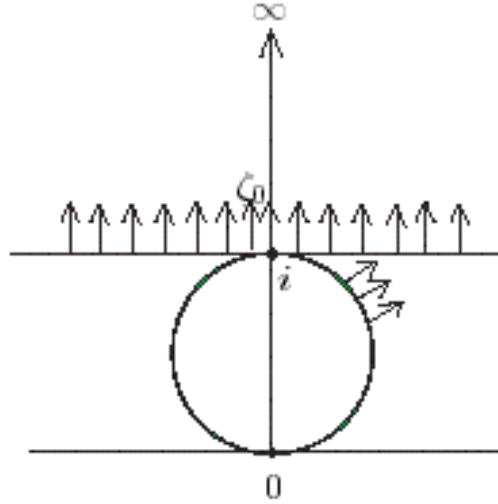
$$(2.7) \quad n_t^+ = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad h_t^+ \leftrightarrow g n_t^+$$

e

$$(2.8) \quad n_t^- = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}, \quad h_t^- \leftrightarrow g n_t^-$$

Ad esempio, partendo dal il punto $(z, \zeta_0) \equiv (i, 0)$ ($g = \mathbb{1}$) si trova

$$(2.9) \quad \begin{aligned} g_t &: (i, 0) \mapsto (ie^t, 0) \\ h_t^+ &: (i, 0) \mapsto (t + i, 0) \\ h_t^- &: (i, 0) \mapsto \left(\frac{i}{1 + it}, -2 \tan^{-1}(t) \right) \end{aligned}$$



Più in generale, partendo da $z = g(i)$ con $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ valgono le formule:

$$(2.10) \quad \begin{aligned} g_t &: \left(\frac{b + ia}{d + ic}, \tan^{-1}\left(\frac{c}{d}\right) \right) \mapsto \left(\frac{b + ia e^t}{d + i c e^t}, -2 \tan^{-1}\left(\frac{c}{d} e^t\right) \right) \\ h_t^+ &: \left(\frac{b + ia}{d + ic}, \tan^{-1}\left(\frac{c}{d}\right) \right) \mapsto \left(\frac{at + b + ia}{ct + d + ic}, -2 \tan^{-1}\left(\frac{c}{ct + d}\right) \right) \\ h_t^- &: \left(\frac{b + ia}{d + ic}, \tan^{-1}\left(\frac{c}{d}\right) \right) \mapsto \left(\frac{b + i(a + bt)}{d + i(c + dt)}, -2 \tan^{-1}\left(\frac{ct + d}{d}\right) \right) \end{aligned}$$

da cui si ottengono facilmente le seguenti regole di commutazione

Lemma 2.1.

$$g_t \circ g_s = g_s \circ g_t, \quad g_t \circ h_s^+ = h_{se^{-t}}^+ \circ g_t, \quad g_t \circ h_s^- = h_{se^t}^- \circ g_t$$

Ad ogni punto $v = (z, \theta) \in S\mathbb{H}$ possiamo associare lo spazio tangente tridimensionale $T_v(S\mathbb{H})$ la cui base è data dai vettori tangenti in $t = 0$ alle curve in $S\mathbb{H}$ formate da $g_t(v)$, $h_t^+(v)$ e

$h_t^-(v)$. I tre sottospazi unidimensionali E_v^0, E_v^+ e E_v^- generati da questi vettori sono invarianti per g_t e dal Lemma 2.1 si ha

$$(2.11) \quad \|T_{g_t}|E_v^0\| = 1, \quad \|T_{g_t}|E_v^+\| = e^{-t}, \quad \|T_{g_t}|E_v^-\| = e^t$$

Infine, la misura di Liouville su $S\mathbb{H}$ è data dal prodotto della misura di area $d\mu$ su \mathbb{H} e della misura di lunghezza $d\theta$ sul cerchio unitario:

$$(2.12) \quad dm = d\mu d\theta = \frac{dx dy d\theta}{y^2}$$

Essendo invariante per moltiplicazione da destra di a_t, n_t^+ e n_t^- la misura m è preservata dai flussi g_t, h_t^+ e h_t^- .

IL FLUSSO ELLITTICO. Possiamo definire anche il *flusso ellittico* $e_t : S\mathbb{H} \rightarrow S\mathbb{H}$ che semplicemente ruota un vettore in senso antiorario di un angolo t . Con l'identificazione fatta sopra l'azione di e_t su $PSL(2, \mathbb{R})$ corrisponde alla moltiplicazione da destra per il sottogruppo ad un parametro delle matrici

$$(2.13) \quad k_{t/2} = \begin{pmatrix} \cos t/2 & \sin t/2 \\ -\sin t/2 & \cos t/2 \end{pmatrix}, \quad e_t \leftrightarrow g k_{t/2}$$

Evidentemente anche e_t conserva la misura m . Inoltre, considerando un punto $v = (x + i, 0)$ sull'orociclo orizzontale $\{z : \Im z = 1\}$ e tracciando la geodetica passante per v e $h_s^+(v) = (x + s + i, 0)$ ci si convince rapidamente che vale la relazione

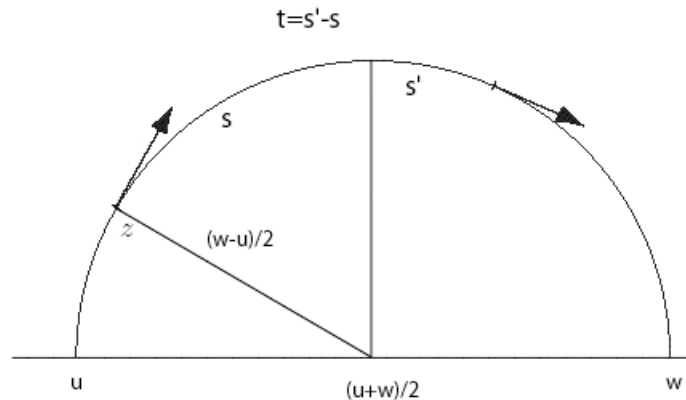
$$(2.14) \quad h_s^+ = e_{\pi-r} \circ g_t \circ e_{-r}$$

con $t(s) \rightarrow \infty$ e $r(s) \rightarrow 0$ quando $s \rightarrow \infty$.

ULTERIORE PARAMETRIZZAZIONE.

Osserviamo infine che dato un punto $v = (z, \theta) \in S\mathbb{H}$ esiste un'unica geodetica $z(t)$ tale che $z = z(0)$ e $\zeta = z'(0)$. Se poniamo $u = z(-\infty)$ e $w = z(\infty)$ allora possiamo parametrizzare $v = (z, \theta)$ nella forma $v = (u, w, s)$ dove s è la lunghezza d'arco misurata dal centro della geodetica al punto z . In queste coordinate la misura di Liouville diviene

$$(2.15) \quad dm = \frac{du dw ds}{(u - w)^2}$$



Con questa parametrizzazione il flusso geodetico diviene semplicemente

$$(2.16) \quad g_t : (u, w, s) \mapsto (u, w, s + t)$$

da cui è immediato verificare che preserva la misura m .

3. GRUPPI FUCHSIANI E LORO DOMINI FONDAMENTALI

Abbiamo visto nella sezione precedente come $S\mathbb{H}$ possa essere identificato con il gruppo $PSL(2, \mathbb{R})$. Oltre ad essere un gruppo $PSL(2, \mathbb{R})$ è uno spazio topologico. La convergenza $g_n \rightarrow g$ in tale spazio significa che esistono delle matrici $m_n \in SL(2, \mathbb{R})$ che rappresentano g_n e tali che $\lim \|m_n - m\| = 0$, dove $\| \cdot \|$ è una norma su $SL(2, \mathbb{R})$ indotta da \mathbb{R}^4 .

Definizione 3.1. Un sottogruppo Γ di $\text{Isom}(\mathbb{H})$ si dice *discreto* se la topologia indotta su Γ è una topologia discreta, ovvero se Γ è un insieme discreto nello spazio topologico $\text{Isom}(\mathbb{H})$. Pertanto Γ è discreto se e solo se $g_n \rightarrow I$, $g_n \in \Gamma$ implica $g_n = I$ per n abbastanza grande.

Definizione 3.2. Un sottogruppo discreto Γ di $\text{Isom}(\mathbb{H})$ si dice *gruppo Fuchsiano* se consiste unicamente in trasformazioni che preservano l'orientazione. In altre parole, un gruppo Fuchsiano è un sottogruppo discreto di $PSL(2, \mathbb{R})$.

Definizione 3.3. Sia Γ un gruppo che agisce su uno spazio topologico X . L'azione di Γ si dice *discontinua* se per ogni $x \in X$ e per ogni successione infinita (g_i) di elementi distinti di Γ l'insieme $\{g_i x\}$ non ha punti di accumulazione; si dice *propriamente discontinua* se per ogni coppia di punti $x, y \in X$ esistono due intorni U_x e U_y tali che l'insieme $\{g \in \Gamma : g(U_x) \cap U_y \neq \emptyset\}$ sia finito.

Poincaré ha dimostrato che per un sottogruppo Γ di $PSL(2, \mathbb{R})$ l'essere Fuchsiano, agire in modo discontinuo ed agire in modo propriamente discontinuo sono proprietà equivalenti. Poniamo infine la

Definizione 3.4. Un punto $z \in \mathbb{H}$ si dice *punto ellittico* di un gruppo Fuchsiano Γ se è il punto fisso di un elemento ellittico di Γ . Un punto $x \in \mathbb{R} \cup \{\infty\}$ si dice *cuspidale* se è punto fisso di qualche elemento parabolico di Γ .

Essendo un gruppo discreto l'orbita $\Gamma z := \{g(z) : g \in \Gamma\}$ di un punto $z \in \mathbb{H}$ non ha punti di accumulazione in \mathbb{H} . Tuttavia può averne sulla retta reale. Indichiamo con $\Lambda(\Gamma)$ l'*insieme limite* di Γ , cioè l'insieme dei punti di accumulazione delle orbite Γz con $z \in \mathbb{H}$. Allora $\Lambda(\Gamma) \subset \hat{\mathbb{R}}$. L'insieme limite può essere vuoto, può contenere uno o due punti, o può contenerne infiniti. Nell'ultimo caso si distinguono i gruppi Fuchsiani del primo tipo, il cui insieme limite è $\hat{\mathbb{R}}$, e del secondo tipo, il cui insieme limite è un insieme di Cantor.

DOMINI FONDAMENTALI

Un gruppo Fuchsiano può essere visualizzato attraverso il suo dominio fondamentale. Due punti $z, w \in \hat{\mathbb{C}}$ si dicono *equivalenti* se $w \in \Gamma z$ e in tal caso si scrive $z \equiv w \pmod{\Gamma}$. Un sottoinsieme $\mathbb{F}_\Gamma \subset \mathbb{H}$ si chiama *dominio fondamentale* per Γ se

- \mathbb{F}_Γ è aperto e connesso,
- punti distinti di \mathbb{F}_Γ non sono equivalenti,
- ogni orbita di Γ contiene un punto in $\overline{\mathbb{F}_\Gamma}$.

Evidentemente, l'azione di Γ su \mathbb{F}_Γ ricopre tutto \mathbb{H} (tessellazione):

$$(3.1) \quad \cup_{g \in \Gamma} g(\mathbb{F}_\Gamma) = \mathbb{H}$$

Ogni gruppo Fuchsiano ammette un dominio fondamentale, anche se non unico. Tuttavia, tutti i domini fondamentali hanno la stessa area

$$(3.2) \quad |\mathbb{F}_\Gamma| = \int_{\mathbb{F}_\Gamma} d\mu$$

Inoltre, il dominio fondamentale di un gruppo Fuchsiano del primo tipo ha sempre volume finito e può essere scelto come un poligono convesso del tipo (dominio di Dirichlet):

$$(3.3) \quad D(w) := \{z \in \mathbb{H} : d(z, w) < d(z, g(w)), \quad \forall g \in \Gamma, \quad g \neq I\}$$

GRUPPI MODULARI.

Esempi importanti di gruppi Fuchsiani del primo tipo sono dati dai *gruppi modulari*. Il principale di essi è $\Gamma(1) := PSL(2, \mathbb{Z})$ cioè

$$(3.4) \quad \Gamma(1) \ni g : z \mapsto \frac{az + b}{cz + d}, \quad a, b, c, d, \in \mathbb{Z}, \quad ad - bc = 1$$

Questo gruppo ha due generatori:

$$(3.5) \quad S := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{e} \quad R := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

che agiscono come $S(z) = -1/z$ e $R(z) = z + 1$. Osserviamo che, con le notazioni della Sezione 1.3, si ha $S \in K$ (è la rotazione di angolo π attorno al punto i) e $R \in N$ (è la traslazione di un'unità verso destra). Inoltre si ha

$$(3.6) \quad S^2 = (SR^{-1})^3 = I$$

Pertanto $\Gamma(1)$ non è un gruppo libero³. Inoltre si ha che

Lemma 3.5. *I punti ellittici di $\Gamma(1)$ sono i punti $z = i$ e $z = e^{2\pi i/3} + 1$. Le cuspidi di $\Gamma(1)$ sono esattamente i punti di $\mathbb{Q} \cup \{\infty\}$.*

Dimostrazione. Il punto $i \in \mathbb{H}$ è un punto fisso dell'involutione S , mentre il punto $e^{2\pi i/3} + 1 = \frac{1}{2} + i\frac{\sqrt{3}}{2}$ è un punto fisso di SR^{-1} . In effetti, prendendo S e SR^{-1} come generatori si può riguardare $\Gamma(1)$ come prodotto libero dei due gruppi ciclici⁴ (rispettivamente di ordine 2 e 3) generati da S e SR^{-1} rispettivamente.

Veniamo ora alle cuspidi. Evidentemente ∞ è p.to fisso di R . Sia ora $p/q \in \mathbb{Q}$ con $(p, q) = 1$. Per il teorema cinese del resto⁵ esistono altri due interi r, s tali che $(r, s) = 1$ e $ps - rq = 1$.

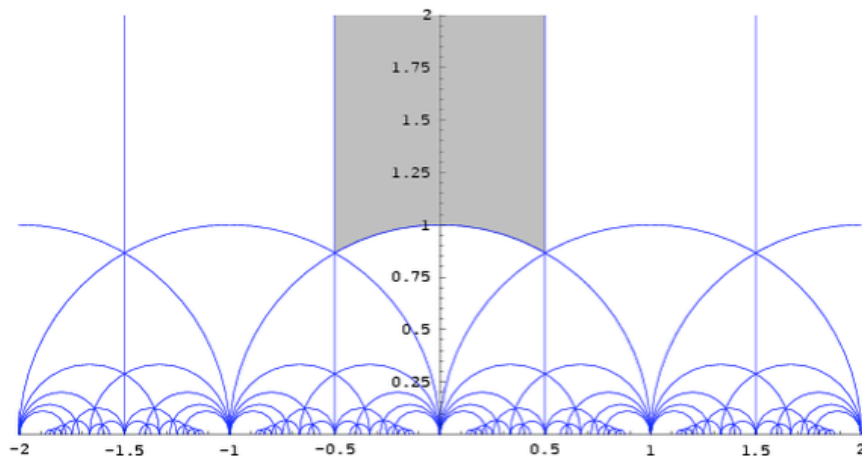
³Un gruppo G si dice *libero* se esiste un sottoinsieme K di G tale che è possibile scrivere ogni elemento di G come prodotto di un numero finito di elementi di K e dei suoi inversi in modo unico.

⁴Un gruppo G si dice *ciclico* se può essere generato da un singolo elemento g , cioè $G = \{g^n : n \in \mathbb{Z}\}$. Il suo *ordine* è il numero dei suoi elementi non equivalenti.

⁵*Teorema cinese del resto.* Se n_1, n_2, \dots, n_k sono interi mutuamente primi tra loro, allora per ogni scelta di interi a_1, a_2, \dots, a_k , esiste un intero x che risolve simultaneamente le congruenze $x \equiv a_1 \pmod{n_1}$, $x \equiv a_2 \pmod{n_2}$, \dots , $x \equiv a_k \pmod{n_k}$. Inoltre tutte le soluzioni x di questo sistema sono congruenti modulo il prodotto $n_1 \cdots n_k$.

Se dunque $g = \begin{pmatrix} p & r \\ q & s \end{pmatrix}$ allora $g(\infty) = p/q$ e dunque p/q è punto fisso di $gRg^{-1} \in \Gamma(1)$. Viceversa, dall'algebra lineare si ha che ogni elemento parabolico di $\Gamma(1)$ è coniugato a R .

Il dominio fondamentale $\mathbb{F}_{\Gamma(1)} := \{z = x + iy : |x| < 1/2, |z| > 1\}$ e la relativa tessellazione di \mathbb{H} per il gruppo $\Gamma(1)$ sono mostrati in figura. Osserviamo che si tratta di un poligono con quattro lati: le due semirette verticali e i due tratti della geodetica che unisce $e^{2\pi i/3}$ a $e^{2\pi i/3} + 1$ separati dal punto fisso ellittico i di ordine 2.



Possiamo poi considerare i sottogruppi principali di congruenza di ordine $N \in \mathbb{N}$ definiti da

$$(3.7) \quad \Gamma(N) = \{g \in \Gamma(1) : g \equiv I \pmod{N}\}$$

Tali sottogruppi sono normali ed hanno indice $\mu_N := [\Gamma(1) : \Gamma(N)]$ finito⁶. Più precisamente

$$(3.8) \quad \mu_N = N^3 \prod_{p|N} (1 - p^{-2})$$

mentre il numero di cuspidi non equivalenti⁷ è μ_N/N .

LE SUPERFICI $\Gamma \backslash \mathbb{H}$.

Per ogni gruppo Fuchsiano Γ si definisce la superficie M_Γ data dall'insieme delle orbite $M_\Gamma := \Gamma \backslash \mathbb{H} = \{\Gamma z : z \in \mathbb{H}\}$, munito della topologia quoziente⁸. Per quanto visto sopra M_Γ può essere rappresentata identificando i lati del poligono \mathbb{F}_Γ che si corrispondono per mezzo dei generatori di Γ . La superficie così ottenuta, non solo è una superficie riemanniana, ma eredita da \mathbb{H} anche la metrica iperbolica ed è dunque una varietà riemanniana iperbolica di curvatura negativa costante $\kappa = -1$. Se poi indichiamo con Γ^* il gruppo di trasformazioni

⁶Un sottogruppo H di un gruppo G si dice *normale* se i suoi *lateral* sinistro e destro, cioè $gH = \{gh : h \in H\}$ e $Hg = \{hg : h \in H\}$, coincidono. L'*indice* di un sottogruppo normale H è la cardinalità dei suoi *lateral*.

⁷Due cuspidi in p/q e p'/q' sono equivalenti sse $\pm \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} p' \\ q' \end{pmatrix} \pmod{N}$.

⁸Nella topologia quoziente $(s_n) \in M_\Gamma$ converge a $s \in M_\Gamma$ sse si può rappresentare ogni s_n con un elemento $z_n \in \mathbb{H}$ nella classe di equivalenza di s_n tale (z_n) converge a qualche $z \in \mathbb{H}$ nella classe di equivalenza di s .

indotto da Γ su $S\mathbb{H}$ allora si ha

$$(3.9) \quad SM_\Gamma = S(\Gamma \backslash \mathbb{H}) = \Gamma^* \backslash S\mathbb{H}$$

Osserviamo che se $v = (z, \theta) \in S\mathbb{H}$, dove si è identificato il vettore tangente $\zeta = \Im(z)e^{i(\theta + \frac{\pi}{2})}$ con l'angolo θ da esso formato con la verticale e misurato in senso antiorario, allora si ha (cf. (2.3))

$$(3.10) \quad \Gamma^* \ni g^* : (z, \theta) \mapsto (g(z), \theta + \beta_g(z))$$

Infine, se Γ un gruppo Fuchsiano del primo tipo la corrispondente varietà riemanniana $M \equiv M_\Gamma$ ha volume finito ma non è necessariamente compatta. Quest'ultimo caso si verifica se e solo se Γ non contiene elementi parabolici.

In particolare, se Γ è un sottogruppo di $\Gamma(1)$ di indice finito μ , la superficie M_Γ prende il nome di *superficie modulare*. Ad esempio, la superficie $M_{\Gamma(1)}$ si ottiene identificando le semirette verticali di $\mathbb{F}_{\Gamma(1)}$ per mezzo di R e i due tratti della geodetica che contiene i per mezzo di S . Si ottiene così una superficie non compatta ma di volume finito $|\mathbb{F}_{\Gamma(1)}| = \pi/3$ e genere $g = 0^9$.

Per $\Gamma = \Gamma(N)$ con $N > 1$ il genere $g(M_{\Gamma(N)})$ della superficie $M_{\Gamma(N)}$ è dato dalla formula

$$(3.11) \quad g = 1 + \frac{N-6}{12N} \mu_N$$

mentre il primo numero di Betti $b^1(M_{\Gamma(N)})$ è dato da

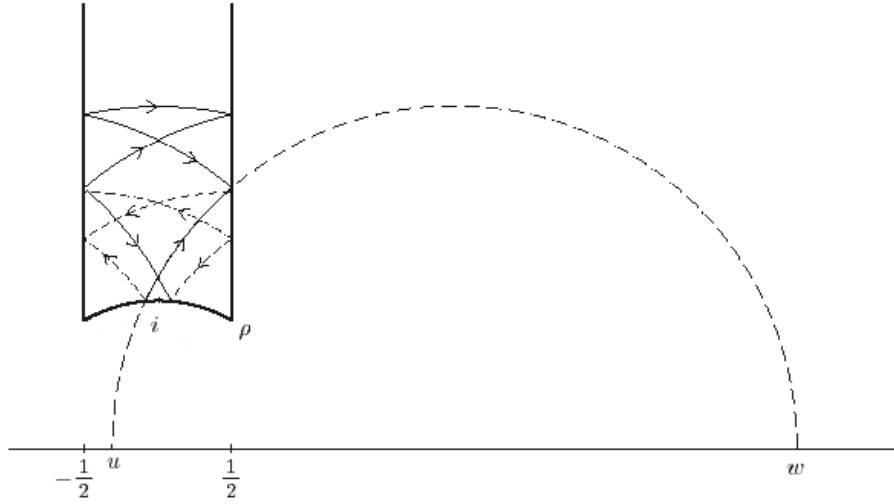
$$(3.12) \quad b^1 = \frac{N+6}{12N} \mu_N$$

Ricordiamo che b^1 fornisce la dimensione del primo gruppo di coomologia $H^1(M_\Gamma, \mathbb{R})$ i cui elementi sono le classi di 1-forme chiuse su M_Γ che differiscono tra loro solo per un differenziale.

4. IL FLUSSO GEODETICO SU $\Gamma \backslash \mathbb{H}$

Sia Γ un gruppo Fuchsiano del primo tipo e $M \equiv M_\Gamma$ la corrispondente varietà riemanniana iperbolica di volume finito (ma non necessariamente compatta) e curvatura negativa costante $\kappa = -1$. Se $\pi : \mathbb{H} \rightarrow M$ indica la proiezione sul quoziente si ha che se γ è una geodetica su \mathbb{H} allora $\pi(\gamma)$ è una geodetica su M e la mappa π induce il flusso geodetico sul fibrato tangente unitario SM di M , che indichiamo con il simbolo φ_t - in altre parole $\varphi_t := \pi_* g_t \pi_*^{-1}$ dove $\pi_* : S\mathbb{H} \rightarrow SM$ è il sollevamento di π ai fibrati tangenti. In modo analogo, essendo gli orocicli preservati dalle isometrie, otteniamo i flussi orociclici $\psi_t^+ := \pi_* h_t^+ \pi_*^{-1}$ e $\psi_t^- := \pi_* h_t^- \pi_*^{-1}$ su SM .

⁹Il *genere* di una superficie connessa ed orientabile è il massimo numero di tagli lungo curve chiuse che si possono effettuare senza rendere la superficie disconnessa. In modo equivalente, è il numero di manici su di essa.



Geodetiche sulla superficie modulare $M_{\Gamma(1)}$

La misura di Liouville m , essendo invariante sotto l'azione di Γ , fornisce una misura finita su SM , che indichiamo ancora con m , normalizzata ad avere massa totale 1, ed invariante per i flussi $\varphi_t, \psi_t^+, \psi_t^-$.

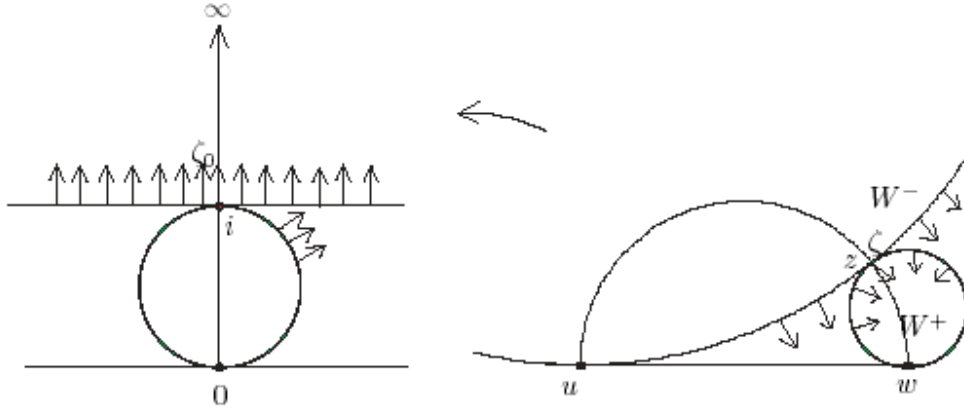
Tutte le proprietà discusse sopra, in particolare il Lemma 2.1 e le (2.11) si trasportano su SM . Più precisamente si ha il seguente

Teorema 4.1. *Il flusso geodetico φ_t è un flusso di Anosov (o iperbolico), cioè esiste una decomposizione C^∞ e φ_t -invariante $T(SM) = E^+ \oplus E^0 \oplus E^-$ tale che*

- (1) *Le curve integrali di E^0 sono le orbite del flusso geodetico.*
- (2) *Le proiezioni delle curve integrali di E^+ ed E^- su SM - chiamate varietà fortemente stabile e fortemente instabile, indicate con i simboli W^+ e W^- , rispettivamente - sono orocicli ortogonali alle orbite di φ_t .*
- (3) *Esistono due costanti positive C e λ tali che per ogni coppia di punti $v_1, v_2 \in SM$ giacenti sulla stessa foglia di W^+ (resp. W^-) si ha*

$$d^{W^\pm}(\varphi_t(v_1), \varphi_t(v_2)) \leq C e^{-\lambda|t|} d^{W^\pm}(v_1, v_2) \quad \text{per } t \geq 0 \quad (\text{resp. } t \leq 0)$$

Dimostrazione. Possiamo prima ragionare su $S\mathbb{H}$ e poi usare la proiezione π . Facendo riferimento alla parte destra della figura seguente, sia $z(t)$ l'unica geodetica che passa per $v = (z, \zeta)$ e tale che $z(-\infty) = u$ e $z(\infty) = w$. Possiamo poi trasportare $z(t)$ sull'asse immaginario per mezzo di una trasformazione $g \in G$ tale che $z = g(i)$ e fare i calcoli nella situazione della parte sinistra della figura. La varietà stabile $W^+(v)$ viene mappata nel campo vettoriale dei vettori unitari rivolti verso l'alto e normali all'orociclo $\mathbb{R} + i$, mentre $W^-(v)$ è mappato nel campo vettoriale dei vettori unitari normali esterni all'orociclo passante per i e 0 mostrato in figura.



Prendendo $v_1 = (i, \zeta_0)$ e $v_2 = (i + x, \zeta_0)$ si ha $d^{W^+}(v_1, v_2) = x$ e, in accordo con (2.9) si ha $d^{W^+}(g_t(v_1), g_t(v_2)) = x e^{-t}$ e dunque la stima vale con $C = \lambda = 1$ (cfr. (2.11)). La stima per la varietà instabile si ottiene invertendo la direzione del flusso. \square

La decomposizione descritta sopra è una caratteristica comune ai flussi geodetici su SH e SM . Vediamo ora alcune proprietà peculiari del flusso su SM .

Lemma 4.2. (*Anosov*) Supponiamo che $v \in SM$ sia tale che $d(v, \varphi_T(v)) < \epsilon$ per qualche $T > 0$ e $\epsilon > 0$. Allora esiste $v_0 \in SM$ la cui orbita è chiusa, cioè $\varphi_{T_0}(v_0) = v_0$ per qualche $T_0 > 0$, e tale che $d(v, v_0) < C\epsilon$, $|T - T_0| < C\epsilon$, e per $0 \leq t \leq T$ vale

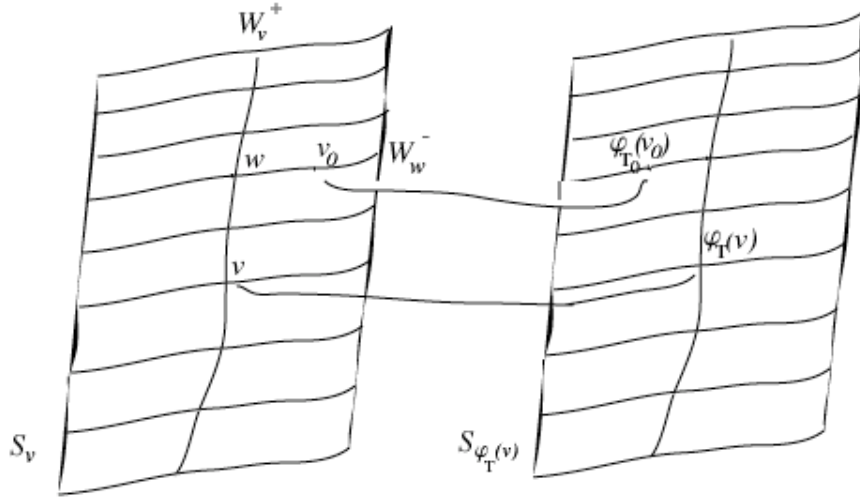
$$d(\varphi_t(v), \varphi_t(v_0)) \leq C'\epsilon e^{-\lambda \min\{t, T-t\}}$$

dove C, C' sono due costanti positive opportune.

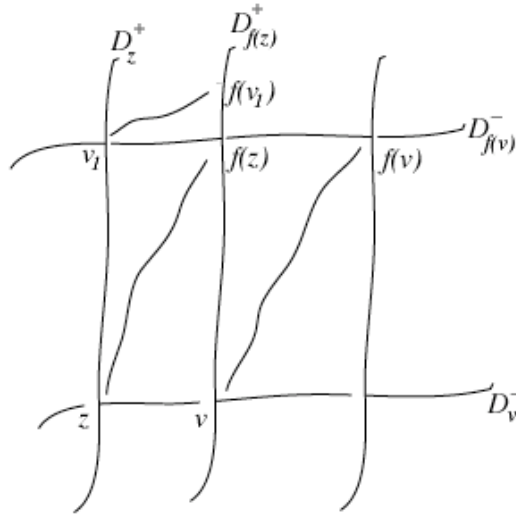
Dimostrazione. Introduciamo preliminarmente alcuni strumenti ausiliari.

Dato $v \in SM$ sia W_v^+ l'elemento della foliazione W^+ che contiene v . Indichiamo con D_v^+ l'insieme di tutti i punti $w \in W_v^+$ con $d^{W^+}(v, w) < \delta_0$, per qualche δ_0 che sceglieremo più avanti. Per ogni punto $w \in D_v^+$ indichiamo con W_w^- l'elemento della foliazione W^- che contiene w e con D_w^- l'insieme di tutti i punti $z \in W_w^-$ con $d^{W^-}(z, w) < \delta_0$.

Sia poi $S_v = \{z : z \in D_w^- \text{ per qualche } w \in D_v^+\}$. Per δ_0 abbastanza piccolo, S_v è una sotto-varietà di dimensione due trasversale all'orbita $\{\varphi_t(v)\}$. Ciò fornisce un modo conveniente di parametrizzare localmente SM per mezzo delle coordinate (s, t, u) in modo che s misuri la lunghezza lungo la foglia stabile D_v^+ , u lungo la foglia instabile D_w^- , e t lungo l'orbita di v .



Sia ora $S_v \ni v$ come sopra, dove scegliamo δ_0 in modo tale che se t_0 soddisfa $|t_0| < C_1\epsilon$ allora esiste $\gamma \in \Gamma$ tale che $g_{T+t_0}S_v \ni d\gamma(v)$. Consideriamo la mappa $f := (d\gamma)^{-1}g_{T+t_0} : S_v \rightarrow S_v$. Essendo $d\gamma$ un'isometria si ha $d(v, f(v)) = d(v, g_{T+t_0}(v)) < C_2\epsilon$. Sia $z \in D_v^-$ tale che $f(z) \in D_v^+$. Essendo $d^{W^-}(f(z), f(v)) \leq \delta_0$ per la terza proprietà del Teorema 4.1 si ha $d^{W^-}(z, v) \leq C_3\epsilon e^{-\lambda T}$. Inoltre, se prendiamo $v_1 \in D_{f(z)}^- \cap D_z^+$ si ha $f(v_1) \in D_{f(z)}^+$ e $d^{W^+}(f(v_1), f(z)) < C_4\epsilon e^{-\lambda T}$ e quindi $d(v_1, f(v_1)) < C_5\epsilon e^{-\lambda T}$. Iterando questa procedura arriviamo ad un punto fisso $v_0 = f(v_0)$ ovvero $(d\gamma)(v_0) = g_{T_0}(v_0)$ per qualche T_0 tale che $|T - T_0| < C_1\epsilon$.

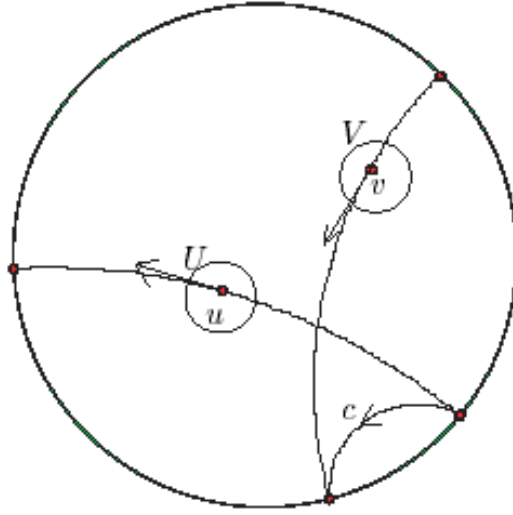


Si ha $d(v, v_0) < C\epsilon$. Per costruzione $v_0 \in S_v$. Sia $w \in D_v^+$ un punto tale che $v_0 \in D_w^-$. Allora si ha $d^{W^+}(v, w) < C_6\epsilon$ e $d^{W^-}(v_0, w) < C_7\epsilon$. Per la stessa ragione, essendo $d(\varphi_T(v), \varphi_{T_0}(v_0)) < C\epsilon$, possiamo concludere che per qualche t_1 con $|t_1| < C'_1\epsilon$ si abbia $\varphi_{T_0+t_1}(v_0) \in S_{\varphi_T(v)}$ con $d^{W^+}(\varphi_T(v), \varphi_T(v_0)) < C_8\epsilon$ e $d^{W^-}(\varphi_T(v_0), \varphi_T(w)) < C_8\epsilon$. Invocando

ancora la terza proprietà del Teorema 4.1 otteniamo che $d^{W^+}(\varphi_t(v), \varphi_t(w)) < C_9 e^{-\lambda t}$ e $d^{W^-}(\varphi_t(v_0), \varphi_t(w)) < C_9 e^{-\lambda(T-t)}$ e dunque $d(\varphi_t(v), \varphi_t(v_0)) < C_{10} e^{-\lambda \min\{t, T-t\}}$. \square

Teorema 4.3. *Sia $M = \Gamma \backslash \mathbb{H}$ con Γ Fuchsiano del primo tipo (e dunque $\mu(M) < \infty$). Allora il flusso geodetico $\{\varphi_t\}$ è topologicamente transitivo su SM , cioè esiste almeno un'orbita densa su SM .*

Dimostrazione. La proprietà di transitività topologica definita sopra è equivalente al fatto che per ogni coppia di aperti $U, V \in SM$ esista $t \in \mathbb{R}$ tale che $\varphi_t(U) \cap V \neq \emptyset$. Per mostrare ciò facciamo uso del modello \mathbb{D} del piano iperbolico.



Essendo la misura μ invariante per flusso geodetico e $\mu(M) < \infty$ possiamo applicare il teorema del ritorno di Poincaré¹⁰ per affermare che esistono due tempi T e T' tali che $\varphi_T(U) \cap U \neq \emptyset$ e $\varphi_{T'}(V) \cap V \neq \emptyset$. Per il Lemma 4.2 esistono allora due punti periodici $u \in U$ e $v \in V$ (i cui sollevamenti su $S\mathbb{D}$ indichiamo ancora con u e v). Siano c_u e c_v le geodetiche in \mathbb{D} tali che $\dot{c}_u = u$ e $\dot{c}_v = v$. Possiamo assumere che $c_u(-\infty) \neq c_v(\infty)$ (se così non fosse possiamo sostituire c_u con $\gamma(c_u)$ per qualche $\gamma \in \Gamma$). Consideriamo poi la geodetica c tale che $c(\infty) = c_v(\infty)$ e $c(-\infty) = c_u(-\infty)$. Per il Teorema 4.1, per ogni $t \in \mathbb{R}$ esistono t_u e t_v tali che $d(\dot{c}_u(t_u), \dot{c}(t)) < e^{-|t|}$ per $t \rightarrow -\infty$ e $d(\dot{c}_v(t_v), \dot{c}(t)) < e^{-t}$ per $t \rightarrow \infty$. Siccome c_u e c_v si proiettano su geodetiche chiuse su M ciò mostra che esistono t_1 e t_2 tali che la proiezione di $\dot{c}(t_1)$ su SM appartiene ad U e la proiezione di $\dot{c}(t_2)$ su SM appartiene ad V . \square

Mettendo insieme il Teorema 4.3 e il Lemma 4.2 otteniamo il

Corollario 4.4. *Le orbite periodiche del flusso geodetico sono dense in SM .*

¹⁰Teorema del ritorno di Poincaré (1890). Sia g una trasformazione continua e biunivoca che mandi un dominio limitato D in sé e che conservi il volume. Allora in ogni sottoregione $U \subset D$ di volume positivo esiste almeno un punto che ritorna in U .

Dimostrazione. Le immagini successive $U, g(U), g^2(U), \dots$ hanno tutte lo stesso volume e sono tutte contenute in D . Se fossero tutte disgiunte D avrebbe volume infinito, contrariamente all'ipotesi. Dunque esistono l e k con $m = k - l > 0$ tali che $g^l(U) \cap g^k(U) \neq \emptyset$. Applicando g^{-l} a questo insieme si vede che $U \cap g^m(U) \neq \emptyset$.

Il *Teorema ergodico di Birkhoff* per il flusso $\varphi_t : SM \rightarrow SM$ che conserva la misura di Liouville $dm = d\mu d\theta$ afferma che per ogni funzione $f \in L^1$ esiste una funzione $f^+ \in L^1$ tale che

$$(4.1) \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\varphi_t(v)) dt = f^+(v), \quad m - q.o.$$

Inoltre la media temporale a primo membro converge a f^+ anche in L^1 e $f^+ \circ \varphi_t = f^+$, m -q.o. Infine, se f^- indica il limite m -q.o. della media ergodica nel passato $\frac{1}{T} \int_{-T}^0 f(\varphi_t(v)) dt$ allora $f^+ = f^-$, m -q.o. Infine, diremo che il flusso è *ergodico* se f^+ è costante m -q.o. per ogni $f \in L^1$. Tale costante dovrà poi essere necessariamente uguale a $\int f dm$. In particolare ciò è equivalente al fatto che se $f \in L^1$ ha media nulla ed è φ_t -invariante, cioè $\int f dm = 0$ e $f = f \circ \varphi_t^+$, $\forall t \in \mathbb{R}$, allora $f = 0$ m -q.o. A sua volta ciò è equivalente al fatto che se $U \subset SM$ è un sottoinsieme misurabile e φ_t -invariante cioè tale che $\varphi_t(U) = U$, $\forall t \in \mathbb{R}$, allora $m(U) = 0$ oppure $m(SM \setminus U) = 0$.

Si ha il seguente

Teorema 4.5. *Il flusso geodetico φ_t su SM è ergodico rispetto alla misura di Liouville.*

Dimostrazione. Usiamo un classico argomento in dinamica iperbolica dovuto a Hopf, che sfrutta la struttura di decomposizione locale in varietà stabile, instabile e centrale. Innanzitutto, osserviamo che sarà sufficiente dimostrare che f^+ è costante m -q.o. per una funzione f continua a supporto compatto. Tali funzioni sono infatti dense in $L^1(SM, m)$. Sia allora f una funzione siffatta. Essendo uniformemente continua per ogni ϵ possiamo trovare un δ tale che $|f(v_1) - f(v_2)| < \epsilon$ ogni volta che v_1 e v_2 distano meno di δ . Se poi $se^{-t_0} < \delta$ allora $\varphi_t(v)$ e $\varphi_t \circ \psi_s^+(v) = \psi_{se^{-t}}^+ \circ \varphi_t(v)$ distano meno di δ se $t \geq t_0$. Pertanto se $\frac{1}{T} \int_0^T f(\varphi_t(v)) dt$ converge a $f^+(v)$ lo stesso fa $\frac{1}{T} \int_0^T f(\varphi_t(\varphi_{t_0}(v))) dt$, che dista meno di ϵ da $\frac{1}{T} \int_0^T f(\varphi_t(\varphi_{t_0} \circ \psi_s^+(v))) dt$. Da ciò si evince che anche la media temporale lungo la φ_t -orbita di $\psi_s^+(v)$ converge a $f^+(v)$. Dunque la convergenza a f^+ ha luogo sull'unione degli orocicli positivi ed f^+ è costante su ciascuno di essi.

La *varietà stabile* $W^S(v)$ è definita dall'unione delle φ_t -orbite dei punti sulla varietà fortemente stabile, ovvero dall'insieme $\{\varphi_t \circ \psi_s^+(v) : s, t \in \mathbb{R}\}$. In altre parole W^S è l'insieme delle curve integrali di $E^+ \oplus E^0$. Per quanto visto, la media temporale di f converge a f^+ sull'insieme dato dall'unione di tutte le varietà stabili (ed è costante su ciascuna di esse) e tale insieme ha misura piena. In modo simile, la media temporale di f nel passato converge a f^- sull'insieme (di misura piena) dato dall'unione delle *varietà instabili* $W^I(v) = \{\varphi_t \circ \psi_u^-(v) : t, u \in \mathbb{R}\}$ e inoltre f^- è costante su ciascuna di queste. Ricordiamo inoltre che $f^+ = f^-$, m -q.o.

Osserviamo che f^+ e f^- non potranno essere costanti ovunque, potendo ad esempio assumere valori diversi su diverse geodetiche chiuse. Dobbiamo però far vedere che assumono un unico comune valore su un insieme di misura piena. Per quanto abbiamo visto (cf. Teorema 4.1), ogni punto $v \in SM$ ha un intorno diffeomorficamente equivalente ad un intorno cubico di 0 in \mathbb{R}^3 via la mappa

$$(s, t, u) \mapsto \psi_s^+ \circ \varphi_t \circ \psi_u^-(v)$$

Ora, la misura m in un intorno cubico di v è equivalente alla misura di Lebesgue $ds dt du$, nel senso che hanno gli stessi insiemi di misura nulla. In termini di queste coordinate, nell'insieme di misura piena in cui f^+ e f^- esistono e sono uguali, vi sarà un insieme di misura $ds dt$ piena

in $W^S(\psi_u^-(v))$ per ogni u in un insieme U di misura du piena. Presi $u_1, u_2 \in U$, f^+ avrà un valore su $W^S(\psi_{u_1}^-(v))$ e uno su $W^S(\psi_{u_2}^-(v))$. Inoltre l'insieme delle coppie (s, t) su cui tali valori sono uguali a f^- sono entrambi di misura piena, dunque hanno intersezione non vuota e i due valori di f^+ in realtà coincidono. Pertanto f^+ è quasi ovunque costante nell'intorno di ogni $v \in SM$ ed essendo SM connesso deduciamo che è costante quasi ovunque. \square

5. IL FLUSSO OROICICLICO SU $\Gamma \backslash \mathbb{H}$

Analogo risultato vale per il flusso orociclico.

Teorema 5.1. *Il flusso orociclico ψ_t^+ su SM è ergodico rispetto alla misura di Liouville.*

Dimostrazione. Supponiamo che $f \in L^1$ abbia media nulla e sia invariante per il flusso orociclico, cioè $\int f dm = 0$ e $f = f \circ \psi_s^+$ per ogni $s \in \mathbb{R}$. Applicando la (2.14) si ha che

$$f = \frac{1}{T} \int_0^T f(\psi_s^+(v)) ds = \frac{1}{T} \int_0^T f(e_{\pi-r} \circ \varphi_t \circ e_{-r}(v)) ds$$

con $t(s) \rightarrow \infty$ e $r(s) \rightarrow 0$ quando $s \rightarrow \infty$. Siccome per t grande $e_{\pi-r} \circ \varphi_t \circ e_{-r}$ è molto vicino a $e_\pi \circ \varphi_t$, che è identico al flusso geodetico a parte il fatto che trasporta lungo le geodetiche i vettori 'rovesciati', usando l'ergodicità del flusso geodetico si deduce facilmente che $f = 0$ m-q.o. \square

Vediamo ora alcune proprietà più fini del flusso orociclico (vedi [Sa]).

Abbiamo visto che M non è compatta se Γ contiene elementi parabolici. In questo caso Γ ha un numero finito, diciamo r , di cuspidi non equivalenti in $\mathbb{R} \cup \{\infty\}$ che indichiamo con c_1, \dots, c_r . I corrispondenti gruppi di stabilità $\Gamma_{c_1}, \dots, \Gamma_{c_r}$ sono sottogruppi parabolici di Γ . Ad esempio per la cuspidi standard all'infinito si ha

$$(5.1) \quad \Gamma_\infty = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} : n \in \mathbb{Z} \right\}$$

Scegliamo dei numeri $-\lambda_i$, $i = 1, \dots, r$, tali che sotto la trasformazione $g_i \in PSL(2, \mathbb{R})$ definita da

$$(5.2) \quad g_i : z \mapsto \frac{-\lambda_i}{z - c_i}$$

c_i viene mandata nella cuspidi standard all'infinito. Se ad esempio $c_i = p/q$ allora si trova $\lambda_i = q^{-2}$. Vediamo adesso come ogni cuspidi dia luogo ad una famiglia ad un parametro di orbite periodiche per ψ_t^+ . Supponiamo innanzitutto che c_1 sia la cuspidi standard all'infinito e dunque $\Gamma_1 = \Gamma_\infty$. Dato $z = x + iy \in \mathbb{H}$ e $v = (z, 0) \in S\mathbb{H}$ si ha

$$(5.3) \quad h_{1/y}^+(v) = (z + 1, 0) = (z, 0) \pmod{\Gamma_\infty^*}$$

e dunque $\psi_{1/y}^+(v) = v$ in SM . Inoltre $1/y$ è il periodo primo di v rispetto a ψ_t^+ perchè se $\psi_t^+(v) = v$ allora $(z + ty, 0) = (z, 0) \pmod{\Gamma_\infty^*}$ che implica $ty \in \mathbb{Z}$. Pertanto, ogni punto di SM che giace sotto un punto della forma $(x + iy, 0)$ in $S\mathbb{H}$ (rispetto alla proiezione canonica π_*) è periodico di periodo $1/y$. Viceversa, tutti i punti in $S\mathbb{H}$ che giacciono sopra un punto periodico di periodo $1/y$ in SM hanno la stessa y . Sia C^+ il cilindro $[0, 1] \pmod{1} \times (0, \infty)$ e $h_1 : C^+ \rightarrow SM$ l'applicazione definita da $h_1(x, y) = \pi_*(x, y, 0)$. Tale applicazione è continua e biunivoca. Infatti se $h_1(x, y) = h_1(\xi, \eta)$ allora $g^*(x, y, 0) = (\xi, \eta, 0)$ per qualche $g \in \Gamma$ e se

$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ allora $-2 \arg(c(x + iy) + d) = 0$ e dunque $c = 0$ (cf. (3.10)). Sappiamo già che deve essere $y = \eta$ e dunque essendo per la (1.8) $\eta = y/|d|^2$ e $\xi + i\eta = (a(x + iy) + b)/d$ si trova $a = d = 1$ e $\xi = x + b$ con $b \in \mathbb{Z}$, perchè il gruppo di stabilità dell'infinito è $\Gamma_1 = \Gamma_\infty$. Dunque $(x, y) = (\xi, \eta)$ come elementi di C^+ . Dunque h_1 definisce un'immersione (liscia) di C^+ in SM . La sua immagine, che indichiamo con P_1 è una sottovarietà di SM che consiste unicamente di punti periodici per ψ_t^+ . La stessa cosa si può fare per le altre cuspidi c_i ripetendo l'argomento appena mostrato dopo aver riportato c_i alla cuspide standard all'infinito con la trasformazione g_i definita sopra (osserviamo che $g_i \notin \Gamma$ per la supposta inequivalenza delle cuspidi c_i). Per la stessa ragione le sottovarietà (cilindri immersi) P_1, \dots, P_r ottenuti in questo modo soddisfano $P_i \cap P_j = \emptyset$ per $i \neq j$.

Viceversa, se v è un punto fisso per ψ_t^+ allora $h_t^+(v) = v \pmod{\Gamma^*}$ e dunque esiste $g^* \in \Gamma^*$ tale che $g^*(h_t^+(v)) = v$ ovvero g lascia fisso l'orociclo positivo associato a v . Pertanto g deve essere parabolico ed è il generatore del gruppo di stabilità del punto di tangenza del suddetto orociclo. Tale punto di tangenza è dunque una cuspide per Γ .

Per quanto visto, un punto $v \in P_i$ corrisponde a tutti i $v' = \pi_*^{-1}(v)$ tali che il punto di tangenza dell'orociclo associato a v' è Γ -equivalente a c_i . Ma se Γ è un gruppo Fuchsiano del primo tipo tali punti formano un insieme denso in \mathbb{R} . Ad esempio i punti $SL(2, \mathbb{Z})$ -equivalenti al punto all'infinito sono tutti i razionali. Abbiamo dunque provato il seguente

Teorema 5.2. *Se P è l'insieme dei punti periodici di $\psi_t^+ : SM \rightarrow SM$ allora P è l'unione disgiunta di r sottovarietà bidimensionali P_1, \dots, P_r di SM . Ciascuna di esse è densa in SM .*

Dalla discussione fatta sopra segue anche che quando M è compatta l'insieme P è vuoto (non avendo Γ elementi parabolici). Il comportamento delle altre orbite di ψ_t^+ è descritto dal seguente risultato la prima parte del quale (densità) ottenuta da Hedlund nel 1936, la seconda (distribuzione uniforme) da Dani per il gruppo modulare nel 1982 e poi per ogni gruppo Fuchsiano da Dani e Smillie nel 1984 ([Da], [DS]).

Teorema 5.3. *Se $v \notin P$ allora l'orbita $\psi_{\mathbb{R}^+}^+(v) := \{\psi_t^+(v)\}_{t \in \mathbb{R}^+}$ è densa in SM . Più precisamente, se $v \notin P$ allora l'orbita $\psi_{\mathbb{R}^+}^+(v)$ è uniformemente distribuita rispetto a m , ovvero per ogni funzione $f : SM \rightarrow \mathbb{R}$ continua e limitata si ha*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\psi_t(v)) dt = \int f dm$$

In particolare se M è compatta il flusso $\psi_{\mathbb{R}}^+$ è *minimale* su SM .

6. ALTRI FLUSSI

Vediamo brevemente come alcuni dei comportamenti riscontrati per i flussi $\varphi_t, \psi_t^+, \psi_t^-$ su SM si ritrovino già in altri flussi definiti su spazi euclidei.

Il primo esempio è il *flusso irrazionale* $\phi_t : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ dove $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ è il toro bidimensionale. Dato $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ poniamo $\phi_t(x, y) = (x + t, y + \alpha t)$ le cui orbite in \mathbb{R}^2 sono rette parallele di pendenza α . Siccome ϕ_t commuta con l'azione di \mathbb{Z}^2 su \mathbb{R}^2 esso induce un flusso su \mathbb{T}^2 , indicato con lo stesso simbolo, che conserva la misura di Lebesgue su \mathbb{T}^2 . Su \mathbb{T}^2 ogni orbita

di ϕ_t è densa, come si comprende chiaramente osservando che le sue intersezioni con il cerchio $\{(x, y) \in \mathbb{T}^2 : x = 0\}$ formano la sequenza uniformemente distribuita $\{(0, q + m\alpha(\text{mod}1))\}$ per qualche $q \in \mathbb{R}$. Questo primo esempio ha dunque alcune caratteristiche in comune con i flussi orociclici ψ_t^+ e ψ_t^- su $SM \setminus P$.

Nel secondo esempio $\phi_t : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ è il flusso verticale dato da $\phi_t(x, y, z) = (x, y, z + t)$, il quale induce un flusso $\phi_t : N \rightarrow N$ sulla varietà compatta tridimensionale N ottenuta quotizzando \mathbb{R}^3 rispetto all'azione del gruppo generato dalle tre mappe che mandano (x, y, z) in $(x + 1, y, z)$, $(x, y + 1, z)$ e $(x + y, x + 2y, z - 1)$ rispettivamente. Questo flusso si chiama *flusso sospeso* perchè ottenuto dalla *sospensione* (con funzione di ritorno costante) della mappa $A : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ indotta dalla matrice $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Si tratta dunque del flusso nella direzione del

secondo fattore su $N = \mathbb{T}^2 \times [0, 1]/(x, y, 1) \sim (x + y, x + 2y, 0)$. Le orbite di $\phi_t : N \rightarrow N$ si possono comprendere studiando le loro intersezioni successive con $\mathbb{T}^2 \times \{0\}$, le quali formano un'orbita della mappa $A : \mathbb{T}^2 \rightarrow \mathbb{T}^2$, ovvero $A(x, y) = (x + y, x + 2y) \pmod{1}$.

Come il flusso geodetico $\varphi_t : SM \rightarrow SM$, anche $\phi_t : N \rightarrow N$ è un flusso di Anosov, ergodico rispetto alla misura di Lebesgue ϕ_t -invariante su N . Il sottospazio lineare $E^0 = \text{Span}(0, 0, 1)$ di TN genera le orbite di ϕ_t , mentre i sottospazi E^+ e E^- corrispondono agli autospazi contraente ed espandente di A , ovvero $E^\pm = \text{Span}(1, \lambda^{\pm 1} - 1, 0)$ dove $\lambda = (3 - \sqrt{5})/2$ è l'autovalore < 1 di A . Le varietà (fortemente) stabili e instabili sono dunque le strisce $W^\pm = \Lambda^\pm \times [0, 1]$ dove Λ^\pm sono le rette $x = s, y = (\lambda^{\pm 1} - 1)s, s \in \mathbb{R}$. Due punti qualsivoglia su W^+ (risp. W^-) si avvicinano (risp. si allontanano) con velocità esponenziale sotto l'azione di ϕ_t . Essendo poi $\lambda^{\pm 1} - 1 \in \mathbb{R} \setminus \mathbb{Q}$ tali varietà sono dense in N . Osserviamo a questo punto che nel caso del flusso geodetico su SM la sola possibilità per ottenere una discrepanza tra medie temporali f^+ e f^- e la media spaziale $\int f dm$ rispetto alla misura di Liouville era quella di calcolare le medie temporali lungo geodetiche periodiche. Mostriamo ora che in questo caso invece si può ottenere tale discrepanza anche lungo un'orbita densa in N . Tale orbita è in questo caso proprio quella di un punto generico di W^+ . Infatti se $\xi = (x, y, z) \in W^+$ allora $\phi_t(x, y, z) = (\lambda^n x, \lambda^n y, z + t) \pmod{1}$ dove $n = [t]$. E quindi $\lim_{t \rightarrow \infty} \phi_t(x, y, z) = \{0, 0\} \times [0, 1]$. Se ora prendiamo una funzione (analitica) della forma $f(x, y, z) = e^{2\pi i x}$ allora si ha

$$\frac{1}{T} \int_0^T f(\phi_t(\xi)) dt = \frac{1}{T} \sum_{n=0}^{[T]-1} e^{2\pi i \lambda^n x} \rightarrow 1, \quad T \rightarrow \infty$$

e d'altra parte

$$\int_N f(x, y, z) dx dy dz = \int_0^1 e^{2\pi i x} dx = 0$$

Vediamo infine alcune caratteristiche delle orbite periodiche. Osserviamo innanzitutto che se (x, y) è un punto periodico di periodo n della mappa $A : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ allora ogni punto $(x, y, z) \in N$ fa parte di un'orbita ϕ_t -periodica di periodo n . D'altra parte si ha che

Lemma 6.1. *L'insieme $\text{Per}_n A$ dei punti periodici di periodo n della mappa $A : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ costituisce un reticolo regolare su \mathbb{T}^2 e il loro numero $\#\text{Per}_n A$ è dato da*

$$\#\text{Per}_n A = \lambda^{-n} + \lambda^n - 2$$

Dimostrazione. Consideriamo la sequenza numerica (u_n) data da

$$u_n = \frac{\lambda^{-n} - \lambda^n}{\sqrt{5}}$$

che inizia con $0, 1, 3, 8, 21, 55, 144, 377, 987, 2584, 6765, 17711, \dots$. Usando l'equazione caratteristica $\det(A - \lambda I) = \lambda^2 - 3\lambda + 1 = 0$ si vede che i suoi termini soddisfano la formula ricorsiva

$$u_0 = 0, u_1 = 1, u_n = 3u_{n-1} - u_{n-2}, \quad n > 1$$

Inoltre, si verifica facilmente per induzione che

$$A^n = \begin{pmatrix} u_n - u_{n-1} & u_n \\ u_n & 2u_n - u_{n-1} \end{pmatrix}$$

L'asserto segue ora risolvendo l'equazione lineare $(A^n - I)\eta = k$ con $\eta \in \mathbb{T}^2$ e $k \in \mathbb{Z}^2$, con l'ausilio della formula scritta sopra, ed osservando che il numero dei vertici del reticolo è uguale al reciproco dell'area di una sua cella elementare, cioè $|\det(A^n - I)| = (\lambda^{-n} - 1)(1 - \lambda^n)$ (per ulteriori dettagli vedi [Is1]). \square

Vediamo dunque come anche per questo flusso le orbite periodiche formino un insieme denso nello spazio delle fasi.

7. LA SUPERFICIE MODULARE: FRAZIONI CONTINUE, SEZIONI TRASVERSALI E MAPPE

Consideriamo ora la superficie modulare $M = PSL(2, \mathbb{Z}) \backslash \mathbb{H}$ e ricordiamo innanzitutto che il gruppo $PSL(2, \mathbb{Z})$ ha come generatori $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ e $R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Vediamo innanzitutto in che modo due punti $u, w \in \mathbb{R}$ possono essere mandati l'uno nell'altro usando S e R . A questo scopo ricordiamo che ogni numero $x \in \mathbb{R}$ ammette un unico sviluppo in frazione continua della forma

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\ddots}}}}} \equiv [a_0; a_1, a_2, a_3, \dots]$$

con $a_0 \in \mathbb{Z}$ e $a_i \geq 1$. Se poi $x \in \mathbb{Q}$ allora lo sviluppo si arresta, cioè $x = [a_0; a_1, \dots, a_n]$ per qualche $n \geq 1$. Tale sviluppo è unico se scegliamo $a_n > 1$.

Osserviamo ora che se $x = [a_0; a_1, a_2, \dots]$ allora

$$R^{-a_0}(x) = [0; a_1, a_2, \dots] \quad \text{e} \quad SR^{-a_0}(x) = -[a_1; a_2, a_3, \dots]$$

e poi

$$R^{a_1}SR^{-a_0}(x) = -[0; a_2, a_3, \dots] \quad \text{e} \quad SR^{a_1}SR^{-a_0}(x) = [a_2; a_3, a_4, \dots]$$

Consideriamo adesso due punti $u, w \in \mathbb{R}$ espressi dagli sviluppi $u = [n_0; n_1, n_2, n_3, \dots]$ e $w = [m_0; m_1, m_2, m_3, \dots]$. Supponiamo poi che u e w abbiano la stessa 'coda', cioè che esistano $p, q \in \mathbb{N}$ tali che $n_{p+r} = m_{q+r}, \forall r > 0$. Se $u, v \in \mathbb{Q}$ allora hanno sempre la stessa coda, nel senso che $n_{p+r} = m_{q+r} = \infty$ per qualche $p, q \in \mathbb{N}$ e $r > 0$. Allora per quanto visto sopra, la trasformazione $g \in \Gamma$ data da

$$(7.1) \quad g = R^{m_0} S R^{-m_1} S \dots R^{(-)^q m_q} R^{(-)^{p+1} n_p} S \dots R^{n_1} S R^{-n_0}$$

manda u e w l'uno nell'altro, ossia $w = g(u)$. Se ad esempio

$$w = -\frac{3}{2} = -2 + \frac{1}{2} = [-2; 2] \quad , \quad u = \frac{4}{7} = \frac{1}{1 + \frac{1}{1 + \frac{1}{3}}} = [0; 1, 1, 3]$$

allora si trova $w = g(u)$ con

$$g = R^{-2}SR^{-2}R^3SR^{-1}SRS = \begin{pmatrix} -8 & 5 \\ 3 & -2 \end{pmatrix}$$

Ovviamente è vero anche il viceversa: se $w = g(u)$ per qualche $g \in PSL(2, \mathbb{Z})$ allora u e w devono avere la stessa coda.

7.1. Mappe di primo ritorno. Costruiamo adesso un sottoinsieme di SM che ogni φ_t -orbita incontra infinite volte. A questo scopo indichiamo con I l'asse immaginario su \mathbb{H} , cioè

$$I = \{z = x + iy : x = 0, y \in \mathbb{R}^+\} \subset \mathbb{H}$$

e consideriamo la sezione C fatta da tutte le π_* -proiezioni dei vettori di $S\mathbb{H}$ con base su I ed orientati verso destra, cioè i vettori della forma $v = (z, \theta)$ con $z \in I$ e $\theta \in (\pi, 2\pi)$. Osserviamo che gli elementi in tal modo selezionati sono tutti distinti. Infatti, se v ha base in $I \cap S(\mathbb{F})$ ¹¹ allora $S^*(v)$ ha base in $I \cap \mathbb{F}$ ed è orientato verso sinistra. C'è tuttavia una difficoltà: vi sono orbite di φ_t che non visitano C infinite volte. Tali orbite sono precisamente le proiezioni delle geodetiche che iniziano o terminano in una cuspide di $PSL(2, \mathbb{Z})$, cioè in un punto razionale e pertanto su SM convergono verso (o provengono da) la cuspide all'infinito e prendono il nome di *geodetiche di scattering*. Nella costruzione che faremo non saremo dunque in grado di trattare tali orbite, che tuttavia formano un insieme di misura nulla (vedi più avanti). Ora, come già visto, un vettore $v \in S\mathbb{H}$ la cui proiezione sta in C può essere semplicemente descritto dalle due coordinate asintotiche u e w che individuano la geodetica $\gamma = \gamma(v, t)$. In altre parole possiamo scrivere semplicemente

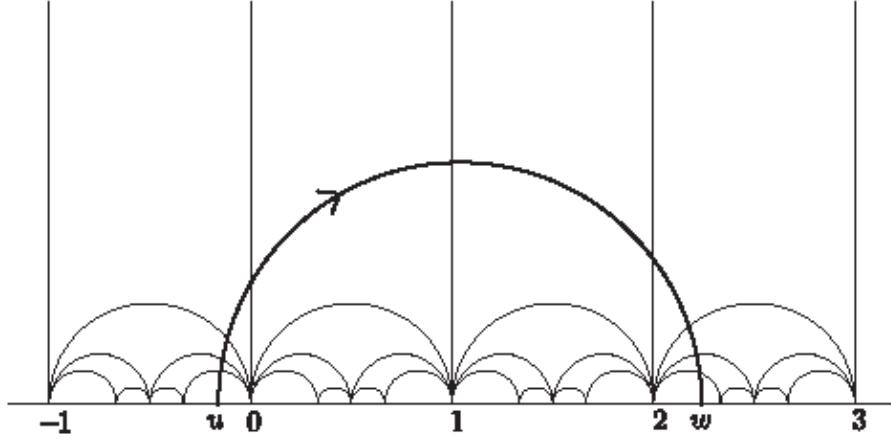
$$C := \{(u, w) : u < 0 < w\}$$

A sua volta C può essere decomposto come $C = C_1 \cup C_2$ dove

$$C_1 = \{(u, w) : u < 0 < w < 1\} \quad , \quad C_2 = \{(u, w) : u < 0, w > 1\}$$

Nella figura seguente è mostrata una geodetica γ tale che la proiezione su SM di $\gamma \cap I$ appartiene a C_2 .

¹¹Ricordando che $\mathbb{F} = \{z = x + iy : |x| < 1/2, |z| > 1\}$ è il dominio fondamentale si ha $S(\mathbb{F}) = \{z \in \mathbb{C} : |z + 1| > 1, |z - 1| > 1, |z| < 1\}$



Vogliamo ora costruire la *mappa di primo ritorno* $T_C : C \rightarrow C$ che manda ogni intersezione di una φ_t -orbita con C nella successiva. A questo scopo consideriamo il triangolo geodetico \mathbb{G} di vertici $0, 1$ e ∞ , cioè

$$(7.2) \quad \mathbb{G} = \left\{ z \in \mathbb{H} \mid 0 < \Re z < 1, \left| z - \frac{1}{2} \right| > \frac{1}{2} \right\}$$

I suoi tre lati sono equivalenti rispetto a $PSL(2, \mathbb{Z})$: i lati $0\hat{1}$ e $1\hat{\infty}$ vengono mandati in I rispettivamente dalle trasformazioni $SRS : z \mapsto z/(1-z)$ e $R^{-1} : z \mapsto z-1$. Sia di nuovo γ la geodetica determinata da $v \in S\mathbb{H}$.

- Se (la proiezione di) v è in C_2 allora γ lascia \mathbb{G} attraverso $1\hat{\infty}$. Se (u, w, s) sono le coordinate del vettore tangente unitario nel punto di uscita da \mathbb{G} allora γ è identificata da R^{-1} con la geodetica che entra in \mathbb{G} attraversando I con vettore tangente unitario $(u-1, w-1, s)$.
- Se (la proiezione di) v è in C_1 allora γ lascia \mathbb{G} attraverso $0\hat{1}$ con vettore tangente unitario (u, w, s) . Ma γ è identificata da SRS con la geodetica che entra in \mathbb{G} attraversando I con vettore tangente unitario $(\frac{u}{1-u}, \frac{w}{1-w}, s)$.

La mappa di primo ritorno su $C = C_1 \cup C_2$ è dunque

$$(7.3) \quad T_C : (u, w) \mapsto \begin{cases} \left(\frac{u}{1-u}, \frac{w}{1-w} \right) & \text{su } C_1 \\ (u-1, w-1) & \text{su } C_2 \end{cases}$$

La misura invariante per la mappa T_C si ottiene dalla misura di Liouville nella forma (2.15) semplicemente elidendo la variabile s :

$$(7.4) \quad dm_C = \frac{du dw}{(u-w)^2}$$

L'azione di T_C sulla seconda coordinata corrisponde alla mappa (detta *mappa fattore*) $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ data da

$$(7.5) \quad F : x \mapsto \begin{cases} \frac{x}{1-x}, & 0 \leq x \leq 1 \\ x-1, & x > 1 \end{cases}$$

la cui misura invariante ν si ottiene da m_C per semplice integrazione

$$(7.6) \quad \nu(dx) = h(x)dx \quad , \quad h(x) = \int_{-\infty}^0 \frac{du}{(u-x)^2} = \frac{1}{x} \quad , \quad x > 0$$

Si può anche verificare direttamente che h è la densità di una misura assolutamente continua ed F -invariante. Infatti l'invarianza di ν si esprime con il fatto che $\nu = \nu F^{-1}$ dove νF^{-1} è la misura che assegna ad ogni insieme misurabile $A \subset \mathbb{R}^+$ il numero $\nu(F^{-1}(A))$. D'altra parte, se $\nu(dx) = h(x)dx$ ciò significa che

$$h(x) = \sum_{y:F(y)=x} \frac{h(y)}{|F'(y)|} = \frac{1}{(1+x)^2} h\left(\frac{x}{1+x}\right) + h(x+1) = \frac{1}{x(1+x)} + \frac{1}{1+x} = \frac{1}{x}$$

Osserviamo che si tratta di una misura infinita: $h \notin L^1(\mathbb{R}^+, dx)$!

Un altro modo di riguardare l'azione di F è attraverso lo sviluppo in frazione continua: è infatti immediato verificare che se $x = [a_0; a_1, a_2, \dots]$ allora

$$(7.7) \quad F : [a_0; a_1, a_2, \dots] \mapsto \begin{cases} [0; a_1 - 1, a_2, \dots] , & a_0 = 0 \\ [a_0 - 1; a_1, a_2, \dots] , & a_0 > 0 \end{cases}$$

dove nel primo caso, se $a_1 = 1$ si deve intendere $[0; a_1 - 1, a_2, \dots] = [a_2; a_3, a_4, \dots]$.

Osservazione 7.1. Consideriamo la trasformazione $Z := SR^{-1}$. Dalla (3.6) si ha che $Z^3 = I$. Pertanto $\Gamma_0 := \Gamma(1)/\langle Z \rangle$ è un sottogruppo di $\Gamma(1)$ di indice 3. È inoltre facile verificare che \mathbb{G} è un dominio fondamentale per Γ_0 . Infatti, indicando con \mathbb{F}' il quadrilatero con vertici in $i, i+1, \infty$ e $(1+i\sqrt{3})/2$, cioè il dominio fondamentale per $\Gamma(1)$ ottenuto da \mathbb{F} applicando R alla sua metà di sinistra, si vede facilmente che $\mathbb{G} = \mathbb{F}' \cup Z(\mathbb{F}') \cup Z^2(\mathbb{F}')$.

7.2. La funzione di Minkowski. Dato $x \in \mathbb{R}^+$ con sviluppo in frazione continua $x = [a_0; a_1, a_2, \dots]$, poniamo

$$(7.8) \quad p(x) := 0 . \underbrace{11\dots 1}_{a_0} \underbrace{00\dots 0}_{a_1} \underbrace{11\dots 1}_{a_2} \dots$$

o, in modo equivalente,

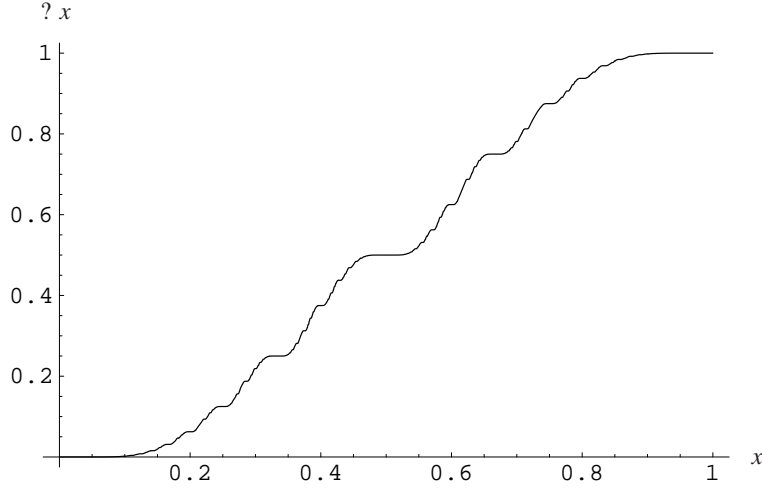
$$p(x) = 1 - \sum_{k \geq 0} (-1)^k 2^{-(a_0 + \dots + a_k)}$$

Ad esempio $p(1/n) = 1/2^n$ e $p(n) = 1 - 1/2^n$ per ogni $n \geq 1$. Poniamo inoltre $p(0) = 0$ e $p(\infty) = 1$.

In figura riportiamo la funzione $p(x/(1-x)) : [0, 1] \rightarrow [0, 1]$, usualmente ricordata come funzione $?$ di Minkowski. La sua azione su $x = [0; a_1, a_2, \dots]$ è

$$(7.9) \quad ?(x) = 0 . \underbrace{00\dots 0}_{a_1-1} \underbrace{11\dots 1}_{a_2} \underbrace{00\dots 0}_{a_3} \dots$$

Si tratta di una funzione singolare (non derivabile q.o.) ma Hölderiana. Inoltre $x \in \mathbb{Q}$ se e solo se $?(x)$ è un razionale diadico, mentre x è irrazionale quadratico se e solo se $?(x)$ è razionale (non-diadico).



Lemma 7.2. $p(x)$ e $p(1-x)$ soddisfano le equazioni funzionali

$$(7.10) \quad p(x) + p(1-x) = 1 \quad , \quad x \in [0, 1]$$

$$(7.11) \quad p(x) + p(1/x) = 1 \quad , \quad x \in \mathbb{R}^+$$

Dimostrazione. Se $x \in [1/2, 1]$ poniamo $x = 1/(y+1)$ con $y \in [0, 1]$ e $1-x = y/(y+1) \in [0, 1/2]$. Posto $y = [a_1, a_2, \dots]$ si ha $x = [1, a_1, a_2, \dots]$ e $1-x = [1 + a_1, a_2, \dots]$. Il primo asserto segue ora applicando direttamente la definizione. Per il secondo basta usare la relazione $p(x) = p(x/(1+x))$ \square

Teorema 7.3. ([BI]) La funzione p di Minkowski stabilisce una coniugazione topologica tra mappa $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ e la mappa $D : [0, 1] \rightarrow [0, 1]$, $D(x) = 2x \pmod{1}$.

Dimostrazione. Sia $\phi : \mathbb{R}^+ \cup \{\infty\} \rightarrow [0, 1]$ la mappa $\phi(x) = x/(x+1)$. Allora non è difficile verificare che $F = \phi^{-1} \circ \tilde{F} \circ \phi$ dove $\tilde{F} : [0, 1] \rightarrow [0, 1]$ è la mappa di Farey modificata :

$$(7.12) \quad \tilde{F}(x) = \begin{cases} \frac{x}{1-x} & \text{se } 0 \leq x < \frac{1}{2} \\ 2 - \frac{1}{x} & \text{se } \frac{1}{2} \leq x \leq 1 \end{cases}$$

Si ha che

$$\tilde{F} : [0; a_1, a_2, \dots] \mapsto \begin{cases} [0; a_1 - 1, a_2, \dots] , & a_1 > 1 \\ 1 - [0; a_2, a_3, \dots] , & a_1 = 1 \end{cases}$$

D'altra parte, usando le identità

$$\frac{1}{a + \frac{1}{b}} + \frac{1}{1 + \frac{1}{a-1+\frac{1}{b}}} = 1 \quad \text{e} \quad \frac{1}{1 + \frac{1}{b+c}} + \frac{1}{1 + b + c} = 1$$

si vede che

$$x = [0; a_1, a_2, \dots] \iff 1-x = \begin{cases} [0; 1 + a_2, a_3, \dots] & \text{se } a_1 = 1 \\ [0; 1, a_1 - 1, a_2, \dots] & \text{se } a_1 > 1 \end{cases}$$

Infine D agisce come una traslazione sullo sviluppo binario. Dunque $\tilde{F} = \phi^{-1} \circ D \circ \phi$ e l'asserto segue mettendo tutto insieme ed osservando che $p = \phi \circ \tilde{F} \circ \phi^{-1}$. \square

Corollario 7.4. *La mappa $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ è caotica, cioè è topologicamente transitiva, le sue orbite periodiche sono dense ed ha dipendenza sensibile dalle condizioni iniziali.*

Da quanto visto, segue anche che la misura $dp(x)$ è F -invariante (mentre $d?$ è \tilde{F} -invariante), e la sua entropia è $\log 2$, cioè uguale all'entropia topologica. In altre parole, $dp(x)$ è la *misura di massima entropia* per F ($d?$ quella per \tilde{F}). Si tratta ovviamente di misure singolari rispetto a Lebesgue. Si può mostrare che ad esempio $d?$ è concentrata su un insieme $X \subset [0, 1]$ con dimensione di Hausdorff $\alpha \approx 0.875$. Sappiamo d'altronde che F e \tilde{F} hanno anche le misure assolutamente continue (non normalizzabili) con densità $1/x$ e $1/x(1-x)$, rispettivamente.

7.3. La mappa di Gauss. La mappa indotta da F su $[0, 1]$, definita da

$$(7.13) \quad \hat{F}(x) = F^n(x), \quad x \in [0, 1], \quad n = \min\{k : F^k(x) \in [0, 1]\}$$

è data da $\hat{F}(x) = \frac{x}{1-x} - \left[\frac{x}{1-x} \right] = \frac{1}{1-x} - \left[\frac{1}{1-x} \right]$ per $x \in [0, 1)$ e possiamo estenderla ponendo $\hat{F}(1) = 0$. La sua misura invariante è evidentemente dx/x . Osserviamo che la mappa

$$(7.14) \quad G(x) = \hat{F}(1-x) = \frac{1}{x} - \left[\frac{1}{x} \right], \quad x \neq 0, \quad G(0) = 0$$

è la celebre *mappa di Gauss*, che opera come una traslazione sullo sviluppo in frazione continua:

$$(7.15) \quad G : [0; a_1, a_2, \dots] \mapsto [0; a_2, a_3, \dots]$$

e dunque, per ogni $x \in [0, 1]$ della forma $x = [0; a_1, a_2, \dots]$ si ha

$$(7.16) \quad x = [0; a_1, a_2, \dots, a_n + G^n(x)] \quad , \quad n \geq 1$$

In altre parole la mappa G può essere usata per generare lo sviluppo in frazione continua. Ad esempio per il numero π si trova

$$\begin{aligned} \pi - 3 &= 0,141592\dots, & \left[\frac{1}{\pi - 3} \right] &= 7 \\ G(\pi - 3) &= 0,062513\dots, & \left[\frac{1}{G(\pi - 3)} \right] &= 15 \\ G^2(\pi - 3) &= 0,996594\dots, & \left[\frac{1}{G^2(\pi - 3)} \right] &= 1 \\ G^3(\pi - 3) &= 0,003417\dots, & \left[\frac{1}{G^3(\pi - 3)} \right] &= 292 \\ G^4(\pi - 3) &= 0,634591\dots, & \left[\frac{1}{G^4(\pi - 3)} \right] &= 1 \end{aligned}$$

e dunque

$$(7.17) \quad \pi = [3; 7, 15, 1, 292, 1, \dots]$$

Si può anche estendere l'azione di G a \mathbb{R}^+ ma la semiretta $x > 1$ viene mappata da G tutta nell'intervallo unitario che a sua volta è G -invariante. In particolare, se $x = [a_0; a_1, a_2, \dots]$ con $a_0 \geq 1$ allora $G^2(x) = [0; a_1, a_2, \dots]$. Dunque sarà sufficiente considerare l'azione di G su $[0, 1]$. La misura G -invariante è dunque tutta concentrata su $[0, 1]$ ed ha la forma $dx/(1+x)$

come si vede anche per calcolo diretto: ogni punto $x_0 \in [0, 1]$ ha infinite immagini inverse, l' n -esima delle quali è $x_n = 1/(n + x_0)$, $n \geq 1$. Pertanto

$$\sum_{n \geq 1} \frac{|dx_n|}{1 + x_n} = \sum_{n \geq 1} \frac{|dx_0|}{(n + x_0)(1 + n + x_0)} = \sum_{n \geq 1} \left(\frac{1}{n + x_0} - \frac{1}{1 + n + x_0} \right) |dx_0| = \frac{|dx_0|}{1 + x_0}$$

Normalizzata con il fattore $1/\log 2$ tale misura diviene una misura di probabilità, detta *misura di Gauss*, che indicheremo con $d\rho$.

7.4. Frazioni continue: convergenti veloci e convergenti di Farey. Richiamiamo ora alcuni fatti della teoria delle frazioni continue.

Dato $x \in \mathbb{R}^+$ poniamo

$$\begin{aligned} p_{-2} &= 0, & p_{-1} &= 1, & p_n &= a_n p_{n-1} + p_{n-2} \\ q_{-2} &= 1, & q_{-1} &= 0, & q_n &= a_n q_{n-1} + q_{n-2} \end{aligned}$$

Da quest'ultima si vede subito che i q_n crescono almeno esponenzialmente:

$$(7.18) \quad q_n \geq 2^{(n-1)/2}$$

Osserviamo poi che $q_n p_{n-1} - q_{n-1} p_n = -(q_{n-1} p_{n-2} - q_{n-2} p_{n-1})$ ed essendo $q_0 p_{-1} - q_{-1} p_0 = 1$ si ha la *formula di Lagrange*:

$$(7.19) \quad q_n p_{n-1} - q_{n-1} p_n = (-1)^n, \quad n \geq 0$$

Osserviamo che

$$\frac{p_0}{q_0} = a_0, \quad \frac{p_1}{q_1} = \frac{a_1 a_0 + 1}{a_1} = a_0 + \frac{1}{a_1}, \quad \frac{p_2}{q_2} = \frac{a_2(a_1 a_0 + 1) + a_0}{a_2 a_1 + 1} = a_0 + \frac{1}{a_1 + \frac{1}{a_2}}$$

e più in generale

$$(7.20) \quad \frac{p_n}{q_n} = [a_0; a_1, a_2, \dots, a_n], \quad n \geq 0$$

Chiameremo i rapporti p_n/q_n *convergenti veloci* di x . Osserviamo anche che per $n \geq 1$ si ha

$$\frac{q_{n-1}}{q_n} = \frac{1}{a_n + \frac{1}{\frac{q_{n-2}}{q_{n-1}}}} = \frac{1}{a_n + \frac{1}{a_{n-1} + \frac{1}{\frac{q_{n-3}}{q_{n-2}}}}}$$

e così via. Pertanto

$$(7.21) \quad \frac{q_{n-1}}{q_n} = [0; a_n, a_{n-1}, a_{n-2}, \dots, a_1].$$

La frazione p_n/q_n fornisce la migliore approssimazione razionale di x il cui denominatore non superi q_n . In effetti, dalle formule ricorsive scritte sopra segue la formula più generale

$$(7.22) \quad \left[a_0; a_1, a_2, \dots, a_n + \frac{1}{r} \right] = \frac{r p_n + p_{n-1}}{r q_n + q_{n-1}}, \quad n \geq 0, \quad r \geq 1$$

e inoltre dalla formula di Lagrange si ottiene

$$(7.23) \quad \frac{r p_n + p_{n-1}}{r q_n + q_{n-1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(r q_n + q_{n-1})}$$

che ponendo $r = a_{n+1}$ dà

$$(7.24) \quad \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n q_{n+1}}$$

Ora, è evidente che, in virtù della (7.22) e della (7.16), se $x \in [0, 1]$ allora si può scrivere¹²

$$(7.25) \quad x = \frac{(G^n(x))^{-1} p_n + p_{n-1}}{(G^n(x))^{-1} q_n + q_{n-1}}$$

Osservando che $[(G^n(x))^{-1}] = a_{n+1}$ e dunque $a_{n+1} \leq (G^n(x))^{-1} < a_{n+1} + 1$, ed usando la (7.23) e le formule ricorsive per p_n e q_n otteniamo l'importante stima

$$(7.26) \quad \frac{1}{q_{n+1} + q_n} < (-1)^n (q_n x - p_n) < \frac{1}{q_{n+1}}$$

da cui si evince che la successione p_n/q_n approssima x alternatamente da sopra e da sotto, cioè $p_{2n}/q_{2n} < x < p_{2n+1}/q_{2n+1}$. In modo equivalente la successione a termini positivi

$$(7.27) \quad f_n := (-1)^n (q_n x - p_n)$$

è strettamente decrescente. Che cosa succede se nella (7.22) si prendono tutti i valori intermedi $r = 1, \dots, a_{n+1}$? Otteniamo in tal caso i cosiddetti *convergenti di Farey* [AO]:

$$(7.28) \quad \frac{t_{n+1,r}}{s_{n+1,r}} := \frac{r p_n + p_{n-1}}{r q_n + q_{n-1}} \quad , \quad r = 1, \dots, a_{n+1}$$

Dalla (7.23) si vede che $t_{n,r}/s_{n,r}$ fornisce la migliore approssimazione razionale *unilatera* di x il cui denominatore non superi $r q_{n-1} + q_{n-2}$, nel senso che se $r < a_n$ ci può essere un convergente con denominatore più piccolo e più vicino ad x , ma dall'altra parte di x .

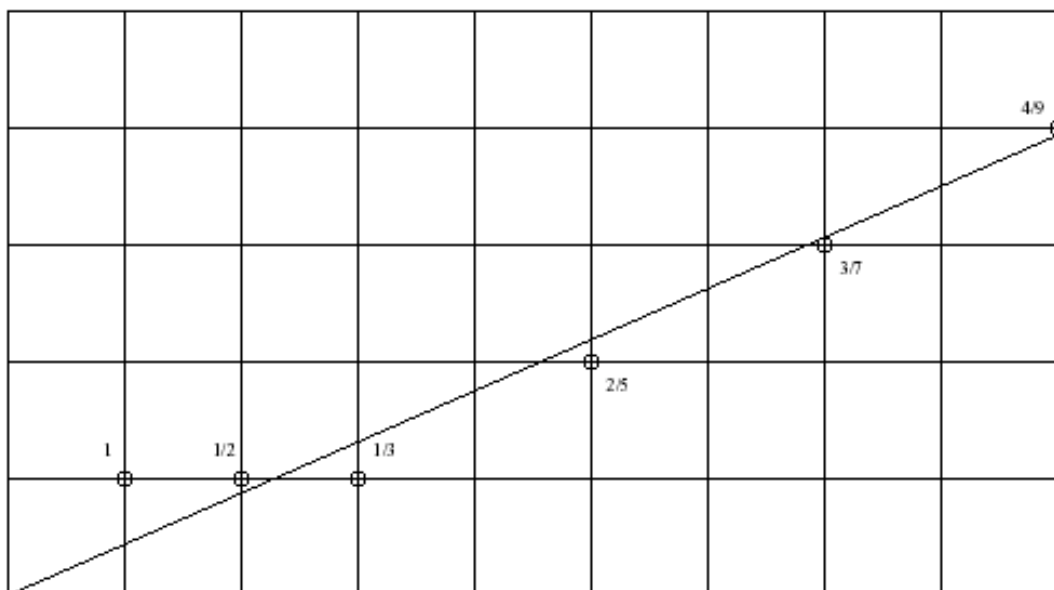
ESEMPIO. Sia $x = [0; 2, 3, 2, 4, 2, 5, \dots]$ I primi quattro convergenti veloci sono

$$\begin{aligned} n = 1 & \quad \frac{p_1}{q_1} = \frac{1}{2} \\ n = 2 & \quad \frac{p_2}{q_2} = \frac{1}{2 + \frac{1}{3}} = \frac{3}{7} \\ n = 3 & \quad \frac{p_3}{q_3} = \frac{1}{2 + \frac{1}{3 + \frac{1}{2}}} = \frac{7}{16} \\ n = 4 & \quad \frac{p_4}{q_4} = \frac{1}{2 + \frac{1}{3 + \frac{1}{2 + \frac{1}{4}}}} = \frac{31}{71} \end{aligned}$$

¹²Se $x > 1$ vale una formula analoga con $G^{n+2}(x)$ al posto di $G^n(x)$, ma le conclusioni che trarremo sono identiche.

Entro lo stesso grado di accuratezza vi sono $2 + 3 + 2 + 4 = 11$ convergenti di Farey:

$$\begin{array}{ll}
 n = 1, r = 1, & \frac{t_{1,1}}{s_{1,1}} = \frac{p_1 + p_0}{q_1 + q_0} = 1 \\
 n = 1, r = 2, & \frac{t_{1,2}}{s_{1,2}} = \frac{2p_1 + p_0}{2q_1 + q_0} = \frac{1}{2} \\
 n = 2, r = 1, & \frac{t_{2,1}}{s_{2,1}} = \frac{p_2 + p_1}{q_2 + q_1} = \frac{1}{3} \\
 n = 2, r = 2, & \frac{t_{2,2}}{s_{2,2}} = \frac{2p_2 + p_1}{2q_2 + q_1} = \frac{2}{5} \\
 n = 2, r = 3, & \frac{t_{2,3}}{s_{2,3}} = \frac{3p_2 + p_1}{3q_2 + q_1} = \frac{3}{7} \\
 n = 3, r = 1, & \frac{t_{3,1}}{s_{3,1}} = \frac{p_3 + p_2}{q_3 + q_2} = \frac{4}{9} \\
 n = 3, r = 2, & \frac{t_{3,2}}{s_{3,2}} = \frac{2p_3 + p_2}{2q_3 + q_2} = \frac{7}{16} \\
 n = 4, r = 1, & \frac{t_{4,1}}{s_{4,1}} = \frac{p_4 + p_3}{q_4 + q_3} = \frac{10}{23} \\
 n = 4, r = 2, & \frac{t_{4,2}}{s_{4,2}} = \frac{2p_4 + p_3}{2q_4 + q_3} = \frac{17}{39} \\
 n = 4, r = 3, & \frac{t_{4,3}}{s_{4,3}} = \frac{3p_4 + p_3}{3q_4 + q_3} = \frac{24}{55} \\
 n = 4, r = 4, & \frac{t_{4,4}}{s_{4,4}} = \frac{4p_4 + p_3}{4q_4 + q_3} = \frac{31}{71}
 \end{array}$$



7.5. Ergodicità e mescolamento. Ricordiamo le definizioni di ergodicità e di mescolamento per una trasformazione T che conserva la misura μ di uno spazio di probabilità (X, \mathcal{F}, μ) . Il sistema dinamico (X, \mathcal{F}, μ, T) è *ergodico* se per ogni $F \in \mathcal{F}$ invariante, cioè tale che $T^{-1}F = F$, si ha $\mu(F) = 0$ oppure $\mu(F^c) = 0$. Un altro modo di caratterizzare l'ergodicità è dato dalla seguente

Proposizione 7.5. *Il sistema dinamico (X, \mathcal{F}, μ, T) è ergodico se e solo se per ogni $E, F \in \mathcal{F}$ si ha*

$$(7.29) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(E \cap T^{-k}F) = \mu(E)\mu(F)$$

Dimostrazione. L'ergodicità implica che $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \rightarrow \int_X f d\mu$, μ -q.o. Se integriamo questa relazione su E con $f = 1_F$ si ottiene la (7.29), osservando che $f \circ T^k = 1_{T^{-k}F}$. Viceversa, supponiamo che $F = T^{-1}F$ e poniamo $E = F^c$. Allora la (7.29) dà subito $\mu(F)\mu(F^c) = 0$, ossia l'ergodicità. \square

L'ergodicità è dunque equivalente alla convergenza secondo Cesàro della successione $\mu(E \cap T^{-k}F)$. Se tale successione converge a zero termine a termine, ovvero se per ogni $E, F \in \mathcal{F}$ vale

$$(7.30) \quad \lim_{n \rightarrow \infty} \mu(E \cap T^{-n}F) = \mu(E)\mu(F)$$

allora il sistema dinamico (X, \mathcal{F}, μ, T) si dice *mescolante*. È evidente che il mescolamento implica l'ergodicità.

Infine, dati n e k_1, k_2, \dots, k_n interi positivi definiamo un *intervallo fondamentale di rango n* , indicato con $I_n(k_1, k_2, \dots, k_n)$, come il sottointervallo di $[0, 1]$ dato da $I_n = \{x \in [0, 1] : a_1 = k_1, \dots, a_n = k_n\}$. Usando le (7.18) e (7.24) si vede che tali intervalli hanno lunghezza $|I_n| \leq C 2^{-n}$. Possiamo ora dimostrare il seguente risultato

Teorema 7.6. *Il sistema dinamico $([0, 1], G, \rho)$ è ergodico e mescolante.*

Dimostrazione. Adottiamo per semplicità la notazione $[a_1, a_2, \dots] \equiv [0; a_1, a_2, \dots]$. Fissati n e k_1, k_2, \dots, k_n si ha che $x \in I_n$ se e solo se $x = [k_1, \dots, k_n + r_n(x)]$ per qualche $0 \leq r_n(x) \leq 1$. In particolare $|I_n| = r_n^{-1}(1) - r_n^{-1}(0)$. Sia $A = [\xi, \eta]$ con $0 \leq \xi \leq \eta \leq 1$. Allora

$$|I_n \cap G^{-n}A| = r_n^{-1}(\eta) - r_n^{-1}(\xi)$$

D'altra parte, per quanto visto sopra si ha

$$r_n^{-1}(t) = \frac{p_n + tp_{n-1}}{q_n + tq_{n-1}}$$

Pertanto si ha

$$\begin{aligned}
\frac{|I_n \cap G^{-n}A|}{|I_n|} &= \frac{r_n^{-1}(\eta) - r_n^{-1}(\xi)}{r_n^{-1}(1) - r_n^{-1}(0)} \\
&= \frac{\frac{p_n + \eta p_{n-1}}{q_n + \eta q_{n-1}} - \frac{p_n + \xi p_{n-1}}{q_n + \xi q_{n-1}}}{\frac{p_n + p_{n-1}}{q_n + q_{n-1}} - \frac{p_n}{q_n}} \\
&= |A| \frac{q_n(q_n + q_{n-1})}{(q_n + \eta q_{n-1})(q_n + \xi q_{n-1})} \asymp |A|
\end{aligned}$$

dove abbiamo usato la notazione $a \asymp b$ per dire che esistono due costanti positive c_1 e c_2 tali che $c_1 b \leq a \leq c_2 b$. D'altra parte, essendo la misura di Gauss comparabile rispetto a quella di Lebesgue, abbiamo che per ogni $A \in \mathcal{F}$

$$(7.31) \quad \frac{\rho(I_n \cap G^{-n}A)}{\rho(I_n)} \asymp \rho(A)$$

Supponiamo ora che A sia G -invariante, cioè che $G^{-1}A = A$ con $\rho(A) > 0$. Allora

$$(7.32) \quad \frac{\rho(I_n \cap A)}{\rho(A)} \asymp \rho(I_n) \quad , \quad \forall n.$$

Essendo $\{I_n\}$ una famiglia generante¹³ ciò implica

$$\frac{\rho(A \cap B)}{\rho(A)} \asymp \rho(B)$$

per ogni B boreliano. Se $B = A^c$ allora $\rho(B) = 0$ e dunque A ha misura piena. Da ciò segue l'ergodicità.

Ma possiamo dire qualcosa in più. Indichiamo con \mathcal{G}_n la σ -algebra generata da insiemi della forma $\{x : a_k(x) = a\}$ con $k \geq n$ e chiamiamo $\mathcal{G}_\infty = \bigcap_{n \geq 1} \mathcal{G}_n$ la σ -algebra all'infinito (i suoi elementi dipendono solo dal 'futuro infinitamente lontano'). Mostriamo innanzitutto che ogni insieme invariante appartiene necessariamente a \mathcal{G}_∞ . Infatti, se $A \in \mathcal{F}$ è G -invariante allora, essendo $A = G^{-n}A \in \mathcal{G}_n$ per ogni n e dunque $A \in \mathcal{G}_\infty$. Tuttavia in \mathcal{G}_∞ ci sono anche insiemi che non sono invarianti (ad es. l'insieme degli x tali che $a_n(x) = 1$ per infiniti n), e dunque in linea di principio \mathcal{G}_∞ può contenere un insieme A con $0 < \rho(A) < 1$, anche se G è ergodica. Sia dunque $A \in \mathcal{G}_\infty$ con $\rho(A) > 0$. Per ogni n , quindi, A è del tipo $A = G^{-n}B$ per qualche $B \in \mathcal{F}$. D'altra parte, la (7.31) dà

$$\rho(A) = \rho(G^{-n}B) = \rho(B) \asymp \frac{\rho(I_n \cap G^{-n}B)}{\rho(I_n)} = \frac{\rho(I_n \cap A)}{\rho(I_n)}$$

per ogni intervallo fondamentale I_n di rango n . Infine, essendo $\rho(A) > 0$, l'ultima identità dà nuovamente la (7.32) e come sopra possiamo concludere che $\rho(A) = 1$. Abbiamo dunque mostrato che la σ -algebra all'infinito è banale, nel senso che contiene solo insiemi di misura nulla o di misura piena. Tale condizione evidentemente implica l'ergodicità. Usando la nozione di aspettazione condizionata si mostra in effetti che essa implica il mescolamento (vedi [Bil], p.121). \square

¹³Cioè che genera la σ -algebra dei boreliani \mathcal{F} . Ciò è equivalente al fatto che ogni sequenza monotona di intervalli converge ad un punto.

OSSERVAZIONE. Gauss in una lettera a Laplace afferma (senza dimostrarlo) che

$$\lim_{n \rightarrow \infty} |\{x \in [0, 1] : G^n(x) < t\}| = \frac{\log(1+t)}{\log 2} = \rho([0, t])$$

per ogni $0 < t \leq 1$, e si domanda quale possa essere l'errore. Se poniamo $A = [0, t)$ possiamo riscrivere questa relazione nella forma

$$\lim_{n \rightarrow \infty} \int_{G^{-n}A} (\log 2)(1+x)\rho(dx) = \rho(A)$$

Ora, la proprietà di mescolamento implica che

$$\lim_{n \rightarrow \infty} \int_{G^{-n}A} f(x)\rho(dx) = \rho(A) \int_0^1 f(x)\rho(dx)$$

per ogni funzione indicatrice f e ogni $A \in \mathcal{F}$. Essendo la funzione $(\log 2)(1+x)$ approssimabile con funzioni semplici con precisione arbitraria, si vede che l'affermazione di Gauss segue dal fatto che G è mescolante.

Un'indagine più accurata, mostra poi che il mescolamento vale in modo *uniforme ed esponenziale*. Ciò significa che se n e m sono interi positivi e $A \in \mathcal{F}$ si ha

$$(7.33) \quad \rho(I_n \cap G^{-n-m}A) = \rho(I_n)\rho(A)(1 + O(q^m))$$

uniformemente in n, m, A e I_n , dove q è un numero in $(0, 1)$. Pertanto anche l'errore indagato da Gauss è esponenzialmente ed uniformemente piccolo.

7.6. Alcune conseguenze sulle frazioni continue. Vediamo innanzitutto due semplici applicazioni dell'ergodicità della misura di Gauss, mostrando che le medie geometrica ed armonica dei quozienti parziali sono ben definite e calcolabili quasi ovunque rispetto a ρ .

Proposizione 7.7. *La funzione $\log a_1(x)$ appartiene a $L_1(\rho)$ e si ha*

$$(7.34) \quad \lim_{n \rightarrow \infty} \left(\prod_{k=1}^n a_k \right)^{\frac{1}{n}} = e^{\rho(\log a_1)} = e^{K_1} \simeq 2,68554 \quad \rho - \text{q.o.}$$

dove la costante K_1 , detta costante di Khinchin, è data da

$$(7.35) \quad e^{K_1} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k(k+2)} \right)^{\log_2 k}$$

Dimostrazione. Essendo ρ una misura di probabilità G -invariante, la ρ -probabilità che un quoziente parziale qualsiasi $a_i(x)$ sia uguale all'intero k non dipende da i ed è data da

$$(7.36) \quad p(k) := \rho(\{x \in [0, 1] : a_i(x) = k\}) = \int_{1/(k+1)}^{1/k} \rho(dx) = \log_2 \left(1 + \frac{1}{k(k+2)} \right)$$

Si ha

$$\rho(\log a_1) = \sum_{k=1}^{\infty} \log k \cdot p(k) = \sum_{k=1}^{\infty} \log k \cdot \log_2 \left(1 + \frac{1}{k(k+2)} \right) = K_1 < \infty$$

Questo calcolo mostra allo stesso tempo che $\log a_1 \in L_1(\rho)$ e l'ultima identità in (7.37). La prima (7.37) segue dall'osservazione che $a_k(x) = a_1(G^{k-1}(x))$ insieme al teorema ergodico e il teorema 7.6. \square

Proposizione 7.8. *La funzione $1/a_1(x)$ appartiene a $L_1(\rho)$ e si ha*

$$(7.37) \quad \lim_{n \rightarrow \infty} \frac{n}{\sum_{k=1}^n \frac{1}{a_k}} = \frac{1}{\rho(1/a_1)} = \frac{1}{K_2} \simeq 1,74540 \quad \rho - \text{q.o.}$$

dove la costante K_2 è data da

$$(7.38) \quad K_2 = \sum_{k=1}^{\infty} \frac{1}{k} \cdot \log_2 \left(1 + \frac{1}{k(k+2)} \right)$$

Dimostrazione. Come sopra, notando che

$$\rho(1/a_1) = \sum_{k=1}^{\infty} \frac{p(k)}{k} = \sum_{k=1}^{\infty} \frac{1}{k} \cdot \log_2 \left(1 + \frac{1}{k(k+2)} \right) = K_2 < \infty \quad \square$$

Che ne è della media aritmetica dei quozienti parziali? Innanzitutto osserviamo che

$$\rho(a_1) = \sum_{k=1}^{\infty} k \cdot p(k) = \sum_{k=1}^{\infty} k \cdot \log_2 \left(1 + \frac{1}{k(k+2)} \right) = \infty$$

La media aritmetica dei quozienti parziali riflette questo risultato in accordo con il teorema ergodico (senza tuttavia che i quozienti parziali stessi appartengano a $L_1(\rho)$). Più precisamente vale il seguente risultato di Khinchin.

Proposizione 7.9. *Posto $S_n := \sum_{i=1}^n a_i$ si ha*

$$(7.39) \quad \lim_{n \rightarrow \infty} \frac{S_n}{n} = \infty \quad \rho - \text{q.o.}$$

e

$$(7.40) \quad \lim_{n \rightarrow \infty} \frac{S_n}{n \log_2 n} = 1 \quad \text{in probabilità}$$

Dimostrazione. L'ergodicità di $([0, 1], G, \rho)$ implica che per ρ -quasi ogni $x = [a_1, a_2, \dots]$ e per ogni $\epsilon > 0$, si può trovare un n_k tale che per ogni $n \geq n_k$ la frequenza con cui il numero k si ripete tra i primi n quozienti parziali di x soddisfa

$$(7.41) \quad \left| \frac{\#\{a_i = k; 1 \leq i \leq n\}}{n} - p(k) \right| < \epsilon$$

Sia ora $N \geq 1$ un intero e poniamo $\tilde{n} = \max(n_1, \dots, n_{N-1})$. Sia poi $x^{(N)} = [b_1, b_2, \dots]$ definito da $b_i = a_i$ se $a_i \leq N$ e $b_i = N$ se $a_i > N$. La sequenza (b_i) prende valori in $\{1, \dots, N\}$ e indicando con $S_n^{(N)}$ la somma dei primi n quozienti b_i si ha evidentemente $S_n \geq S_n^{(N)}$ per ogni n . Se x soddisfa la (7.41) allora per ogni $n \geq n_k$ si avrà

$$(7.42) \quad \left| \frac{\#\{b_i = k; 1 \leq i \leq n\}}{n} - p^{(N)}(k) \right| < \epsilon$$

dove $p^{(N)}(k) = p(k)$ per ogni $k < N$, mentre

$$p^{(N)}(N) = \sum_{k \geq N} p(k) = \log_2 \left(1 + \frac{1}{N} \right)$$

Pertanto, se $n \geq \max(\tilde{n}, n_N)$, si può scrivere

$$\begin{aligned} \frac{S_n^{(N)}}{n} &\geq \sum_{k=1}^{N-1} k(p(k) - \epsilon) + N(p^{(N)}(N) - \epsilon) \\ &= \log_2 \prod_{k=1}^{N-1} \left(\frac{k+1}{k} \cdot \frac{k+1}{k+2} \right)^k \left(\frac{N+1}{N} \right)^N - \epsilon \frac{N(N+1)}{2} \\ &= \log_2(N+1) - \epsilon \frac{N(N+1)}{2} \end{aligned}$$

Infine, se per un dato intero positivo N scegliamo $\epsilon = 2/N(N+1)$, possiamo trovare $\hat{n}(N)$ tale che per ogni $n \geq \hat{n}$

$$\frac{S_n}{n} \geq \frac{S_n^{(N)}}{n} \geq \log_2(N+1) - 1$$

e ciò implica il primo asserto. Per la dimostrazione del secondo asserto rimandiamo a [Kh2] (vedi anche [CI]). \square

Osservazione 7.10. Possiamo chiederci se (7.40) vale anche in senso forte, cioè se

$$\rho \left(\lim_{n \rightarrow \infty} \frac{S_n}{b_n} = 1 \right) = 1$$

con $b_n = n \log_2 n$. Mostriamo che così non è e che in realtà

$$(7.43) \quad \rho \left(\lim_{n \rightarrow \infty} \frac{S_n}{b_n} = 1 \right) = 0$$

In effetti, essendo $\rho(a_n > N) \geq C/N$, per ogni numero $\ell > 1$ e per n abbastanza grande si ha

$$\rho(a_n > \ell b_n) \geq \frac{C}{\ell b_n}$$

Quindi

$$\sum_{n \geq 1} \rho(a_n > \ell b_n) = \infty$$

Usando l'estensione del lemma di Borel-Cantelli ad eventi dipendenti si vede che

$$\rho \left(\frac{a_n}{b_n} > \ell \text{ infinitamente spesso} \right) = 1$$

e a maggior ragione

$$\rho \left(\frac{S_n}{b_n} > \ell \text{ infinitamente spesso} \right) = 1$$

Infine

$$\rho \left(\limsup_{n \rightarrow \infty} \frac{S_n}{b_n} = \infty \right) = 1$$

che implica la (7.43). Ma possiamo dire di più: dal momento che (7.40) implica la convergenza quasi ovunque su una sottosuccessione, (7.43) è valida *per ogni* sequenza di costanti b_n .

Il seguente risultato di Diamond and Vaaler [DV] mostra che l'ostacolo alla convergenza quasi ovunque è dovuto all'occorrenza di un *singolo grande valore* di a_i .

Proposizione 7.11. Per quasi ogni $x \in [0, 1]$ esiste $n_0 = n_0(x)$ t.c. per ogni $n \geq n_0$

$$S_n = n \log_2 n (1 + o(1)) + \vartheta \max_{1 \leq k \leq n} a_k(x)$$

con $\vartheta = \vartheta(n, x) \in [0, 1]$.

La dimostrazione di questo fatto si basa su un semplice ma interessante risultato.

Lemma 7.12. Sia $d > 1/2$. Sotto le ipotesi fatte sopra per quasi ogni $x \in [0, 1]$ si può trovare un numero $n_0 = n_0(x)$ t.c. $\forall n \geq n_0$ esiste al più un intero k tra $\{1, \dots, n\}$ tale che $a_k > n' \equiv n(\log n)^d$

Dimostrazione. Fissiamo $i < j$. Una forma debole della proprietà di mescolamento è

$$\begin{aligned} \rho(a_i > n'', a_j > n') &\asymp \rho(a_i > n') \cdot \rho(a_j > n') \\ &= (\rho(a_1 > n'))^2 \asymp (n')^{-2} \end{aligned}$$

Da ciò segue che la misura dell'insieme in cui $a_i > n'$ e $a_j > n'$ per qualche coppia di indici distinti $i, j \leq 2n$ è dell'ordine al più $(\log n)^{-2d}$. Per $K = 1, 2, \dots$ sia

$$U_K = \cup_{k \geq K} \{a_i > (2^k)', a_j > (2^k)', i, j \leq 2^{k+1}, i \neq j\}$$

Allora

$$\rho(U_K) \asymp \sum_{k \geq K} k^{-2d} \rightarrow 0 \quad \text{quando } K \rightarrow \infty$$

Infine, per $x \notin U_K$ e $n \geq 2^K$ vi è al più un indice $k \leq n$ t.c. $a_k > n'$. \square

Dimostrazione della Proposizione 7.11. Poniamo $S'_n(E, x) = \sum_{i \leq n} a'_i$ dove $a'_i = a_i$ se $a_i \leq n'$ and $a'_i = 0$ altrimenti, e inoltre

$$M'_n := \rho(S'_n) \quad , \quad V'_n := \rho((S'_n - M'_n)^2)$$

Si ha

$$M'_n = n\rho(a'_1) = n \sum_{\ell \leq n'} \ell \rho(a_1 = \ell) \asymp n \log n' \asymp n \log n$$

Inoltre $\rho((S'_n)^2) = \sum_{i,j=1}^n \rho_{ij}$ con

$$\begin{aligned} \rho_{ij} &:= \rho(a'_i a'_j) = \sum_{\ell, k \leq n'} \ell \cdot k \rho(\{a'_i = \ell, a'_j = k\}) \\ &= \sum_{\ell, k \leq n'} \ell \cdot k \rho(\{a_1 = \ell\}) \rho(\{a_1 = k\}) (1 + O(q^{j-i})) \\ &= (M'_n)^2 n^{-2} (1 + O(q^{j-i})) \end{aligned}$$

dove abbiamo usato (7.33). In particolare $\rho_{jj} = \sum_{\ell \leq n'} \ell^2 \rho(\{a_1 = \ell\}) \asymp n'$, e quindi

$$V'_n = \sum_{i,j=1}^n \rho_{ij} - (M'_n)^2 = (M'_n)^2 n^{-2} \sum_{i < j \leq n} O(q^{j-i}) + n n' \asymp (M'_n)^2 n^{-1} + n n' \sim n n'$$

Riassumendo abbiamo trovato che

$$M'_n \asymp n \log n \quad \text{e} \quad V'_n \asymp n^2 (\log n)^d$$

Siano $0 < a < 1$ e $b > 1$ due numeri da fissare più avanti e definiamo le sequenze $n_k := \exp(k^a)$ e k^{-b} . Per quanto visto si ha

$$\rho \left(\sum_{k \geq 1} \frac{(S'_{n_k} - M'_{n_k})^2}{n_k n'_k} k^{-b} \right) \asymp \sum_{k \geq 1} k^{-b} < \infty$$

e dunque

$$S'_{n_k} - M'_{n_k} = o \left(\frac{n_k n'_k}{k^{-b}} \right)^{1/2} \quad q.o.$$

D'altra parte si ha $n_k n'_k / k^{-b} = o(M_{n_k}^2)$ perchè $n_k n'_k / k^{-b} = (n_k \log n_k)^2 r_n$ con $r_n = (\log n_k)^{d-2} k^b = o(1)$ purchè $a(2-d) > b$. Quindi $S'_{n_k} = (1 + o(1))M'_{n_k}$ per quasi ogni $x \in [0, 1]$. È infine facile vedere che $M'_{n_{k-1}}/M'_{n_k} \sim 1$ quando $k \rightarrow \infty$ e per n grande

$$S'_n = (1 + o(1)) M'_n \quad q.o.$$

La tesi segue ora da quanto visto e dal Lemma 7.12, per cui $0 \leq S_n - S'_n \leq \max_{1 \leq i \leq N} \sigma_i$. \square

7.7. Crescita dei denominatori. Mostriamo ora che tipicamente i denominatori dei convergenti veloci crescono con velocità esponenziale mentre quelli dei convergenti di Farey crescono con velocità subesponenziale. Sia $\frac{t_m}{s_m} \equiv \frac{t_{n,r}}{s_{n,r}}$ con $m = \sum_{i=0}^{n-1} a_i + r$ l' m -esimo convergente di Farey.

Teorema 7.13.

$$(7.44) \quad \frac{\log q_n}{n} \rightarrow \frac{\pi^2}{12 \log 2} \quad \text{quasi ovunque}$$

mentre

$$(7.45) \quad \frac{\log s_m}{m} \sim \frac{\pi^2}{12 \log m} \quad \text{in misura}$$

Dimostrazione. Dalla (7.25) si ha

$$(7.46) \quad G^n(x) = -\frac{q_n x - p_n}{q_{n-1} x - p_{n-1}} = \frac{f_n}{f_{n-1}}$$

e dunque

$$(7.47) \quad f_n = \prod_{k=0}^n G^k(x)$$

Pertanto, l'ergodicità di G rispetto alla misura di Gauss ρ (e l'equivalenza di quest'ultima rispetto alla misura di Lebesgue) dà

$$(7.48) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log f_{n-1} = \int_0^1 \log x \rho(dx) = -\frac{\pi^2}{12 \log 2} \quad \text{quasi ovunque}$$

Inoltre per quanto visto sopra si ha

$$(7.49) \quad \frac{1}{a_{n+1} + 2} < \frac{q_n}{q_n + q_{n+1}} < q_n f_n < \frac{q_n}{q_{n+1}} < \frac{1}{a_{n+1}}$$

Se $\frac{p}{q}$ e $\frac{p'}{q'}$ sono frazioni consecutive ad un qualunque stadio della costruzione allora vale la relazione di unimodularità $qp' - pq' = 1$. Inoltre, ogni $x \in \mathbb{Q}^+$ compare esattamente una volta e corrisponde ad un unico cammino finito che parte dalla radice $\frac{1}{1}$ ed il cui numero d di vertici è la *profondità* di x , cioè la generazione a cui appartiene (poniamo uguale a zero la profondità degli antenati $\frac{0}{1}$ e $\frac{1}{0}$). Evidentemente vi sono 2^{d-1} elementi di profondità $d \geq 1$.

Lemma 7.14. *I 2^{d-1} elementi dell'albero di Stern-Brocot di profondità $d \geq 1$ sono tutti e soli quei numeri razionali $x \in \mathbb{Q}_+$ il cui sviluppo in frazione continua $x = [a_0; a_1, \dots, a_n]$ con $a_0 \geq 0$, $a_i \geq 1$ ($0 < i < n$) e $a_n > 1$, soddisfa*

$$\sum_{i=0}^n a_i = d$$

Dimostrazione. Per la radice $\frac{1}{1}$ è ovvio. Se la profondità di $x = [a_0; a_1, \dots, a_n]$ è $d > 1$ allora per raggiungere x da $\frac{1}{1}$ ci dobbiamo innanzitutto muovere a_0 passi verso destra, giungendo così sul nodo $a_0 + \frac{1}{1}$. Quindi, muovendosi a_1 passi verso sinistra giungiamo sul nodo $a_0 + \frac{1}{a_1 + \frac{1}{1}}$. Altri a_2 passi verso destra portano sul nodo $a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{1}}}$ e così via. Vediamo in questo modo che il cammino che raggiunge x compie esattamente n deviazioni e la lunghezza del blocco tra la $(i-1)$ -esima e la i -esima deviazione è dato dal quoziente parziale a_i per $1 \leq i < n$, mentre l'ultimo blocco ha lunghezza $a_n - 1$. D'altra parte, il numero di soluzioni dell'equazione $\sum_{i=0}^n a_i = d$ con $0 \leq n \leq d-1$, $a_0 \geq 0$ e $a_i \geq 1$, $1 \leq i < n$ e $a_n > 1$ è uguale al numero di soluzioni con $a_i \geq 1$, $0 \leq i \leq n$. Ora, per ogni $n \geq 0$ fissato, il numero di tali soluzioni è $\binom{d-1}{n}$. Infatti, ogni scelta di $a_0 \geq 1, \dots, a_n \geq 1$ con $a_0 + \dots + a_n = d$ è in corrispondenza biunivoca con la sequenza di n numeri

$$a_0 + \frac{1}{2}, a_0 + a_1 + \frac{1}{2}, \dots, a_0 + \dots + a_{n-1} + \frac{1}{2} = d - a_n - \frac{1}{2}$$

Tale sequenza può sua volta essere riguardata come un'estrazione di n elementi dalla sequenza di $d-1$ numeri

$$\frac{1}{2}, \frac{3}{2}, \dots, \frac{1}{2} + d - 1 = d - \frac{1}{2}$$

E vi sono appunto $\binom{d-1}{n}$ possibili estrazioni di n elementi da un insieme di $d-1$ elementi. Dunque il numero cercato è $\sum_{n=0}^{d-1} \binom{d-1}{n} = 2^{d-1}$. \square

ESEMPIO: $\frac{8}{5}$ ha profondità $d = 5$ e $\frac{8}{5} = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{2}}}}$

Questo argomento può essere sviluppato ulteriormente. A questo scopo costruiamo una rappresentazione matriciale dei razionali positivi: dato $z \in \mathbb{C}$ e $X = \begin{pmatrix} n & m \\ t & s \end{pmatrix} \in SL(2, \mathbb{Z})$ poniamo $X(z) := (nz + m)/(tz + s)$ e identifichiamo

$$(7.53) \quad X \iff X(1) = \frac{n+m}{t+s} \in \mathbb{Q}^+$$

Osservando che ogni $x \in \mathbb{Q}^+$ può essere decomposto in modo unico come

$$x = \frac{m}{s} \oplus \frac{n}{t} \quad \text{con} \quad ns - mt = 1$$

vediamo che le colonne della matrice $X \in SL(2, \mathbb{Z})$ che rappresenta x sono i genitori di x sull'albero di Stern-Brocot in ordine invertito (sulla colonna sinistra c'è quello più grande e su quella destra il più piccolo). In questo modo, la radice dell'albero non è altro che la matrice identità:

$$\frac{1}{1} = \frac{0}{1} \oplus \frac{1}{0} \iff \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Inoltre, se $X \in SL(2, \mathbb{Z})$ rappresenta la frazione $x \in \mathbb{Q}^+$, la matrice UMU rappresenta la frazione simmetrica $1/x$, dove

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

In particolare

$$(7.54) \quad \frac{1}{2} \iff \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} =: A \quad \text{e} \quad \frac{2}{1} \iff \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} =: B$$

e più in generale, per ogni $k \in \mathbb{N}$,

$$A^k = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix} \iff \frac{1}{k+1} \quad \text{e} \quad B^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \iff k+1$$

Ora, l'elemento x considerato sopra ha a sua volta un unico paio di figli (sinistro e destro), dati rispettivamente da

$$\frac{m}{s} \oplus \frac{m+n}{s+t} \quad \text{e} \quad \frac{m+n}{s+t} \oplus \frac{n}{t}$$

Inoltre,

$$\begin{pmatrix} n & m \\ t & s \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} m+n & m \\ s+t & s \end{pmatrix} \iff \frac{m}{s} \oplus \frac{m+n}{s+t}$$

e

$$\begin{pmatrix} n & m \\ t & s \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} n & m+n \\ t & s+t \end{pmatrix} \iff \frac{m+n}{s+t} \oplus \frac{n}{t}$$

Detto altrimenti, le matrici A e B , fatte agire da destra, muovono sull'albero in basso a sinistra e in basso a destra, rispettivamente.

Osservazione 7.15. Una rapida riflessione mostra che le matrici A e B si esprimono in termini dei generatori S e R di $PSL(2, \mathbb{Z})$ come $A = SR^{-1}S$ e $B = R$.

Tutto ciò insieme al Lemma precedente dà la seguente

Proposizione 7.16. *Ad ogni elemento x dell'albero di Stern-Brocot, con profondità $d > 1$, corrisponde un unico elemento $X \in SL(2, \mathbb{Z})$, per il quale vi sono due sole possibilità:*

- $x = [a_0; a_1, \dots, a_n]$, n pari $\implies X = B^{a_0} A^{a_1} \dots A^{a_{n-1}} B^{a_n-1}$
- $x = [a_0; a_1, \dots, a_n]$, n dispari $\implies X = B^{a_0} A^{a_1} B^{a_2} \dots A^{a_n-1}$

Corollario 7.17. *Sia $x = [a_0; a_1, \dots, a_n]$ con $a_n > 1$ e n pari. Allora i suoi figli sinistro e destro sono dati rispettivamente da $x' = [a_0; a_1, \dots, a_n - 1, 2]$ e $x'' = [a_0; a_1, \dots, a_n + 1]$. Se invece n è dispari gli sviluppi di x' e x'' sono scambiati.*

Dimostrazione. Per n pari e maggiore di uno si ha $x = [a_0; a_1, \dots, a_n] \iff X = B^{a_0} A^{a_1} B^{a_2} \dots B^{a_{n-1}}$. Quindi $x' \iff X' = B^{a_0} A^{a_1} B^{a_2} \dots B^{a_{n-1}} A$ e $x'' \iff X'' = B^{a_0} A^{a_1} B^{a_2} \dots B^{a_n}$. Analogamente si ragiona se $n = 0$ e n dispari. \square

ESERCIZIO. Verificare che se $\mathbb{Q}^+ \ni x \equiv \prod_{1 \leq i \leq k} X_i$, allora

$$X_i = \begin{cases} A, & \text{se } F^{i-1}(x) < 1 \\ B, & \text{se } F^{i-1}(x) \geq 1 \end{cases}$$

dove F è la mappa introdotta in (7.5).

Vediamo ora una maniera ‘dinamica’ di generare l’albero di Stern-Brocot. Dato $\frac{p}{q}$ chiamiamo le frazioni $\frac{p}{p+q}$ e $\frac{p+q}{q}$ i suoi *discendenti*.¹⁴ Osserviamo che i discendenti di x altro non sono che le pre-immagini $x/(x+1)$ e $x+1$ di x con la mappa F introdotta in (7.5), e possono essere ottenuti con l’azione modulare delle matrici A e B su x :

$$A(x) = \frac{x}{x+1}, \quad B(x) = x+1$$

Ciò implica tra l’altro che se $x \iff X$ in accordo con la Proposizione 7.16 allora $x = X(1)$. Osserviamo inoltre che se $x = [a_0; a_1, \dots, a_n]$ allora $A(x) = [0; 1, a_0, a_1, \dots, a_n]$ e $B(x) = [a_0 + 1; a_1, \dots, a_n]$. Dunque si ha

Lemma 7.18. *L’insieme di tutti i discendenti degli elementi di profondità d sono esattamente gli elementi di profondità $d+1$.*

Usando questo fatto si dimostra facilmente la seguente curiosa proprietà, che può essere interpretata nel quadro della teoria musicale [Is2].

Lemma 7.19. *Dato $p/q \in \mathbb{Q}^+$ definiamo la sua semplicità come $1/pq$. Allora la semplicità totale degli elementi di data profondità è sempre uguale a 1, ovvero*

$$\sum \frac{1}{pq} = 1$$

dove, per ogni assegnato $d \geq 1$, la somma è fatta su tutte le frazioni in $p/q \in \mathbb{Q}^+$ della forma $p/q = [a_0; a_1, \dots, a_n]$ con $\sum_{i=0}^n a_i = d$.

Dimostrazione. Per induzione in d . La proprietà è vera per $d = 1$. Usando il Lemma 7.18 basta osservare che la semplicità di p/q è uguale alla somma di quelle dei suoi discendenti $\frac{p}{p+q}$ e $\frac{p+q}{q}$. \square

Osservazione 7.20. Abbiamo visto che le matrici A e B , fatte agire da destra, muovono verso i figli sinistro e destro, rispettivamente. Se invece, applichiamo A e B da sinistra si ottiene

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} n & m \\ t & s \end{pmatrix} = \begin{pmatrix} n & m \\ n+t & m+s \end{pmatrix} \iff \frac{n+m}{n+m+t+s}$$

e

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} n & m \\ t & s \end{pmatrix} = \begin{pmatrix} n+t & m+s \\ t & s \end{pmatrix} \iff \frac{n+m+t+s}{s+t}$$

¹⁴Viceversa, se vogliamo risalire da una frazione p/q alla sua *antecedente* p'/q' (ovvero la frazione dalla quale discende) allora se $p > q$ si ha $q' = q$ e $p' = p - q$; se invece $q > p$ allora $p' = p$ e $q' = q - p$.

Quindi in questo modo le matrici A e B muovono verso i discendenti sinistro e destro, rispettivamente.

7.9.1. *La codifica infinita.* Si può estendere la costruzione appena illustrata associando ad ogni $x \in \mathbb{R}^+$ un unico cammino infinito sull'albero \mathcal{o} , in modo equivalente, un'unica parola semi-infinita $\pi(x) \in \{A, B\}^{\mathbb{N}}$, nel modo naturale. Innanzitutto, ad $x \in \mathbb{R}^+ \setminus \mathbb{Q}^+$ ha sviluppo (infinito) $x = [a_0; a_1, a_2, a_3, \dots]$ corrisponderà l'unica sequenza $\pi(x) = B^{a_0} A^{a_1} B^{a_2} A^{a_3} \dots$, dove adesso A e B vanno intesi come meri simboli di un alfabeto binario. Se invece x è razionale possiamo procedere come segue. Poniamo innanzitutto $\pi(\frac{0}{1}) = A^\infty$ e $\pi(\frac{1}{0}) = B^\infty$. Notiamo poi che ciascun $x \in \mathbb{Q}^+$ ammette esattamente due cammini infiniti che convergono al nodo x : sono quelli che iniziano con la sequenza finita necessaria a raggiungere x in accordo con la Proposizione 7.16 e terminano con BA^∞ o AB^∞ . Conveniamo poi che $\pi(x)$ termini con BA^∞ o AB^∞ a seconda che il numero dei quozienti parziali di x sia pari o dispari. Riassumendo si ha

- $x = [a_0; a_1, \dots, a_n]$, n pari $\implies \pi(x) = B^{a_0} A^{a_1} \dots B^{a_n} A^\infty$
- $x = [a_0; a_1, \dots, a_n]$, n dispari $\implies \pi(x) = B^{a_0} A^{a_1} \dots A^{a_n} B^\infty$
- $x = [a_0; a_1, a_2, a_3, \dots]$ $\implies \pi(x) = B^{a_0} A^{a_1} B^{a_2} A^{a_3} \dots$

Si verifica facilmente che se \succ denota l'ordine lessicografico su $\{A, B\}^{\mathbb{N}}$ allora

$$x > y \implies \pi(x) \succ \pi(y)$$

Infine, da quanto visto segue che se $x \in \mathbb{R}^+ \setminus \mathbb{Q}^+$ esiste un unico cammino $\{x_m\}_{m \geq 1}$ sull'albero che inizia in $x_1 = 1$ e converge a x . Se $x = [a_0; a_1, a_2, \dots]$ allora

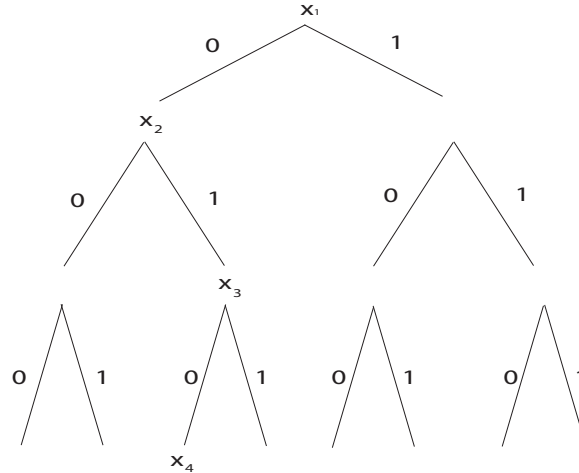
$$(7.55) \quad x_m = \frac{r p_{n-1} + p_{n-2}}{r q_{n-1} + q_{n-2}} \quad , \quad m = \sum_{i=0}^{n-1} a_i + r \quad , \quad 1 \leq r \leq a_n$$

Osserviamo che per $m \leq a_0 + 1$ si ha $x_m = m$. In altre parole, dato un numero positivo ed irrazionale x il cammino sull'albero che converge a x coincide con la successione dei convergenti di Farey descritta sopra.

Al cammino $\{x_m\}_{m \geq 1}$ possiamo poi associare una sequenza binaria $\sigma \in \{0, 1\}^{\mathbb{N}}$ ottenuta codificando i salti lungo il cammino (ovvero con la ricodifica $\{A, B\} \rightarrow \{0, 1\}$). Così facendo si ottiene

$$(7.56) \quad x = [a_0; a_1, a_2, \dots] \iff \sigma(x) = 1^{a_0} 0^{a_1} 1^{a_2} \dots$$

e vediamo che vale la relazione $p(x) = 0.\sigma(x)$, dove p è la funzione di Minkowski introdotta in (7.9).



7.9.2. *La distribuzione delle frazioni di Stern-Brocot.* Riprendendo la funzione p di Minkowski e ricordando il Lemma 7.2 osserviamo che $p(0/1) = 0$, $p(1/1) = 1/2$ e $p(1/0) = 1$ e dunque

$$(7.57) \quad p\left(\frac{0+1}{1+0}\right) = \frac{1}{2} \left[p\left(\frac{0}{1}\right) + p\left(\frac{1}{0}\right) \right]$$

e, più in generale, per ogni paio di frazioni consecutive a/b e a'/b' si ha

$$(7.58) \quad p\left(\frac{a+a'}{b+b'}\right) = \frac{1}{2} \left[p\left(\frac{a}{b}\right) + p\left(\frac{a'}{b'}\right) \right]$$

Cioè la funzione di Minkowski del figlio di due frazioni è uguale alla media aritmetica dei suoi genitori. Ciò implica che l'immagine con p dell'albero di Stern-Brocot, come insieme ordinato, è l'albero diadico le cui porzioni di 'profondità' crescente sono

$$D_0 = \left(\frac{0}{1}, \frac{1}{1}\right), \quad D_1 = \left(\frac{0}{1}, \frac{1}{2}, \frac{1}{1}\right), \quad D_2 = \left(\frac{0}{1}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{1}{1}\right)$$

e, in generale, $D_k \setminus \{0\}$ è l'insieme delle radici dell'unità di ordine 2^k . Più precisamente, se poniamo

$$T_k := \{a/b : \text{profondità}(a/b) \leq k\}$$

allora $p(T_k) = D_k$ (ciò si poteva dedurre anche dal Teorema 7.3). Vale quindi il seguente risultato

Teorema 7.21. *Si ha*

$$(7.59) \quad x = \lim_{k \rightarrow \infty} \frac{\#\{a/b \in D_k : a/b \leq x\}}{2^k}$$

e dunque

$$(7.60) \quad p(x) = \lim_{k \rightarrow \infty} \frac{\#\{a/b \in T_k : a/b \leq x\}}{2^k}.$$

Quale ulteriore immediata conseguenza, possiamo calcolare i coefficienti di Fourier-Stieltjes di $p(x)$.

(3) Vale la relazione di commutazione $R \circ F = F \circ R^2$.

Dimostrazione. Per la facile prova delle prime due rimandiamo a ([BI], Prop. 2.1). L'ultima segue da un'ispezione diretta. \square

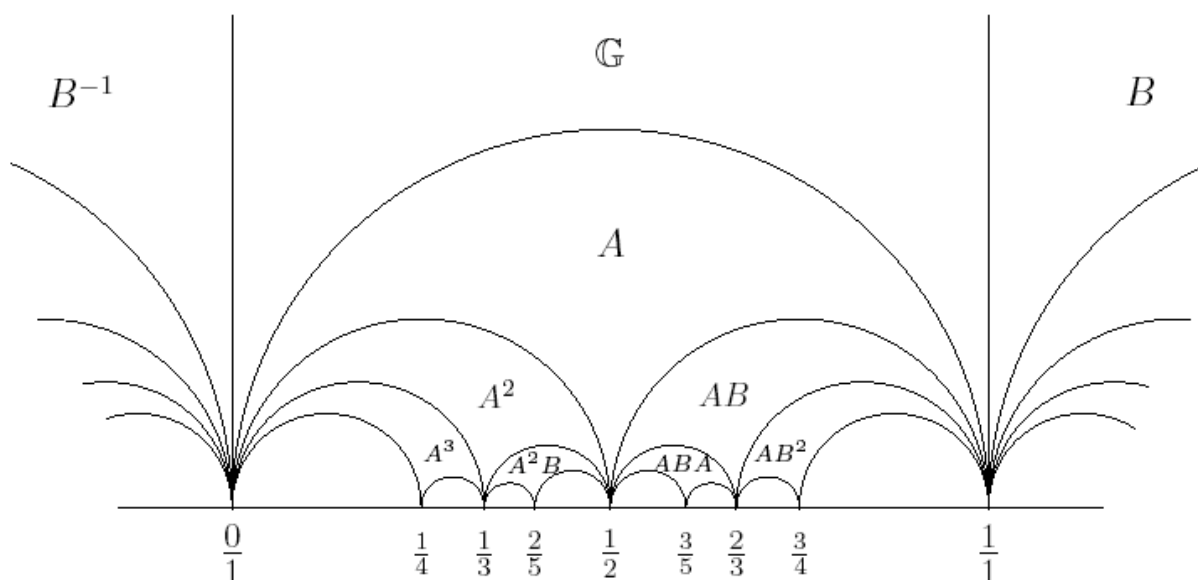
Osservazione 7.24. Sebbene la R -orbita di $\frac{1}{0}$ sia densa in \mathbb{R}^+ , essa 'diffonde' solo con velocità logaritmica. Infatti $R^i(\frac{1}{0}) = n$ per $i = 2^n$ e quindi

$$\sup_{0 < i \leq n} T^i(1/0) = O(\log n)$$

Si dimostra poi che (R, \mathbb{R}^+) è unicamente ergodico, e dunque ogni orbita $\{R^i(x) : i \geq 0\}$, $x \in \mathbb{R}_+$, è densa in \mathbb{R}_+ ([BI], Cor. 2.5).

Osservazione 7.25. Se \hat{x}_i l' i -esimo elemento dell'albero di Stern-Brocot permutato allora si ha la rappresentazione [CW]: $\hat{x}_{i+1} = b(i-1)/b(i)$, $i > 1$, dove $b(n)$ è il numero di rappresentazioni iperbinarie di n . Ad esempio $8 = 2^3 = 2^2 + 2^2 = 2^2 + 2 + 2 = 2^2 + 2 + 1 + 1$ e dunque $b(8) = 4$.

7.10. I triangoli di Farey e i cerchi di Ford. Abbiamo visto che le immagini del dominio fondamentale \mathbb{F} del gruppo discreto $\Gamma(1)$ formano una tassellazione di \mathbb{H} . Ricordiamo che $\Gamma(1)$ non è un gruppo libero. Le isometrie di \mathbb{H} associate alle matrici (7.54) sono $z \mapsto \frac{z}{z+1}$ e $z \mapsto z + 1$, rispettivamente, generano un'altra tassellazione di \mathbb{H} , per mezzo dei cosiddetti *triangoli di Farey*, a partire dal triangolo geodetico \mathbb{G} definito in (7.2), che contiene esattamente tre immagini di \mathbb{F} , e dunque ha tre volte la sua area, cioè $|\mathbb{G}| = \pi$.



Definiamo ora la *sequenza di Farey di ordine n* F_n come l'insieme delle frazioni p/q tali che $(p, q) = 1$ e $|p| \leq n$, $|q| \leq n$, ordinate secondo la loro grandezza crescente [Fa]. Ad esempio,

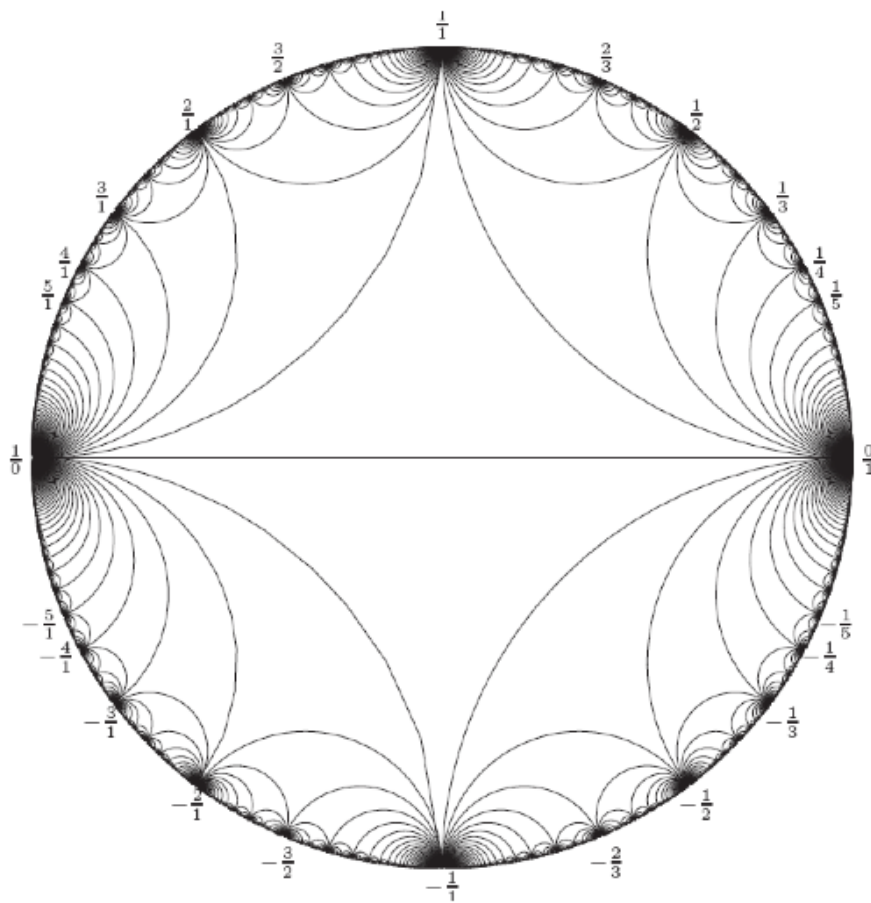
gli elementi non negativi delle prime quattro sequenze sono

$$\begin{aligned}
 F_1 &: \left\{ \frac{0}{1}, \frac{1}{1}, \frac{1}{0} \right\} \\
 F_2 &: \left\{ \frac{0}{1}, \frac{1}{2}, \frac{1}{1}, \frac{2}{1}, \frac{1}{0} \right\} \\
 F_3 &: \left\{ \frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}, \frac{2}{2}, \frac{3}{1}, \frac{1}{0} \right\} \\
 F_4 &: \left\{ \frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1}, \frac{1}{1}, \frac{1}{1}, \frac{1}{0} \right\}
 \end{aligned}$$

Per ogni coppia p/q e p'/q' di frazioni contigue in una qualunque di queste sequenze si ha $p'q - pq' = 1$, inoltre F_n si ottiene da F_{n-1} inserendo tra due frazioni contigue la loro somma di Farey $(p + p')/(q + q')$ (ogni volta che $\max\{p + p', q + q'\} \leq n$).

Inoltre si ha $|F_n| = 1 + 2 \sum_{k=1}^n \varphi(k) \sim 6n^2/\pi^2$ dove $\varphi(k) = |\{0 < i \leq k : m.c.d.(i, k) = 1\}|$ è la funzione di Eulero.

Una rappresentazione alternativa dei triangoli di Farey che usa il disco di Poincaré è riportata nella figura seguente



Osservazione 7.26. Posto $N = (|F_n| - 1)/2 = |\{x \in F_n : x < 1\}|$, indichiamo con x_l , $l = 1, \dots, N$ la sequenza (ordinata) delle frazioni in $F_n \cap (0, 1]$, l'equidistribuzione di tale sequenza, nel senso che

$$\sum_{l=1}^N \left| x_l - \frac{l}{N} \right| = O(N^{\frac{1}{2} + \epsilon}) \quad , \quad \forall \epsilon > 0$$

è equivalente all'ipotesi di Riemann (vedi [Kn]).

Se invece togliamo il vincolo di avere numeratore e denominatore inferiori ad un intero dato e si prendono *tutte* le frazioni ottenute con la procedura suindicata invece di F_n otteniamo esattamente l'insieme, che nella sezione precedente abbiamo chiamato T_n , delle $2^n + 1$ frazioni di profondità non superiore a n che compaiono nell'albero di Stern-Brocot (detta anche *sequenza di Farey modificata di ordine n*).

Sia ora $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$ e poniamo di nuovo $X(z) = (az + b)/(cz + d)$. Allora si ha che $X(0) = b/d$ e $X(\infty) = a/c$ sono frazioni contigue in F_n dove $n = \max\{|a|, |b|, |c|, |d|\}$. In particolare, $X(\mathbb{G})$ è un triangolo geodetico la cui cuspide centrale è $X(1) = p/q = (a+b)/(c+d)$. Viceversa, l'insieme delle immagini di \mathbb{G} con le matrici $X \in SL(2, \mathbb{Z})$ tali che $\max\{|a|, |b|, |c|, |d|\} \leq n$, ricopre un insieme convesso di \mathbb{H} le cui cuspidi formano esattamente la sequenza F_n .

Abbiamo visto che applicando la trasformazione $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ alla geodetica verticale coincidente con l'asse immaginario I si ottiene la geodetica che connette le due cuspidi $X(0)$ e $X(\infty)$. Se invece che alla geodetica verticale I , applichiamo X agli orocicli positivo e negativo di $v = (i, 0) \in S\mathbb{H}$, ovvero alla retta orizzontale $H^+ = \{z = i + t, t \in \mathbb{R}\}$ e al cerchio $H^- = \{z = \frac{i}{1+it}, t \in \mathbb{R}\}$ tangente ad \mathbb{R} in 0, otteniamo due cerchi: il primo tangente a \mathbb{R} in a/c e di diametro (euclideo) $1/c^2$, il secondo tangente a \mathbb{R} in b/d e di diametro (euclideo) $1/d^2$. I due cerchi sono tangenti nel punto $X(i)$.

ESEMPIO. Se $X = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$ allora $X(I) = \{z = \frac{ie^t + 2}{2ie^t + 5} : t \in \mathbb{R}\}$ è la geodetica che connette $2/5$ con $1/2$, $X(H^-) = \{z = \frac{i(2t+1)+2}{i(5t+2)+5} : t \in \mathbb{R}\}$ è il cerchio di diametro $1/25$ e tangente a \mathbb{R} in $2/5$ e $X(H^+) = \{z = \frac{i+t+2}{2i+2t+5} : t \in \mathbb{R}\}$ è il cerchio di diametro $1/4$ e tangente a \mathbb{R} in $1/2$. Tutte e tre queste curve si incontrano nel punto $X(i) = \frac{i+2}{2i+5}$ corrispondente a $t = 0$.

Al variare di $X \in SL(2, \mathbb{Z})$ le immagini di H^+ (o H^-) formano la famiglia dei *cerchi di Ford* [Fo] ciascuno dei quali, indicato con $C_{\frac{p}{q}}$, è tangente a \mathbb{R} in $\frac{p}{q}$ ed ha diametro $\frac{1}{q^2}$.

Lemma 7.27. *Due cerchi di Ford $C_{\frac{p}{q}}$ e $C_{\frac{p'}{q'}}$, con $\frac{p}{q} < \frac{p'}{q'}$, o non si intersecano, oppure sono tra di loro tangenti in $\frac{ip+p'}{iq+q'}$ se $p'q - pq' = 1$. Tre cerchi $C_{\frac{p}{q}}$, $C_{\frac{p'}{q'}}$ e $C_{\frac{p''}{q''}}$ con $\frac{p}{q} < \frac{p''}{q''} < \frac{p'}{q'}$ sono tangenti tra di loro se e solo se $\frac{p''}{q''} = \frac{p}{q} \oplus \frac{p'}{q'}$.*

Dimostrazione. Il quadrato della distanza D tra i centri di $C_{\frac{p}{q}}$ e $C_{\frac{p'}{q'}}$ è

$$D^2 = \left(\frac{p}{q} - \frac{p'}{q'}\right)^2 + \left(\frac{1}{2q^2} - \frac{1}{2q'^2}\right)^2$$

mentre il quadrato della somma dei loro raggi r e r' è

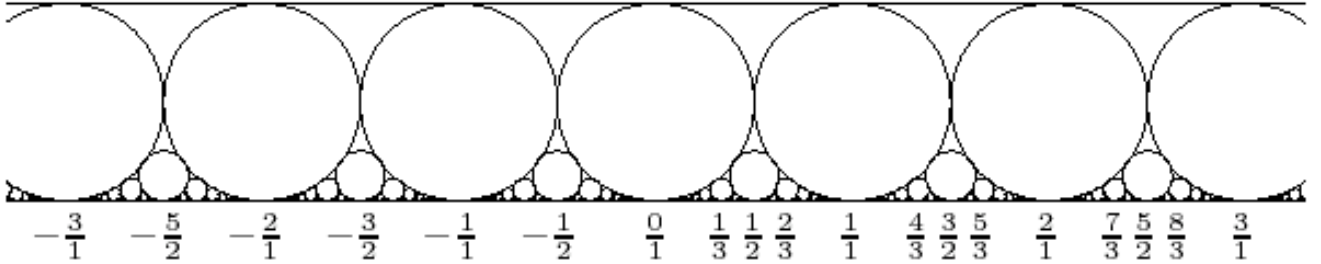
$$(r + r')^2 = \left(\frac{1}{2q^2} + \frac{1}{2q'^2}\right)^2$$

Da ciò si ricava

$$D^2 - (r + r')^2 = \frac{(pq' - qp')^2 - 1}{q^2q'^2} \geq 0$$

con uguaglianza se e solo se $pq' - qp' = 1$. Riguardo alla seconda affermazione osserviamo che se $C_{\frac{p}{q}}$ e $C_{\frac{p'}{q'}}$ sono tangenti allora $(q + q')p' - (p + p')q' = 1$ e $(p + p')q - (q + q')p = 1$ e quindi tra di loro ne esiste un altro tangente ad entrambi e tangente in \mathbb{R} in $\frac{p''}{q''}$. Il viceversa è immediato. \square

In particolare, ogni cerchio di Ford ha infiniti cerchi di Ford tangenti. In questo modo otteniamo il cosiddetto *impacchettamento di Farey* della striscia unitaria $\{z \in \mathbb{H} : \Im z \leq 1\}$.



Osservazione 7.28. Dal fatto che i cerchi di Ford $C_{\frac{p}{q}}$ non si intersecano segue che la loro area totale A al variare di $\frac{p}{q}$ in $[0, 1)$ si può calcolare per mezzo di una serie convergente. Più precisamente si ha

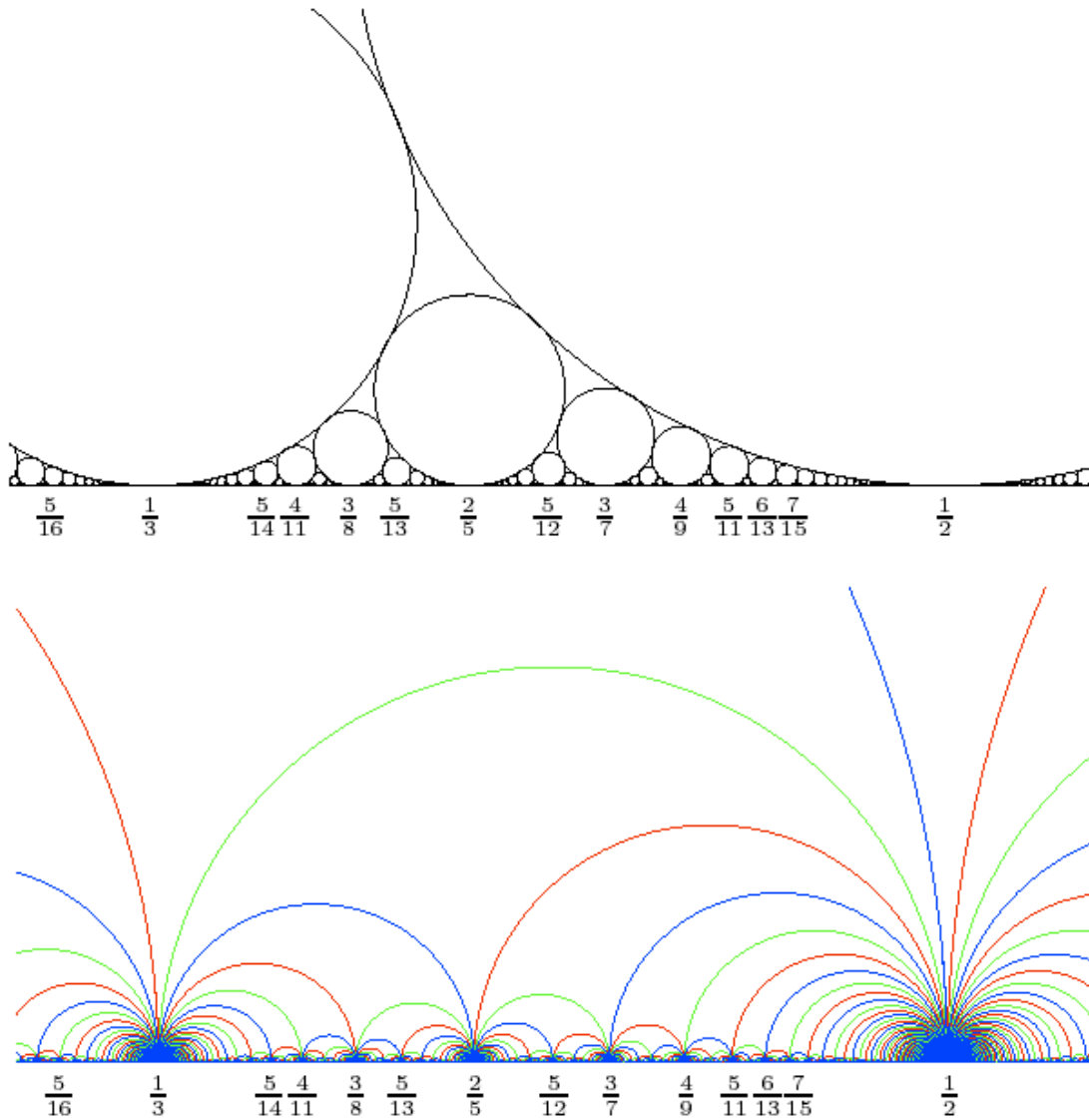
$$A = \sum_{q \geq 1} \sum_{\substack{(p,q)=1 \\ 1 \leq p < q}} \pi \left(\frac{1}{2q^2}\right)^2 = \frac{\pi}{4} \sum_{q \geq 1} \frac{1}{q^4} \sum_{\substack{(p,q)=1 \\ 1 \leq p < q}} 1 = \frac{\pi}{4} \sum_{q \geq 1} \frac{\varphi(q)}{q^4} = \frac{\pi}{4} \frac{\zeta(3)}{\zeta(4)} \simeq 0,872284041$$

dove $\varphi(q)$ è la *funzione di Eulero* che dà il numero di interi minori di q e ad esso coprimi (ad es. $\varphi(9) = 6$ e $\varphi(11) = 10$), a cui risulta associata la serie di Dirichlet

$$\sum_{n \geq 1} \frac{\varphi(n)}{n^s} = \frac{\zeta(s-1)}{\zeta(s)}$$

dove $\zeta(s) := \sum_{n \geq 1} n^{-s}$ è la *funzione zeta di Riemann*.

Nelle prossime due figure vediamo un ingrandimento della regione intorno a $z = 2/5$ con i cerchi di Ford (orocicli) e i triangoli di Farey (geodetiche).



7.11. **Geodetiche di scattering.** Infine, se ad X associamo la *geodetica di scattering* $c_{p/q}$ in \mathbb{H} proveniente da ∞ e convergente nella cuspidale centrale di $X(\mathbb{G})$ (cioè in p/q) allora, in accordo con quanto visto sopra (cf. (7.54) e Proposizione 7.16), a tale geodetica corrisponde il cammino sull'albero di Stern-Brocot che parte da 1 e termina in p/q (che per semplicità supponiamo positivo) e viceversa. Sebbene le geodetiche di scattering abbiano lunghezza infinita, è possibile confrontarne la lunghezza rispetto ad una di riferimento, come ad esempio $c_0 = I$. A questo scopo serve la nozione di *ritardo* [Kn].

Definizione 7.29. Il ritardo $\tau_{p/q}$ della geodetica di scattering $c_{p/q}$ è il tempo necessario per andare lungo di essa dal punto d'intersezione $z_{p/q}^- := c_{p/q} \cap C_{p/q}$ in \mathbb{H} fino a $z_{p/q}^+ := c_{p/q} \cap H^+$.

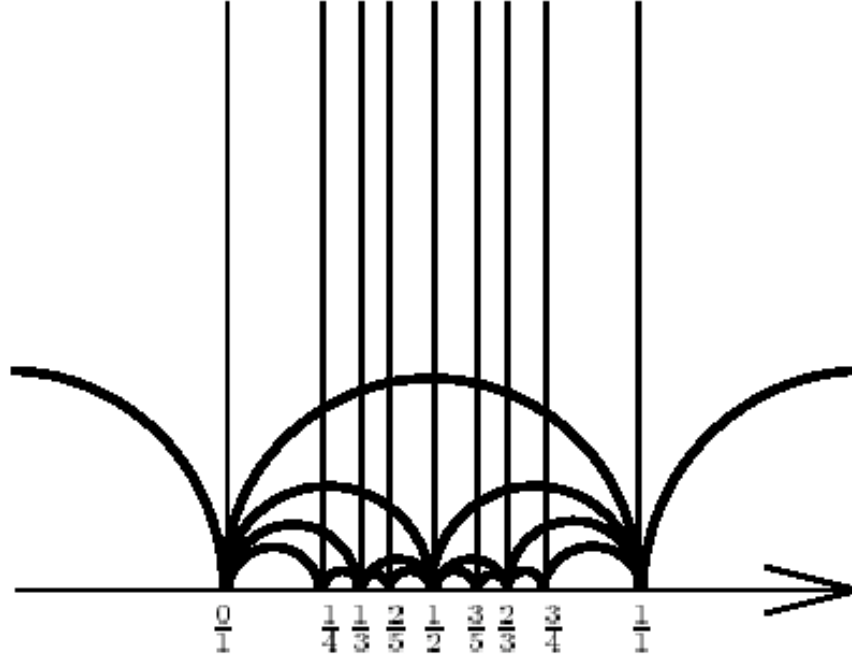
Lemma 7.30.

$$\tau_{p/q} = 2 \log q$$

Dimostrazione. Si ha $z_{p/q}^- := c_{p/q} \cap C_{p/q} = \frac{p}{q} + \frac{i}{q^2}$ e $z_{p/q}^+ := c_{p/q} \cap H^+ = \frac{p}{q} + i$. Pertanto, usando la (1.4) si vede che

$$(7.64) \quad \tau_{p/q} = d(z_{p/q}^-, z_{p/q}^+) = 2 \log q \quad \square$$

Osserviamo che $\tau_n = 0, \forall n \in \mathbb{Z}$.

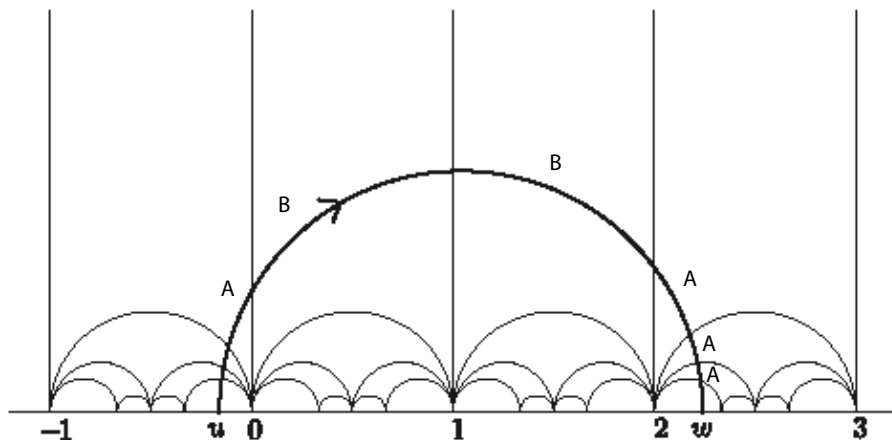


La nozione di ritardo assume un'importanza centrale nella connessione tra moto illimitato 'classico' (lungo le geodetiche di scattering appunto) e scattering 'quantistico' descritto dalle autofunzioni non appartenenti a L^2 dell'operatore laplaciano iperbolico.

7.12. Geodetiche periodiche. Discuteremo ora il comportamento delle geodetiche chiuse. Il materiale esposto è liberamente tratto da [DP] e dall'articolo di Caroline Series in [BKS]. Consideriamo nuovamente la tassellazione di Farey, generata dalle trasformazioni $A = SR^{-1}S$ e $B = R$ introdotte in (7.54). Una geodetica γ orientata con estremi $u < 0 < w, u, w \in \mathbb{R} \setminus \mathbb{Q}$ interseca un'infinità di triangoli di Farey. Più precisamente, posto $L_0 = \mathbb{G}$ indichiamo con L_1, L_2, \dots , (rispettivamente L_{-1}, L_{-2}, \dots) la successione dei triangoli di Farey attraversati da γ se percorsa nel senso positivo (rispettivamente nel senso negativo) a partire dalla sua intersezione con I . Sia Y_n l'elemento di $PSL(2, \mathbb{Z})$ tale che $Y_n(\mathbb{G}) = L_n, n \in \mathbb{Z}$. Evidentemente $Y_0 = Id$ e, in accordo con quanto visto più sopra, $Y_n = \prod_{1 \leq k \leq n} X_k$ per $n \geq 1$, con $X_k \in \{A, B\}$. Inoltre, identificando $Y \in PSL(2, \mathbb{Z})$ con il numero razionale $Y(1)$ (vedi (7.53)) allora si ha $Y_n \rightarrow w$ quando $n \rightarrow \infty$. Non è difficile rendersi conto che $X_k = A$ se e solo se la geodetica γ interseca L_k lasciando il vertice del triangolo geodetico da essa tagliato alla sua destra, mentre $X_k = B$ se lo lascia alla sua sinistra. In altre parole possiamo codificare la geodetica γ con una sequenza simbolica (*cutting sequence*) del tipo $\dots A^{m_{-2}} B^{m_{-1}} A^{m_0} B^{m_1} A^{m_2} B^{m_3} \dots$ dove $m_i \in \mathbb{N}$.

Lemma 7.31. *Sia $w > 1$ e sia $\gamma^+ := \gamma \cap \{\Re z > 0\}$ il segmento di geodetica che unisce un punto su I con w , codificato dalla sequenza $B^{m_1} A^{m_2} B^{m_3} \dots$. Allora $w = [m_1, m_2, m_3, \dots]^{-1}$.*

Se invece $0 < w < 1$ allora la sequenza inizia con A e $w = [m_1, m_2, \dots]$.¹⁵ Se infine $w < 0$ allora lo stesso vale per la corrispondente γ^- con $-w$ al posto di w ed A e B scambiati.



Dimostrazione. Prendiamo il caso $w > 1$. È chiaro che $[w] = m_1$ è il numero di rette verticali della tassellazione di Farey attraversate da γ^+ prima che inizi a scendere verso w . Se $q = \gamma^+ \cap \{\Re z = m_1\}$ allora $SR^{-m_1}(q) = -(q - m_1)^{-1} =: q_1 \in I$ e SR^{-m_1} manda γ^+ nel segmento di geodetica γ_1^- orientato verso l'asse negativo che connette q_1 con il punto $w_1 = -(w - m_1)^{-1}$. Gli m_2 tratti di tipo A che seguono gli m_1 tratti di tipo B in γ^+ , appaiono ora come m_2 strisce verticali attraversate da γ_1^- prima di iniziare a discendere verso w_1 . Dunque $m_2 = [(w - m_1)^{-1}]$ e dunque $w = m_1 + 1/(m_2 + r)$ con $0 < r < 1$. Possiamo ora applicare SR^{m_2} a γ_1^- e procedere come sopra. \square

OSSERVAZIONE. Un semplice corollario è che se γ e γ' sono due geodetiche in \mathbb{H} convergenti ad uno stesso punto w , le loro sequenze simboliche coincidono da un certo punto in poi.

Per capire come la rappresentazione descritta sopra sia utile per classificare le geodetiche chiuse sulla superficie modulare M , consideriamo una geodetica γ su \mathbb{H} e sia c_0 un suo *punto di cambiamento*, ossia un punto giacente sulla frontiera di un triangolo di Farey in cui la sequenza simbolica di γ cambia da A a B , o da B ad A . D'altra parte, possiamo sempre trasportare il punto c_0 su I con un elemento di $PSL(2, \mathbb{Z})$ e riportarsi così alla situazione descritta nel lemma appena visto. Supponendo quindi che $c_0 \in I$, la geodetica descritta connette due punti $u, w \in \mathbb{R}$ tali che $u < 0 < w$ e $(1 + u)(1 - w) < 0$, e che possono dunque essere rappresentati nella forma

$$u = S([m_0, m_{-1}, m_{-2}, \dots]^{-\epsilon}) \quad , \quad w = [m_1, m_2, m_3, \dots]^{-\epsilon}$$

con $\epsilon \in \{\pm 1\}$. Se dunque alla geodetica 'indicizzata' (γ, c_0) associamo la quadrupletta $([m_0, m_{-1}, m_{-2}, \dots], [m_1, m_2, m_3, \dots], \epsilon, c_0)$ allora la geodetica indicizzata (γ, c_1) , dove c_1 è il punto di cambiamento successivo, cioè $c_1 = g(c_0)$ dove $g = B^{m_1}$ se $\epsilon = 1$ e $g = A^{m_1}$ se $\epsilon = -1$, sarà associata a

$$([m_1, m_0, m_{-1}, \dots], [m_2, m_3, m_4, \dots], -\epsilon, c_1) =: \sigma([m_0, m_{-1}, m_{-2}, \dots], [m_1, m_2, m_3, \dots], \epsilon, c_0)$$

Osserviamo che l'azione di σ sulla seconda coordinata corrisponde alla mappa di Gauss G .

¹⁵La distinzione tra i due casi $w > 1$ e $0 < w < 1$ è conseguenza del fatto che stiamo qui usando la notazione abbreviata $[0; m_1, m_2, \dots] \equiv [m_1, m_2, \dots]$, che ci sarà utile nel seguito, per rappresentare punti nell'intervallo unitario.

Sia ora g un elemento iperbolico di $PSL(2, \mathbb{Z})$ e denotiamo con $Fix(g)$ la geodetica fissata da g orientata dal punto repulsivo a quello attrattivo, e c_0 un suo punto di cambiamento. Se $c_0 \notin I$ indichiamo con g_0 l'elemento di $PSL(2, \mathbb{Z})$ tale che $g_0(I)$ è la frontiera di Farey su cui giace c_0 . Rappresentiamo quindi la geodetica indicizzata $(Fix(g), c_0)$ con la quadrupletta $(x_-, x, \epsilon, g_0) \in [0, 1]^2 \times \{\pm 1\} \times PSL(2, \mathbb{Z})$, dove se $g_0 \neq Id$ avremo $u = g_0(S(x_-^{-\epsilon}))$ e $w = g_0(x_-^{-\epsilon})$. Essendo $(g(Fix(g)), g(c_0)) = (Fix(g), g(c_0))$ la coppia $(Fix(g), g(c_0))$ sarà rappresentata da (x_-, x, ϵ, gg_0) . D'altra parte $g(c_0)$ è necessariamente un altro punto di cambiamento, diciamo c_k per qualche $k \geq 1$, di $Fix(g)$. Dunque, per quanto visto sopra, $(x_-, x, \epsilon, gg_0) = \sigma^k(x_-, x, \epsilon, g_0)$. Da ciò si deduce agevolmente il

Lemma 7.32. *Sia g un elemento iperbolico di $PSL(2, \mathbb{Z})$. Allora esiste $g_0 \in PSL(2, \mathbb{Z})$, $\epsilon \in \{\pm 1\}$ e una sequenza di k interi (m_1, \dots, m_k) strettamente positivi, con $k \geq 1$ pari, tali che $g = g_0 M_\epsilon^{m_1} \dots M_{-\epsilon}^{m_k} g_0^{-1}$ dove $M_{-1} = A$ e $M_1 = B$. Inoltre gli estremi di $Fix(g)$ sono dati da $u = g_0(S([\overline{m_k, \dots, m_k}]^{-\epsilon}))$ e $w = g_0([\overline{m_1, \dots, m_k}]^{-\epsilon})$, dove $[\overline{m_1, \dots, m_k}]$ indica la frazione continua periodica di periodo $[m_1, \dots, m_k]$.*

Determiniamo ora il modulo dell'autovalore $|\lambda_g| > 1$ di g .

Lemma 7.33. *L'autovalore λ con modulo superiore a 1 di $M_\epsilon^{m_1} \dots M_{-\epsilon}^{m_k}$ è uguale a*

$$\prod_{l=0}^{k-1} (G^l([\overline{m_1, \dots, m_k}]))^{-1}$$

Dimostrazione. Dal Lemma 7.32 si ha che¹⁶

$$M_\epsilon^{m_1} \dots M_{-\epsilon}^{m_k} \begin{pmatrix} [\overline{m_1, \dots, m_k}]^{-\epsilon} & \\ & 1 \end{pmatrix} = \lambda \begin{pmatrix} [\overline{m_1, \dots, m_k}]^{-\epsilon} & \\ & 1 \end{pmatrix}$$

Ora, le matrici M_1^m e M_{-1}^m si scrivono anche nella forma $M_1^m = JD_m$ e $M_{-1}^m = D_m J$, dove $D_m = \begin{pmatrix} 0 & 1 \\ 1 & m \end{pmatrix}$ e $J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Essendo $J^2 = Id$ la relazione scritta sopra si riscrive nella forma

$$D_{m_1} \dots D_{m_k} \begin{pmatrix} 1 & \\ [\overline{m_1, \dots, m_k}]^{-1} & \end{pmatrix} = \lambda \begin{pmatrix} 1 & \\ [\overline{m_1, \dots, m_k}]^{-1} & \end{pmatrix}$$

L'identità $D_m \begin{pmatrix} 1 \\ x \end{pmatrix} = x \begin{pmatrix} 1 \\ m + \frac{1}{x} \end{pmatrix}$ dà

$$D_{m_1} \dots D_{m_k} \begin{pmatrix} 1 \\ [\overline{m_1, \dots, m_k}]^{-1} \end{pmatrix} = [\overline{m_1, \dots, m_k}]^{-1} D_{m_1} \dots D_{m_{k-1}} \begin{pmatrix} 1 \\ [\overline{m_k, m_1, \dots, m_{k-1}}]^{-1} \end{pmatrix}$$

coscicchè iterando l'identità k volte il termine a destra diviene

$$[\overline{m_1, \dots, m_k}]^{-1} [\overline{m_2, \dots, m_k, m_1}]^{-1} \dots [\overline{m_k, m_1, \dots, m_{k-1}}]^{-1} \begin{pmatrix} 1 \\ [\overline{m_1, \dots, m_k}]^{-1} \end{pmatrix} \quad \square$$

Siamo ora pronti per effettuare il passaggio al quoziente. Una geodetica $\tilde{\gamma}$ su M è indicizzata da un punto di cambiamento \tilde{c}_0 se si solleva su \mathbb{H} in una geodetica indicizzata da c_0 . I diversi sollevamenti su \mathbb{H} di $(\tilde{\gamma}, \tilde{c}_0)$ corrispondono alle immagini di (γ, c_0) con $PSL(2, \mathbb{Z})$. Così,

¹⁶È facile verificare che se $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$ con $|a + d| \equiv |\rho| > 2$ allora l'autovalore di modulo > 1 è $\lambda = (\rho + \sqrt{\rho^2 - 4})/2$ e un suo autovettore è $\begin{pmatrix} w \\ 1 \end{pmatrix}$ dove $w = (\lambda - d)/c$ è la soluzione più grande di $X(w) = w$.

se (γ, c_0) è codificata da (x_-, x, ϵ, g_0) , l'insieme di tutti i sollevamenti è rappresentato da $\{(x_-, x, \epsilon, gg_0), g \in PSL(2, \mathbb{Z})\}$. Per codificare $(\tilde{\gamma}, \tilde{c}_0)$ scegliamo dunque il rappresentate con $gg_0 = Id$, indicato semplicemente con (x_-, x, ϵ) (come già visto si tratta dell'unica geodetica γ su \mathbb{H} che si proietta su $\tilde{\gamma}$, che connette due punti $u, w \in \mathbb{R}$ tali che $u < 0 < w$ e $(1+u)(1-w) < 0$ e che ha il suo punto di cambiamento c_0 su I).

Se ora $\tilde{\gamma}$ è una geodetica primitiva chiusa di M , essa si solleva su \mathbb{H} nell'asse $Fix(g)$ di qualche elemento iperbolico di $PSL(2, \mathbb{Z})$ primitivo (cioè $g \neq \tau^n$ con $\tau \in PSL(2, \mathbb{Z})$ e $n > 1$). La lunghezza di $\tilde{\gamma}$ sarà uguale alla distanza tra un punto arbitrario $z \in Fix(g)$ e la sua immagine $g(z) \in Fix(g)$, ovvero $2 \log |\lambda_g|$ dove λ_g è l'autovalore di modulo > 1 di g . Sia (x_-, x, ϵ) la tripletta che codifica $(\tilde{\gamma}, \tilde{c}_0)$. Per quanto visto sopra, si può trovare un intero pari $k > 1$ e k interi strettamente positivi (m_j) tali che $x_- = [\overline{m_k, \dots, m_1}]$ e $x = [\overline{m_1, \dots, m_k}]$. Dunque per codificare $(\tilde{\gamma}, \tilde{c}_0)$ possiamo usare solo (x, ϵ) (visto che x_- si deduce immediatamente da quello). Siccome $\sigma([\overline{m_1, \dots, m_k}], \epsilon) = ([\overline{m_2, \dots, m_k, m_1}], -\epsilon)$ codifica le geodetica indicizzata dal punto di cambiamento successivo a \tilde{c}_0 , la geodetica $\tilde{\gamma}$ contiene esattamente k punti di cambiamento diversi. Mettendo insieme quanto visto sopra possiamo dire che la lunghezza di una geodetica chiusa ed indicizzata $(\tilde{\gamma}, \tilde{c}_0)$ su M , codificata dalla coppia $([\overline{m_1, \dots, m_k}], \epsilon)$, è uguale a $-2 \log \prod_{l=0}^{k-1} G^l([\overline{m_1, \dots, m_k}])$.

Abbiamo già visto in precedenza che orbite periodiche del flusso geodetico sono dense in SM . Possiamo ora dare una stima precisa del loro numero. Denotiamo con \mathcal{C} l'insieme delle geodetiche chiuse primitive su M e se $\gamma \in \mathcal{C}$, sia $l(\gamma)$ la sua lunghezza. Si chiama *spettro delle lunghezze* di M l'insieme $\{l(\gamma), \gamma \in \mathcal{C}\}$, dove le varie lunghezze sono ripetute in accordo con la loro molteplicità. Osservando che se $G^k(x) = x$ allora $\prod_{l=0}^{k-1} (G^l(x))^2 = |(G^k)'(x)|^{-1}$ si ha dunque che

Proposizione 7.34. *Esiste una corrispondenza biunivoca tra lo spettro delle lunghezze di M e l'insieme dei valori $\log |(G^{2p})'(x)|$ dove x è un punto periodico di periodo p di $G^2 : [0, 1] \rightarrow [0, 1]$, ovvero $x \in \{y, G^2(y), \dots, (G^2)^{p-1}(y)\}$ con $G^{2p}(y) = y$.*

Corollario 7.35. *Sia $\pi_0(L)$ il numero di geodetiche chiuse primitive di lunghezza non superiore ad L . Allora*

$$\pi_0(L) = \sum_{p=1}^{\infty} \frac{1}{p} \#\{x \in]0, 1] : G^k(x) \neq x, 1 \leq k < 2p, G^{2p}(x) = x \text{ e } \log |(G^{2p})'(x)| \leq L\}$$

Le informazioni sullo spettro delle lunghezze possono essere tutte incorporate in una sorta di funzione generatrice, detta *funzione zeta di Selberg* [LZ], e definita per $\Re s > 1$ dal prodotto infinito

$$(7.65) \quad Z(s) = \prod_{\gamma \in \mathcal{C}} \prod_{k \geq 0} (1 - e^{-(s+k)l(\gamma)})$$

Si può mostrare che $Z(s)$ ammette un prolungamento analitico al semipiano $\Re s > 1/2$ ed è ivi priva di zeri ad eccezione di uno zero semplice in $s = 1$. Usando questa funzione in modo analogo a quello in cui la funzione zeta di Riemann è stata usata da Hadamard e La Vallée-Poussin per dimostrare il teorema dei numeri primi, è possibile mostrare che il comportamento asintotico di $\pi_0(L)$ è

$$(7.66) \quad \pi_0(L) = Li(e^L) + O(e^{\beta L})$$

dove $Li(x) = \int_2^x (\log t)^{-1} dt$ è l'integrale logaritmico standard e $0 < \beta < 1$ (vedi [BKS]). Non abbiamo qui lo spazio per addentrarci nella prova di questo risultato. Limitiamoci ad osservare che in virtù della Proposizione 7.34 possiamo manipolare $Z(s)$, prendendo $\Re s$ grande abbastanza da assicurare la convergenza assoluta di ogni termine, e usando l'identità $\log(1/(1-x)) = \sum_{n \geq 1} x^n/n$, come segue

$$\begin{aligned}
Z(s) &= \exp \left(- \sum_{m \geq 1} \frac{1}{m} \sum_{k \geq 0} \sum_{\gamma \in \mathcal{C}} e^{-m(s+k)l(\gamma)} \right) = \exp \left(- \sum_{m \geq 1} \frac{1}{m} \sum_{\gamma \in \mathcal{C}} \frac{e^{-msl(\gamma)}}{1 - e^{-ml(\gamma)}} \right) \\
&= \exp \left(- \sum_{m \geq 1} \sum_{p \geq 1} \frac{1}{mp} \sum_{\substack{G^{2p}(x)=x \\ G^k(x) \neq x, 1 \leq k < 2p}} \frac{|(G^{2p})'(x)|^{-sm}}{1 - |(G^{2p})'(x)|^{-m}} \right) \\
(7.67) \quad &= \exp \left(- \sum_{n \geq 1} \frac{1}{n} \sum_{G^{2n}(x)=x} \frac{|(G^{2n})'(x)|^{-s}}{1 - |(G^{2n})'(x)|^{-1}} \right)
\end{aligned}$$

Dunque il comportamento di $Z(s)$ è interamente determinato dalla distribuzione dei punti periodici di G^2 . Osserviamo infine che un classico teorema di Lagrange afferma che un numero reale ha uno sviluppo in frazione continua (eventualmente) periodico se e solo se è un *irrazionale quadratico*, cioè se e solo se è soluzione di un'equazione quadratica con coefficienti razionali, e dunque è della forma $(a + b\sqrt{c})/d$ con $a, b, c, d \in \mathbb{Z}$, $d \neq 0$ e $c > 1$ e non quadrato. L'insieme di tali numeri ha la stessa cardinalità delle quadruplette ordinate di interi ed è pertanto numerabile.

REFERENCES

- [AO] H APPELGATE, H ONISHI, The slow continued fraction algorithm via 2×2 integer matrices, *Amer. Math. Monthly* **90** (1983), 443–455
- [BI] C BONANNO, S ISOLA, Orderings of the rationals and dynamical systems, *Colloquium Mathematicum* **116**, No. 2 (2009), 165-189
- [Bil] P BILLINGSLEY, *Ergodic Theory and Information*, New York: Wiley, 1965
- [BKS] T BEDFORD, M KEANE, C SERIES EDS., *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, Oxford Univ. Press, 1991
- [Br] A BROCOLT, Calcul des rouages par approximation, nouvelle méthode, *Revue Chronométrique* **6** (1860), 186-194
- [CW] N CALKIN, H S WILF, Recounting the rationals, *Amer. Math. Monthly* **107** (2000), 360-363
- [CI] M CAMPANINO, S ISOLA, Infinite invariant measures for non-uniformly expanding transformations of $[0, 1]$: weak law of large numbers with anomalous scaling, *Forum Mathem.* **8** (1996), 71-92.
- [DP] F DAL'BO, M PEIGNÉ, Comportement asymptotique de nombre de géodésiques fermés sur la surface modulaire en courbure non constante, *Astérisque* **238** (1996), 129-177.
- [Da] S G DANI, On uniformly distributed orbits of certain horocycle flows, *Ergod. Th. Dynam. Syst.* **2** (1982), 139-158.
- [DS] S G DANI, J D SMILLIE, Uniform distribution of horocycle orbits for Fuchsian groups, *Duke Math. J.* **51** (1984), 185-194.
- [DV] H G DIAMOND, J D VAALER, Estimates for partial sums of continued fraction partial quotients, *Pacific J. Math.* **122** (1986), 73-82.
- [Fa] J FAREY, On a Curious Property of Vulgar Fractions, *London, Edinburgh and Dublin Phil. Mag.* **47** (1816), 385.
- [Fo] L R FORD, Fractions, *Amer. Math. Monthly*, 45, n. 9 (1938), 586601.
- [GKP] R L GRAHAM, D E KNUTH, O PATASHNIK, *Concrete Mathematics*, Addison-Wesley 1990
- [Is1] S ISOLA, ζ -function and distribution of periodic orbits of toral automorphisms *Europhysics Letters* **11** (1990), 517-522.
- [Is2] S ISOLA, *Temperamenti: aritmetica e teoria musicale*, <http://unicam.it/stefano.isola/materiale.html>
- [Kh1] A YA KHINCHIN, *Continued Fractions*, The University of Chicago Press, 1964
- [Kh2] A YA KHINCHIN, Metrische Kettenbruchprobleme, *Compositio Math.*, **1** (1935), 361-382.
- [LZ] J B LEWIS, D ZAGIER, Period functions and the Selberg zeta function for the modular group, in *The Mathematical Beauty of Physics*, 83–97, *Adv. Series in Math. Phys.* **24**, World Sci. Publ., River Edge, NJ, 1997
- [Kn] A KNAUF, Number theory, dynamical systems and statistical mechanics, *Reviews in Mathematical Physics* **11** (1999), 1027–1060
- [N] M NEWMAN, Recounting the rationals, Continued, *Amer. Math. Monthly* **110** (2003), 642-643
- [Sa] P SARNAK, Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series, *Comm. Pure Appl. Math.* **34** (1981), 719–739.
- [St] M STERN, Über eine zahlentheoretische Funktion, *Journal für die reine und angewandte Mathematik* **55** (1858), 193-220
- [Ve] L VEPSTAS, The Minkowski Question Mark and the Modular Group $SL(2, \mathbb{Z})$, <http://linas.org/>